



HAL
open science

Remaniements chromosomiques complexes : de la caractérisation aux conséquences fonctionnelles

Nicolas Chatron

► **To cite this version:**

Nicolas Chatron. Remaniements chromosomiques complexes : de la caractérisation aux conséquences fonctionnelles. Génétique humaine. Université de Lyon, 2019. Français. NNT : 2019LYSE1311 . tel-02613417

HAL Id: tel-02613417

<https://theses.hal.science/tel-02613417>

Submitted on 20 May 2020

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

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

N°d'ordre NNT : xxx



THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N°476
Neurosciences et Cognition

Spécialité de doctorat : Sciences de la vie
Discipline : Génétique Humaine

Soutenue publiquement le 20/12/2019, par :
Nicolas CHATRON

Remaniements chromosomiques complexes : de la caractérisation aux conséquences fonctionnelles

Devant le jury composé de :

SCHAEFFER Laurent	PU-PH	Université Claude Bernard Lyon I	Président
MAGDINIER Frédérique	DR2 INSERM	Aix-Marseille Université	Rapporteuse
PELLESTOR Franck	PU-PH	Université de Montpellier	Rapporteur
REYMOND Alexandre	Professeur	Université de Lausanne	Examineur
SANLAVILLE Damien	PU-PH	Université Claude Bernard Lyon I	Directeur
SCHLUTH-BOLARD Caroline	MCU-PH	Université Claude Bernard Lyon I	Co-directrice

UNIVERSITE CLAUDE BERNARD LYON 1

Président	Pr Frédéric FLEURY
Président du Comité de Coordination Des Etudes Médicales	Pr Pierre COCHAT
Directeur Général des services	M. Damien VERHAEGHE
Secteur Santé :	
Doyen de l'UFR de Médecine Lyon Est	Pr Gilles RODE
Doyenne de l'UFR de Médecine Lyon-Sud Charles Mérieux	Pr Carole BURILLON
Doyenne de l'Institut des Sciences Pharmaceutiques (ISPB)	Pr Christine VINCIGUERRA
Doyenne de l'UFR d'Odontologie	Pr Dominique SEUX
Directrice du département de Biologie Humaine	Pr Anne-Marie SCHOTT
Secteur Sciences et Technologie :	
Directeur de l'UFR Sciences et Technologies	M. Fabien DE MARCHI
Directeur de l'UFR Sciences et Techniques des Activités Physiques et Sportives (STAPS)	M. Yanick VANPOULLE
Directeur de Polytech	Pr Emmanuel PERRIN
Directeur de l'IUT	Pr Christophe VITON
Directeur de l'Institut des Sciences Financières Et Assurances (ISFA)	M. Nicolas LEBOISNE
Directrice de l'Observatoire de Lyon	Pr Isabelle DANIEL
Directeur de l'Ecole Supérieure du Professorat et de l'Education (ESPé)	Pr Alain MOUGNIOTTE

A Maxime et Matthieu,

Remerciements

Au Professeur Damien Sanlaville, pour votre indéfectible soutien et vos précieux conseils depuis ma première visite dans votre bureau il y a 8 ans. Vous êtes à l'origine de ce travail et plus largement de ma passion pour notre discipline sous toutes ses facettes. Vos qualités managériales et humaines sont une grande source d'inspiration. Merci de votre gentillesse et de votre écoute dans les moments de doute qui m'ont traversé pendant ce travail de thèse. Je ne l'oublierai pas.

A la Docteure Caroline Schluth-Bolard, tu as poussé un peu plus chaque jour ma réflexion autour de ce travail et m'a soutenu pour aller jusqu'au bout. C'est avec toi que j'ai appris la persévérance nécessaire à toute activité de recherche et je t'en remercie. Tes intuitions ont toujours été les bonnes au cours de ce travail. J'espère développer ce même talent en travaillant avec toi dans le futur. Je suis ravi du chemin parcouru à tes côtés depuis les premiers essais douloureux de séquençage mate-pair jusqu'à aujourd'hui et me réjouit des années à venir.

Au Professeur Alexandre Reymond, pour ton accueil chaleureux à Lausanne pendant cette année pour moi et ma famille. Nos discussions scientifiques ont largement contribué à améliorer le travail présenté ici et à ouvrir de nouvelles voies, ensemble je l'espère, pour les années à venir. Nos discussions non scientifiques sont également d'une grande valeur à mes yeux. Tes qualités humaines et scientifiques sont une source d'inspiration parfaite pour la suite de ma vie professionnelle.

Au Professeur Laurent Schaeffer, pour avoir accepté de juger ce travail et présider le jury de soutenance. J'espère que vous trouverez ici les premières pierres de collaborations possibles au sein de l'Institut NeuroMyoGène sur ces problématiques cytogénétiques, contribuant ainsi à transformer nos excellentes relations de voisinage hospitalier en succès scientifiques.

Au Professeur Franck Pellestor, l'un de mes premiers maîtres en cytogénétique dès la deuxième année de médecine. Merci d'avoir accepté de juger ce travail et donc de boucler la boucle de ma formation initiale. Votre travail sur les remaniements chromosomiques complexes a largement servi d'appui à celui présenté ici. En espérant que vous y trouviez matière à collaboration future.

A la Docteure Frédérique Magdinier, pour avoir accepté de juger ce travail sans que nous ne nous connaissions. J'espère que les réflexions présentées ici nous mèneront à des échanges pertinents et des collaborations fructueuses.

A toute l'équipe du service de génétique des Hospices Civils de Lyon, et plus particulièrement au laboratoire de cytogénétique qui m'a accueilli interne, m'a appris mon métier et m'a permis de partir en Suisse pendant un an. Le soutien et les encouragements de ma voisine de bureau Audrey Labalme et des Docteurs Marianne Till et Gaetan Lesca m'ont été indispensables. Je suis heureux de pouvoir vous présenter le fruit de mon travail, qui est aussi un peu celui de l'équipe. Je me fais une joie de pouvoir revenir parmi vous sur le long terme pour continuer à rendre service à tous les patients qui nous font confiance et mener à bien plein de beaux projets.

A toute l'équipe GENDEV, pour m'avoir fait découvrir le monde de la recherche et encouragé dans mon projet. J'espère que nous garderons des liens actifs dans notre nouvelle configuration.

To the Reymond Lab in Lausanne, for an amazing year with all of you girls (and Kostya). It has been an insightful journey both science-wise and for personal life. It was a new world for me from chromosomes (of course) to mendelian randomization through zebrafish, mice, polygenic risk scores, iron metabolism, psychiatry, Italian and Sardinian (!) culture, Lavaux vineyards, Gothenburg, ... Your advices have been extremely helpful to this work. I really hope that the first stones we set together during this year will be used to build larger projects on the long term between Lyon, Lausanne and every possible destination you will take for your brilliant careers.

A l'équipe de la cellule bioinformatique de la plateforme NGS du CHU de Lyon pour m'avoir supporté ces dernières années sans penser à celles qui vont venir !

A la Fondation Thérèse et René Planiol pour la recherche sur le cerveau, le Réseau de Régional de Rééducation et de Réadaptation Pédiatrique en Rhône Alpes, l'IDEX Lyon, la Fondation Maladies Rares et les Hospices Civils de Lyon pour leur soutien financier à ce travail.

A l'ensemble de praticiens des autres services de génétique en France impliqués dans les publications présentées ici pour leur collaboration et leur réactivité.

A mes confrères et consœurs du Centre de Biologie Pathologie Est et plus particulièrement à Yohann, Florence, Alexandre, Anthony qui de très près ou de plus loin se sont intéressés ou impliqués dans ce travail et m'ont soutenu.

Aux patients et leurs familles qui acceptent de participer aux études que nous menons au bénéfice de la communauté scientifique et médicale dans l'espoir d'une meilleure prise en soins.

A ma famille, pour leurs encouragements toutes ces années et leur présence aujourd'hui.

A Judith, pour tout, tout simplement.

Résumé

Titre : Remaniements chromosomiques complexes : de la caractérisation aux conséquences fonctionnelles

La cytogénétique humaine est une discipline visant à l'étude de la structure et de la fonction des chromosomes de notre espèce. Au début des années 2010, le séquençage de génome a révélé des remaniements chromosomiques d'une complexité encore inconnue, dénommés *chromoanagenesis*. Si une importante proportion de génomes tumoraux présentent ce type d'anomalies, les descriptions de cas constitutionnels sont rares et les mécanismes sous-jacents mal compris. Nous rapportons ici le séquençage de génome de 20 nouveaux cas de *chromoanagenesis* constitutionnels (6 équilibrés et 14 déséquilibrés), constituant la plus large cohorte à ce jour. Chez plusieurs patients, des loci de moins d'un kilobase apparaissent plusieurs fois dans le remaniement dans ce que nous nommons un « hub ». L'étude de la distribution des points de cassure de ces *chromoanagenesis* et de ceux de la littérature a montré que la réplication tardive de la chromatine était le « facteur de risque » principal de cassure chromosomique. Ce résultat vient apporter une démonstration orthogonale à l'hypothèse de condensation prématurée d'un chromosome à l'origine de ces remaniements et montre pour la première fois une origine commune aux *chromoanagenesis* constitutionnels et tumoraux. Parallèlement, la distribution des points de cassure de remaniements simples apparaît biaisée vers le centre du noyau ouvrant d'importantes voies de recherche. Pour mieux comprendre les conséquences de ces remaniements complexes sur le fonctionnement du génome nous avons étudié le transcriptome des 6 cas de *chromoanagenesis* équilibrés. Nous n'avons pas mis en évidence de dérégulation massive du génome. Même localement, à proximité des points de cassure, il n'apparaît pas de dérégulation de l'expression génique posant là encore d'importantes questions sur les mécanismes de « résistance » aux remaniements.

La détection de remaniements chromosomiques n'est aujourd'hui (presque) plus limitée par la technologie (séquençage *short-read*, *long-read*, *linked-read*, cartographie optique). La compréhension des mécanismes à leur origine, la connaissance de leur pouvoir pathogène et plus généralement la maîtrise de la biologie du génome sont les nouvelles limites à la corrélation génotype/phénotype. Par l'étude des *chromoanagenesis*, d'un exceptionnel transcrit de fusion chez un patient hémophile et la description d'éléments transposables méconnus (rétrocopies) nous apportons d'importants nouveaux éléments.

Mots-clés: cytogénétique, *chromoanagenesis*, rétrocopies, séquençage, génome, transcriptome

Abstract

Title: Complex chromosomal rearrangement: from precise molecular characterization to functional consequences

Human cytogenetics is a discipline aimed at studying the structure and function of the chromosomes of our species. In the early 2010's, genome sequencing revealed chromosomal rearrangements of as yet unknown complexity, termed *chromoanagenesis*. While a significant proportion of tumour genomes present such anomalies, descriptions of constitutional cases are rare and the underlying mechanisms poorly understood. We report here the genome sequencing of 20 new cases of constitutional *chromoanagenesis* (6 balanced and 14 unbalanced), constituting the largest cohort to date. In several patients, loci of less than one kilobase appear several times in the reshuffled chromosome in what we call a "hub". The analysis of the distribution of breakpoints of these *chromoanagenesis* and those in the literature showed that late chromatin replication was the main "risk factor" for chromosomal breakage. This result provides an orthogonal demonstration of the premature condensation of a chromosome hypothesis at the origin of these rearrangements and shows for the first time a common origin of constitutional and tumoral *chromoanagenesis*. At the same time, breakpoint distribution of simple rearrangements appears to be biased towards the center of the nucleus, opening up important avenues of research. To better understand the consequences of these complex reshuffles on the functioning of the genome, we studied the transcriptome of the 6 balanced *chromoanagenesis*. We have not detected any massive deregulation of the genome. Even locally, near the breakpoints, there does not appear to be any deregulation of gene expression, again raising important questions about the mechanisms of "resistance" to structural variants.

The detection of chromosomal rearrangements is (almost) no longer limited by technology (short-read , long-read, linked-read sequencing, optical mapping). Understanding the mechanisms at their origin, knowledge of their pathogenicity and more generally the mastering of genome biology are the new limits to the genotype/phenotype correlation. By studying *chromoanagenesis*, an exceptional fusion transcript in a hemophiliac patient and describing poorly known transposable elements (retrocopies) we are bringing important new information to the field.

Keywords: cytogenetics, chromoanagenesis, retrocopies, sequencing, genome, RNA-Seq

Table des matières

Résumé	9
Abstract	10
Liste des figures	13
Liste des tableaux	15
Liste des abréviations	16
I - Introduction	17
I-1/ Introduction générale	17
I – 2/ Organisation du génome	19
I – 2/ A- Chromosomes	19
I – 2/ B- Territoires chromosomiques	21
I – 2/ C- Compartiments A et B	23
I – 2/ D- <i>Topologically Associated Domains</i>	25
I – 2/ E- Domaines d'hétérochromatine	26
I – 2/ F- Association entre organisation et fonctionnement du génome	30
I – 3/ Techniques d'étude des remaniements chromosomiques	33
I – 3/ A- Historique	33
I – 3/ B- Séquençage de génome en <i>short-read</i>	37
I – 3/ C- Technologies émergentes	46
I – 4/ Variant de structure chromosomique	55
I – 4/ A- Réparation des cassures double brin de l'ADN	55
I – 4/ B- Remaniements simples	63
I – 4/ C- Remaniements complexes	66
I – 4/ D- Insertion d'éléments mobiles	75
I – 5/ Conséquences des variants de structure chromosomique constitutionnels	78

I – 5/ A- Variants affectant la séquence codante	78
I – 5/ B- Effet de position	82
I – 5/ C- Pathogénicité des éléments transposables	86
II - Résultats	89
II – 1/ Article 1 : Severe hemophilia A caused by an unbalanced chromosomal rearrangement identified using nanopore sequencing	89
II – 2/ Article 2 : Identification of mobile retrocopies during genetic testing: Consequences for routine diagnosis	102
II – 3/ Article 3 : <i>Chromoanagenesis</i> breakpoints are enriched in late-replicating chromatin.	114
II – 4/ Conséquences fonctionnelles des <i>chromoanagenesis</i>	234
II – 4/ A- Introduction.....	234
II – 4/ B- Matériel et Méthodes.....	235
II – 4/ C- Résultats	239
II – 4/ D- Discussion	243
III - Discussion	247
III – 1/ Complexité des remaniements chromosomiques complexes	247
III – 2/ Remaniements simples : vraiment ?	250
III – 3/ Nouveaux outils : nouvelles questions !	252
IV - Conclusion	255
V - Bibliographie.....	256

LISTE DES FIGURES

FIGURE 1 : SCHEMA ET SEQUENCE PROTEIQUE DE L'EXTREMITÉ N-TERMINAL DE L'HISTONE H3 AVEC TOUTES LES POSSIBILITÉS DE MODIFICATIONS DES ACIDES AMINES	20
FIGURE 2 : SCHEMA DE LA PREPARATION D'ADN POUR LES TECHNIQUES "C" DE CAPTURE DE CONFORMATION DE LA CHROMATINE.....	24
FIGURE 3 : SCHEMA DES DIFFERENTS NIVEAUX D'ORGANISATION DU GENOME HUMAIN.	26
FIGURE 4 : REPARTITION DES NADS SUR LE GENOME HUMAIN.....	29
FIGURE 5 : SCHEMA DE L'ORGANISATION INTRANUCLEAIRE DE L'HETEROCHROMATINE.	29
FIGURE 6 : SCHEMA DE SYNTHESE DES DIFFERENTS NIVEAUX D'ORGANISATION TRIDIMENSIONNELLE DE LA CHROMATINE AU SEIN D'UN TERRITOIRE CHROMOSOMIQUE.....	30
FIGURE 7 : CORRELATION ENTRE PROGRAMME DE REPLICATION ET L'ORGANISATION TRIDIMENSIONNELLE DU GENOME.....	32
FIGURE 8 : CARYOTYPE HUMAIN FEMININ EN BANDES RHG (DENATURATION THERMIQUE).....	34
FIGURE 9 : IMAGE DE FISH SUR METAPHASES	35
FIGURE 10 : PROFIL D'ANALYSE CHROMOSOMIQUE SUR PUCE A ADN	36
FIGURE 11 : SCHEMA DE SYNTHESE DES CARACTERISTIQUES DES PRINCIPALES TECHNIQUES UTILISEES POUR LA DETECTION DE VARIANTS DE STRUCTURE CHROMOSOMIQUE.	38
FIGURE 12 : SCHEMA DE L'ETAPE D'ALIGNEMENT DE 4 PAIRES DE LECTURES.....	39
FIGURE 13 : VISUALISATION SCHEMATIQUE D'UN ALIGNEMENT DE SEQUENCES PAIREES « SHORT-READ ».	41
FIGURE 14 : COMPARAISON DES PREPARATIONS DE LIBRAIRIE POUR SEQUENÇAGE SHORT READ "PAIRED-END" ET « MATE-PAIR ».	44
FIGURE 15 : SCHEMA DE L'ALIGNEMENT DES LECTURES OBTENUES PAR SEQUENÇAGE..	46
FIGURE 16 : PRINCIPE DU SEQUENÇAGE SMRT	48
FIGURE 17 : PRINCIPE DU SEQUENÇAGE NANOPORE.....	50
FIGURE 18 : SCHEMA RECAPITULATIF DES TROIS APPROCHES DE SEQUENÇAGE PRESENTEES.....	51
FIGURE 19 : SCHEMATISATION DES RESULTATS DE SEQUENÇAGE HI-C.	53

FIGURE 20 : PRINCIPE DE LA CARTOGRAPHIE OPTIQUE	54
FIGURE 21 : MECANISMES DE RECOMBINAISON HOMOLOGUE POUR LA REPARATION DES CASSURES DOUBLE BRIN DE L'ADN.	57
FIGURE 22 : SCHEMA DES DIFFERENTS TYPES DE REARRANGEMENTS CONSECUTIFS A UNE RECOMBINAISON HOMOLOGUE NON ALLELIQUE ENTRE DEUX LCR.	58
FIGURE 23 : COMPARAISON DES MECANISMES DE NHEJ CLASSIQUE ET ALTERNATIF.....	60
FIGURE 24 : SCHEMA DU MECANISME DE FOSTES.	62
FIGURE 25 : EXEMPLES DE REMANIEMENTS DESEQUILIBRES VISIBLES AU CARYOTYPE.....	64
FIGURE 26 : EXEMPLES DE REMANIEMENTS CHROMOSOMIQUES EQUILIBRES VISIBLES AU CARYOTYPE	65
FIGURE 27 : SCHEMA DES 3 TYPES DE REMANIEMENTS CHROMOSOMIQUES COMPLEXES DITS CHROMOANAGENESIS.	68
FIGURE 28 : RISQUES DE DESEQUILIBRES ASSOCIES A UNE TRANSLOCATION RECIPROQUE EN FONCTION DU TYPE DE SEGREGATION CHROMOSOMIQUE	80
FIGURE 29 : SCHEMA DES RISQUES DE DESEQUILIBRES LIES A UNE TRANSLOCATION ROBERTSONIENNE.	81
FIGURE 30 : SYNTHESE DES DIFFERENTES POSSIBILITES D'EFFET DE POSITION EN LIEN AVEC L'ORGANISATION DE LA CHROMATINE EN TADS.	84
FIGURE 31 : VISUALISATION IGV DES SEQUENCES LONG-READ OBTENUES PAR SEQUENÇAGE DE GENOME DU PATIENT HEMOPHILE A ETUDIE DANS L'ARTICLE 2..	100
FIGURE 32 : EXPRESSION GENIQUE EN FONCTION DE LA DISTANCE AU POINT DE CASSURE (A) OU DE SON APPARTENANCE AU MEME TAD, TAD VOISIN OU ESPACE INTERTAD (B).	235
FIGURE 33 : HEATMAP DES DISTANCES INTER-ECHANTILLONS.....	239
FIGURE 34 : ANALYSE EN COMPOSANTE PRINCIPALE DES DISTANCES INTER-ECHANTILLONS..	240
FIGURE 35 : DEGRE DE DEREGULATION GENIQUE, EXPRIME EN FOLD-CHANGE EN FONCTION DE LA DISTANCE LINEAIRE A UN POINT DE CASSURE CHROMOSOMIQUE.	241
FIGURE 36 : COMPARAISON DES TADS AUX CRDS SUR UN INTERVALLE DU BRAS COURT DU CHROMOSOME 8.....	253

Liste des tableaux

TABLEAU 1 : SYNTHÈSE DES RESULTATS OBTENUS LORS DE LA CARACTERISATION DU VARIANT STRUCTURAL DU PATIENT HEMOPHILE A	100
TABLEAU 2 : RESUME DES RESULTATS OBTENUES PAR SEQUENÇAGE SHORT-READ ET LINKED-READ.....	101
TABLEAU 3 : LISTE DES PATIENTS ANALYSES PAR RNA-SEQ SUR LIGNEES LYPHOBLASTOÏDES.....	238
TABLEAU 4 : VALEUR DE FOLD-CHANGE (LOG-BASE2) POUR LES DIFFERENTS GENES RETENUS COMME RESPONSABLES DE TOUT OU PARTIE DU PHENOTYPE DES PATIENTS.....	242
TABLEAU 5 : LISTE DES GENES IDENTIFIES COMME DEREGULES PAR RNA-SEQ.....	245

Liste des abréviations

ACLF	Association des Cytogénéticiens de Langue Française
ACMG	American College of Medical Genetics and Genomics
ACPA	Analyse Chromosomique sur Puce à ADN
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
BAC	Bacterial Artificial Chromosome
CGH	Comparative Genomic Hybridization
CNV	Copy Number Variant
CRD	Cis-Regulatory Domain
CTR	Constant Timing Region
FISH	Fluorescent <i>in situ</i> Hybridization
FoSTeS	Fork Stalling and Template Switching
kb	kilobase
LAD	Lamina-Associated Domain
LCR	Low-Copy Repeat
LINE	Long Interspersed Nuclear Elements
Mb	Mégabase
MMBIR	Microhomology-Mediated Breakage Induced Repair
NAD	Nucleolus-Associated Domain
NAHR	Non Allelic Homologous Recombination
NHEJ	Non Homologous End Joining
PCC	Premature Chromosome Condensation
SINE	Short Interspersed Nuclear Elements
SNV	Single Nucleotide Variant
SV	Structural Variant
SVA	SINE-R, VNTR, and Alu
TAD	Topologically-Associated Domain
TTR	Timing Transition Region
TPRT	Target-Primed Reverse Transcription
YAC	Yeast Artificial Chromosome

I - Introduction

I-1/ Introduction générale

L'association des cytogénéticiens de langue française (ACLF) définit la cytogénétique comme l'étude de la structure et du fonctionnement normal et pathologique des chromosomes (condensation, recombinaison, réparation, ségrégation, transmission) et de la chromatine (organisation et rôle dans la régulation de l'expression des gènes).

S'agissant d'applications médicales, la cytogénétique a pour but de détecter les anomalies chromosomiques constitutionnelles ou acquises grâce à des techniques microscopiques (techniques de bandes, techniques de cytogénétique moléculaire) ou de biologie moléculaire afin d'établir un diagnostic biologique et d'assurer un conseil génétique (Guide de bonnes pratiques en cytogénétique ALCF 2014).

Les outils de séquençage dits « à haut débit » révolutionnent la discipline en décuplant le nombre de variations de structure chromosomique identifiées par individu. Pour autant le travail de corrélation génotype / phénotype reste majeur pour être en mesure d'impliquer ces « nouveaux » variants en pathologie.

Trois questions se posent alors :

- Face à une maladie monogénique, en l'absence de variant nucléotidique identifié, comment mettre en évidence un remaniement de structure pathogène ?
- Une fois identifié, par quel(s) mécanisme(s) un remaniement de structure peut-il être responsable d'un phénotype ?
- Quels sont les mécanismes à l'origine de ces « nouveaux » remaniements de structure chromosomique ?

Sur ce troisième point, nous nous sommes particulièrement intéressés aux remaniements chromosomiques complexes. Au début des années 2010, le séquençage de génome « entier » de tumeurs puis de génomes constitutionnels a permis de révéler des remaniements extrêmement complexes affectant principalement ou exclusivement un seul chromosome. Les

mécanismes menant à ces remaniements, dénommés *chromoanagenesis*, restent cependant mal compris avec deux questions principales dans ce cas :

- Les *chromoanagenesis* tumoraux et constitutionnels sont-ils issus des mêmes mécanismes ?
- Remaniements simples et remaniements complexes sont-ils les deux extrêmes d'un même continuum biologique simplement séparés par un seuil arbitraire de nombre de points de cassure ? Ou bien sont-ils le résultat de mécanismes biologiques distincts ?

La connaissance de la structure physiologique de notre génome et des mécanismes qui permettent le maintien de cette organisation est nécessaire pour répondre à ces questions. D'importants progrès sur la compréhension de l'organisation physique de notre génome dans le noyau ont été faits récemment et sont présentés ici en introduction. Les différentes technologies d'étude du génome, utilisées pour répondre à ces questions sont également présentées en introduction.

Ainsi ce travail couvre aussi bien la condensation, la recombinaison, la réparation des chromosomes que l'impact d'une réorganisation chromatinienne sur l'expression du génome dans le but de contribuer à améliorer nos connaissances sur la biologie de notre génome et nos chances de diagnostic pour les patients *in fine*.

I – 2/ Organisation du génome

I – 2/ A- Chromosomes

Le génome humain est diploïde, chaque chromosome d'une même paire étant hérité de l'un des parents. Il est composé de 46 chromosomes, 22 paires d'autosomes et une paire de gonosomes, chromosomes sexuels : 2 chromosomes X chez la femme, un chromosome X et un chromosome Y chez l'homme.

Les chromosomes sont le support physique de l'information génétique. Ils sont composés d'une part d'acide désoxyribonucléique (ADN), molécule composée elle-même d'une chaîne de riboses et associés à une combinaison de 4 bases nucléotidiques (adénine, cytosine, guanine, thymine) responsables du code génétique ; et d'autre part, d'un ensemble de protéines permettant la condensation et la décondensation de la double hélice d'ADN. Ainsi les 205 centimètres d'ADN du génome humain peuvent être contenus dans chacun des noyaux cellulaires de l'organisme (Piovesan et al., 2019) comme une pelote contenu dans une boîte plus ou moins sphérique.

La structure de base de condensation de la chromatine est le nucléosome. Il s'agit de l'enroulement de 146 paires de bases sur 1,65 tours autour d'un complexe protéique faits d'un octamère de protéines histones (paires de protéines H2A, H2B, H3 et H4). Cet enchaînement en collier de perles va pouvoir être plus ou moins super-enroulé pour donner la fibre de chromatine dite de 30 nm. Ce degré de condensation est en partie modulé par des modifications, dites épigénétiques, des protéines histones, dont principalement l'histone H3. La triméthylation des lysines 9 et 27 de l'histone H3 est observée au niveau des gènes réprimés de manière constitutionnelle ou facultative respectivement (Barski et al., 2007). A l'inverse, l'acétylation des mêmes résidus est observée au niveau des sites les plus actifs. En réalité, le « code histone » est beaucoup plus complexe. La présence de 19 lysines dans la protéine H3 et la possibilité de mono-, di- ou tri-méthylation multiplie en plus des acétylations, phosphorylations, sumoylation permet 280 milliards de combinaisons soit plus que le nombre d'histones présentes dans un noyau (Nowak and Corces, 2004; Wang et al., 2008) (Figure 1).

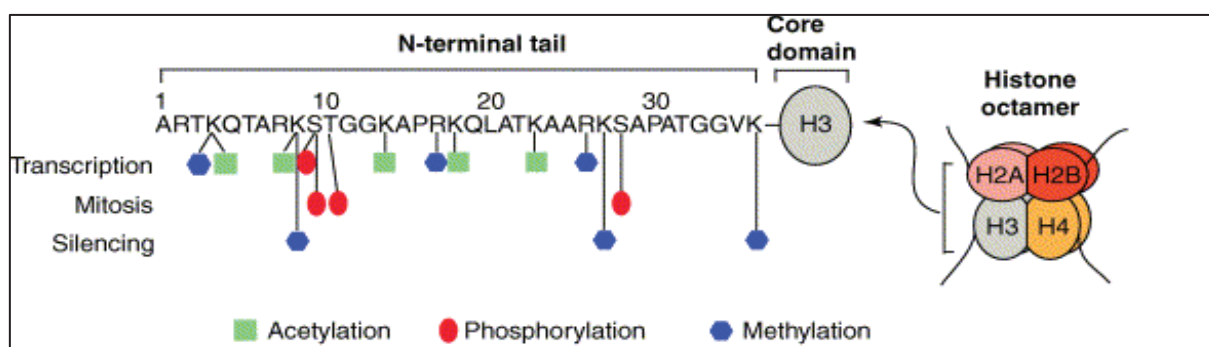


Figure 1 : Schéma et séquence protéique de l'extrémité N-terminale de l'histone H3 avec toutes les possibilités de modifications des acides aminés (adapté de Nowak et Corces, 2004)

Chaque chromosome est composé de deux chromatides unis au niveau du centromère. Le centromère est une structure nucléoprotéique servant de point de contact entre les deux chromatides d'un chromosome et de point d'attache du fuseau mitotique. Les séquences, dites alpha-satellites qui le composent sont une répétition d'un monomère de 171 paires de bases sur une longueur variable de 0,1 à 4 Mb (Schueler and Sullivan, 2006). L'enchaînement des motifs diffère selon les chromosomes et rend les séquences centromériques spécifiques d'une paire chromosomique. Un variant spécifique de l'histone H3 (CENP-A) est exclusivement localisé sur ces séquences et permet la formation d'un nucléosome variant. Ce nucléosome est reconnu par le complexe protéique CENP-A^{NAC} (CENP-A nucleosome associated complex) puis CENP-A^{CAD} (CENP-A distal complex) (Black and Bassett, 2008). Au total, 196 protéines ont été localisées au kinétochore et permettent la fixation des microtubules du fuseau de division cellulaire (Tipton et al., 2012).

Les extrémités des chromosomes sont appelées télomères. Il s'agit là encore d'un complexe nucléoprotéique composé de séquences hautement répétitives (répétition d'un hexanucléotide TTAGGG sur 2 à 15 kb) associées à un complexe protéique « shelterin ». La fonction de ce complexe est de « protéger » la séquence codante et notamment les régions subtélomériques riches en gènes pour que l'extrémité chromosomique ne soit pas reconnue comme une cassure double-brin au risque de faire l'objet d'un mécanisme de réparation. Le complexe protéique organise le repliement des 300 derniers nucléotides en *T-loop* pour masquer l'extrémité libre de la double hélice en utilisant l'extrémité 5' simple brin pour fermer le nœud.

Le reste de la longueur du chromosome peut être distingué en euchromatine, où se concentre le contenu génique codant, et l'hétérochromatine dans laquelle l'ADN est principalement inactif sur le plan transcriptionnel mais permet de maintenir la structure du chromosome.

Outre les protéines histones autour desquelles la double hélice d'ADN s'enroule et dont les modifications contribuent à moduler le degré de condensation de la chromatine, deux autres familles de protéines sont présentes. Les condensines sont les actrices de la condensation de l'ADN. En se fixant à lui, elles sont capables de générer des boucles et des superenroulements. Les topoisomérases sont, elles, nécessaires à la résolution des super enroulements en étant capable de couper l'ADN pour laisser une boucle se défaire avant de reconstruire la jonction native à l'identique.

Stricto sensu, le chromosome ne désigne que le stade le plus condensé de cette structure nucléoprotéique observable en cours de division cellulaire. Nous n'aborderons pas ici la question du chromosome mitochondrial, chromosome circulaire de 16 kb contenu dans les mitochondries.

I – 2/ B- Territoires chromosomiques

Si les chromosomes sont visibles au moment de la mitose, ils sont décondensés aux autres phases du cycle cellulaire de façon, notamment, à ce que les régions codantes puissent être transcrites. En effet, ce mécanisme nécessite que la fibre chromatinienne s'ouvre pour que la machinerie transcriptionnelle ait physiquement accès à la séquence codante. Cette décondensation se fait de manière organisée dans les 3 dimensions du noyau, chaque chromosome occupant un « territoire chromosomique » c'est-à-dire un espace préférentiel du noyau en interphase. Initialement proposé en 1885 par Carl Rabl, il a fallu près d'un siècle pour pouvoir confirmer ce modèle et infirmer celui, plus anarchique, du « plat de spaghetti » où toutes les fibres chromatiniennes sont mélangées au hasard dans le noyau. C'est par l'utilisation de sondes de peinture de FISH que les territoires chromosomiques ont pu être identifiés (Cremer et al., 1982) avant d'être confirmés bien plus tard par séquençage Hi-C (Lieberman-Aiden et al., 2009). S'agissant des cellules humaines, il a été montré que la position radiale des chromosomes dans le noyau était corrélée à leur taille : les petits chromosomes étant

préférentiellement situés au centre et les grands en périphérie (Sun et al., 2000). Il semble cependant que le contenu génique soit également un facteur prépondérant de cette distribution (Boyle et al., 2001). Les chromosomes 18 et 19 humains illustrent bien cet élément. Ils ont une longueur de 85 et 67 Mb respectivement, le chromosome 18 étant pauvre en gènes alors que le chromosome 19 est lui très riche. Le chromosome 19 est habituellement central alors que le chromosome 18 est repoussé en périphérie nucléaire (Croft et al., 1999). Aussi, à l'exception du chromosome 18, les « petits » chromosomes (de 16 à 22) présentent une fréquence de contacts interchromosomiques plus élevée que ce qui serait observé en cas de répartition anarchique (Lieberman-Aiden et al., 2009).

En revanche il n'existe pas de colocalisation particulière des territoires chromosomiques de deux chromosomes homologues (Boyle et al., 2001), les contacts observés étant pour la plupart le fruit du hasard. Une exception a été décrite pour le locus 15q11q13 pour lequel il existe une association significative. Ce locus étant soumis à empreinte parentale et responsables en cas de défaut de l'empreinte de syndromes de Prader-Willi ou d'Angelman. Il est possible qu'une étape de reconnaissance physique soit nécessaire à la bonne mise en place et/ou au contrôle de l'empreinte (LaSalle and Lalonde, 1996). En dehors de ce cas il n'existe pas de véritable association entre une fonction et l'agencement territorial observé à cette échelle.

Au niveau infrachromosomique, les bandes observées suite aux différentes techniques de marquage révèlent un deuxième niveau d'organisation du génome. Les bandes G (Giemsa) sont plutôt pauvres en gènes et contiennent préférentiellement les gènes d'expression tissu spécifique alors que les bandes R (reverse) sont plus riches en gènes et de réplication précoce.

Entre les territoires chromosomiques, il existe un espace libre de chromatine, conséquences des forces électromagnétiques qui font se repousser les fibres chromatiniennes (Cremer et al., 2000). Ces espaces larges de quelques nanomètres à quelques micromètres définissent le compartiment interchromatinien, véritable réseau ramifié qui part du pore nucléaire, sépare les territoires chromosomiques et pénètrent ces derniers. Contrairement à l'hypothèse historique de territoires chromosomiques sphériques imperméables, l'observation d'activité transcriptionnelle et répliquative au cœur des territoires permet de dessiner une organisation très festonnée augmentant ainsi la surface de contact entre territoire et espace interchromatinien (Verschure et al., 1999). S'il s'agissait simplement de permettre l'assemblage

des complexes de transcription ou de réplication de l'ADN, un espace de 30 nm serait suffisant (Singer and Green, 1997). En étant localement plus large, le compartiment interchromatinien permet la décondensation des boucles chromatiniennes lorsque cela est nécessaire à leur transcription et contribue à évacuer les molécules d'ARN produites vers la membrane nucléaire et les pores nucléaires (Visser et al., 2000). En revanche, les régions peu ou pas transcrites peuvent être tenues à distance en permanence de cet espace. Ainsi, la transcription et la régulation d'un gène ne se feraient que lorsque celui-ci est au contact de l'espace interchromatinien, les mouvements chromatiniens étant donc un possible levier de régulation dans le temps de l'expression d'un gène. C'est probablement en partie cette accessibilité à l'espace interchromatinien qui explique l'observation d'alternances de régions riches et pauvres en gènes et le profil de bandes cytogénétiques (Cremer and Cremer, 2001).

I – 2/ C- Compartiments A et B

L'identification d'une organisation plus fine du génome nucléaire a été permise par la technique de séquençage Hi-C.

Cette technique est la version pangénomique des technologies utilisant une approche de capture de conformation de la chromatine, dites « C ». L'ensemble de ces techniques « C » débute en effet par une première phase de fixation de la conformation de la chromatine (*crosslink*) avant que celle-ci ne soit digérée par une enzyme de restriction. Les molécules sont ensuite circularisées par ligation avant de reverser le *crosslink*. Ainsi ces molécules circulaires unissent deux fragments d'ADN qui étaient proches dans l'espace au sein du noyau mais qui peuvent être distantes lors de l'alignement sur le génome de référence du fait du repliement tridimensionnel du génome. Alors que le séquençage 3C permet d'analyser l'intensité d'une interaction entre deux loci et que le séquençage 4C permet d'identifier l'ensemble des interactions existantes et leur intensité depuis un locus par inverse-PCR, le séquençage Hi-C permet l'analyse de l'ensemble des contacts chromatiniens dans le génome (Figure 2).

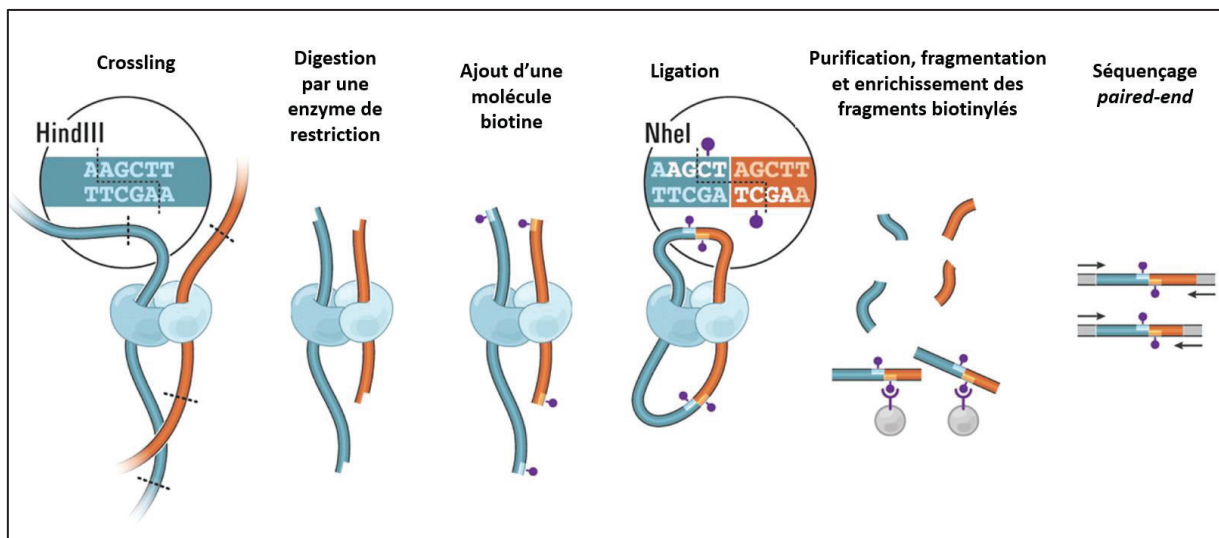


Figure 2 : Schéma de la préparation d'ADN pour les techniques "C" de capture de conformation de la chromatine. Les fragments obtenus *in fine* sont analysés (i) en PCR quantitative utilisant deux loci partenaires déjà connus pour le séquençage 3C ; (ii) par PCR inverse à partir d'un point de vue pour identifier l'ensemble des points de contact dans le séquençage 4C et (iii) tous séquencés par séquençage *paired-end* dans la technique Hi-C.

Une des premières observations de ces données a été, à la résolution d'un mégabase, la distinction au sein de chaque chromosome de deux compartiments au sein desquels les contacts chromatiniens étaient plus nombreux (Lieberman-Aiden et al., 2009) (Figure 3). En dehors des chromosomes 4 et 5, cette distinction en compartiment est le facteur le plus fort pour expliquer la fréquence observée de contacts chromatiniens au sein d'un chromosome, plus fort même que l'appartenance à un bras chromosomique.

S'agissant des contacts interchromosomiques, ils se feront également préférentiellement entre loci d'un même compartiment. Les compartiments sont des ensembles de régions génomiques de plusieurs mégabases. Les contacts au sein du compartiment B sont plus intenses qu'au sein du compartiment A. Cette 'compartimentalisation' est en partie variable selon les types cellulaires. Elle varie également au cours du développement avec 36% du génome qui change de compartiment (Dixon et al., 2015). En 2009, au moment de sa publication la technique de Hi-C ne disposait d'une résolution que de l'ordre de la mégabase du fait de coûts de séquençage encore prohibitifs. Les auteurs expliquaient alors qu'une amélioration de la résolution d'un facteur n nécessiterait d'augmenter le nombre de lectures de séquençage haut

débit par un facteur n^2 mais pourrait permettre d'identifier des structures plus fines d'organisation (Lieberman-Aiden et al., 2009).

I – 2/ D- *Topologically Associated Domains*

Avec la réduction des coûts de séquençage, une meilleure résolution Hi-C a effectivement été possible pour atteindre aujourd'hui celle du kilobase (Figure 3). Cette amélioration a permis d'identifier un nouveau niveau d'organisation tridimensionnelle de notre génome : les *topologically associated domains* (TADs) (Dixon et al., 2012). Les TADs sont des domaines de contacts préférentiels au sein de la chromatine séparés par des frontières nettes dites « éléments insulateurs ». Comme les autres niveaux de repliement il ne s'agit pas d'une organisation figée mais il s'agit là d'associations préférentielles sur des millions de cellules analysées en Hi-C. Ils s'intègrent comme un niveau supplémentaire d'organisation, une partie de leurs frontières étant communes avec celles des compartiments A/B mais aussi des *lamina-associated domains* (LADs, traité plus loin) (Dixon et al., 2012). Le mécanisme de formation des TADs aujourd'hui admis est celui de *loop extrusion* (Fudenberg et al., 2016; Ganji et al., 2018). Un ensemble de protéines cohésines vont former un anneau au sein duquel la chromatine va s'engager formant ainsi une boucle. La progression de la boucle dans l'anneau se fait jusqu'au blocage de l'anneau aux sites de liaison CTCF lorsque ceux-ci sont orientés en sens inverse vers l'intérieur de la boucle. Ainsi, ces sites CTCF constituent de véritables frontières pour les contacts chromatiniens. Les TADs, d'une taille moyenne de l'ordre du mégabase, sont relativement conservés d'un type cellulaire à l'autre (Jost et al., 2017; Szabo et al., 2019). Ils occupent environ 40% du génome humain (Rao et al., 2014) mais ce chiffre est discuté en fonction du programme utilisé pour détecter les domaines (Zufferey et al., 2018) et de la résolution obtenue par séquençage Hi-C. En effet, il semble que les TADs soient organisés en différents niveaux allant du méta-TAD au sous-TAD, les plus petits éléments étant moins fortement délimités et donc plus difficiles à mettre en évidence (Bonev and Cavalli, 2016).

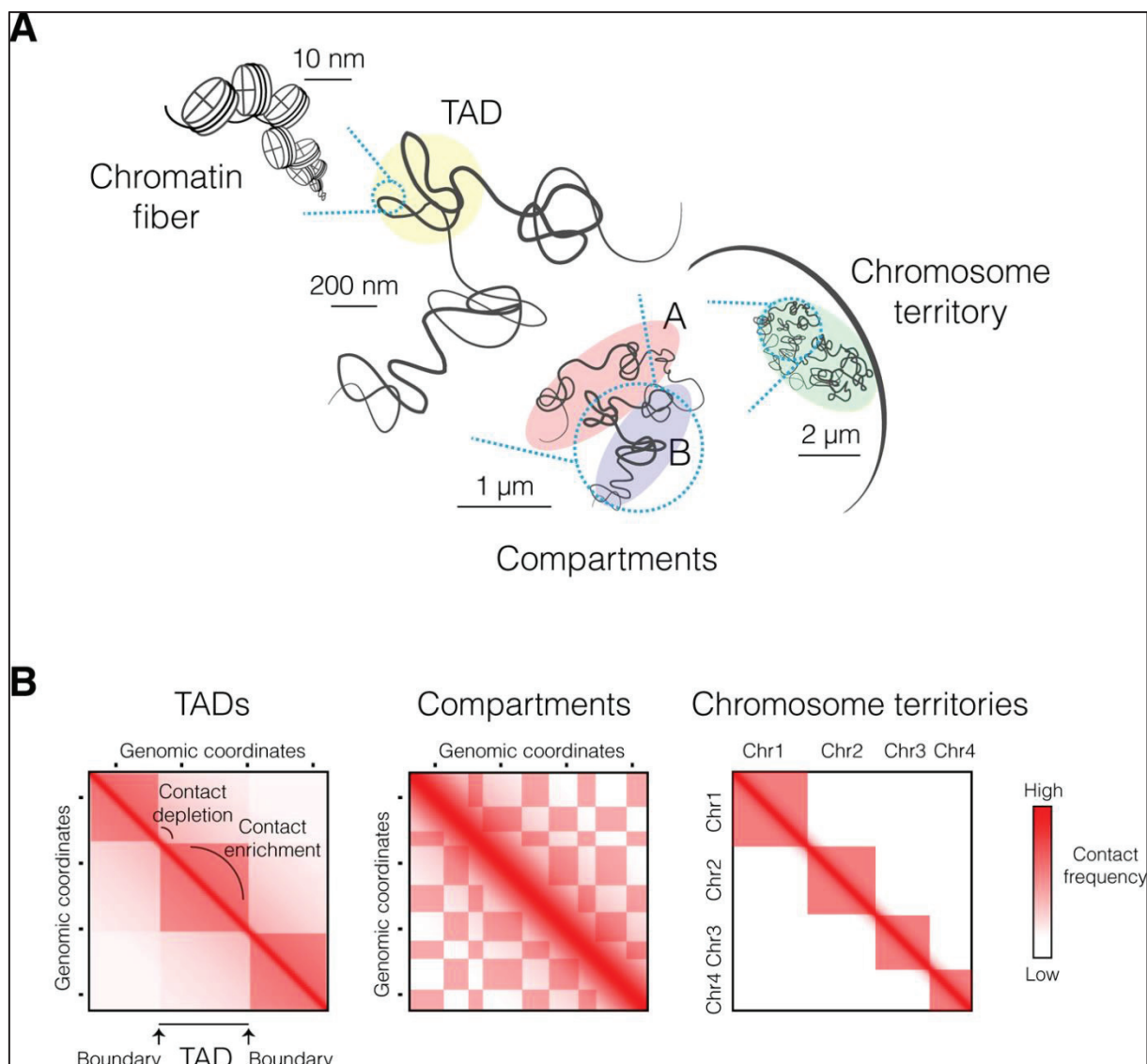


Figure 3 : Schéma des différents niveaux d'organisation du génome humain (A), visualisés sur les résultats de séquençage Hi-C. (B). (Figure extraite de Szabo et al. Science Advances 2019).

I – 2/ E- Domaines d'hétérochromatine

Ensemble les territoires chromatinien, les compartiments A/B et les TADs définissent les différents niveaux de repliement et organisation tridimensionnelle de la chromatine. Cette organisation se fait en grande partie en l'absence de matrice ou équivalent cytosquelette dans le noyau. Cependant, comme il existe des chromosomes dont le territoire est préférentiellement périphérique dans le noyau interphasique selon les résultats de microscopie, il existe des

domaines chromatiniens qui sont identifiés comme périphériques et associés au réseau de lamines nucléaires. La surface intranucléaire de la membrane nucléaire est en effet sous-tendue par un réseau de protéines filamenteuses, les lamines. La mise en évidence de ces domaines chromatiniens a été réalisée par la technique DamID venant confirmer de nombreuses observations microscopiques (Vogel et al., 2007). Le principe est d'introduire dans des cellules en culture une protéine chimérique composée d'une adénine méthyltransférase (Dam) et de la protéine d'intérêt, ici la lamine B1. L'ADN au contact de cette protéine est donc méthylé. Il faut ensuite détecter la méthylation pour identifier les régions associées aux lamines (*lamina associated domains* (LADs)). Les cartes de LADs obtenues en utilisant les différents types de lamines sont très proches entre elles (lamine C non testée) (Meuleman et al., 2013). Chez l'Homme, il existe 1300 LADs dont la taille varie de 0,1 à 10 Mb avec une médiane de 553 kb représentant environ 40% du génome au total (Guelen et al., 2008). Comme les TADs, les LADs sont délimités fréquemment par des sites de liaison CTCF. La densité en gènes est deux fois plus faible dans les LADs que dans le reste du génome. De plus, les gènes présents dans les LADs sont 5 à 10 fois moins exprimés que les autres. Sur le plan fonctionnel il s'agit donc de domaines d'hétérochromatine, particulièrement réprimés ou inactifs. L'hétérochromatine centromérique et certaines régions télomériques font notamment partie de ces LADs. Il existe une certaine variabilité entre type cellulaire des cartes de LADs. Une partie des LADs sont constants dans tous les types cellulaires et donc dits constitutifs (cLADs). Il s'agit des LADs les plus pauvres en gènes. Leur caractère constitutif pourrait être un élément majeur d'organisation du génome en interphase. Le reste des LADs, près de la moitié, sont dits facultatifs (fLADs) et varient d'un type cellulaire à l'autre (Meuleman et al., 2013). Un mouvement chromatiniens vers le centre du noyau s'accompagne généralement de l'activation du contenu génique concerné (Figure 5). La forme la plus visible d'une telle relocalisation est observée dans les bâtonnets rétiens des animaux à vision nocturne avec même dans ce cas le déplacement de cLADs vers le centre du noyau (Solovei et al., 2009). Dans ce cas particulier extrême, ceci aurait pour bénéfice une meilleure transmission de la lumière avec un effet de lentille optique et ne serait cependant pas en lien avec l'activation du contenu génique.

Les acteurs de l'interaction entre chromatine et lamines ne sont pas encore clairement identifiés. Le fait que de longs segments de chromatines soient accolés aux lamines laisse à penser que de multiples modes d'interaction entrent en jeu. Côté chromatine, l'interaction plus forte

chromatine-lamine pour les chromosomes riches en LADs suggère une certaine coopération entre les LADs (Kind et al., 2015). La méthylation de la lysine 9 sur l'histone H3 semble un élément majeur. La suppression des méthyltransférases chargées de cette méthylation réduit fortement l'association entre LAD et lamines (Chen et al., 2014b). Côté membrane nucléaire les protéines lamines jouent bien sur un rôle prépondérant, très largement souligné par l'impact des variants nucléotidiques impliqués dans les laminopathies (*LMNA*, *LMNB1* et *LMNB2*) et leurs conséquences sur l'architecture nucléaire (Robin and Magdinier, 2016).

Par ailleurs, les éléments qui promeuvent l'interaction LAD-lamine sont en équilibre avec les forces chargées de localiser les régions interLAD (et fLADs) vers l'intérieur du noyau. Diverses structures au centre du noyau sont en effet visibles en microscopie. Parmi elles, les corps de Cajal sont impliqués dans le contrôle de la longueur des télomères (télomérase), les corps nucléaires PML sont impliqués dans la réponse au stress cellulaire, la défense face aux virus, le maintien de l'intégrité génomique... (Mao et al., 2011). Pour leur part, les nucléoles sont les sites de production des ARN ribosomiques et de l'assemblage des ribosomes. Ils participent également à la régulation du cycle cellulaire en séquestrant/ modifiant de nombreuses protéines. De manière intéressante, la chromatine associée aux nucléoles présente des caractéristiques similaires aux LADs. Ceci a mené à définir des NADs, *nucleolus-associated domains* (van Koningsbruggen et al., 2010; Dillinger et al., 2017) (Figure 4). La proximité entre LADs et NADs est telle qu'il a été montré qu'il pouvait y avoir permutation de l'un vers l'autre suite à une mitose. Ainsi, une partie au moins des LADs peut se situer au contact des nucléoles.



Figure 4 : Répartition des NADs sur le génome humain (lignée fibroblastiques IMR90). Figure extraite de Dillinger et al., 2017.

Au total, l'hétérochromatine semble être répartie entre trois domaines partiellement chevauchants en fonction des types cellulaires : LADs, NADs et chromatine péri-centromérique. Ils sont en « compétition » et le défaut d'un élément nucléolaire ou du réseau de lamines va entraîner une redistribution de la chromatine vers les autres domaines. A eux trois, ils constituent le compartiment B précédemment discuté (Pombo and Dillon, 2015).

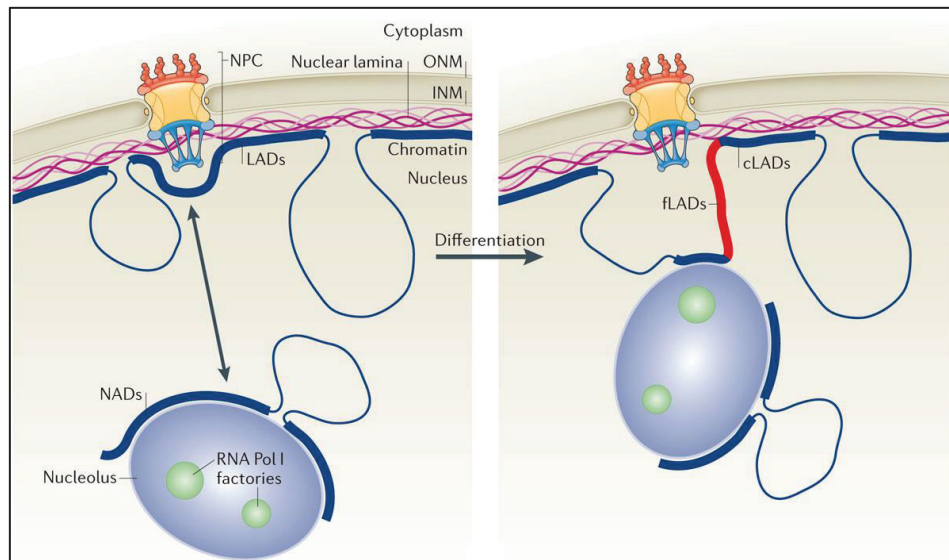


Figure 5 : Schéma de l'organisation intranucléaire de l'hétérochromatine. Lors de la différenciation cellulaire, certains LADs se décrochent du réseau de lamines situés sous la membrane nucléaire interne (INM : Inner Nuclear Membrane) et sont ainsi activés. (NPC : Nuclear Pore Complex, ONM : Outer Nuclear Membrane). (Figure extraite de Pombo and Dillon, 2015)

Au total, les connaissances sur l'organisation tridimensionnelle du génome ont largement progressé au cours de la dernière décennie. En dessous du territoire chromosomique, identifié de longues dates, les différents niveaux de repliement sont synthétisés dans la Figure 6 issue de (Robson et al., 2019).

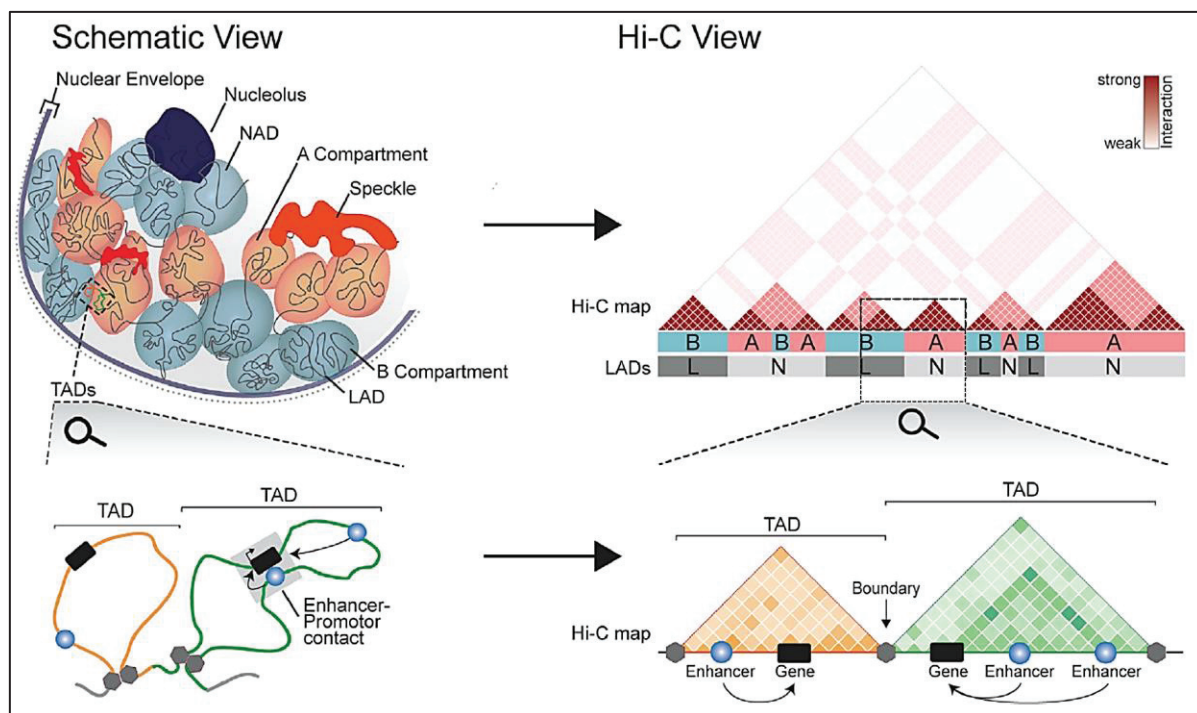


Figure 6 : Schéma de synthèse des différents niveaux d'organisation tridimensionnelle de la chromatine au sein d'un territoire chromosomique. Figure adaptée de Robson et al. 2019.

I – 2/ F- Association entre organisation et fonctionnement du génome

Interactions *enhancers* / promoteurs

La séquence promotrice d'un gène sert de fixation à l'ARN polymérase en vue de la transcription. Elle est située à proximité du site d'initiation de la transcription. En fonction du type cellulaire, du stade de développement et du cycle cellulaire la transcription du gène est en partie modulée par l'action d'*enhancers* et de *silencers* (Andrey et al., 2017). Ces séquences situées à distance viennent interagir avec la séquence promotrice par le biais du repliement chromatinien. Ces contacts sont limités en distance par la présence des éléments insulateurs

délimitants les TADs de telle sorte que les TADs constituent l'intervalle maximale d'action d'un *enhancer* (Dixon et al., 2012; Bonev et al., 2017). Une modification de délimitation d'un TAD sera responsable de nouveaux contacts chromatiniques et potentiellement du « piratage » de gènes par un ou plusieurs nouveaux éléments régulateurs. Ce sont ces modifications pathologiques qui ont confirmé l'importance de ces structures physiologiques à certains loci mais leur importance reste encore contestée à l'échelle pangénomique. En effet, l'altération isolée d'un *enhancer* par un variant structural reste habituellement sans conséquence du fait d'une importante redondance existante entre ces éléments (Will et al., 2017) protégeant en partie le gène cible d'une dérégulation. Savoir si l'organisation en TADs permet de favoriser les contacts longue distance entre un promoteur et un *enhancer* en les pré-positionnant et/ou s'ils existent plutôt pour délimiter des terrains d'action des *enhancers* et contrôler les expressions aberrantes est une question encore débattue (Lonfat and Duboule, 2015; Andrey et al., 2017).

A plus grande échelle, le positionnement de la chromatine dans le noyau est associée à son activité transcriptionnelle. Le compartiment A regroupe les régions les plus actives, les gènes les plus exprimés, où la chromatine est ouverte alors que c'est l'inverse pour le compartiment B.

Programme de réplication du génome

L'organisation en LADs est, elle, associée au programme de réplication du génome au cours du cycle cellulaire. En effet, la taille du génome humain et son organisation en de nombreux chromosomes rend impossible une réplication simultanée ou désordonnée du matériel génétique en amont d'une division cellulaire. Les principes du programme ne sont pas détaillés ici et simplement synthétisés dans la Figure 7 extraite de la publication de Marchal et al, faisant le point sur ce sujet (Marchal et al., 2019).

Brièvement, il existe des domaines de réplifications synchrones composés de plusieurs réplicons adjacents. Les domaines de réplication constantes, tardive ou précoce, (CTR) sont séparés entre eux par des zones de transition intermédiaire d'environ 800kb (TTR)(Marchal et al., 2019). La réplication tardive est notamment associée à un risque plus important de variations

nucléotidiques *de novo* sans que l'on sache si ceci est une cause ou une conséquence de ces différentes associations.

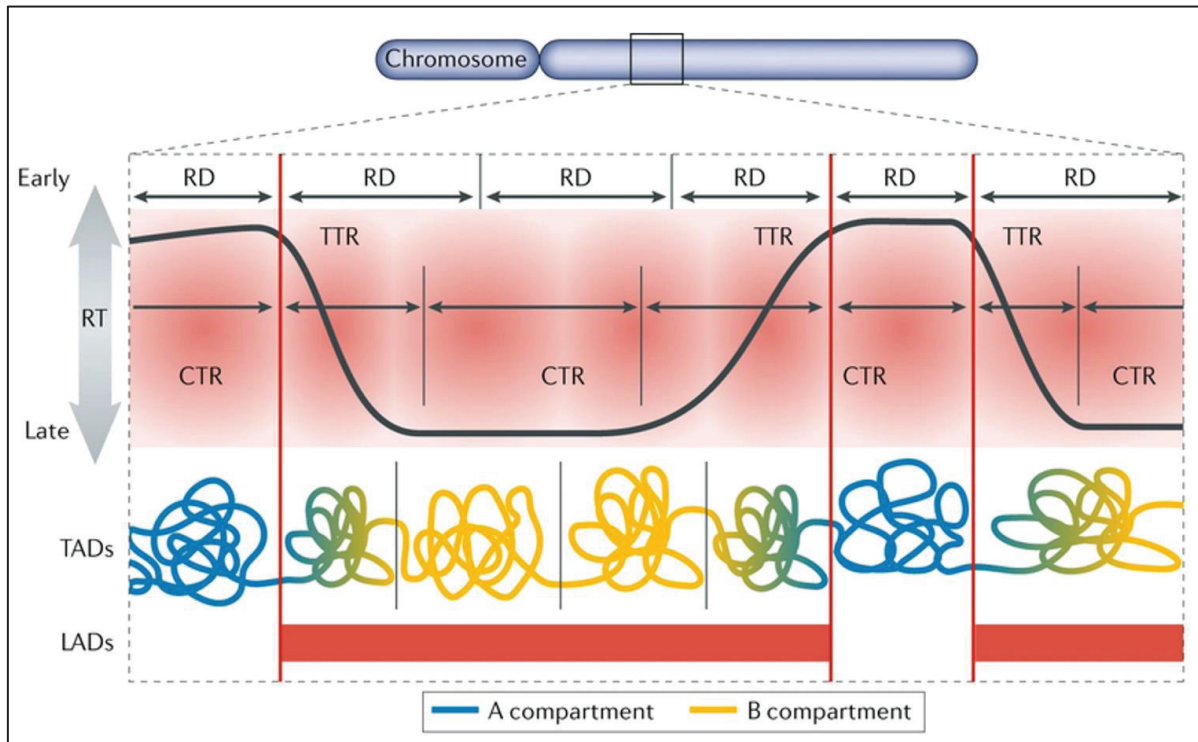


Figure 7 : Corrélation entre programme de réplication et l'organisation tridimensionnelle du génome. Figure adaptée de Marchal et al. 2019. RD : Replication Domain, CTR : Constant Timing Region TTR : Transition Timing Region

I – 3/ Techniques d'étude des remaniements chromosomiques

Avant de pouvoir expliquer les mécanismes à l'origine des variants de structure et les raisons pour lesquels ils pourraient être pathogènes, il faut d'abord être en mesure de les identifier. Nous présentons ici une mise à jour sur les principales technologies utilisées pour la détection de variants de structure chromosomique. Face à la diversité des situations et des solutions qui ont été trouvées il ne nous est cependant pas possible d'être exhaustif.

I – 3/ A- Historique

Longtemps restée hasardeuse, l'observation des chromosomes n'a été rendu fiable qu'après la découverte par sérendipité des bénéfices du choc hypotonique lors de la préparation cellulaire (Hsu, 1952) et son amélioration ensuite avec l'utilisation de la colchicine (Tijo and Levan, 1956). Par convention les 22 paires d'autosomes sont numérotées en fonction de leur taille du plus grand au plus petit. Examen cytogénétique de référence et même de première intention jusque dans les années 2000, le caryotype après techniques de *banding* permet une analyse pangénomique du nombre et de la structure des chromosomes avec une résolution comprise entre 5 et 10 mégabases (Mb).

Ce n'est qu'en 1959 que l'étude des chromosomes humains, la cytogénétique, devient une discipline médicale. Une équipe parisienne rapporte la présence d'un chromosome 21 surnuméraire chez les patients présentant un syndrome de Down (Lejeune et al., 1959). Dans les mois qui suivent d'autres anomalies du nombre de chromosomes sont décrites comme responsables du syndrome de Turner (monosomie X), du syndrome de Klinefelter (XXY), du syndrome d'Edwards (trisomie 18), du syndrome de Patau (trisomie 13). En 1960, la première anomalie cytogénétique acquise est identifiée. Une équipe de Philadelphie décrit la présence d'un chromosome minute chez les patients présentant une leucémie myéloïde chronique (Nowell et al., 1960). Ce n'est que dans un deuxième temps avec l'arrivée des techniques de bandes que ce remaniement sera caractérisé comme étant une translocation réciproque apparemment équilibrée entre les bras longs d'un chromosome 9 et les bras longs d'un

chromosome 22 (t(9;22)(q34;q11)) à l'origine d'un transcrit de fusion *bcr-abl* à activité oncogénique.

Dans les années 1960, le rendement diagnostique du caryotype est limité par sa résolution et de rares grands réarrangements déséquilibrés sont identifiés comme responsables de phénotypes reconnaissables (délétion du bras court du chromosome 5 dans le syndrome de Cri-du-chat ; délétion du bras court du chromosome 4 dans le syndrome de Wolf-Hirschhorn,...).

Au tournant des années 1970, l'utilisation d'agents alkylants au cours de la préparation chromosomique permet l'identification de bandes chromosomiques (Caspersson et al., 1970), alternance de zones marquées et non marquées réalisant un motif spécifique d'une paire chromosomique. Les techniques de bandes permettent à la fois de mieux identifier les chromosomes d'une part et les anomalies subchromosomiques d'autre part. Elles se simplifient rapidement avec l'utilisation d'une dénaturation enzymatique (bandes G) ou thermique (bandes R) contribuant à la démocratisation de l'étude du caryotype (Figure 8). Ce sont ces techniques qui ont aujourd'hui cours dans les laboratoires de cytogénétique.

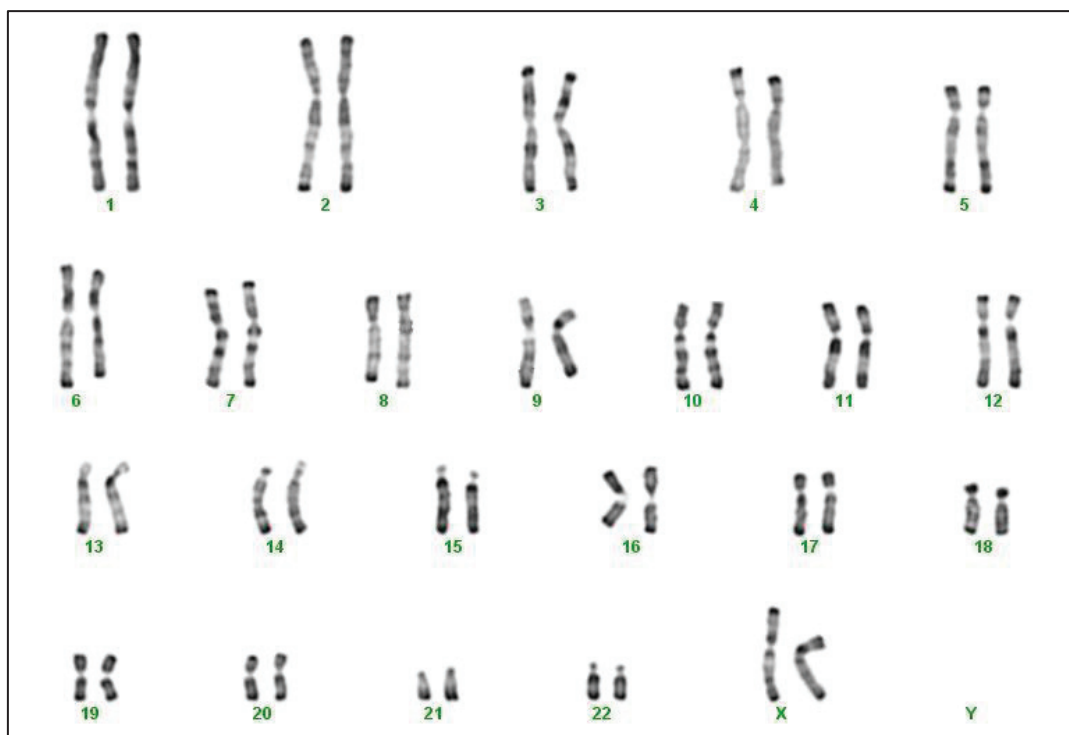


Figure 8: Caryotype humain féminin en bandes RHG (dénaturation thermique)

Avec le développement de sondes radioactives à la fin des années 1960, la cytogénétique devient « moléculaire ». Depuis les années 1980 les chromosomes bactériens artificiels (Bacterial Artificial Chromosomes : BACs) correspondants à un locus du génome humain et marqués par un fluorochrome sont utilisés en routine dans la technique d'hybridation *in situ* en fluorescence (FISH). 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) et a permis le diagnostic de très nombreuses anomalies chromosomiques déséquilibrées non visibles au caryotype, et généralement récurrentes (del/dup 7q11.2 ; del/dup 22q11.2 ; del/dup 15q11q13) (Figure 9) mais aussi la caractérisation d'anomalies chromosomiques équilibrées visibles au caryotype (sondes de peinture). Les techniques de M-FISH, SKY, Fiber-FISH ont également contribuées à étudier les remaniements chromosomiques mais ne sont pas présentées ici.

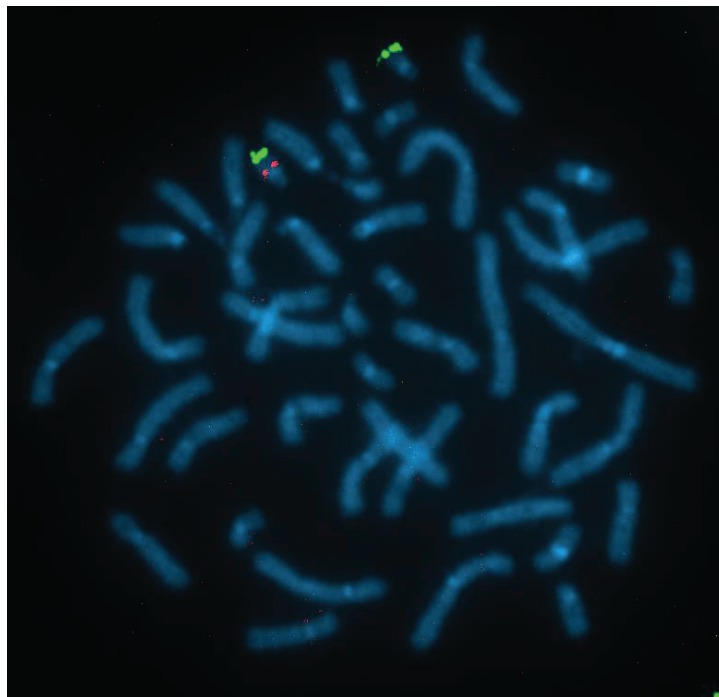


Figure 9 : Image de FISH sur métaphases avec la sonde TBX1 (rouge) ciblant le locus 22q11.2 et la sonde 22qter contrôle (vert). Présence d'une délétion hétérozygote 22q11.2.

En 1992, la technique d'hybridation génomique comparative (CGH) est publiée (Kallioniemi et al., 1992). Le principe est de cohybrider sur un étalement chromosomique normal la même quantité de deux ADN (celui testé et un ADN contrôle) marqué chacun par un fluorochrome différent. Un déséquilibre entre les deux fluorochromes signe une perte ou un gain de matériel

chromosomique de l'ADN testé. Si cette technique quantitative présente l'intérêt de ne plus être ciblé mais bien pangénomique elle est limitée par la condensation des chromosomes et sa résolution se rapproche de celle du caryotype. Il faut attendre 3 ans pour que la préparation chromosomique qui sert de support à l'hybridation soit remplacée par une puce à ADN : une lame de verre recouverte de milliers de fragments d'ADN simple brin (array) sur lesquels les 2 ADN en compétition vont s'hybrider (CGH-array) (Schena et al., 1995). Ces fragments d'ADN sont initialement des BACs ou des YACs (Yeast Artificial Chromosomes) déjà utilisés en FISH et dans le cadre du projet de séquençage du génome humain. Pour faciliter l'examen des gonosomes, l'ADN du patient est co-hybridé avec un ADN d'un témoin de même sexe. L'intensité de chaque fluorescence est lue par un scanner avant qu'un ratio d'intensité de fluorescence ne soit calculé pour chaque locus étudié. Les résultats sont représentés sous forme graphique avec des rapports d'intensités exprimés en log ratio de base 2. Plus tard, l'amélioration de la technologie a permis de remplacer ces séquences de 100-200 kb par des oligonucléotides d'une taille de 20 à 70 bases. La résolution de l'analyse dépend maintenant du nombre de fragments d'ADN présents sur la puce et peut aller jusqu'à 1 kb. La CGH-array est donc une analyse quantitative qui n'informe pas sur la structure des chromosomes et la mécanique chromosomique mais bien sur le nombre de copies de chaque locus étudié dans un échantillon par rapport à un témoin en mettant en évidence des variations du nombre de copie (*Copy Number Variant CNV*) (Figure 10).

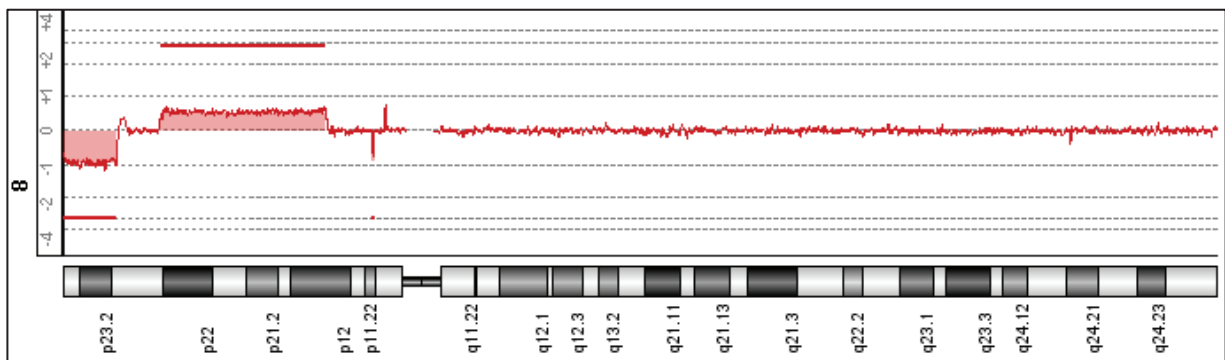


Figure 10 : Profil d'analyse chromosomique sur puce à ADN du chromosome 8 montrant une perte terminale du bras court du chromosome 8 suivie d'une région au nombre de copie normale puis d'un gain d'une copie. L'étude cytogénétique complémentaire a montré qu'il s'agissait d'une inversion délétion duplication affectant le bras court d'un chromosome 8.

Ces limitations justifient la réalisation de techniques de FISH et/ou caryotype pour confirmer une anomalie et connaître la mécanique chromosomique. De plus la technique ne permet pas la

détection des remaniements chromosomiques équilibrées ni la caractérisation au nucléotide près des points de cassure chromosomique. Ce n'est qu'au milieu des années 2000 que les premières puces ont été commercialisées. En France, les techniques de CGH-array et de SNP-array (utilisant les données de génotypage pour la détection de CNVs) ont été regroupées sous l'appellation « Analyse Chromosomique sur Puce à ADN : ACPA ».

I – 3/ B- Séquençage de génome en *short-read*

Principe

En 2001, la séquence du génome humain est publiée dans ses premières versions après des efforts internationaux extrêmement coûteux (Lander et al., 2001; Venter et al., 2001). Elle permet une explosion des connaissances en génétique et l'amélioration des techniques existantes mais l'étude de génome d'individu à des fins médicales reste excessivement chère et techniquement irréaliste. A la fin des années 2000, l'arrivée des technologies de séquençage massif en parallèle (initialement dénommé « de nouvelle génération ») a entraîné une réduction brutale des coûts de séquençage rendant accessible le séquençage de génome humain individuel (Levy et al., 2007). Aujourd'hui, la technologie dominante « *Sequencing By Synthesis* SBS ® » commercialisée par la société Illumina, Inc (San Diego, CA, USA) produit des séquences de très bonne qualité d'une longueur habituelle de 100-150 paires de bases dit « *short-read* ».

Le principe de la technologie « *Sequencing by Synthesis* » est de fragmenter l'ADN en morceaux de 300-600 paires de bases avant de venir ajouter des adaptateurs universels à chaque extrémité pour permettre aux fragments de se fixer sur le support de séquençage. Chaque brin attaché est amplifié pour former une colonie de brins d'ADN identiques. Ensuite la réaction de séquençage proprement dite peut démarrer. A partir d'une amorce fixée sur l'adaptateur universel, une ADN polymérase vient ajouter des nucléotides marqués par un fluorochrome un par un et une caméra détecte la fluorescence de chaque colonie à chaque temps. Une fois l'image capturée, le nucléotide suivant peut être ajouté. Le « haut débit » est permis par le séquençage de millions de colonies en parallèle sur le support de séquençage. L'approche la plus couramment utilisé est celle du séquençage en *paired-end*. Dans ce processus, après avoir

séquencé les 100-150 bases à une extrémité du fragment, la molécule d'ADN est retournée et c'est l'autre extrémité qui est séquencée selon le même principe. Les deux lectures sont donc appariées et doivent normalement pouvoir être alignées ensemble sur le génome de référence à une distance correspondant à la taille du fragment d'origine (300-600 bases).

Après avoir généré les séquences, le « puzzle » de notre génome doit être reconstitué à partir de très nombreuses petites pièces (« short-read ») d'ADN nécessitant d'importantes étapes bioinformatiques dont la plus importante est celle de l'alignement sur le génome de « référence ». C'est sur ces données alignées que se fera la détection de variants.

Application cytogénétique

Les caractéristiques des différentes techniques utilisées en routine pour le diagnostic cytogénétique sont résumées en Figure 11. Le séquençage permet une résolution à la base près pour l'étude des remaniements motivant ainsi le transfert actuellement en cours de cette technologie en diagnostique.

	Technique	Pangénomique ?	Remaniement équilibré?
1	Séquençage de génome	+	+
10			
100	MLPA/qPCR	-	-
1000			
10000	ACPA	+	-
100000			
1000000	FISH	-	+
10000000	Caryotype	+	+
Résolution ↓ (bases)			

Figure 11 : Schéma de synthèse des caractéristiques des principales techniques utilisées pour la détection de variants de structure chromosomique.

Le séquençage de génome est à la jonction entre biologie moléculaire et cytogénétique puisque la comparaison à ce génome de référence permet la détection des variants nucléotidiques

ponctuels (*Single Nucleotide Variants* SNVs et indels) d'une part et variants de structure chromosomique d'autre part (équilibrés ou non) (*Structural Variants*) et ceci avec la résolution ultime d'un nucléotide.

En effet, s'il existe un ensemble de lectures alignées au même locus qui ont leurs paires toutes alignées à un autre locus incompatible avec la taille des fragments d'origine (voire sur un autre chromosome) et/ou orientées improprement, c'est qu'il existe un variant de structure mettant en contact les deux loci chez le patient et dont les points de cassure chromosomique sont très proches de ces alignements aberrants (approche *read pair*). Différentes possibilités sont présentées dans la Figure 12.

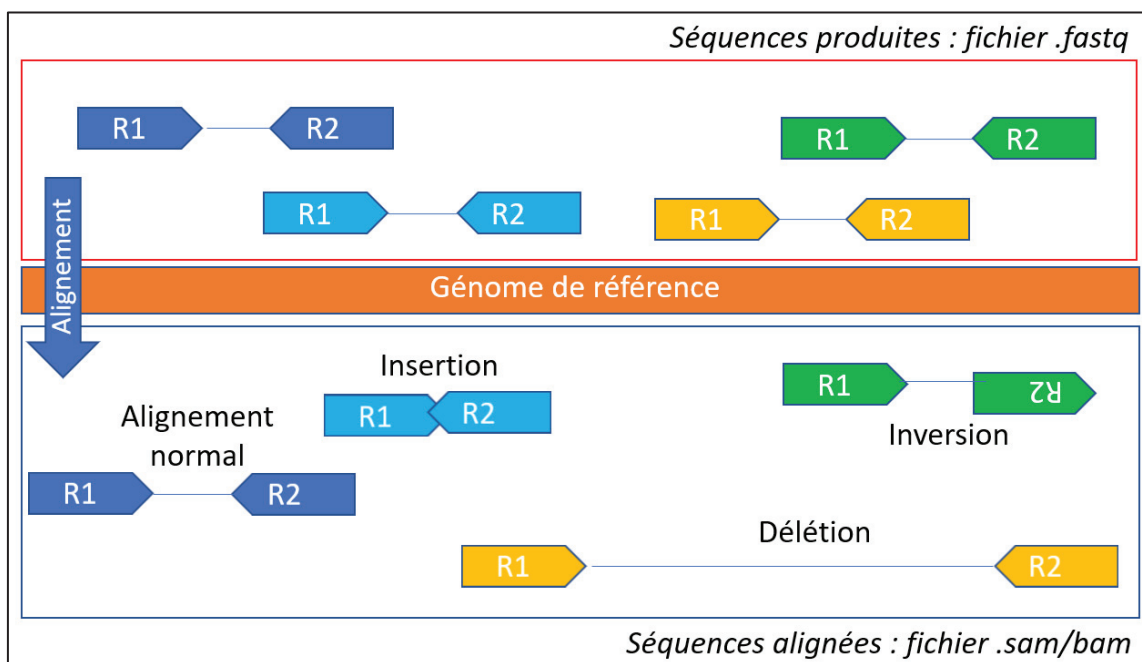


Figure 12 : Schéma de l'étape d'alignement de 4 paires de lectures. En cas d'insertion, une séquence est présente entre les deux éléments de la paire alors qu'elle est absente dans l'échantillon analysé. Les lectures s'alignent donc très proches l'une de l'autre. C'est l'inverse en cas de délétion : dans l'échantillon les deux éléments de la paire sont à une distance compatible avec la taille des fragments préparés mais s'alignent loin l'un de l'autre car dans le génome de référence ils sont séparés par la séquence perdue.

C'est par cette approche que sont identifiées les jonctions des variants de structure. Si la quantité de données générées est suffisante, on peut également observer une ou plusieurs lectures qui chevauchent directement le(s) point(s) de cassure chromosomique (approche *split-read*). Dans ce cas, soit la lecture est fragmentée et alignée aux deux loci partenaires ; soit une seule partie

de la lecture est alignée et le reste de la séquence est « clippée » - c'est-à-dire non prise en compte dans le reste de l'analyse. Le point de cassure est situé à la jonction entre la séquence alignée et la séquence « clippée ».

S'agissant particulièrement des variations du nombre de copies, la quantité de séquences produites pour un locus donné reflète la quantité d'ADN de départ. On appelle profondeur de séquençage le nombre de fois qu'une même position génomique est couverte par une lecture. Si la profondeur vient à être divisée par deux sur un intervalle génomique par rapport aux régions avoisinantes, ceci signe habituellement la perte d'une des deux copies chromosomiques, une délétion hétérozygote (approche *read-depth*) (Abyzov et al., 2011). Les bornes de ce changement de couverture correspondent aux points de cassure détectés avec les approches précédemment citées. La balance allélique des variants nucléotidiques, c'est à dire la proportion de lectures portant un allèle alternatif peut également participer à la détection de CNVs à la manière des puces SNP-array. Dans le cas d'une délétion, une perte d'hétérozygotie sera détectée. A l'inverse, sur un intervalle génomique donné, si pour chaque variant nucléotidique détecté la proportion de lecture portant cet allèle alternatif est de $1/3$ ou $2/3$ au lieu de $1/2$ c'est qu'il existe 3 allèles au lieu de 2, gain d'une copie. Multiplier les approches permet d'améliorer les performances des algorithmes de détection de variants. L'ensemble est synthétisé dans la Figure 13.

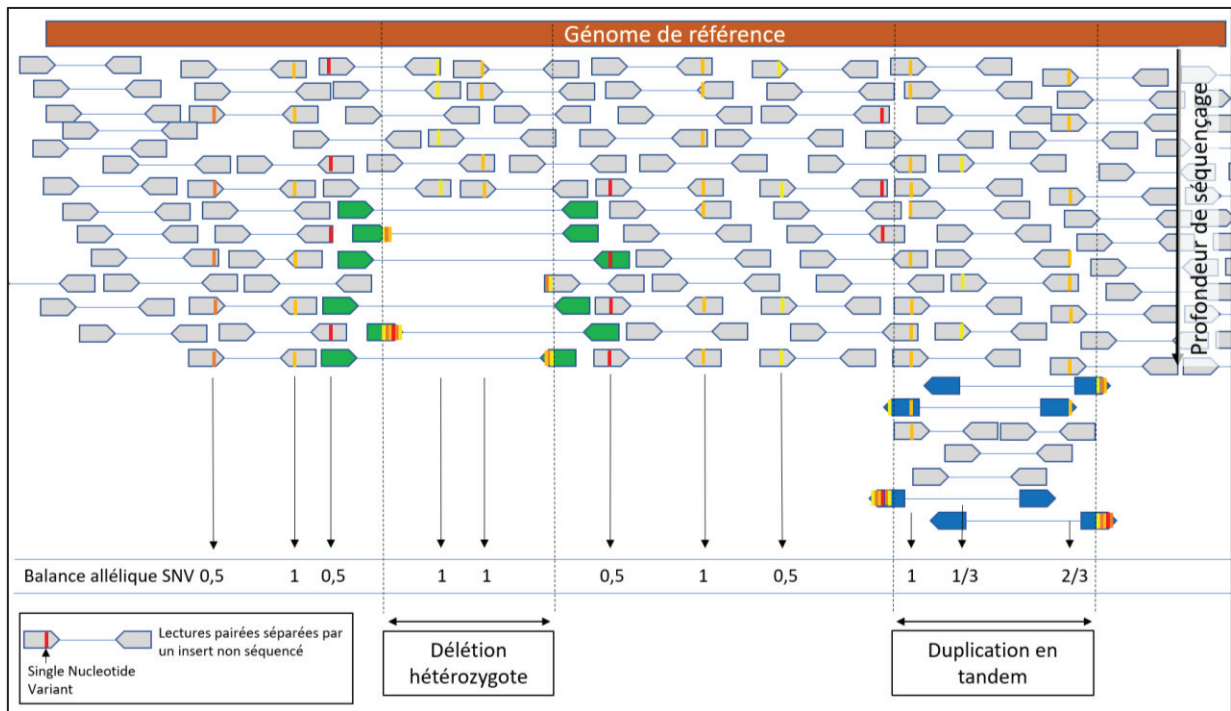


Figure 13: Visualisation schématique d'un alignement de séquences pairées « short-read ». Les paires séparées par une distance anormale ou orientées de façon aberrante sont colorées. Les lectures qui chevauchent directement un point de cassure présentent une extrémité entièrement différente de la séquence de référence, le partenaire du remaniement étant directement séquençé. Pour ce qui est des anomalies équilibrées, la balance allélique des variations nucléotidiques ponctuelles éventuellement concernées peut être utilisée pour détecter ces CNVs.

La profondeur de séquençage communément admise comme nécessaire à l'appel de variants nucléotidiques est de 30x (Bentley et al., 2008). Cette profondeur permettrait de détecter l'ensemble des variations nucléotidiques hétérozygotes. Avec une moyenne de 15 lectures par chromosome le taux d'erreur de séquençage devient quasi nul et la détection de variants donc extrêmement fiable. De leur côté, les variants de structure n'affectant pas une position unique, leur détection ne nécessite pas une profondeur si importante, ce qui permet de réduire les coûts d'analyse. En effet, même s'il subsiste quelques erreurs dans la lecture l'important est de pouvoir correctement l'aligner sur le génome de référence. Plus la lecture est longue et plus l'alignement est fiable. Pour le génome humain, en dehors des régions répétées une lecture de 20 nucléotides est habituellement alignée de façon univoque. L'approche en *paired-end* permet, grâce aux lectures appariées d'aligner des lectures même dans certaines régions répétées en s'aidant de l'autre lecture de la paire qui est, elle, alignée de façon univoque. Ce sont les régions répétées de taille supérieures à celles des fragments d'ADN qui posent problème.

Il n'existe aujourd'hui pas de *pipeline* d'analyse bioinformatique *gold standard* pour la détection de variants de structure. A partir des données *short-reads*, les différentes approches discutées précédemment ont été d'abord utilisées isolément puis combinées ensemble (Jeffares et al., 2017). Ces approches combinées n'étant toujours pas en mesure de capturer l'ensemble des variants pourtant observables dans les données de séquençage. Comme pour les variants nucléotidiques, la plus grande base de données de variants de structure a été établie par le consortium gnomAD à partir de 12,549 individus indépendants et réputés sains dont le génome entier a été séquençé en *short-reads* (Collins et al., 2019). Le pipeline bioinformatique utilisé par le consortium gnomAD a une sensibilité de 97,8% pour les CNVs de plus de 40 kb. Grâce à 4 échantillons séquençés également en *long-reads*, le taux de vrais positifs a pu être évalué à 88% et le taux de violation des lois de transmission mendélienne, utilisé comme approximation de la spécificité était de 4,2%.

Avantages

L'arrivée de la technologie de séquençage massif en parallèle et la chute des coûts de séquençage qui s'en est suivie permet aujourd'hui le séquençage de génome humain (*GS : genome sequencing*) et a donc révolutionné la stratégie de caractérisation d'un remaniement chromosomique (Mills et al., 2011; English et al., 2015) d'une part, et amélioré le rendement diagnostique d'autre part (Gilissen et al., 2014; Soden et al., 2014). Le séquençage massif en parallèle tend donc à supplanter au moins en partie les techniques de séquençage Sanger, exome (SNVs); ACPA (CNVs); Long-range PCR, Southern-blot, caryotype (remaniements équilibrés) pour l'étude du génome humain.

De longue date, la caractérisation des points de cassure de remaniement a en effet été un défi cytogénétique faisant appel à diverses techniques de façon séquentielle. A partir d'un point d'un remaniement vu au caryotype, il fallait avancer de proche en proche jusqu'à identifier un point de cassure avant de pouvoir affiner la résolution pour atteindre la position exacte du point de cassure. La procédure pouvait donc être extrêmement fastidieuse (et coûteuse), affectée au moins en partie par un facteur chance, et quasi incapable de révéler des points de cassure non vus au caryotype.

Cette approche pangénomique permet en effet de ne pas travailler sur les points de cassure les uns après les autres mais de les étudier tous dans un même temps (Chen et al., 2010; Slade et al., 2010; Schluth-Bolard et al., 2013). Elle permet également de s'affranchir d'une mise au point de la technique pour chaque patient, diminuant ainsi le coût et le temps nécessaire. En cas de remaniement complexe, cette technologie est donc d'un intérêt tout particulier. De plus le remaniement étant plus complexe qu'attendu dans 1 cas sur 5 (Chiang et al., 2012), l'analyse du génome entier permet de révéler de potentiels points de cassure non suspectés initialement.

Pour améliorer la sensibilité de la technique, et résoudre en partie la question des éléments répétés, différentes stratégies ont été appliquées, visant principalement à augmenter la taille des fragments d'ADN étudiés et donc la distance entre les deux lectures d'une même paire. Ceci permet d'une part de résoudre un plus grand nombre de régions répétées mais aussi d'augmenter le nombre de paires de lectures qui sont concernées par un alignement anormal et donc en théorie la probabilité qu'un variant structural soit détecté. Ces approches sont décrites dans la littérature sous les appellations *mate-pair sequencing* et *long-insert sequencing* (Korbel et al., 2007; Talkowski et al., 2011) (Figure 14). En pratique des étapes supplémentaires de préparation de l'ADN sont ajoutées pour partir de grands fragments et rapprocher deux séquences lointaines sur un fragment de plus petite taille compatible avec le séquenceur. En effet, la longueur du fragment d'ADN qui doit s'hybrider sur le support de séquençage ne peut être supérieure à 1 kb sous peine de ne plus générer de colonies monoclonales.

Les coûts de séquençage ayant été massivement réduits lors de la dernière décennie l'avantage économique initial de ces librairies modifiées s'est atténué. Pour ce qui est de leur avantage théorique vis-à-vis des éléments répétés, un insert de 10kb permettrait de couvrir de façon univoque 99,43% de la séquence de référence (97,33% pour 1 kb d'insert) contre 96,3% pour les librairies standards (Talkowski et al., 2011). En pratique, le séquençage de génome a permis de caractériser à la base près 90,8% des remaniements chromosomiques vus au caryotype en utilisant ces librairies grands inserts (Redin et al., 2017), contre 89% pour des librairies standards (sur des cohortes de patients différents) (Schluth-Bolard et al., 2019) ($p=0,87$). Considérant l'objectif final de maximiser le rendement diagnostique pour nos patients dans le temps le plus court possible il semble logique de privilégier un test capable d'identifier l'ensemble des variations de l'ADN en un seul temps (SNVs, CNVs, SVs).

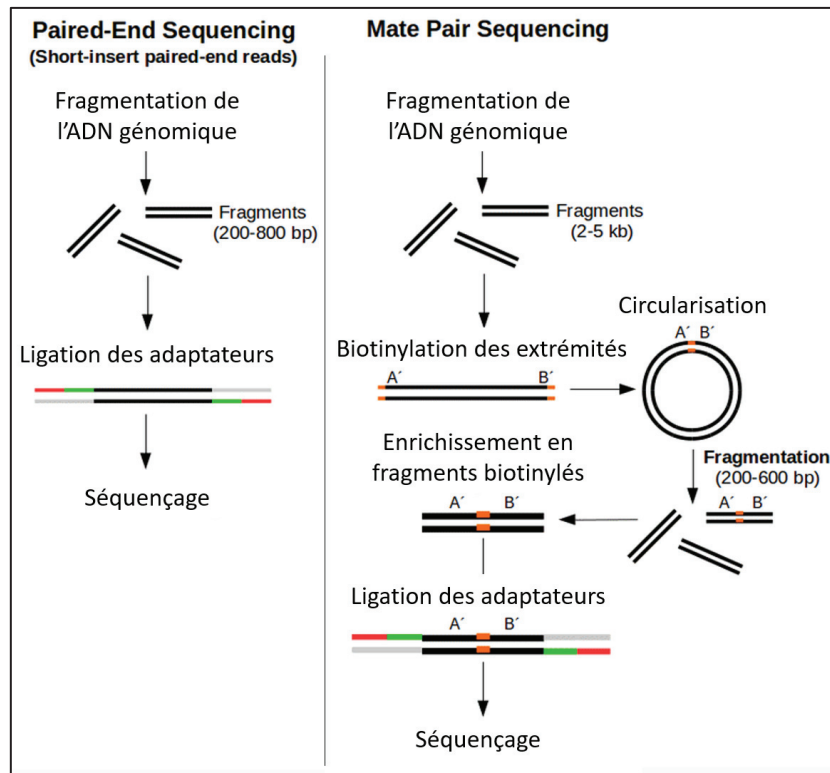


Figure 14 : Comparaison des préparations de librairie pour séquençage short read "paired-end" et « mate-pair ». Les deux peuvent être séquençés sur les mêmes appareils mais la seconde approche a pour intérêt de rapprocher des fragments distants de plusieurs kb.

De plus, ne pas utiliser de préparations de librairies spécifiques permet de bénéficier de l'ensemble des données de séquençage déjà produites avec une approche standard pour faciliter l'interprétation des variants identifiés (filtres de fréquence) (The 1000 Genomes Project Consortium, 2015; Collins et al., 2019).

Après avoir participé à faire la preuve de principe de cette méthode (Schluth-Bolard et al., 2013) et avoir mené à bien un projet national de recherche translationnelle montrant (pilote par le Dr Schluth-Bolard) un rendement diagnostique de plus de 40% (Schluth-Bolard et al., 2019) le test est maintenant disponible en routine diagnostique au laboratoire de cytogénétique des Hospices Civils de Lyon pour la caractérisation de remaniements chromosomiques apparemment équilibrés vus au caryotype chez des patients présentant un phénotype anormal. Outre les diagnostics nouvellement posés, cette approche a également permis l'identification de nouveaux gènes candidats dont certains ont pu être démontrés comme responsables de nouveaux syndromes (*BCL11B* (Lessel et al., 2018), *TLK2* (Reijnders et al., 2018), *ZMIZ1* (Carapito et al., 2019)) et d'autres en cours d'étude.

Limites

L'augmentation du nombre de génomes humains analysés dans diverses populations contribue à améliorer le génome dit « de référence », pour le rendre plus complet et plus représentatif mais elle reste un facteur limitant aujourd'hui. En effet, c'est cette séquence de référence qui sert de support à l'alignement des séquences obtenues lors du séquençage d'un individu. La version en vigueur au moment de la rédaction du manuscrit (GRCh38p.13) se compose de 3,272,116,950 bases (génome haploïde) dont 3,110,748,599 ne sont pas des N (nucléotides non identifiés) et environ un millier de « trous » persistent dans cette carte. De fait chaque trou dans la carte va empêcher l'alignement de séquences produites et donc l'identification d'éventuels variants.

De plus, 52,5% du génome humain sont constitués de séquences répétées (Lander et al., 2001). Ces séquences sont en partie le résultat de l'accumulation au cours de l'évolution d'évènements de rétrotranspositions et plus localement de grands blocs dupliqués pour donner naissance à des duplications segmentaires. Ces séquences répétées peuvent être à l'origine de variants de structure récurrents par des mécanismes discutés plus loin dans ce travail.

Ces séquences répétées sont problématiques pour le séquençage *short-read* puisque les séquences produites ne peuvent pas être positionnées de façon univoque sur la carte. Sur les puces à ADN, il n'y a habituellement pas d'oligonucléotides dans ces régions. Pour la même raison, lors du séquençage de génome les fragments d'ADN séquencés qui ne pourront pas être alignés sur le génome de référence de façon univoque seront perdus. Ces régions ne sont donc habituellement pas étudiées et les variants impossibles à détecter (Figure 15). L'impossibilité de les analyser a incité l'ACMG (American College of Medical Geneticists) à supprimer l'adjectif « complet » pour désigner le séquençage du génome puisque même si le séquençage est complet, l'analyse ne l'est pas. Par ailleurs, pour réduire les biais de couverture, ont été développés des protocoles de préparation de librairie sans utilisation de PCR. En effet, la PCR est connue pour amplifier plus facilement certaines séquences que d'autres (très riches ou très pauvres en GC) introduisant ainsi un biais dans le volume de données généré par locus. Il n'en reste pas moins que même sans PCR lors de la préparation de librairie, l'étape d'amplification pour générer une colonie de séquences identiques sur le support de séquençage introduit le

même type de biais avec des conséquences sur la détection de variants de structure notamment (Chen et al., 2013).

En analysant les données de 10 séquençages de génome Illumina d'une profondeur moyenne de 37x, une équipe a dénombré plus de 36000 régions géniques (dont 748 gènes avec des régions exoniques) non ou insuffisamment séquencées. 76 gènes déjà impliqués en pathologie contiennent au moins une partie de leur séquence codante non accessible par séquençage Illumina (Ebbert et al., 2019).

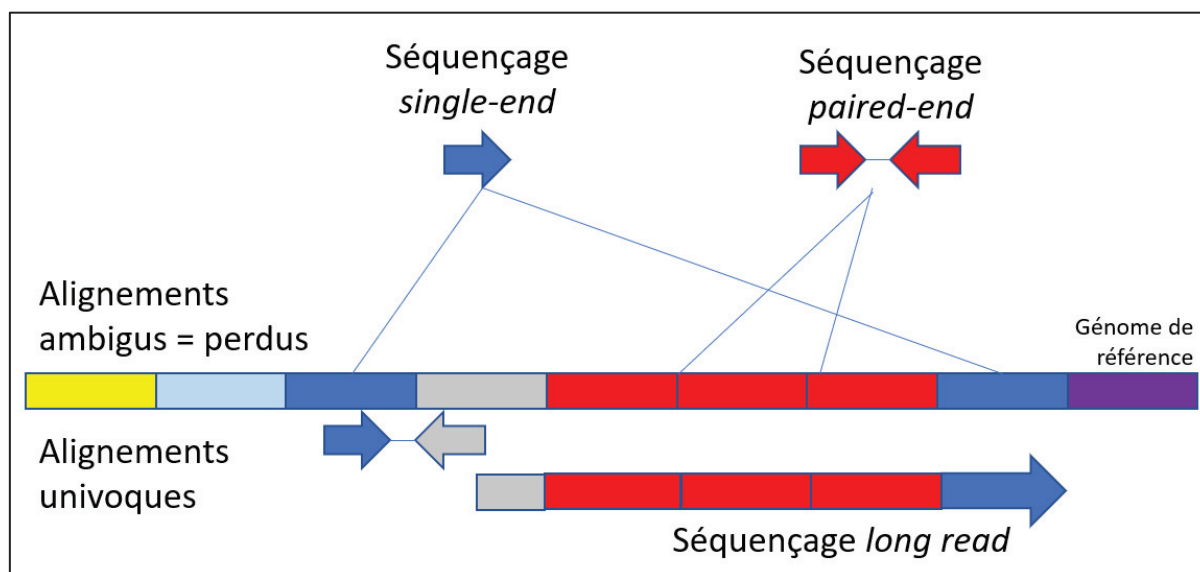


Figure 15 : Schéma de l'alignement des lectures obtenues par séquençage. Le génome de référence est schématisé par l'enchaînement de blocs de couleurs différentes. Les blocs de même couleur représentent des séquences répétées. Si le séquençage en paired-end permet de résoudre les éléments répétés les plus petits, le séquençage long-read va lui pouvoir solutionner de plus grandes répétitions.

I – 3/ C- Technologies émergentes

Pour dépasser les limites du séquençage short-read différentes technologies sont actuellement développées.

Séquençage *long-read*

De nouvelles technologies de séquençage dites *long-reads* permettent d'obtenir moins de pièces de « puzzle » facilitant ainsi la reconstruction et l'étude des régions répétées dans le génome humain (Figure 15). Deux technologies principales coexistent aujourd'hui.

La première, développée par la société Pacific Biosciences est déposée sous le nom de « SMRT sequencing ® » pour « Single Molecule Real Time sequencing ». Il s'agit de préparer l'ADN de départ en y ajoutant des adaptateurs comme précédemment mais cette fois les adaptateurs sont en épingles et permettent de former une molécule d'ADN circulaire. Ensuite chaque molécule va se positionner dans un puits du support de séquençage avec un objectif d'une seule molécule par puits. Au fond du puits est fixée une polymérase qui va synthétiser le brin complémentaire de la molécule présentée à partir de nucléotides fluorescents. C'est ce signal de fluorescence qui est, comme pour le *Sequencing by Synthesis* (Illumina), interprété par l'instrument pour établir la séquence d'ADN (Figure 16). En revanche, ici, le marqueur fluorescent est clivé dès l'incorporation du nucléotide sans qu'il n'y ait de pause permettant l'acquisition du signal. Ceci contribue au séquençage de fragments beaucoup plus longs qu'avec la technologie Illumina en ayant pour conséquence une qualité de séquence moindre par lecture. Classiquement la longueur des lectures obtenues est d'environ 10 kb et peut aller jusqu'à 50 kb. La qualité de séquence obtenue a un score de Phred¹ habituellement compris entre 10 et 15. En séquençant plusieurs fois la même molécule circulaire, une séquence consensus de bien meilleure qualité peut cependant être obtenue (lectures HiFi, Phred > 25). De plus, l'interprétation de paramètres additionnels du signal brut, comme le délai séparant l'incorporation du nucléotide suivant peut théoriquement être utilisé pour analyser la méthylation de l'ADN. Les modules optique et fluide de l'instrument le rendent très onéreux. Son débit est progressivement en train d'augmenter avec le développement de nouveaux instruments (Sequel II) et commence à rendre

¹ Le score Phred est une échelle logarithmique correspondant à la probabilité d'erreur de séquençage. Un score de 10 correspond à 1 erreur pour 10 bases séquencées. Un taux d'erreur de 1% donne un score de Phred à 20. Un score de Phred de 30 (1 erreur pour 1000) est le minimum recommandé pour la recherche de variants nucléotidiques sur données Illumina.

réaliste économiquement le séquençage de génomes humains au moins à une profondeur permettant la détection de variants de structure.

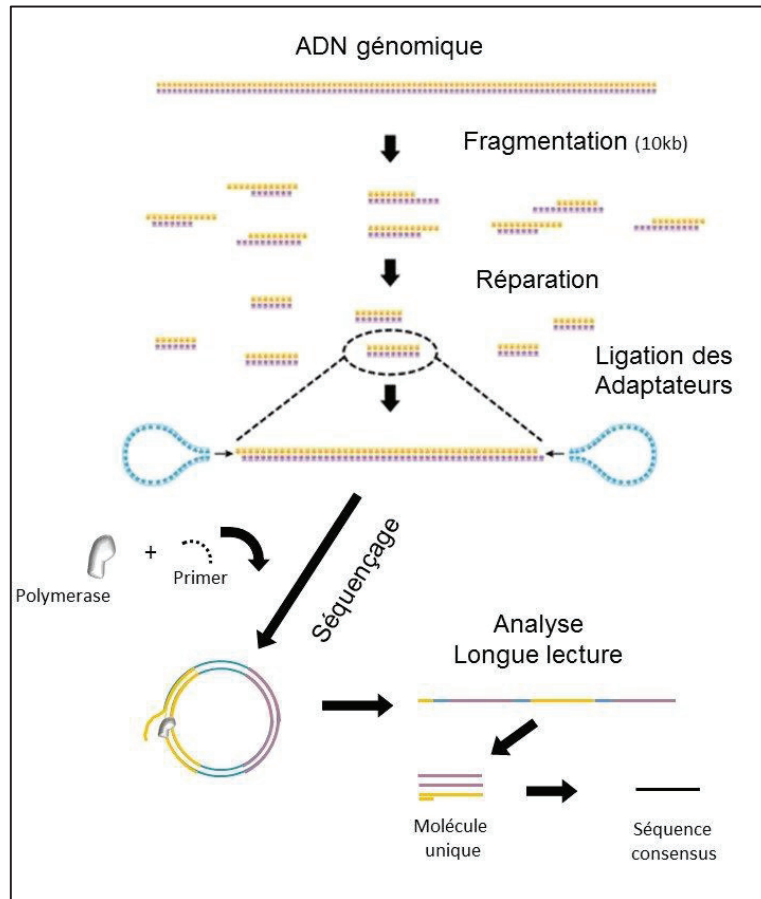


Figure 16 : Principe du séquençage SMRT : préparation de bibliothèques d'ADN génomiques grands fragments avec des adaptateurs en épingles à cheveux générant des molécules circulaires. Dans chaque puits la polymérase va synthétiser le brin complémentaire de la molécule présente avec la possibilité de faire plusieurs fois le tour et donc d'établir une séquence consensus entre les brins forward et reverse de chaque tour. Figure adaptée de www.francegenomique.org

La seconde firme importante sur ce marché est Oxford Nanopore Technologies qui développe une technologie de séquençage par nanopore. Cette fois le support de séquençage est constitué d'une membrane séparant deux espaces et recouvertes de pores biologiques (protéines bactériennes modifiées) les séparant. La préparation de l'ADN consiste cette fois à ajouter à l'extrémité des fragments d'ADN un adaptateur, séquence d'ADN standardisée couplée à une protéine dite « motrice ». L'adaptateur va permettre de fixer la molécule d'ADN au pore rapidement pour maximiser le taux d'occupation des pores. La protéine motrice qui est en fait une hélicase va ensuite ouvrir la double hélice pour faire passer l'un des deux brins d'ADN dans le pore. Entre les deux secteurs du support de séquençage est appliquée une différence de

potentiel électrique. Ce courant électrique va être modifié par le passage de la molécule d'ADN de manière différente en fonction de l'encombrement stérique de chaque nucléotide (Figure 17). C'est cette variation de signal électrique qui est interprétée pour obtenir la séquence de la molécule d'ADN. En fait, 5 nucléotides sont présents en même temps au niveau de la constriction où est mesuré le signal électrique ce qui rend difficile l'analyse des homopolymères aujourd'hui. Contrairement aux autres technologies précédemment décrites il n'y a donc pas de synthèse d'ADN et c'est la molécule native qui est analysée. En entraînant l'algorithme d'interprétation du signal ceci permet par exemple d'analyser la méthylation de l'ADN ou l'ensemble des modifications de l'ARN sur ARN natif. De plus, cela signifie que la seule limite à la taille du fragment séquencé est la taille du fragment d'ADN préparé. Des équipes ont rapporté des lectures de plus d'1Mb. La qualité de séquençage est stable tout au long de la lecture avec un score Phred compris entre 8 et 14. Celui-ci tend à s'améliorer avec les améliorations logicielles pour l'appel des bases et de nouvelles versions du pore utilisé (R10 : deux constriction permettant deux lectures du signal électrique). De plus, une variante de la préparation de librairie permet d'ajouter un adaptateur en épingle (*hairpin*) à l'autre extrémité de la double hélice pour que le brin antisens soit lu à son tour dans le même pore immédiatement après le brin sens (lecture 2D).

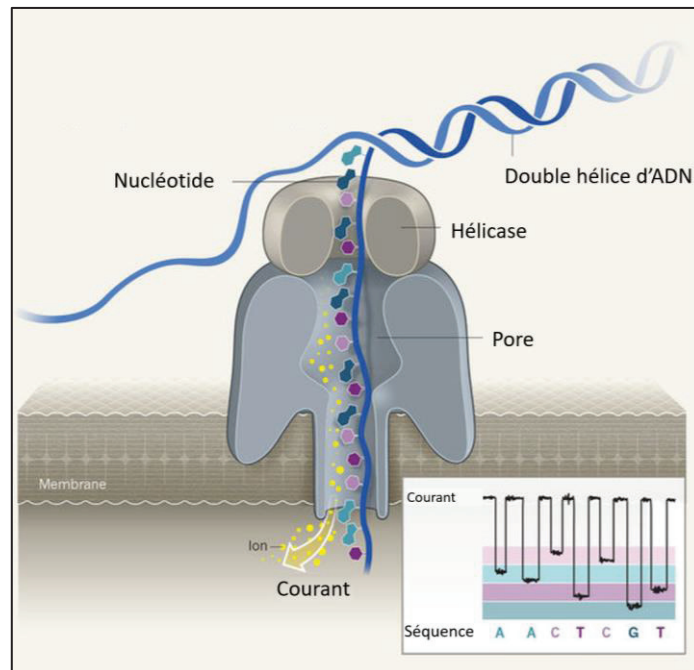


Figure 17 : Principe du séquençage nanopore. Le passage de la molécule d'ADN dans le pore entraîne une modification du courant électrique dépendante de la séquence d'ADN. Figure adaptée de <http://blogs.nature.com/naturejobs/2017/10/16/techblog-the-nanopore-toolbox/>

Dans tous les cas, le débit de 15-20Gb de données par *run* de séquençage annoncé par le fournisseur reste un débit insuffisant pour l'analyse de génome humain. Le fournisseur a développé deux plateformes de plus grande capacité qui parallélisent les *runs* et offrent un débit théorique de plusieurs Térabases par 24 heures. En France, les premières machines commencent à être installées sur les plateformes de génomique (GenoToul à Toulouse, Centre National de Recherche en Génomique Humaine à Evry).

Librairies alternatives

Linked-reads

Outre la possibilité déjà discutée de préparer une librairie pour augmenter la taille de l'insert entre les deux éléments d'une paire, d'autres stratégies compatibles avec le *paired-end sequencing* existent. La dernière étape étant le séquençage Illumina, ces approches bénéficient d'une grande qualité de séquençage. Nous présentons ici deux approches déjà utilisées pour la détection de variants de structure.

La première utilise une émulsion en amont de la préparation standard de librairie. L'appareil Chromium®, commercialisé par la société 10x Genomics, est capable de générer plusieurs milliers de gouttelettes contenant chacune un grand fragment d'ADN. Au sein de chaque gouttelette, le fragment va être fragmenté et les nouveaux fragments vont tous recevoir le même « tag » spécifique de la gouttelette. Celles-ci sont ensuite lysées avant une préparation de librairie standard. Au moment de l'alignement de séquences, les « tags » sont utilisés pour regrouper les lectures de fragments qui étaient liées dans une gouttelette (*linked-reads*) et recréer de manière synthétique les grands fragments présents initialement (Weisenfeld et al., 2017).

La technologie a été utilisée avec succès dans plusieurs études pour la mise en évidence de variants de structure (Elyanow et al., 2017; Eisfeldt et al., 2019; Marks et al., 2019).

Les lectures obtenues par séquençage *short-read*, *linked-read* et *long-read* sont schématisés dans la Figure 18, adapté de Lappalainen et al., 2019.

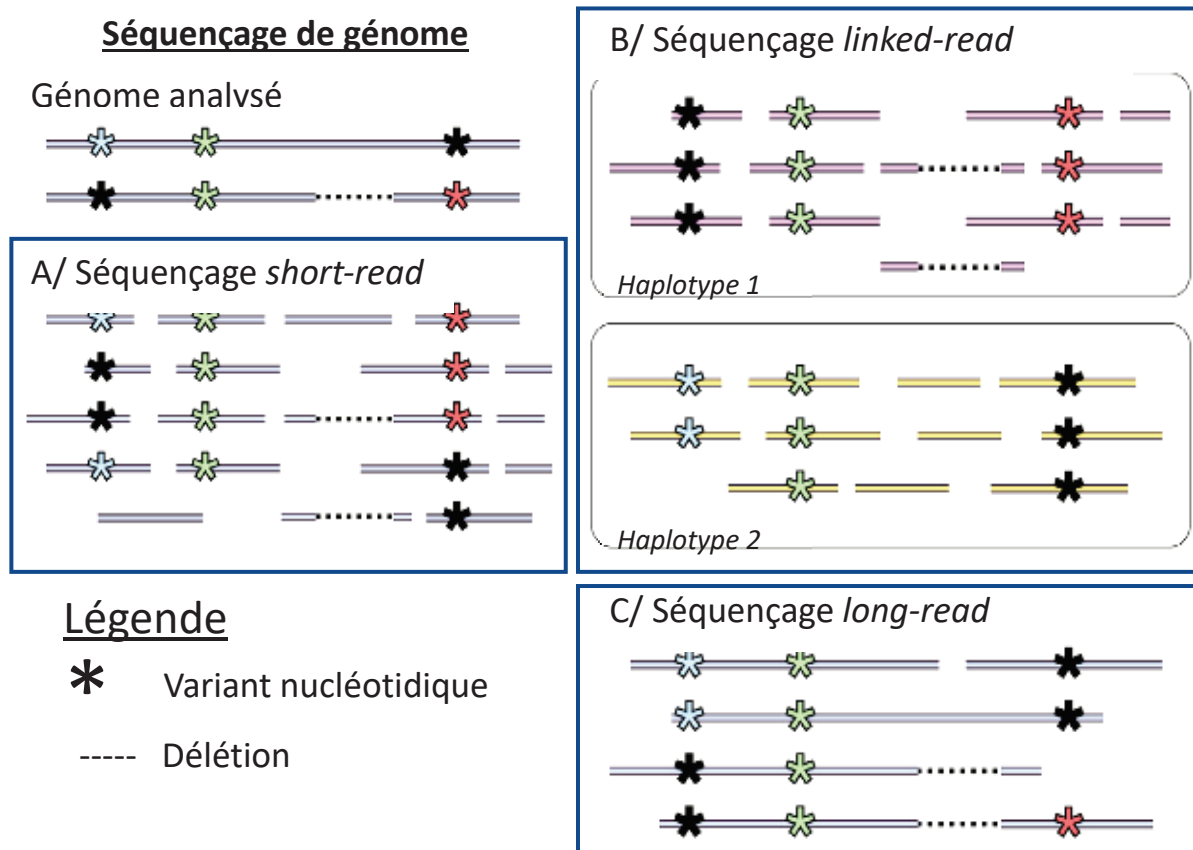


Figure 18 : Schéma récapitulatif des trois approches de séquençage présentées, adapté de Lappalainen et al., 2019.

Hi-C sequencing

Cette fois, il ne s'agit plus d'étudier le génome de manière linéaire mais d'étudier les repliements de la chromatine décrits précédemment et d'identifier les variations dues à des remaniements de structure. Le principe de la technique a été présenté dans le chapitre précédent. La détection de variant de structure est en quelque sorte un détournement de la technique initialement développé. En effet, en cas de variant de structure chromosomique les deux partenaires d'un réarrangement sont adjacents et l'intensité de contact est donc beaucoup plus élevée qu'elle ne l'est habituellement (Chakraborty and Ay, 2018; Dixon et al., 2018) (Figure 19). En s'affranchissant des distances linéaires cette technologie semble être en mesure de résoudre la problématique de la longueur des fragments d'ADN à préserver. Le fragment séquencé à ces deux extrémités est ici indépendant de la distance linéaire séparant les deux lectures (même si proximité tridimensionnelle et proximité linéaire vont logiquement de pair). Aussi, il semble possible de parvenir à détecter les événements encore indétectables par les autres technologies de séquençage parce qu'impliquant de très grandes séquences répétées comme c'est le cas pour les translocations robertsoniennes. Dans ces remaniements l'implication des bras courts de deux chromosomes acrocentriques est aujourd'hui impossible à analyser d'un bout à l'autre et ne peut être identifiée que par le caryotype. Néanmoins ce réarrangement venant perturber l'organisation physiologique du noyau pourrait être détecté en Hi-C du fait d'une plus grande proximité entre les deux chromosomes impliqués et de la redéfinition des territoires chromosomiques.

Si cette approche a l'avantage de révéler en parallèle les remaniements de structure et leurs potentielles conséquences sur l'organisation tridimensionnelle du génome, sa principale limitation est la résolution. En effet, si les translocations et les variations du nombre de copie de plus de 1Mb sont détectées de façon performante, il s'agit approximativement de la résolution obtenue en combinant caryotype et analyse chromosomique sur puce à ADN.

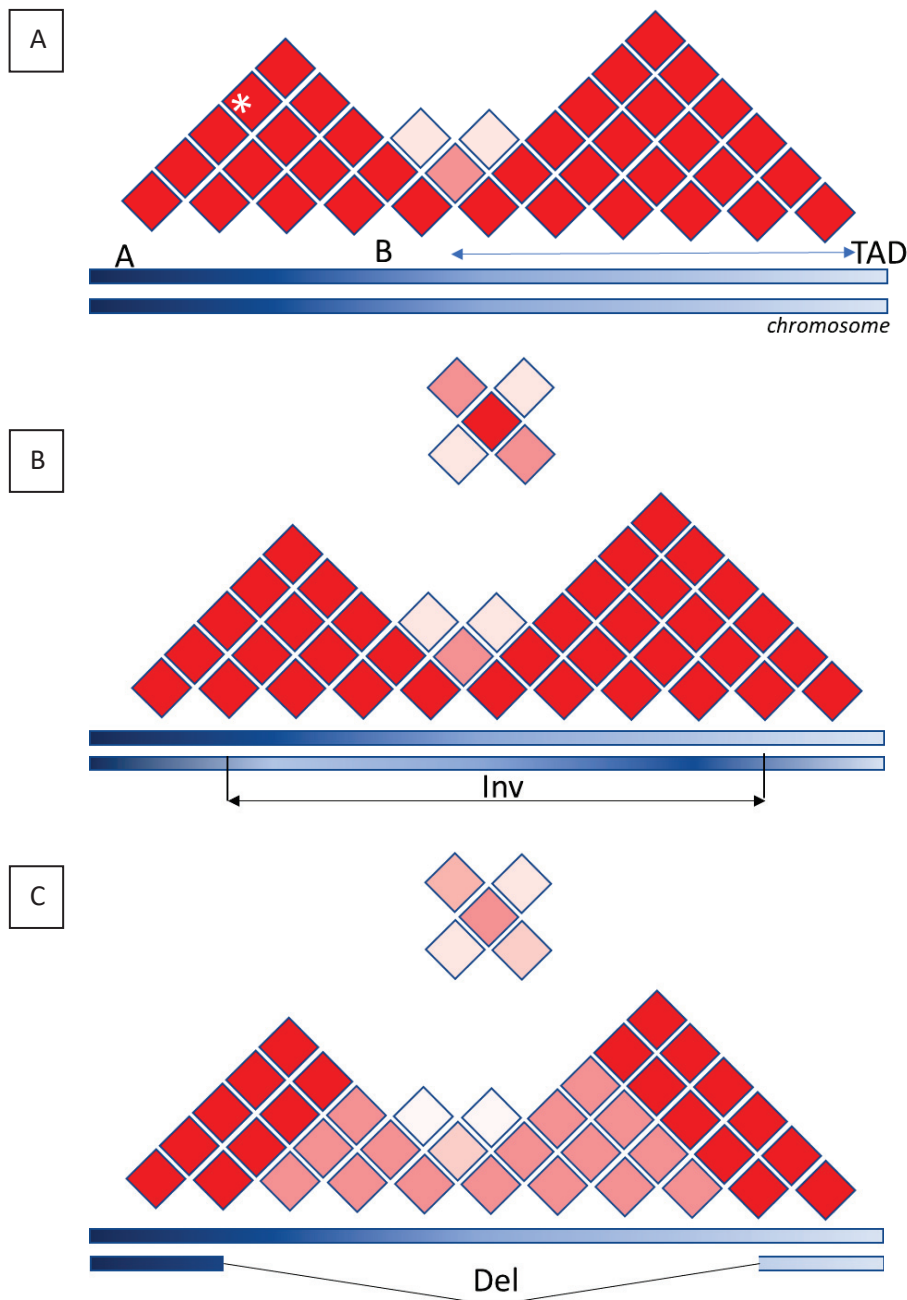


Figure 19 : Schématisation des résultats de séquençage Hi-C. Chaque case correspond à l'intensité de contact (le nombre de lectures) entre deux loci : plus la case est rouge plus le contact est intense. (A) L'astérisque signale la case correspondant aux contacts entre les loci A et B. Les TADs sont des domaines de fortes interactions chromatinienne bien délimitées qui prennent l'aspect de triangles sur ces données. (B) Dans le cas d'une inversion va apparaître un contact aberrant entre les deux points de cassure. (C) Ce contact aberrant sera également observé entre les bornes d'une délétion avec une réduction des intensités de contact dans l'intervalle délété en l'absence d'une copie. Ces remaniements peuvent également faire apparaître/disparaître des TADs, cette possibilité n'est pas schématisée ici.

Cartographie optique

La dernière technologie présentée ici, dite cartographie optique (optical mapping), n'est pas une technologie de séquençage et ne peut donc pas identifier de variations nucléotidiques mais uniquement des variants de structure. Le principe rappelle celui des cartes de restriction établies au début du projet de séquençage du génome humain. Il s'agit d'isoler de très grands fragments d'ADN pour les faire passer dans des canaux individuels après qu'un motif spécifique d'ADN ait été marqué par une enzyme. C'est la fréquence de ce motif qui détermine la résolution de la technique. Chaque molécule d'ADN présente un profil unique de marquage. C'est l'assemblage des molécules entre elles qui va permettre de reconstruire des contigs et de détecter des variants de structure en comparant cet assemblage *de novo* au génome de référence (Figure 20). Cette technologie a également été utilisée pour l'assemblage de génome *de novo* de nombreuses espèces (Kronenberg et al., 2018; Nowoshilow et al., 2018). Outre l'absence de séquençage et donc d'étude des SNVs, sa principale limitation pour la détection de variants de structure est logiquement sa résolution. Le principal fournisseur actuel, Bionano Genomics, annonce une sensibilité de 95% pour les insertions/délétions de plus de 500 bases, 97% pour les CNVs de plus de 500 kb, 95% pour les remaniements interchromosomiques.

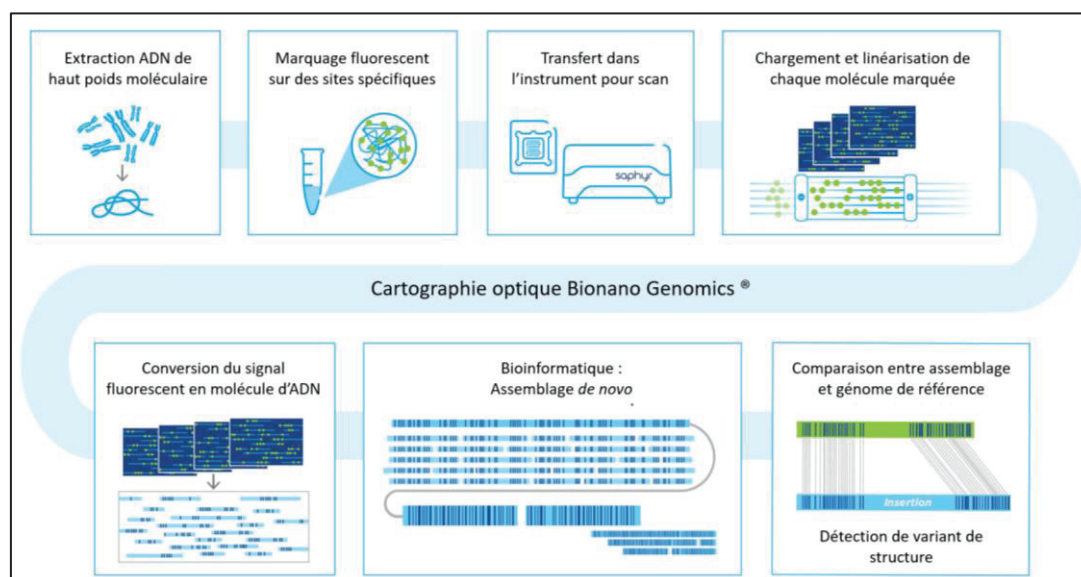


Figure 20 : Principe de la cartographie optique depuis l'extraction d'ADN à la détection de variants de structure. Figure adaptée de bionanogenomics.com

I – 4/ Variant de structure chromosomique

Les anomalies chromosomiques peuvent être distinguées entre anomalies du nombre de chromosomes (aneuploïdies) comme la trisomie 21, ou anomalie de la structure des chromosomes. Les anomalies constitutionnelles du nombre de chromosomes résultent le plus souvent d'une non-disjonction méiotique. Les deux chromosomes d'une même paire ou les deux chromatides d'un chromosome échouent à se séparer et sont présents dans la même cellule fille. Elles ne font pas l'objet de ce travail.

Les anomalies de structures peuvent être déséquilibrées ou équilibrées, sans perte ni gain de matériel génétique. Elles résultent d'un recollement anormal entre au moins deux points de cassure.

I – 4/ A- Réparation des cassures double brin de l'ADN

L'observation de chacun de ces remaniements est la conséquence d'une mauvaise réparation de cassures double brin de l'ADN. Ces cassures peuvent être le fait d'agents exogènes, principalement en cause dans le cancer (radiations ionisantes, virus) ou endogènes (défaut de la fourche de réplication, dérivé réactif de l'oxygène, nucléases, sites fragiles de l'ADN,...). L'impact de ces différents facteurs a été estimé à 0,05 cassure double brin par cellule pour un vol en avion de 10 heures, 0,3 pour une examen tomodensitométrie corps entier, 12 en cas d'accident nucléaire équivalent à l'explosion de Tchernobyl et 80 en cas de radiothérapie externe (Ciccia and Elledge, 2010). Les acteurs de la détection et de la réparation des cassures double brin sont donc sollicités en permanence. La chromatine avoisinante doit d'abord être remodelée pour permettre l'action des mécanismes de réparation (Liu et al., 2012). La protéine JNK va phosphoryler SIRT6 pour permettre sa mobilisation au site de cassure et le recrutement de PARP1, le tout en environ une seconde (Van Meter et al., 2016). Acl1 va pouvoir se fixer à PARP1 et commencer les modifications chromatiniennes nécessaires (Sellou et al., 2016). S'ensuivra le recrutement de la première enzyme de réparation MRE11 (Haince et al., 2008). Parallèlement la décondensation de la chromatine est également permise par le remplacement des histones H2 par un variant phosphorylé H2AX qui va recruter RNF8, responsable de la

décondensation (Rogakou et al., 1998). Ce remplacement se fait dans la minute après la survenue de la cassure et concerne une région allant jusqu'à 2 Mb du point de cassure. Les cassures peuvent ensuite être réparées par différents mécanismes.

Recombinaison homologue

Le mécanisme le plus fidèle de réparation des cassures double brin utilise une séquence homologue pour reconstruire la séquence cassée à l'identique. Le complexe MRN va se fixer aux extrémités double brin libres et réséquer une partie de chaque extrémité 5' pour libérer une extrémité 3' simple brin de chaque côté sur environ 800 nucléotides (Zakharyevich et al., 2010). La protéine RPA a une forte affinité pour ces extrémités 3' libres et va recruter la protéine Rad51 en charge de la formation d'un filament nucléoprotéique recouvert par RPA (Wold, 1997). S'ensuit la recherche de la séquence similaire à l'extrémité 3' libre qui, une fois trouvée sera « envahie ». La structure ainsi formée, D-loop, est ensuite modifiée par l'action d'une polymérase qui va venir allonger l'extrémité 3' libre en utilisant la séquence homologue comme modèle, dans une structure nommée jonction Holliday (Sung and Klein, 2006). Deux modèles existent ensuite. Dans le premier, « *double strand break repair* » (DSBR), l'autre extrémité 3' libre forme elle aussi une jonction Holliday et la jonction est résolue par l'action d'endonucléases (Sung and Klein, 2006) (Figure 21). Dans le second modèle, « *synthesis-dependent strand annealing* » (SDSA), la polymérisation s'arrête et les extrémités 3' ainsi allongées produites peuvent s'hybrider de part et d'autre de la cassure avant synthèse du brin complémentaire (Figure 21).

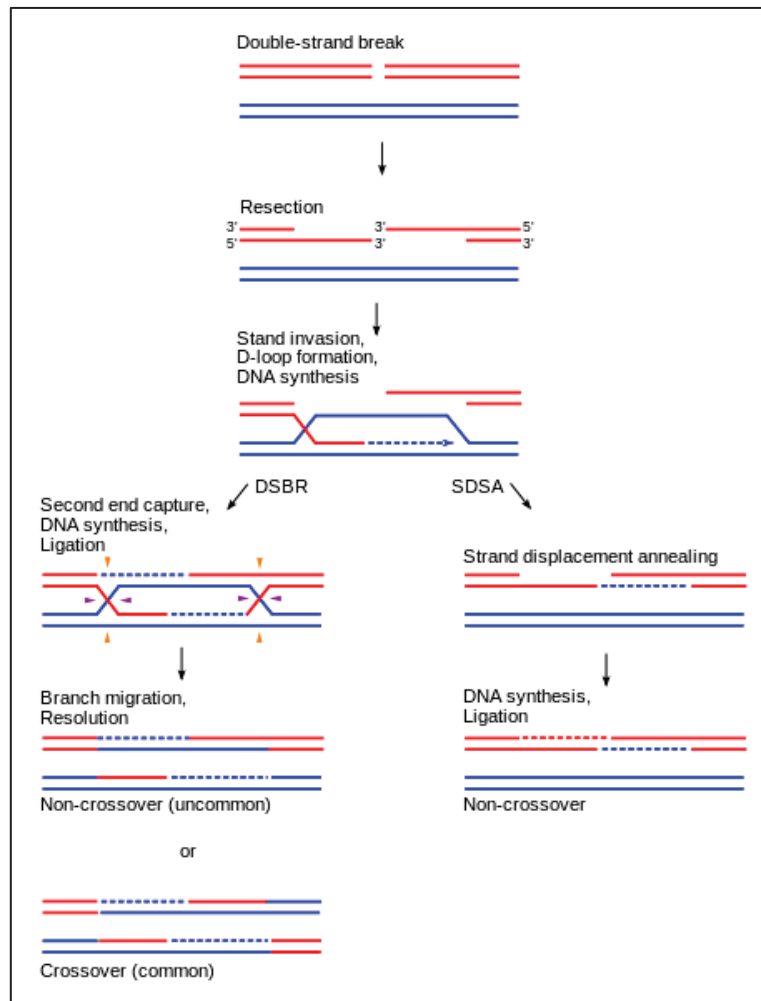


Figure 21 : Mécanismes de recombinaison homologue pour la réparation des cassures double brin de l'ADN. DSBR : double strand break repair. SDSA : synthesis-dependent strand annealing. Figure extraite de Sung and Klein, 2006)

Cette recombinaison homologue (*homologous recombination*) peut cependant être « trompée » par la séquence d'ADN. Nous avons évoqué précédemment la présence d'éléments répétés de grande taille dans le génome dénommés *low copy repeats* (LCR). Ils représentent environ 10% de la séquence du génome humain (Bailey et al., 2001). La présence de ces éléments très homologues sur un même chromosome peut entraîner la survenue d'un *crossing-over* inégal. Lors des multiples recombinaisons survenant entre deux séquences homologues, un décalage va se produire et mettre en rapport deux LCR non alléliques. Ce mécanisme de recombinaison homologue non allélique (*Non allelic Homologous Recombination NAHR*) est responsable de délétions et duplications des régions comprises entre les LCR et donc récurrentes (Beckmann et al., 2007). Cette mauvaise recombinaison peut survenir entre deux chromosomes homologues, entre les deux chromatides d'un même chromosome ou au sein même d'une

chromatide. Cette dernière possibilité ne pouvant donner lieu qu'à des délétions (les fragments acentriques étant perdus) explique la plus grande incidence des microdélétions récurrentes face aux microduplications récurrentes (Turner et al., 2008). Une *crossing-over* entre deux LCR d'une même chromatide disposés en sens inverse peut également promouvoir des inversions récurrentes. La Figure 22 schématise les différentes possibilités (Chen et al., 2014a).

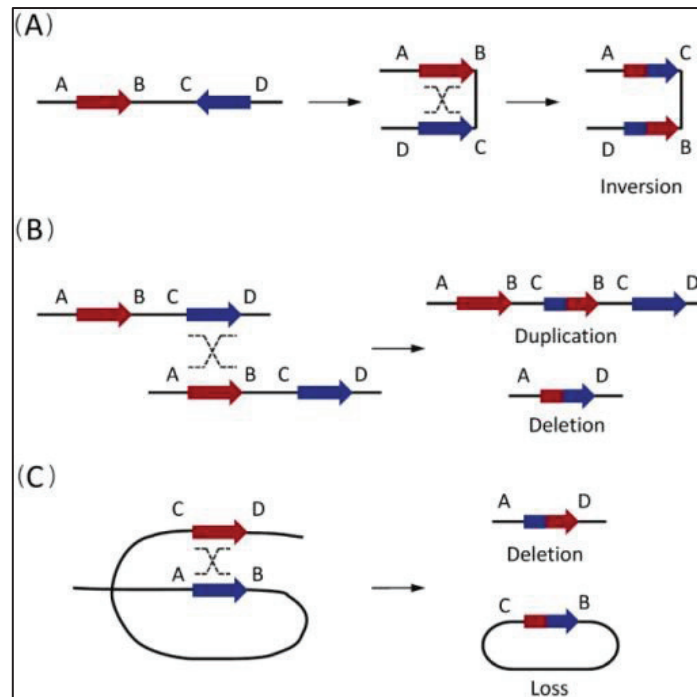


Figure 22 : Schéma des différents types de réarrangements consécutifs à une recombinaison homologue non allélique entre deux LCR (flèches rouges et bleues). (A) 2 LCR en sens inverse favorisent les inversions. Les recombinaisons entre chromosomes homologues, chromatides d'un même chromosome (B) ou intrachromatidiennes (C) entraînent des délétions et duplications récurrentes.

Si la majorité des syndromes liés à ces remaniements récurrents (syndrome de DiGeorge (del 22q11.2), syndrome de Williams et Beuren (del 7q11.2), ...) étaient définis cliniquement et diagnostiqués en FISH, les techniques d'ACPA ont révélé de nouvelles entités cliniques (délétion et duplication 16p11.2, délétion 17q21.31 du syndrome de Koolen et de Vries) (Schluth-Bolard et al., 2008).

Non-homologous end-joining classique

La jonction d'extrémités non homologues (*Non homologous End Joining NHEJ*) a été le deuxième mécanisme décrit (Wilson et al., 1982). Aujourd'hui renommé NHEJ classique (*classical NHEJ*) par opposition au mécanisme décrit dans le paragraphe suivant, elle ne nécessite pas de séquence répétée ou d'homologie de séquence entre les deux partenaires de la jonction même si des microhomologies de quelques nucléotides peuvent être utilisées pour faciliter la réparation (Roth and Wilson, 1986). C'est un mécanisme non-conservatif. Le dimère Ku70/Ku80 va se placer au niveau des cassures double brin et ainsi les protéger rapidement. De nombreuses protéines (XRCC4, Artemis, ligase IV, WRN, BLM,...) sont recrutés par le dimère Ku probablement avec des fonctions en partie redondantes (Deriano and Roth, 2013). Les extrémités peuvent être directement rejointes sous l'action d'une ligase ou nécessiter d'être « nettoyées » notamment par Artemis (Figure 23). Celle-ci a une activité nucléase qui va permettre de rétablir des extrémités franches aux cassures en produisant des extrémités 5'-phosphate et 3'-OH compatibles avec l'action d'une ADN ligase. De courts cycles de polymérisation-suppression aléatoire de séquences peuvent également avoir lieu pour tenter de raccrocher les deux partenaires d'une cassure, insérant ainsi quelques nucléotides. L'étape final de ligation est pris en charge par le complexe XRCC4-Ligase IV-XLF. Pendant ce temps le complexe RAD50/MRE1 permettrait de maintenir à proximité les deux partenaires d'une cassure pour prévenir la survenue d'un remaniement chromosomique (Stracker and Petrini, 2011).

Il s'agit du « recollement » de fragments non cohésifs qui peut laisser une « cicatrice » au niveau de la jonction en modifiant la séquence d'ADN (Lieber et al., 2008). L'insertion de quelques nucléotides est notamment possible. La NHEJ classique est le mécanisme dominant chez les mammifères. Elle est active tout au long du cycle cellulaire et plus particulièrement aux stades G0/G1 (Rothkamm et al., 2003).

Non-homologous end-joining alternatif

Le dernier mécanisme identifié est alternativement dénommé NHEJ alternatif (*alternative NHEJ*), jonction des extrémités médiée par microhomologie (*microhomology mediated end joining MMEJ*) ou *backup-NHEJ* (Zhu et al., 2002; Simsek and Jasin, 2010) (Figure 23). Initialement découvert dans des cellules déficientes pour les acteurs protéiques de la NHEJ classique il est apparu par la suite que ces deux mécanismes, NHEJ classique et NHEJ alternatif, coexistaient dans les mêmes cellules (Liang and Jasin, 1996; Kabotyanski et al., 1998; Nussenzweig and Nussenzweig, 2007). Les acteurs protéiques impliqués dans ce mécanisme restent mal connus. Un dixième des cassures réparées par jonction des extrémités le seraient par ce mécanisme alternatif (Corneo et al., 2007). Les jonctions générées par le NHEJ alternatif peuvent être considérées comme de moins bonne qualité puisqu'elles contiennent plus de délétions. Des microhomologies sont souvent présentes entre les deux partenaires de la jonction mais ne sont pas obligatoires. Ce mécanisme est plus fréquemment impliqué en cas de remaniements chromosomiques (Yan et al., 2007). Ceci pourrait être la conséquence d'un moins bon maintien des extrémités libres d'une cassure ensemble et d'un temps de latence plus long par rapport au NHEJ classique augmentant le risque de mésappariement (Yan et al., 2007; Zhang et al., 2010).

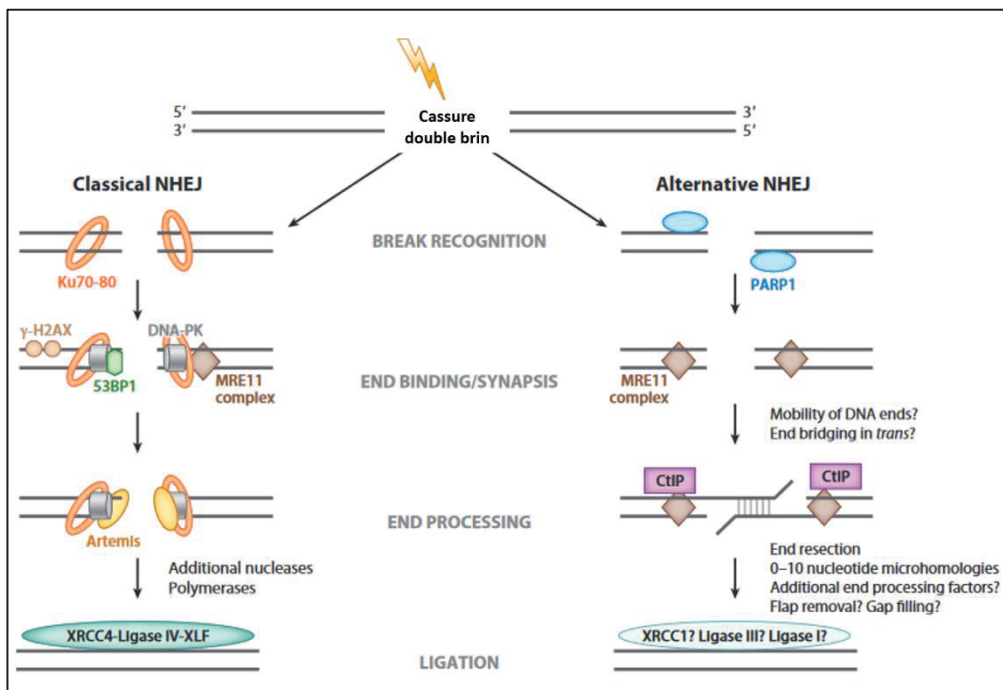


Figure 23 : Comparaison des mécanismes de NHEJ classique et alternatif avec les effecteurs protéiques en jeu. Schéma adaptée de Deriano and Roth, 2013.

Mécanismes réplicatifs

Les mécanismes précédemment évoqués permettent la jonction de deux partenaires après la survenue de deux cassures double-brin, la situation est différente pour le *break induced replication* (BIR) qui survient lors de la réplication de l'ADN. Lorsque l'hélicase qui vient ouvrir la double hélice pour la fourche de réplication rencontre une cassure du brin qui sert de trame à celle-ci, la fourche est brisée. En l'absence de partenaire pour établir une jonction les mécanismes de recombinaisons homologues ou de NHEJ ne peuvent être mis en œuvre. Pour libérer une extrémité 3' simple brin, une exonucléase vient digérer l'extrémité 5' au niveau de la lésion. L'extrémité 3' libre va venir envahir une séquence homologue, généralement la chromatide sœur. En effet, cette invasion se fait via le couple RPA/Rad51 et nécessite une homologie supérieure à 70-100 nucléotides (Ira and Haber, 2002). La perturbation de la fourche de réplication peut entraîner elle-même une nouvelle cassure double brin entraînant une répétition du processus décrit. Plusieurs changements de brin en cours de réplication sont donc possibles au sein d'un même remaniement. Lors de l'emploi du BIR, le risque de remaniement chromosomique est limité par la longueur de séquences effectivement requise pour faire redémarrer la fourche de réplication. Ce n'est pas le cas dans une variante de ce mécanisme, indépendante de RecA/Rad51 et qui utilise des homologies beaucoup plus courtes (29 nucléotides) (Ira and Haber, 2002), dénommé alors *microhomology-mediated BIR* (MMBIR). De fait le risque de remaniement chromosomique est beaucoup plus important en utilisant une séquence d'homologie plus courte.

La mise en évidence de ce mécanisme découle en partie du travail de l'équipe de James Lupski autour de certains remaniements du gène *PLP1* responsables de syndromes de Pelizaeus-Merzbacher (Lee et al., 2007). En étudiant le locus dans une cohorte de patients atteints de ce syndrome ils ont identifié des délétions et des duplications en tandem résultant apparemment de NAHR mais également des profils plus complexes sans point de cassure récurrent. Ces derniers profils présentaient des microhomologies aux points de cassure sans qu'il existe une raison particulière pour qu'une cassure simple brin capable de disloquer la fourche de réplication survienne particulièrement à ce locus *PLP1*. L'analyse des séquences d'ADN avoisinant les sites de microhomologies utilisées, où la fourche se déconnecte et reconnecte, montre une forte probabilité de formation de structures complexes d'ADN : structure

cruciforme, en épingle, ... Le modèle proposé, et aujourd'hui nommé *Fork Stalling and Template Switching (FoSTeS)*, n'impose plus de cassure simple brin mais simplement un décrochage de la fourche de réplication au passage d'une structure complexe. En ce sens, le FoSTeS ne peut être considéré comme un mécanisme de réparation de point de cassure mais plutôt comme générateur de remaniements chromosomiques (Zhang et al., 2009) (Figure 24).

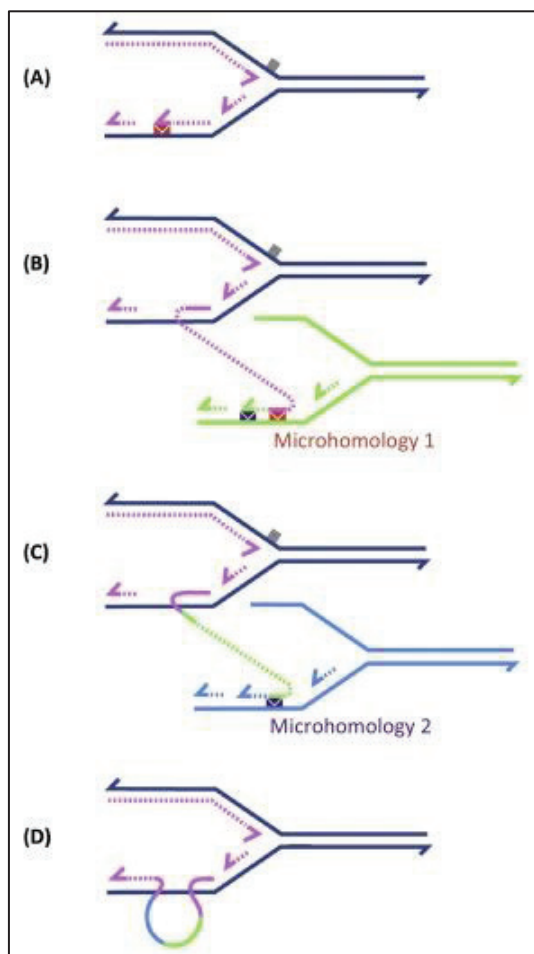


Figure 24 : Schéma du mécanisme de FoSTeS. La fourche de réplication butte sur un "obstacle" laissant libre l'extrémité d'ADN du brin retard (A). Cette extrémité libre va pouvoir trouver une microhomologie et provoquer le « déraillement » de la fourche de réplication vers cette nouvelle séquence (B). Ce procédé peut se répéter une ou plusieurs fois (C) jusqu'à aboutir à la synthèse d'un chromosome remanié (D).

I – 4/ B- Remaniements simples

Les anomalies du nombre de chromosomes (aneuploïdies) sont, par définition, déséquilibrées. Une trisomie est la présence d'un chromosome supplémentaire d'une paire donnée. A l'inverse une monosomie désigne la perte d'un chromosome d'une paire. La présence d'un troisième lot de chromosomes de chaque paire est appelée triploïdie. Les autres situations d'hypoploïdie ou d'hyperploïdie, variation du nombre de lots haploïdes de chromosomes ne sont observés que dans les génomes de tumeurs.

Parmi les remaniements de structure chromosomique on distingue les remaniements équilibrés, sans perte ni gain de matériel, des remaniements déséquilibrés.

Le remaniement déséquilibré le plus simple est la délétion terminale. Elle fait suite à la survenue d'une seule cassure chromosomique suivie de la perte du fragment terminale et capture ou synthèse d'un nouveau télomère. Une délétion interstitielle, elle, nécessite la survenue de deux cassures sur un même chromosome (un même bras chromosomique) suivie de la perte du fragment compris entre les deux points de cassure. Lorsqu'une copie est perdue la délétion est dite hétérozygote, et homozygote lorsque les deux copies sont perdues. On parle de duplication lorsqu'une copie supplémentaire est présente à son locus d'origine. Une duplication est dite en tandem lorsque les deux copies sont orientées dans le même sens et en miroir lorsqu'elles sont orientées en sens inverse. Des exemples de remaniements de ce type visibles au caryotype sont présentés dans la Figure 25.

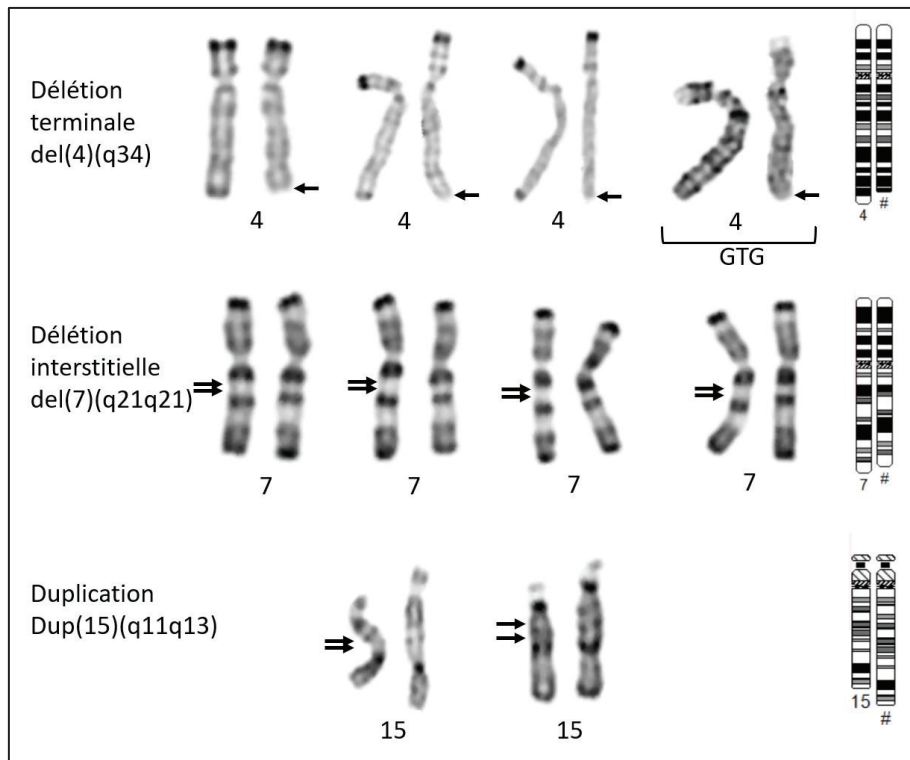


Figure 25 : Exemples de remaniements déséquilibrés visibles au caryotype en bandes RHG (bandes GTG pour la 4^{ème} paire de chromosomes 4) et idéogrammes correspondants. Par convention le chromosome normal est placé à gauche. Les points de cassure sont signalés par une flèche.

Plus rarement peut se former un isochromosome, chromosome qui a remplacé l'un de ces bras par une deuxième copie du bras restant et qui est donc symétrique. Enfin, un chromosome de structure non identifiable est un chromosome dit « marqueur ».

A la frontière entre remaniement équilibré et remaniement déséquilibré se trouve les chromosomes en anneau. Lors de la survenue d'une cassure sur chacun des bras chromosomiques, l'absence de télomère peut, en fonction des points de cassure, entraîner la fusion des régions subtélomériques et former un chromosome en anneau. Si un déséquilibre génomique peut être identifié il est également possible que seules des fragments télomériques, régions répétées non codantes soient perdues. Dans ce cas aucun déséquilibre génomique ne sera observé. C'est le cas par exemple dans une forme particulière d'épilepsie typique lié à la présence d'un chromosome 20 en anneau (Vignoli et al., 2016).

Une translocation réciproque est l'échange de matériel chromosomique entre deux chromosomes non homologues suite à la survenue d'un point de cassure sur chacun d'entre eux. Les translocations robertsoniennes constituent un type de translocation particulier. Elles

concernent les chromosomes qui ont un bras court de très petite taille uniquement composé de séquences d'ADN ribosomiques redondantes entre les chromosomes concernés, dits acrocentriques. Chez l'Homme, les chromosomes acrocentriques sont les chromosomes 13 ; 14 ; 15 ; 21 et 22. En cas de translocation robertsonienne il y a, le plus fréquemment, une cassure juxtacentromérique dans les bras courts puis fusion formant un chromosome dicentrique avant qu'un des deux centromères ne soit inactivé. Par le pachytène qui se forme dans ce cas lors des divisions méiotiques, et du fait des chromosomes concernés, ces remaniements sont particulièrement à risque de trisomie 13 ou 21.

Lors d'une inversion, le fragment chromosomique contenu entre deux cassures survenues sur un même chromosome se retourne. On parle d'inversion péricentrique lorsque le fragment contient le centromère et d'inversion paracentrique lorsque ce n'est pas le cas (cassures sur le même bras chromosomique).

Enfin l'insertion chromosomique consiste au déplacement d'un fragment chromosomique à un autre locus. Nécessitant 3 points de cassure cet évènement est le plus rare dans cette catégorie. Des exemples sont présentés dans la Figure 26.

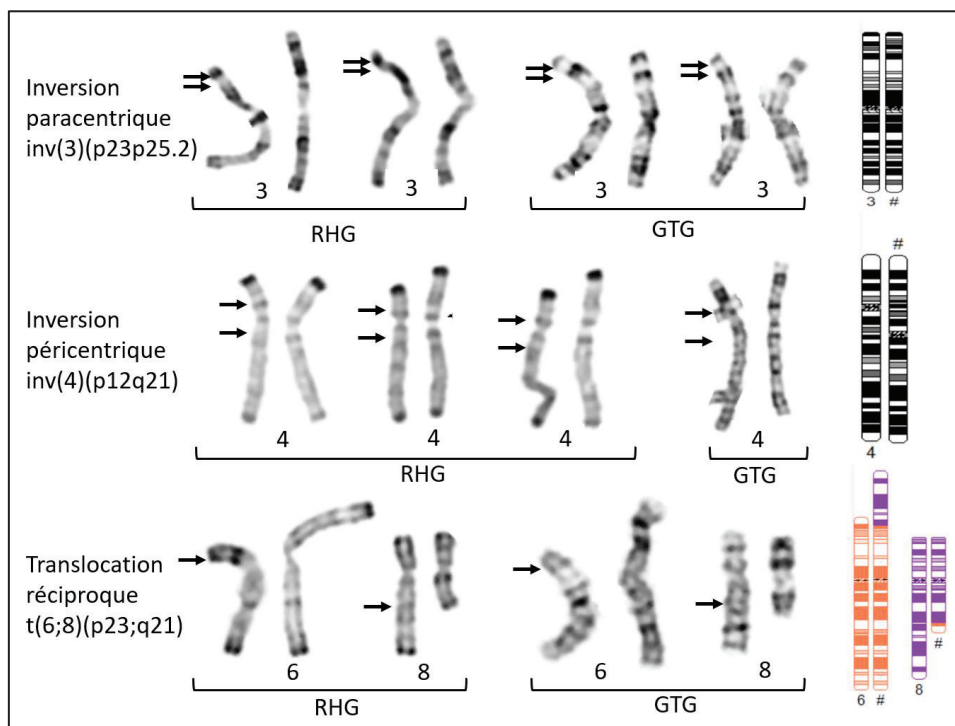


Figure 26 : Exemples de remaniements chromosomiques équilibrés visibles au caryotype et idéogrammes correspondants. Par convention le chromosome normal est placé à gauche. Les points de cassure sont signalés par une flèche.

I – 4/ C- Remaniements complexes

En 1970, les auteurs rapportant un réarrangement chromosomique impliquant 3 (ou 4) chromosomes sont les premiers à utiliser la désignation remaniement complexe (Fredga and Hall, 1970). Après avoir été défini par défaut comme un remaniement plus complexe qu'une simple translocation réciproque (Pai et al., 1980), le caractère complexe d'un réarrangement a ensuite été défini par la présence d'au moins 3 points de cassure sur plus d'un chromosome (Pellestor et al., 2011).

L'arrivée des techniques de séquençage massif en parallèle à la fin des années 2000 révèle des remaniements toujours plus complexes. En 2011, le néologisme « *chromothripsis* » est utilisé pour la première fois pour désigner un réarrangement extrêmement complexe avec 42 points de cassure concentrés sur le bras long d'un chromosome 4 chez un patient présentant une leucémie lymphoïde chronique (Stephens et al., 2011). La définition du remaniement complexe impliquant plusieurs chromosomes ne tient plus. Aussi, un remaniement complexe est aujourd'hui défini en miroir des remaniements dits simples. Est considéré complexe un réarrangement pour lequel le conseil génétique n'est pas maîtrisé, c'est-à-dire un remaniement qui n'est pas une translocation, une inversion ou une insertion (Madan, 2013; McGowan-Jordan, 2016). Le nombre de points de cassure minimum est donc *de facto* de 3 (en excluant les insertions).

Le nombre et le regroupement des points de cassure observé dans ce premier *chromothripsis* va à l'encontre du modèle de Knudson de la tumorigénèse. Dans ce modèle, c'est l'accumulation de changements génétiques au cours du temps qui entraîne le phénomène cancéreux alors qu'il semble ici y avoir eu un évènement catastrophique unique. Lors des descriptions suivantes il est également noté que ce type de remaniement se caractérise par un nombre de copie qui est soit normal soit de 1 (délétion hétérozygote). A l'inverse, dans les remaniements tumoraux qui résultent d'une accumulation de changements sur plusieurs cycles cellulaires on observe des délétions homozygotes et surtout des amplifications, sous forme de marqueur surnuméraire par exemple, procurant un avantage sélectif (Nones et al., 2014).

Enfin, dans le *chromothripsis* est également observé un nombre égal de fragment inversés et non-inversés lui aussi ne pouvant pas être expliqué par l'addition d'évènements successifs

(Korbel and Campbell, 2013). L'identification de ces remaniements à l'état homogène dans les cellules cancéreuses fait dire aux auteurs qu'il s'agit d'un phénomène catastrophique unique survenant lors de la tumorigénèse. L'hypothèse formulée est celle d'une pulvérisation (-thripsis en grec) d'un chromosome (ou d'un bras chromosomique, ou de quelques chromosomes) suivie d'une reconstruction par simple recollement des fragments générés et perte de certains d'entre eux. En effet, l'étude des fragments de jonction de ces remaniements a mis en évidence l'utilisation des mécanismes de *non-homologous end-joining* (NHEJ) classique ou alternatif. Ceci expliquerait le regroupement des points de cassure sur un seul chromosome, un seul bras chromosomique voire une seule bande (Figure 27).

Par la suite, ont été rapportés des profils de remaniements chromosomiques complexes quelque peu différents avec notamment des gains de matériel chromosomique. L'étude des fragments de jonction de ces remaniements a permis d'impliquer des mécanismes de réparation des cassures double-brin différents de ceux habituellement observés dans le *chromothripsis*. Il s'agissait de mécanismes dits réplcatifs, *Microhomology-Mediated Breakage-Induced Repair (MMBIR)/ Fork Stalling and Template Switching (FoSTeS)*. Ainsi le réarrangement serait lié à l'emballement d'un mécanisme réplcatif sans qu'une étape de pulvérisation chromosomique ne soit nécessaire. Ces remaniements ont été désignés par le néologisme *chromoanagenesis* (Liu et al., 2011; Holland and Cleveland, 2012; Luijten et al., 2018) (Figure 27).

Enfin, au cours de leur travail de caractérisation de tumeurs prostatiques, Baca *et al.* rapportent un troisième type de remaniement chromosomique complexe qu'ils nomment *chromoplexie* (-pleko : tisser, entrelacer). Ils observent des remaniements en chaîne pour lesquels les jonctions sont fréquemment flanquées de délétions et qui entraînent la formation d'un transcrit de fusion à fort activité oncogénique (MPRSS2-ERG) (Figure 27). Ces remaniements impliquent au moins 5 chromosomes sans que ne puisse être identifié de focus de points de cassure comme c'est le cas dans le *chromothripsis* (Baca et al., 2013). Les auteurs suggèrent que la formation de ces chaînes dans le remaniement résulte, comme le transcrit de fusion isolé, de l'action du récepteur aux androgènes. C'est lui qui entraînerait une colocalisation d'un ensemble de loci génomiques et des cassures double-brin qui seraient ensuite mal réparées. Ce type de remaniement complexe ayant depuis été également décrit dans des cancers du poumon, oropharyngés et des mélanomes ce mécanisme ne semble pas exclusif.

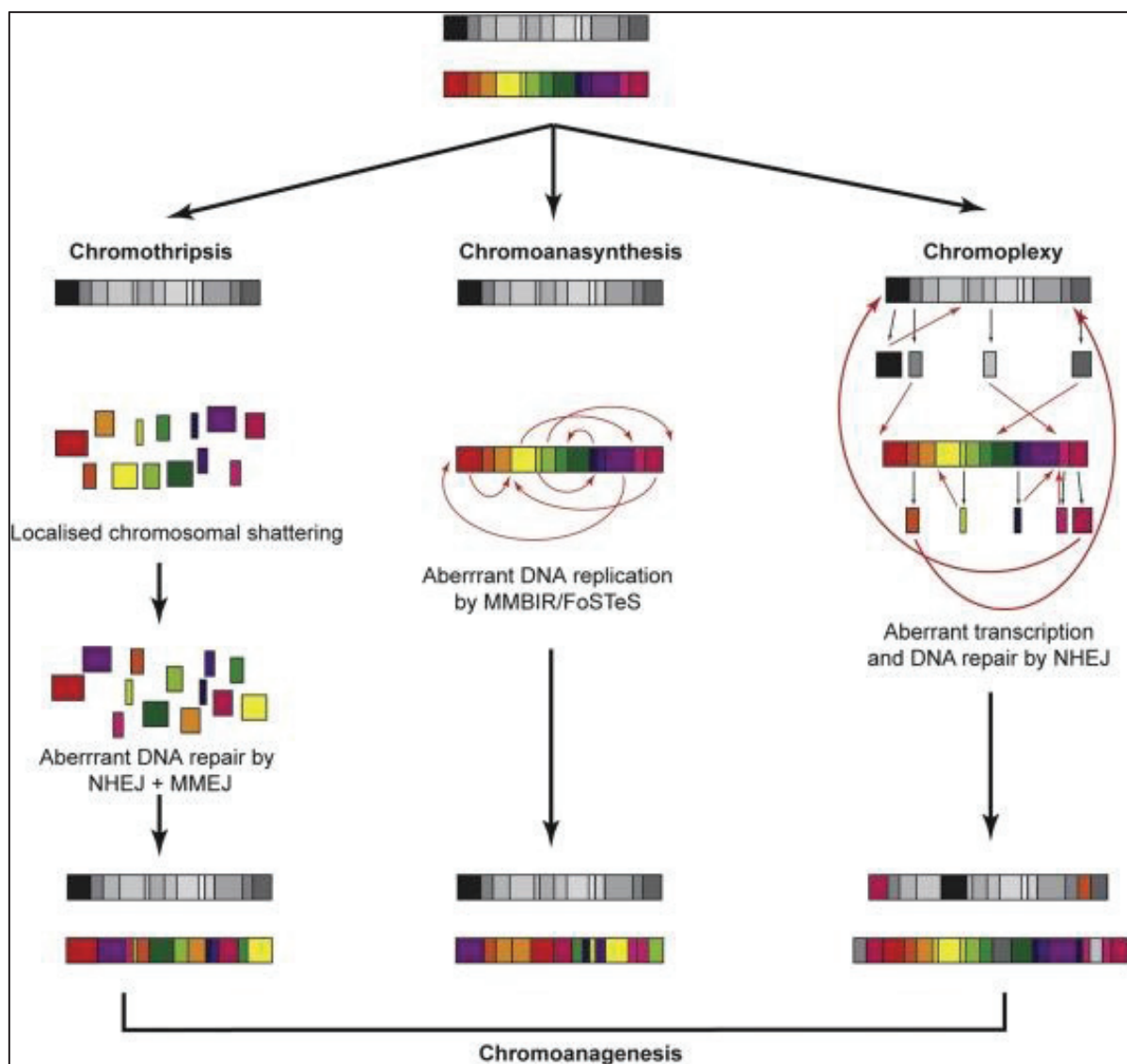


Figure 27 : Schéma des 3 types de remaniements chromosomiques complexes dits chromoanagenesis. Les gains du nombre de copie ne peuvent pas être observées en cas de chromothripsis.

Quelques mois après la description de *chromothripsis* dans un génome tumoral est rapportée la première observation de *chromothripsis* constitutionnel. Il s'agissait d'un jeune garçon présentant une déficience intellectuelle sévère, une dysmorphie faciale et une scoliose pour lequel un remaniement à 12 points de cassure impliquant 3 chromosomes a été identifié (Kloosterman et al., 2011a).

Les trois types de remaniements, *chromothripsis*, *chromoanasythesis* et *chromoplexie* sont aujourd'hui regroupés sous le terme *chromoanagenesis* sans qu'aucun remaniement constitutionnel n'était été considéré *chromoplexie*.

Fréquence

Dans les génomes tumoraux, un *chromothripsis* a pu être identifié dans 2 à 3% des cas (Stephens et al., 2011) avec des disparités en fonction du type de tumeur allant jusqu'à un tiers des cas d'ostéosarcome (Kloosterman et al., 2014; Behjati et al., 2017) ou de glioblastome (Malhotra et al., 2013; Erson-Omay et al., 2017). Ces chiffres, principalement obtenues à partir de données d'analyse chromosomique sur puce à ADN ou de séquençage à faible profondeur pourraient cependant être sous-estimées. Une étude plus récente retrouve des taux de 50% dans plusieurs types de cancer (Cortés-Ciriano et al., 2018).

S'il est probable que la plupart des cellules qui connaissent un réarrangement chromosomique si important entre ensuite en apoptose, il est également possible qu'en affectant des gènes suppresseurs de tumeurs et/ou en activant des oncogènes le *chromothripsis* soit le moteur du processus tumoral (Rausch et al., 2012).

S'agissant des remaniements constitutionnels, il a été montré qu'un *chromothripsis* pouvait être identifié dans 5% des cas de remaniement chromosomique visible au caryotype (Chiang et al., 2012; Redin et al., 2017). Aucun remaniement de ce type n'a cependant été identifié dans les premières études en population générale (Sudmant et al., 2015; Hehir-Kwa et al., 2016) et un seul est présent dans la plus grande base de données de variants structuraux identifiés par séquençage en population générale (Collins et al., 2019) laissant supposer un possible biais de sélection.

Au total, nous avons identifié 53 *chromoanagenesis* constitutionnels dans la littérature (Kloosterman et al., 2011a, 2012; Ochalski et al., 2011; Auger et al., 2013; Gu et al., 2013; Nazaryan et al., 2014; Plaisancié et al., 2014; Zanardo et al., 2014; Fontana et al., 2014; Gamba et al., 2015; Genesio et al., 2015; Macera et al., 2015; McDermott et al., 2015; Bertelsen et al., 2015; Wang et al., 2015; Weckselblatt et al., 2015; de Pagter et al., 2015; Anderson et al., 2016; Masset et al., 2016; Suzuki et al., 2016; Burnside et al., 2016; Del Rey et al., 2016; Redin et al., 2017; Collins et al., 2017, 2019; Kurtas et al., 2018, 2019; Slamova et al., 2018; Eisfeldt et al., 2019). Si 10 cas de *chromoanagenesis* ont été rapportés transmis de la mère à son enfant (Kloosterman et al., 2012; Gu et al., 2013; Bertelsen et al., 2015; de Pagter et al., 2015; Weckselblatt et al., 2015; Kurtas et al., 2019), il n'existe pas à ce jour de cas décrits de

chromoanagenesis de novo affectant les chromosomes maternels. L'étude de l'origine parental du/des chromosome(s) affecté(s) a toujours mis en évidence une provenance paternelle (Kloosterman et al., 2012; Weckselblatt et al., 2015). Trois cas de remaniements hérités du père ont été rapportés (Bertelsen et al., 2015; Weckselblatt et al., 2015; Kurtas et al., 2019).

La distinction entre *chromothripsis* et *chromoanasyntesis* est difficile à la lecture de la littérature dans la mesure où certains remaniements n'ont pas été étudiés par séquençage (Ochalski et al., 2011; Auger et al., 2013; Zanardo et al., 2014; Burnside et al., 2016). De plus, certains auteurs ont préféré utiliser les termes génériques *chromoanagenesis* (Burnside et al., 2016) ou remaniement chromosomique complexe (Ochalski et al., 2011; Auger et al., 2013; Suzuki et al., 2016; Eisfeldt et al., 2019) quand d'autres ont désigné *chromothripsis* un très probable *chromoanasyntesis* (Wang et al., 2015; Collins et al., 2019) ou proposer une nouvelle classe de *chromoanagenesis* (Masset et al., 2016).

La très large majorité des patients décrits présentent une déficience intellectuelle de sévérité variable, le plus souvent syndromique (Fukami et al., 2017) en lien avec les régions affectées par une variation du nombre de copies et/ou la disruption d'un gène sensible à l'haploinsuffisance. Le démasquage d'un variant hétérozygote responsable d'une pathologie récessive est également possible (Auger et al., 2013). Sont maintenant également rapportées plusieurs observations de *chromoanagenesis* chez des sujets sans phénotype, identifiés suite au déséquilibre de ce remaniement dans leur descendance (van Binsbergen et al., 2012; Gu et al., 2013; de Pagter et al., 2015). Bertelsen et al. rapportent eux un enfant présentant une épilepsie syndromique et porteur d'un *chromothripsis* déséquilibré transmis à l'identique par la mère pour lequel un variant ponctuel de *SYNGAPI* explique complètement le phénotype. Les patientes porteuses de remaniements affectant un chromosome X semblent présenter un phénotype moins sévère, avec une dysfonction ovarienne isolée (Ochalski et al., 2011; Suzuki et al., 2016) probablement en lien avec l'inactivation de l'X remanié. Enfin, une seule observation de *chromothripsis* post-zygotique non en lien avec une tumeur a été rapportée chez une patiente ayant présenté un syndrome WHIM (Warts, Hypogammaglobulinemia, Immunodeficiency, Myelokathexis), déficit immunitaire lié à des mutations gain de fonction du gène *CXCR4* (McDermott et al., 2015). Dans la 4^{ème} décennie de vie, l'ensemble des signes cliniques de la pathologie ont disparu amenant la patiente en complète rémission. Cet étrange

scenario a pu être expliqué par la survenue d'un remaniement chromosomique complexe survenu dans une cellule souche hématopoïétique et ayant entraîné la délétion de la copie mutée de *CXCR4*. Ce type de correction par acquisition d'une anomalie chromosomique a été rarement décrit et reflète plus le désavantage procuré par la mutation que l'avantage obtenu par réarrangement chromosomique. A titre d'exemple, des monosomies ou des délétions segmentaires du bras long du chromosome 7 ont été observées chez des patients présentant un syndrome MIRAGE (#MIM : 617053), consécutifs à une mutation gain de fonction du gène *SAMD9* (Roucher-Boulez et al., 2019). Le même mécanisme est décrit dans les anémies de Fanconi (Gregory et al., 2001).

Le faible nombre de *chromoanagenesis* constitutionnels observés en période post-natal pourrait en partie résulter d'une certaine pression de sélection à la manière de ce qu'on observe pour les remaniements plus simples. Les aneuploïdies et plus globalement les déséquilibres génomiques font l'objet d'une importante contre-sélection avec un taux d'anomalies chromosomiques retrouvées proche de 60% dans les fausses couches spontanées précoces et inférieur à 1% en post-natal (Gardner and Amor, 2018). Aussi il est possible que les *chromoanagenesis* observés en postnatal constituent un sous-groupe relativement équilibré de l'ensemble des *chromoanagenesis* effectivement générés. Le taux d'anomalies chromosomiques est connu pour être particulièrement élevé dans les premières divisions cellulaires du zygote. Parmi elles, la proportion de remaniements compatibles avec un *chromoanagenesis* reste inconnue.

Pour confirmer cette hypothèse, compte tenu de la limite de résolution du caryotype en période prénatale sur du matériel biologique possiblement altéré, il faudrait mener une étude par séquençage de génome de produits de fausse couche. A ce jour, à l'exception d'une recherche de variant de structure équilibré sur blastomère (n = 1) (Voet et al., 2013), seuls des recherches de CNVs ont été réalisées (Qi et al., 2018).

Mécanisme de survenue

La survenue de façon concomitante de multiples cassures de l'ADN peut être le fait d'agents exogènes : radiations ionisantes (Morishita et al., 2016), radicaux libres, chimiothérapies. Seules les radiations ionisantes ont montré une capacité à générer des remaniements chromosomiques avec un regroupement des anomalies comme observé pour les *chromothripsis* et *chromoanasythesis*. Ceci peut être une conséquence de l'organisation physiologique en territoires chromosomiques dans le noyau interphasique. En effet, la description de *chromoanagenesis* impliquant plusieurs chromosomes rend très peu probable l'hypothèse d'une explosion chromosomique au moment de la division cellulaire. Les territoires chromosomiques interphasiques sont susceptibles d'être affectés individuellement par un « tir » de radiations ionisantes. L'intégration de séquences virales (HPV) a également été décrite comme générant une instabilité chromosomique et des événements complexes mais ne présentant pas les mêmes caractéristiques (Akagi et al., 2014).

En s'intéressant directement aux *chromoanagenesis* plusieurs hypothèses non exclusives se sont faites jour et plusieurs mécanismes cellulaires ont été investigués. La première hypothèse formulée fut celle d'une apoptose incomplète dans laquelle seule une partie de la chromatine est détruite mais où la cellule reste viable (Tubio and Estivill, 2011). La chromatine la plus accessible serait la plus susceptible à ces événements. Cette chromatine altérée fait ensuite l'objet d'une réparation défectueuse à l'origine du réarrangement observé (Ciavarra and Zacksenhaus, 2011). Cette hypothèse ne permet pas d'expliquer la concentration des points de cassure sur un seul ou un nombre limité de chromosomes.

Par la suite, le rôle de l'attrition télomérique a été discuté (Ernst et al., 2016). L'absence de « protection télomérique » peut engendrer la fusion de deux télomères. Ces phénomènes de fusion peuvent entrer dans un cycle de cassure-pont-fusion connu pour être responsables de remaniements chromosomiques complexes (Gisselsson et al., 2001). En fusionnant il peut se former soit un chromosome en anneau soit un chromosome dicentrique. Dans ce dernier cas, peuvent être observés des ponts chromatiniens entre deux cellules filles au moment de la cytodierèse (Maciejowski et al., 2015). Au moment de la rupture de l'enveloppe nucléaire, ce pont chromatiniens se retrouve exposé à l'action de l'exonucléase TREX1, qui va engendrer de

nombreuses cassures double brin du pont chromatinien. Mieux que le précédent modèle, celui-ci permettrait d'expliquer le regroupement de points de cassure dans un ou quelques points chauds, notamment à proximité des télomères. Cette hypothèse permet également d'expliquer l'excès de mutations ponctuelles parfois observé à proximité immédiate des points de cassure de *chromothripsis* tumoraux (Maciejowski et al., 2015).

Enfin, via différents modèles cellulaires présentant des défauts des points de contrôle de la mitose ou du fuseau mitotique, les aneuploïdies voire une hyperploïdie ont été associés à un risque accru d'instabilité chromosomique et de *chromothripsis* (Mardin et al., 2015).

Mais l'hypothèse qui a été le plus investiguée à ce jour implique la séquestration d'un (ou plusieurs) chromosome(s) au sein d'un micronoyau. En effet, cette hypothèse présente l'avantage majeur de pouvoir expliquer le regroupement des points de cassures sur un ou quelques chromosomes. L'observation fréquente de micronoyaux dans les cellules cancéreuses a contribué à crédibiliser cette hypothèse (Crasta et al., 2012; Hatch et al., 2013a). De plus l'observation de micronoyaux a été associée à l'instabilité génomique (Terradas et al., 2016).

L'apparition d'un micronoyau serait la conséquence d'un défaut de captation d'un centromère par le fuseau de division cellulaire (Crasta et al., 2012). En inactivant le centromère du chromosome Y de manière sélective il a pu être effectivement montré que celui-ci subissait un remaniement chromosomique complexe après avoir été séquestré dans un micronoyau (Ly et al., 2017). Ce micronoyau dispose comme le noyau principal d'une double couche membraneuse mais celle-ci ne s'assemble pas correctement et ne contient notamment pas la même densité de pores nucléaires. Incapable d'importer correctement le matériel nécessaire en son sein, il se retrouve déficient pour les machineries de réplication et de réparation de l'ADN (Hatch et al., 2013a).

La première possibilité est que la conjugaison de ces défauts entraîne par eux-mêmes la fragmentation du chromosome séquestré. C'est la structure du micronoyau qui serait en cause. Chaque cassure double brin va générer un nouveau fragment. L'ensemble des fragments peut ensuite être perdu si le micronoyau est dégradé dans le cytoplasme, et entraîner une aneuploïdie ou bien être réintégré dans une des cellules filles. Dans ce dernier cas, les fragments sont réassemblés pour donner un chromosome remanié.

Une deuxième explication envisagée est celle d'une condensation prématurée du chromosome séquestré dans le micronoyau, ce phénomène étant connu pour entraîner de nombreux points de cassure sur le chromosome concerné (Kürten and Obe, 1975). Récemment cette hypothèse a pu être appuyée expérimentalement. En parvenant à retarder l'entrée en phase M du noyau principal il a été possible de montrer que l'étape d'explosion chromosomique était liée à une asynchronie entre le micronoyau et le noyau principal (Pantelias et al., 2019). Lorsque l'entrée en phase M du noyau principal était retardé (phase G2 allongé) pour « attendre » le micronoyau il n'était pas observé de phénomène d'explosion chromosomique.

Normalement à l'entrée en mitose, la membrane du noyau se rompt et sous l'action de la cycline B1-CDK1 et de la phosphorylation des histones la chromatine se condense. Ici, au moment de l'entrée en phase M, la membrane du micronoyau se rompt avec celle du noyau principal alors même que le chromosome séquestré n'est pas entièrement répliqué. Aussi, plus le retard de réplication est important, plus il y a de fourches de réplifications encore actives et plus le chromosome sera fragmenté. En effet, le stress mécanique porté par le phénomène de condensation de la chromatine sur les fourches de réplication tardives suffit à les faire décrocher de l'ADN (Terzoudi et al., 2015). Le chromosome se retrouve fragmenté. En fonction des mécanismes de réparation qui seront mis en place, chaque extrémité de fragment pourra être détectée comme un point de cassure du remaniement sur le chromosome finalement réassemblé.

S'agissant des mécanismes de réparation il est fort probable que la recombinaison homologue, mécanisme le plus fidèle de réparation soit saturé par le nombre de points de cassure à prendre en charge. Il a pu être montré qu'au-delà de 20 cassures double brin, les mécanismes de *non homologous end joining* (classique et/ou alternatif) devenaient prédominants (Gudjonsson et al., 2012). Bien sûr, la présence d'éléments répétés au niveau des points de cassure va pouvoir favoriser l'emploi de mécanismes impliquant des microhomologies et si un certain enrichissement en éléments répétés au niveau des points de cassure a été évoqué il ne semble cependant pas être la règle (Kloosterman et al., 2011b; Redin et al., 2017). Enfin, le facteur prédominant est probablement la phase du cycle cellulaire au cours de laquelle a lieu cette réparation. Même si un micronoyau peut être maintenu pendant plusieurs cycles cellulaires il semble que l'explosion ait lieu en fin de phase G2 au moment où le noyau principal est prêt pour entrer en phase M. Si les mécanismes de *non homologous end joining* sont mobilisables

tout au long du cycle cellulaire, les mécanismes répliatifs ne sont eux actifs qu'en phase S ou G2.

En impliquant la condensation prématurée d'un chromosome, séquestré ou non dans un micronoyau, ce modèle présente également l'avantage de pouvoir expliquer l'origine paternelle des chromosomes concernés (Kloosterman et al., 2012; Weckselblatt et al., 2015). En effet, à l'inverse des aneuploïdies, le plus souvent consécutives à une non-disjonction méiotique maternelle, les *chromoanagenesis de novo* concernent préférentiellement les chromosomes paternels (Pellestor et al., 2014). La multiplicité des divisions cellulaires au cours de la spermatogénèse notamment entre le stade de spermatogonie et la méiose, fait s'accroître le risque d'erreur poussant un chromosome dans un micronoyau. La réorganisation majeure du repliement chromosomique des spermatides lors de la transition des histones vers les protamines peut aussi contribuer à cette observation (Pellestor et al., 2014).

Aussi, deux questions principales apparaissent :

- Les *chromoanagenesis* tumoraux et constitutionnels sont-ils issus des mêmes mécanismes ?
- Remaniements simples et remaniements complexes sont-ils les deux extrêmes d'un même continuum biologique simplement séparés par un seuil arbitraire de nombre de points de cassure ? Ou bien sont-ils le résultat de mécanismes biologiques distincts ?

I – 4/ D- Insertion d'éléments mobiles

Les éléments mobiles du génome occupent plus de la moitié de la séquence du génome humain (Lander et al., 2001). La mobilisation de ces éléments est à l'origine d'insertions de séquences, pour partie au moins, copie de la séquence d'origine. En ce sens ils peuvent être considérés comme des gains du nombre de copie dont le contenu serait standardisé, à la manière d'un copié-collé, à l'inverse des CNVs habituellement observés dont le contenu est plus aléatoire. De par leur similitude et leur grand nombre de copies ces éléments peuvent être difficiles à assembler pour construire un génome de référence et ont pu ainsi être négligés par le passé. Les techniques de séquençage à haut débit *short-read* mais surtout *long-read* vont permettre de

mieux caractériser ces événements et ainsi connaître leur potentiel pathogène. On distingue différentes familles d'éléments mobiles.

Chez l'Homme, seuls les éléments LINE-1 (*long interspersed element 1*) ou L1 sont autonomes pour se mobiliser et sont nécessaires à la mobilisation des autres familles d'éléments mobiles. Ces éléments ont une longueur de 6 kb chacun et occupent à eux seuls 17% du génome humain (Lander et al., 2001). Seule une centaine d'éléments L1 sont considérés intacts et donc mobilisables, quelques-uns étant hautement actifs (Brouha et al., 2003). Un L1 code un promoteur d'ARN polymérase II. Lorsqu'il est transcrit, il produit un ARN qui code l'ARN binding protein ORF1p (Martin et al., 2005) et une protéine ORF2p aux activités endonucléase et reverse transcriptase (Mathias et al., 1991; Feng et al., 1996). Ces deux éléments sont la clé du mécanisme de « Target Primed Reverse Transcription » (TPRT) au cours duquel l'ARN de l'élément L1 est rétrotranscrit en ADNc avant que celui-ci ne soit inséré dans l'ADN génomique. C'est cette protéine ORF2p qui permet aussi la mobilisation des autres familles d'éléments mobiles du génome.

Les séquences Alu occupent 11% de la séquence du génome humain. Leur longueur n'étant que de 300 bases, leur nombre est beaucoup plus important que celui des L1, 1,07 million de copie. Cette famille d'éléments mobiles est celle qui a le plus proliféré chez les primates (Dewannieux et al., 2003). Ils sont composés de deux monomères distincts de structure proche réunis par une séquence riche en adénine et suivis d'une queue poly-A. Comme les éléments Alu, les SVA (SINE-VNTR-Alu) sont des *short interspersed elements*. Ils sont composés de séquences Alu et d'un nombre variable de répétitions en tandem. Leur taille moyenne est de 2 kb. Comme les éléments Alu ils sont mobilisés par l'intermédiaire de ORF2p (Hancks et al., 2011).

Enfin, la protéine ORF2p peut également être responsable de la rétrotranscription / insertion d'ARNm dans l'ADN génomique. Ceci entraîne la formation de rétrocopies (Esnault et al., 2000), aussi appelés pseudogènes « processés ». Cette dernière famille est la moins abondante et la plus diverse. En effet, si les gènes les plus fortement exprimés sont potentiellement plus sujets à ces événements, chaque transcrit produit est susceptible d'être « rétrocopié ».

On estime la fréquence d'une mobilisation d'Alu à 1 méiose sur 21 (Xing et al., 2009), 1 sur 95-270 pour les éléments L1 (Xing et al., 2009; Ewing and Kazazian, 2010), 1 sur 916 pour les SVA (Xing et al., 2009) et seulement 1 sur 6256 pour les ARNm (Ewing et al., 2013).

Les rétrocopies sont donc à la fois les événements les plus rares et les moins connus dans le domaine de la génétique médicale. Une explication est celle du caractère mono ou oligogénique des investigations menées en diagnostic jusqu'à l'avènement du séquençage haut débit il y a quelques années. En effet, en analysant uniquement les séquences exoniques d'un gène la probabilité d'identifier ce type de remaniement est très faible (1 mobilisation sur 6 256 méioses et environ 1,5% du génome code environ 20 000 gènes soit approximativement 1×10^{-10} probabilité d'observer un événement *de novo* en analysant un seul gène). C'est l'arrivée des techniques de séquençage massif en parallèle, en augmentant le nombre de gènes analysés (panel, exome, génome) qui rend désormais possible ces observations.

I – 5/ Conséquences des variants de structure chromosomique constitutionnels

Les conséquences phénotypiques des remaniements de structure sont extrêmement variables. Elles dépendent avant tout du type de remaniement et des gènes concernés.

I – 5/ A- Variants affectant la séquence codante

Déséquilibre génomique

Les anomalies du nombre de chromosomes (aneuploïdies) homogènes sont responsables de syndrome bien définis. La trisomie 21 (syndrome de Down) est la plus fréquente des aneuploïdies autosomiques viables. Elle est responsable d'une déficience intellectuelle de sévérité variable associée à une dysmorphie faciale avec un visage plat, arrondi, des fentes palpébrales obliques en haut et en dehors. Des anomalies cardiaques (canal atrio-ventriculaire) et thyroïdiennes sont fréquemment présentes. Le risque de leucémie aiguë est augmenté. Les trisomies 13 (syndrome de Patau) et 18 (syndrome d'Edwards) sont elles aussi viables mais responsables de syndromes malformatifs de mauvais pronostic. La monosomie X est responsable d'un syndrome de Turner associant une aménorrhée primaire (infertilité) avec une petite taille et quelques éléments morphologiques particuliers mais habituellement sans déficience intellectuelle. Le syndrome de Klinefelter est associé à une formule XXY et associe des anomalies pubertaires à une azoospermie.

L'appellation variations du nombre de copies recouvre l'ensemble des anomalies déséquilibrées de plus de 1 kb. A la résolution moyenne de 40 kb, une analyse chromosomique sur puce à ADN met en évidence entre 10 et 15 variants par individu. L'interprétation de ces variants repose sur un faisceau d'arguments. Les gènes sensibles à l'haploinsuffisance et connus en pathologie font l'objet du travail de corrélation génotype/phénotype. Le démasquage d'un variant hémizygoté par une délétion d'un gène responsable d'une pathologie autosomique récessive est également possible. Les gènes étant globalement plus sensibles à l'haploinsuffisance qu'à la présence d'une copie supplémentaire les pertes sont habituellement

considérées plus facilement pathogènes que les gains équivalents. L'ensemble de ces éléments permet habituellement de conclure entre le caractère (probablement) pathogène ou (probablement) bénin d'un CNV mais des variants de signification indéterminée persistent.

Au total, les variations du nombre de copie sont considérées responsables d'environ 15% des cas de déficience intellectuelle (Miller et al., 2010), et de 6% des malformations fœtales (Malan et al., 2016; Pons et al., 2017), hors aneuploïdies.

Les régions soumises à empreinte parentale sont un cas particulier. Dans ces régions, seul un allèle est exprimé en fonction de son origine parentale. Une contribution biparentale est donc nécessaire à l'expression adéquate de ces gènes. La délétion d'un allèle aura les mêmes conséquences que la présence de deux allèles d'un même parent (disomie uniparentale). Le locus 15q11q13 est associé aux syndromes de Prader-Willi et d'Angelman en fonction de l'allèle non fonctionnel (paternel pour le Prader-Willi et maternel pour l'Angelman). Outre les délétions récurrentes de tout le locus médiées par NAHR, la perte du locus en charge de la mise en place de l'empreinte est également décrite (Glenn et al., 1996). Enfin, les gènes majeurs du locus sont *SNRPN* pour l'allèle paternel et *UBE3A* pour l'allèle maternel. En fonction de l'allèle touché des variations nucléotidiques ponctuels de ces gènes pourraient être pathogènes. Un cas de translocation réciproque apparemment équilibrée $t(2;15)(q37.2;q11.2)$ dont le point de cassure est situé entre le centre d'empreinte et le gène *SNRPN* a également été décrit chez un patient présentant un syndrome de Prader-Willi atypique (Conroy et al., 1997).

Remaniements équilibrés

Les anomalies équilibrées ne sont, de leur côté, habituellement pas associées à un phénotype anormal. En revanche, elles vont entraîner des anomalies d'appariement des chromosomes sur la plaque équatoriale lors de la méiose. Ces structures anormales peuvent être résolues par différentes ségrégations entre les cellules filles dont certaines peuvent générer un déséquilibre génomique (Figure 28 et Figure 29). Ainsi elles peuvent contribuer à une infertilité ou à la survenue de fausses couches à répétition et doivent être systématiquement recherchées chez les

parents d'un enfant présentant un remaniement déséquilibré. Cette information permettra de connaître le risque de récurrence et donc de donner un conseil génétique fiable.

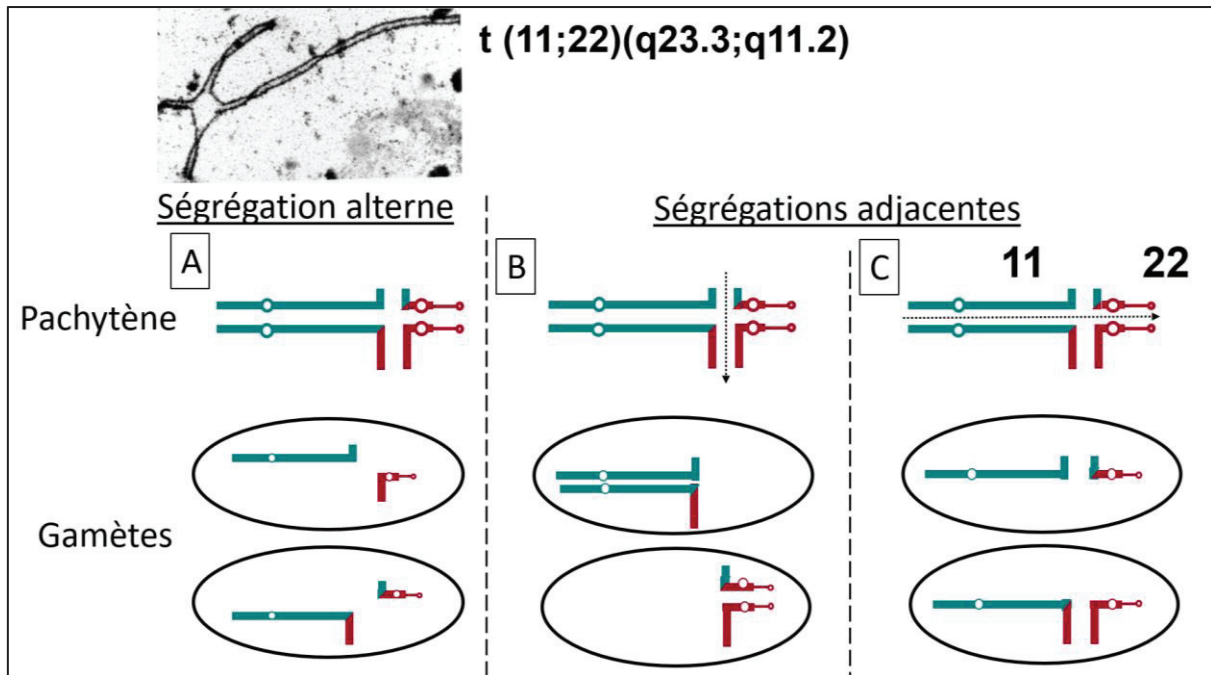


Figure 28 : Risques de déséquilibres associés à une translocation réciproque en fonction du type de ségrégation chromosomique du tétravalent de pachytène (image de microscopie électronique). Seules les ségrégation 2:2 sont présentées. La ségrégation alterne n'entraîne pas de déséquilibre (A). L'utilisation de gamètes issues de ségrégations adjacentes 1 (B) ou 2 (C) vont entraîner des délétions duplications segmentaires. Les ségrégations 3 :1 et 4 :0 sont plus rares mais également possibles .D'après M. Prieur.

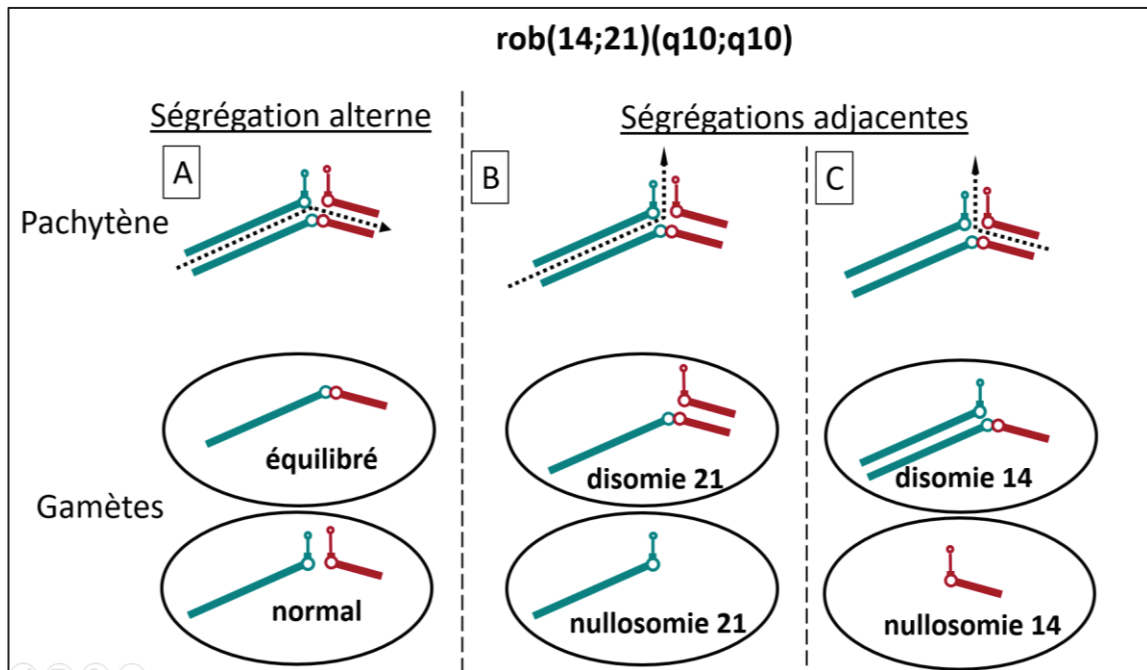


Figure 29 : Schéma des risques de déséquilibres liés à une translocation robertsonienne. D'après M. Prieur

L'association entre une anomalie chromosomique apparemment équilibrée vue au caryotype et un phénotype anormal à 2 ans est observée dans 6,1% des cas de translocation apparemment équilibrée *de novo* et 9,4% des cas d'inversions *de novo* (Warburton, 1991). Plus récemment un suivi à plus long terme des patients porteurs de ce type de remaniements a montré la présence d'un phénotype anormal dans 27% des cas, cette discordance étant principalement dû à la détection plus tardive d'anomalies du neuro-développement (Halgren et al., 2018). Pour 40% de ces cas, il existe en réalité un déséquilibre chromosomique expliquant le phénotype soit au niveau d'un des points de cassure du remaniement vu au caryotype soit ailleurs dans le génome probablement par association fortuite (De Gregori et al., 2007; Schluth-Bolard et al., 2009). Pour les autres cas, une association fortuite reste possible mais le séquençage de génome entier permet d'obtenir la résolution pour investiguer l'hypothèse d'une pathologie du point de cassure. A la manière d'une délétion, un remaniement qui viendra interrompre un gène sensible à l'haploinsuffisance sera responsable du phénotype en rapport. Cette disruption peut également être associée à une autre variation pathogène sur l'autre allèle et causer une pathologie de transmission récessive. Exceptionnellement, cette deuxième variation peut également être un variant de structure. L'observation d'une famille consanguine dans laquelle une translocation réciproque homozygote est associée à une polymicrogyrie/microcéphalie est rapportée (Baala et al., 2007).

En interrompant deux gènes, le remaniement peut également être à l'origine d'un gène de fusion aux conséquences difficiles à prédire. La fusion *bcr/abl* du chromosome dit Philadelphie (translocation entre un chromosome 9 et un chromosome 22) responsable de leucémie myéloïde chronique a été le premier exemple décrit d'une longue liste d'exemple en oncologie (Nowell et al., 1960; ROWLEY, 1973). A ce jour, un seul exemple lié à un remaniement constitutionnel a été décrit. Il s'agit de la fusion des gènes *PKT2* et *THOC2* suite à une translocation entre un chromosome X et un chromosome 8 chez une patiente présentant une ataxie congénitale (Di Gregorio et al., 2013).

I – 5/ B- Effet de position

Sans interrompre d'élément génique, un remaniement chromosomique peut également avoir un « effet de position ». Il s'agit des conséquences du déplacement d'un élément génique à distance de son « paysage » chromatinien habituel. Le gène peut ainsi être déplacé d'un environnement euchromatinien favorable à sa transcription à un environnement hétérochromatinien, condensé et donc défavorable à sa transcription (*position effect variegation*) (Tartof et al., 1984). Plus directement, le gène peut être déconnecté de ses éléments régulateurs physiologiques (promoteurs, *enhancers*) par son déplacement à longue distance ou par l'interposition d'un élément insulateur réduisant les interactions. Il peut tout aussi bien s'agir d'un remaniement équilibré que d'un remaniement déséquilibré affectant un ou plusieurs éléments régulateurs. Trois éléments participent à la régulation de l'expression d'un gène : une région promotrice, un système *enhancer/silencer* et un environnement chromatinien plus ou moins favorable (ouvert/fermé) à l'action du couple *enhancer/promoteur* (Kleinjan and van Heyningen, 1998). C'est en effet l'*enhancer* qui va permettre l'activation du promoteur et l'initiation de la transcription du gène. Si un point de cassure survenait entre les 3 éléments (*enhancer/silencer* ; promoteur ; gène) l'expression du gène peut être dérégulée. Quelques loci bien étudiés ont fait l'objet de la majorité des descriptions et investigations autour de ce mécanisme. C'est le cas du désert génique situé en amont de *SOX9* sur le chromosome 17 où en fonction de la position du remaniement va pouvoir être observé une réversion sexuelle, une dysplasie campomélique ou un phénotype normal (Velagaleti et al., 2005). L'étude des TADs de ce locus a apporté un

niveau de compréhension supplémentaire dans l'effet de position. En effet, des duplications de régions du désert génique étaient responsables de phénotypes alors qu'elles maintenaient la distance linéaire entre le gène, son promoteur et une des deux copies de l'*enhancer* connu sans affecter le niveau d'expression du gène en question. L'analyse de conformation de la chromatine a permis de montrer la formation de néo-TAD et l'adoption de l'*enhancer* dupliqué par un autre gène (Lupiáñez et al., 2015; Franke et al., 2016) (Figure 30-2B et 2C). La multiplicité des possibilités est présentée dans la Figure 30 en fonction du type de variants de structure et de leur contenu (présence d'un élément insulateur ou non). En l'absence d'élément insulateur les manifestations phénotypiques observées résultent d'un simple « effet dose » par délétion ou duplication d'un élément régulateur (Figure 30-1). Les conséquences sont beaucoup plus difficiles à prédire lorsqu'un élément insulateur est concerné par le variant de structure (fusion de TAD, néoTAD) (Figure 30-2 et Figure 30-3).

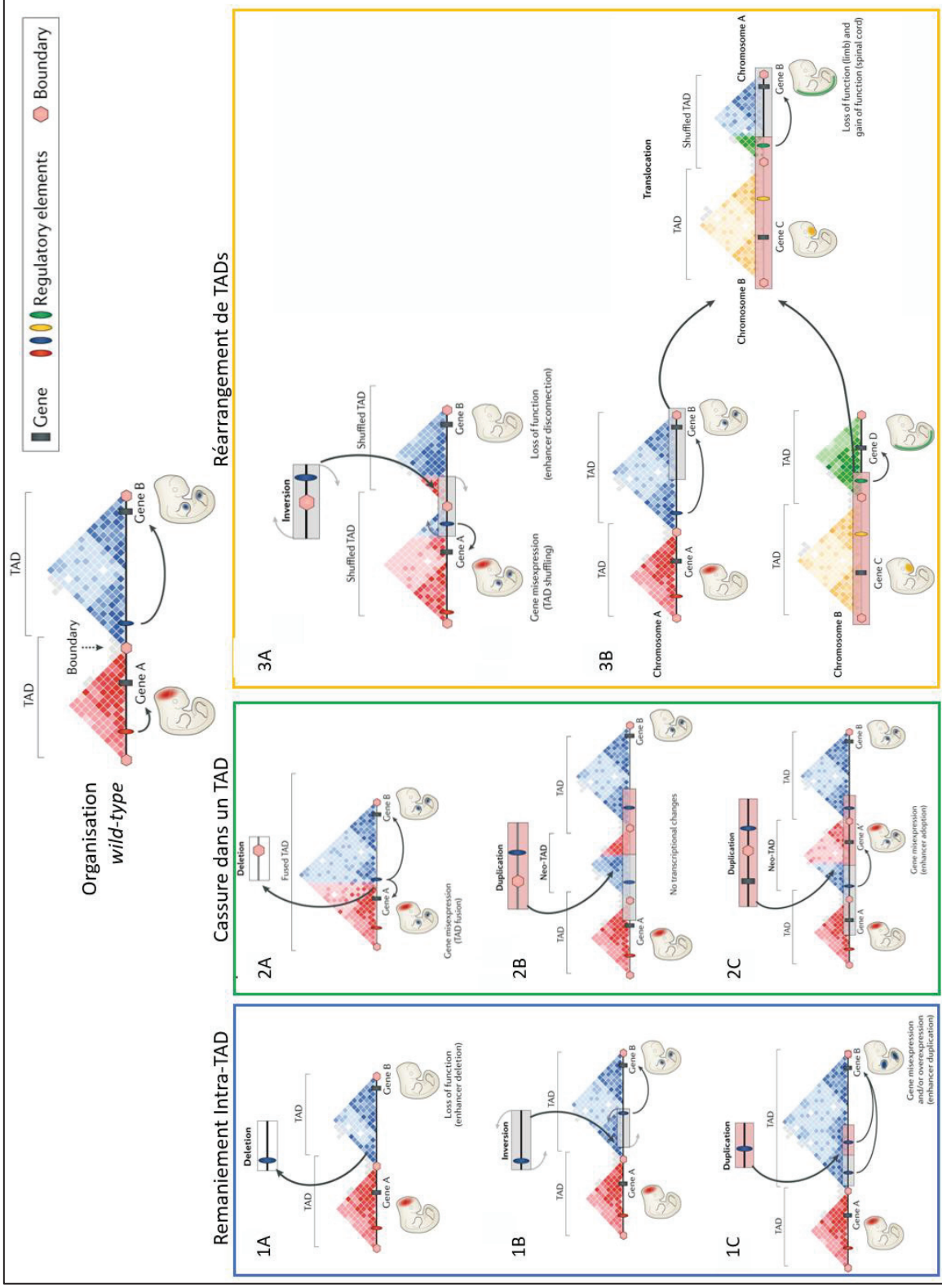


Figure 30 : Synthèse des différentes possibilités d'effet de position en lien avec l'organisation de la chromatine en TADs. Adapté de Spielmann et al. 2018.

Ainsi il semble intéressant de pouvoir distinguer les remaniements intra-TAD, pour lesquels on peut observer des modifications du niveau d'expression d'un gène en fonction de la perte ou de la duplication d'un *enhancer* physiologique ; des remaniements modifiant la conformation des TADs aux effets plus difficiles à prédire car impliquant des interactions inédites.

Les loci pour lesquels de tels mécanismes ont été associés à une pathologie humaine sont encore peu nombreux et concernent généralement des phénotypes très reconnaissables, potentiellement plus simples à étudier via des modèles animaux (Lupiáñez et al., 2015; Will et al., 2017; Kragesteen et al., 2018; Spielmann et al., 2018).

En effet, si les éléments régulateurs d'un gène sont regroupés au sein de TADs et même si ceux-ci sont relativement stables il n'en existe pas moins une certaine variabilité entre les différents types cellulaires, expliquant notamment l'expression des gènes dans un tissu et pas dans un autre. Ainsi l'étude du niveau de régulation d'un gène et de la structure de la chromatine devrait idéalement être réalisée dans un tissu en rapport avec le phénotype du patient. Middelkamp *et al.* ont démontré l'intérêt de cette approche chez un patient présentant une déficience intellectuelle syndromique et porteur d'un *chromothripsis* (Middelkamp et al., 2017). Ils ont utilisé des lymphocytes du patients transformés une première fois en cellule pluripotente induite (iPS) et une seconde fois en neurones pour démontrer la perte de contact entre le gène *TWIST1* et plusieurs de ces *enhancers* non observable dans le sang.

Récemment une approche combinant la digitalisation des phénotypes de patients et les données d'interactions chromatiniennes obtenues dans un grand nombre de tissus a été proposé (Middelkamp et al., 2019). Un large éventail de données a été utilisé : Hi-C, TAD, promoter capture Hi-C (PCHiC), site d'hypersensibilité à la DNase hypersensitivity (DHS), données RNA et ChIP-seq. Dans 25% des cas cette approche a permis de considérer un gène comme affecté par un effet de position et probablement responsable du phénotype. Il est cependant difficile de connaître la part de diagnostic permis par l'outil clinique et l'implication régulière de nouveaux gènes en pathologie humaine d'une part et par les données chromatiniennes d'autre part.

I – 5/ C- Pathogénicité des éléments transposables

Le potentiel pathogène de ces événements de mobilisation dépend en grande partie du site d'insertion. Plus de 24 000 sites d'insertion d'éléments mobiles sont considérés comme polymorphiques, représentant 24 % de l'ensemble des variants de structure connus et 36 % de ces événements ont une fréquence supérieure à 1 %. En dehors de ces polymorphismes on distingue différents mécanismes par lesquels les éléments mobiles peuvent être pathogènes.

La situation la plus claire est celle de l'insertion directement dans une séquence codante. C'était le cas du premier événement L1 pathogène identifié directement dans la séquence du gène *F8* chez un patient hémophile A (Kazazian et al., 1988). Depuis de nombreux autres cas de pathologies monogéniques ont été concernées par ces insertions aboutissant à une protéine tronquée (codon *stop* prématuré directement ou décalage du cadre de lecture aboutissant à un codon *stop*) (Taşkesen et al., 2012; Nakamura et al., 2015). Outre la disruption de la séquence codante, l'élément inséré peut respecter le cadre de lecture, être transcrit au sein du produit normal de transcription mais perturber les mécanismes d'épissage. Ceci a été décrit pour l'insertion d'un élément Alu dans l'exon 11 du gène *CLCN5* entraînant un saut de ce même exon par destruction du site *enhancer* d'épissage à l'origine d'une maladie de Dent (Claverie-Martin et al., 2005). De même, les insertions introniques, surtout lorsqu'elles sont proches des sites d'épissages, peuvent avoir le même effet. En plus de détruire des sites d'épissage, les éléments mobiles peuvent également en faire apparaître. Ceci va entraîner l'*exonisation* de la séquence comprise entre le nouveau site d'épissage et l'exon voisin si un seul site est créé ou l'exonisation de l'élément mobile lui-même s'il crée un couple site donneur-site accepteur d'épissage (Wallace et al., 1991; Gallus et al., 2010; Rodríguez-Martín et al., 2016).

Indépendamment du site d'insertion le phénomène de mobilisation par TPRT peut s'accompagner de délétions interstitielles potentiellement pathogènes. Deux observations distinctes d'insertion d'un SVA dans le gène *SUZ12P* associés à des délétions du gène voisin *NFI* ont été rapportés dans des cas de neurofibromatose (Vogt et al., 2014). Cette situation est à distinguer de cas où l'insertion de l'élément mobile se fait dans le cadre de la réparation d'une cassure double brin. Dans ce cas l'insertion se fait en dehors du mécanisme de TPRT directement dans un point de cassure préexistant (Morrish et al., 2002; Morisada et al., 2010).

S'il s'agit là d'événements privés, identifiés dans des familles uniques, certains éléments transposables présentent également un effet fondateur responsable d'une fréquence accrue d'une pathologie monogénique dans un isolat génétique. C'est par exemple le cas de l'insertion

d'un SVA dans le gène *FKTN* responsable d'une dystrophie musculaire de transmission autosomique récessive (#MIM253800) au Japon (Kobayashi et al., 1998), d'un SVA dans un intron du gène *TAFI* responsable de dystonie parkinsonienne aux Philippines (Bragg et al., 2017; Aneichyk et al., 2018).

En dehors des maladies monogéniques, certains éléments transposables ont également été rapportés comme facteur de prédisposition à des maladies communes via leur effet modulateur sur l'expression d'un gène (eQTL : expression Quantitative Trait Loci) (Payer et al., 2017; Wang et al., 2017). Ce rôle de modulation de l'expression pourrait dans certains cas constituer un avantage sélectif capable d'expliquer la distribution biaisée des éléments Alu vers les régions riches en gènes. Les études GWAS habituellement utilisées pour identifier ce type de prédisposition ne peuvent identifier ces variations de structure et il est probable qu'avec l'utilisation massive du séquençage de génome et l'amélioration de la détection de ces événements l'importance des éléments transposables soit revue à la hausse.

De par leur structure et leur nombre les éléments Alu présentent certaines spécificités. Par la présence de nombreux îlots CpG dans leur séquence ils sont un site majeur de méthylation. On estime que les séquences Alu représentent 30% des sites de méthylation du génome. Des modifications de ce type peuvent affecter le niveau d'expression d'un gène, mais ce mécanisme reste difficile à impliquer en pathologie et aucun exemple n'existe à ce jour.

Parallèlement, les éléments Alu possèdent une forte capacité d'exonisation lorsqu'elles sont présentes sur le brin antisens. Approximativement 5% des exons alternatifs dans le génome résultent d'un site d'épissage lié à un élément Alu (Gal-Mark et al., 2008). Au cours de l'évolution ils auraient contribué à la diversification du protéome mais peuvent aussi à l'échelle individuelle altérer un transcrit et entraîner une pathologie.

Au niveau de l'ADN génomique, l'appariement d'éléments Alu non alléliques peut entraîner l'apparition d'une variation du nombre de copie (délétion ou duplication en tandem) par un mécanisme de recombinaison homologue non allélique (*NAHR*). En effet, du fait du mécanisme de copié-collé les séquences ont un très fort niveau d'homologie entre elles et leur nombre très important (donc leur grande proximité) multiplie le risque de recombinaisons délétères. Ce mécanisme est également possible entre deux éléments d'une autre famille (L1 principalement), comme c'est le cas dans les inversions récurrentes du gène *F8* responsables d'hémophilie (Antonarakis et al., 1995; Startek et al., 2015). Il s'agit ici d'un phénomène « passif » sur le

plan de la mobilité génomique, et plutôt de la conséquence de millions d'années d'accumulation de nouveaux éléments mobiles.

Ceux-ci sont tellement nombreux qu'il n'est d'ailleurs pas exceptionnel d'en identifier deux au sein d'un même transcrit. S'ils sont orientés en sens inverse, ils sont théoriquement en mesure de s'hybrider ensemble et ainsi de créer un ARN double brin. Là encore par un mécanisme « passif », cette structure va perturber le mécanisme d'épissage et pouvoir former un ARN circulaire fermé par les deux éléments Alu d'une part, et un transcrit anormal d'autre part amputé des exons compris entre les deux séquences Alu (Jeck et al., 2013). Cette même structure peut également être à l'origine du recrutement de la protéine ADAR et à l'édition des bases adénine en inosine (Athanasiadis et al., 2004). A ce jour, ces deux mécanismes n'ont pas été décrits en pathologie.

II - Résultats

II – 1/ Article 1 : Severe hemophilia A caused by an unbalanced chromosomal rearrangement identified using nanopore sequencing

Le première question posée en introduction de ce travail était :

- **Face à une maladie monogénique, en l'absence de variant nucléotidique identifié, comment mettre en évidence un remaniement de structure pathogène ?**

Nous prenons ici le cas de l'hémophilie A, qui a été, comme nous l'avons vu précédemment, une des pathologies clés pour l'identification de nouveaux types de variants de structure pathogènes (Kazazian et al., 1988; Antonarakis et al., 1995) possiblement problématique pour le séquençage *short-read* comme discuté en introduction.

Ce qui était connu :

L'hémophilie est une maladie récessive liée à l'X responsable de défauts de la voie intrinsèque de la coagulation. On distingue l'hémophilie A (#MIM 306700) de l'hémophilie B (#MIM 306900) respectivement associées à des déficits en facteur VIII et IX de la coagulation. La mesure de l'activité de ces deux facteurs est réalisée en routine dans les laboratoires d'hémostase après la découverte d'un temps de céphaline activée anormalement long.

Une fois le déficit identifié, après séquençage des exons et des jonctions introns/exons et recherche par Southern-blot des inversions récurrentes médiées par des éléments L1 au locus, 2% des patients restent sans variation génétique identifiée (Swystun and James, 2017). Dès lors, les deux principales hypothèses sont celles d'un variant intronique profond entraînant par exemple un défaut d'épissage d'une part, et un variant de structure avec un point de cassure intragénique.

Dans le cas présent, le patient étudié présentait une hémophilie A sévère. L'étude de l'ARN et l'amplification de l'ensemble du gène (exons et introns) en plusieurs produits de long-range PCR ont été initiées de front. Étonnement, alors qu'il existe bien un transcrit du gène *F8* non dégradé, donc disposant d'une polyadénylation, celui-ci est anormal, dépourvu du dernier exon (26). Sur l'ADN, il n'a pas été possible d'amplifier une région de l'intron 25 ; le tout faisant fortement suspecter un point de cassure chromosomique sans partenaire identifié. Du fait de la présence d'un élément L1 il n'a pas été possible de restreindre la position de ce point de cassure

à moins de 6 kilobases. En revenant à l'ARN, une technique de 3'RACE PCR a été employée en espérant amplifier ce transcrit anormal depuis la queue polyA. Malheureusement, l'amplification n'a pas été spécifique et donc inaccessible au séquençage Sanger.

Ce que la publication apporte :

Nous avons pris l'option d'utiliser le séquençage par nanopore sur les produits de 3'RACE PCR. Dans ce cas, cette solution présente un double intérêt : celui du séquençage de molécules uniques, réglant la question de la non-spécificité d'une part ; et la capacité à générer de longues séquences d'autre part pour ce remaniement inconnu. Après caractérisation du transcrit anormal nous avons mis en évidence une jonction entre l'intron 25 et une région proximale du bras long du chromosome X laissant penser à une inversion chromosomique. Face à l'impossibilité d'amplifier l'autre fragment de jonction théorique, nous avons dû revoir notre hypothèse. Après étude en qPCR et CGH-array il s'est avéré qu'un fragment d'environ 3Mb d'une région proximale du bras court de l'X est en fait dupliquée, inversée et insérée dans l'intron 25 du gène *F8*. Par hasard, il s'avère que l'extrémité de la séquence qui vient désormais à la suite de l'exon 25 est porteuse d'un signal de polyadénylation. Le transcrit observé est donc un transcrit de fusion issu de ce remaniement chromosomique. Ce résultat a pu être confirmé par étude en FISH.

Outre le caractère exceptionnel de ce remaniement chromosomique à l'origine d'un transcrit de fusion constitutionnel, il s'agit du premier remaniement directement détecté par séquençage nanopore rapporté. Ce travail est présenté dans la publication qui suit (Journal of Thrombosis and Haemostasis, impact factor = 4,662).

Received: 14 March 2019 | Accepted: 18 April 2019

DOI: 10.1111/jth.14460

BRIEF REPORT



Severe hemophilia A caused by an unbalanced chromosomal rearrangement identified using nanopore sequencing

Nicolas Chatron^{1,2} | Caroline Schluth-Bolard^{1,2} | Mathilde Frétigny³ | Audrey Labalme¹ | Gaëlle Vilchez⁴ | Sabine-Marie Castet⁵ | Claude Négrier^{3,6} | Damien Sanlaville^{1,2} | Christine Vinciguerra^{3,6} | Yohann Jourdy^{3,6} 

¹Service de génétique, Groupe Hospitalier Est, Hospices Civils de Lyon, Bron, France

²CRNL, équipe GENDEV INSERM

U1028, CNRS UMR5292, Université Claude Bernard, Lyon, France

³Groupe Hospitalier Est, Hospices Civils de Lyon, Bron, France

⁴Groupe Hospitalier Est, Cellule bioinformatique de la plateforme de séquençage NGS du CHU de Lyon, Bron, France

⁵Centre de ressources et compétences-maladies hémorragiques constitutionnelles, Hôpital Universitaire de Bordeaux, Bordeaux, France

⁶EA 4609 Hémostase et cancer, Université Claude Bernard, Lyon, France

Correspondence

Yohann Jourdy, EA 4609 Hémostase et cancer, Université Claude Bernard Lyon 1, 7 rue Guillaume Paradin, Lyon, France. Email: yohann.jourdy@chu-lyon.fr

Abstract

Background: No *F8* genetic abnormality is detected in about 2% of severe hemophilia A patients using conventional genetic approaches. In these patients, deep intronic variation or *F8* disrupting genomic rearrangement could be causal.

Objective: To characterize, in a genetically unresolved severe hemophilia A patient, a new Xq28 rearrangement disrupting *F8* using comprehensive molecular techniques including nanopore sequencing.

Results: Long-range polymerase chain reaction (PCR) performed throughout *F8* identified a nonamplifiable region in intron 25 indicating the presence of a genomic rearrangement. *F8* messenger ribonucleic acid (mRNA) analysis including 3' rapid amplification of complementary deoxyribonucleic acid (cDNA) ends and nanopore sequencing found the presence of a *F8* fusion transcript in which *F8* exon 26 was replaced by a 742-bp pseudoexon corresponding to a noncoding region located at the beginning of the long arm of chromosome X (Xq12; chrX: 66 310 352-66 311 093, GRCh37/hg19). Cytogenetic microarray analysis found the presence of a Xq11.1q12 gain of 3.8 Mb. The PCR amplification of junction fragments and fluorescent in situ hybridization (FISH) analysis found that the Xq11q12 duplicated region was inserted in the *F8* intron 25 genomic region.

Conclusion: We characterized a novel genomic rearrangement in which a 3.8-Mb Xq11.1q12 gain inserted in the *F8* intron 25 led to an aberrant fusion transcript in a patient with severe hemophilia A (HA), using comprehensive molecular techniques. This study highlights the value of single-molecule long-read sequencing technologies for molecular diagnosis of HA especially when conventional genetic approaches have failed.

KEYWORDS

chromosome aberrations, duplication, *F8*, gene fusion, hemophilia A, nanopore sequencing

1 | INTRODUCTION

Hemophilia A (HA; OMIM #306700; <http://www.omim.org/>) is an X-linked recessive bleeding disorder caused by *F8* (OMIM #300841; <http://www.omim.org/>) genetic variations resulting in coagulation factor VIII (FVIII) deficiency. The clinical bleeding phenotype is defined on the basis of residual plasma FVIII activity (FVIII:C) and classified as severe (FVIII:C < 1 IU dL⁻¹), moderate (1 IU dL⁻¹ ≤ FVIII:C < 5 IU dL⁻¹), or mild HA (5 IU dL⁻¹ ≤ FVIII:C < 40 IU dL⁻¹).¹ *F8* spans 186 kb of the long arm of chromosome X (Xq28) and includes 26 exons.

Hemophilia A is caused by a large number of genetic abnormalities affecting *F8*, and about half of the severe HA is caused by two large nonallelic homologous recombinations disrupting *F8* known as intron 22 inversion (Inv22) and intron 1 inversion (Inv1).^{2,3} Nevertheless, conventional genetic approaches including Inv22 and Inv1 detection, direct sequencing of *F8* coding regions, and screening for copy number variation for *F8* exons still fail in approximately 2% of severe HA cases.^{4,5} Genetic counseling is incomplete for these patients, hindering genetic family investigation and prenatal diagnosis. In these genetically unresolved patients, deep intronic variation or genomic rearrangement disrupting the *F8* could be causal.⁶⁻⁸ Herein, we describe an original molecular mechanism in which *F8* was disrupted by a 3.8-Mb Xq11.1q12 inserted duplication leading to the production of an *F8*/Xq12 (noncoding sequence) fusion transcript. This complex rearrangement was characterized using nanopore sequencing technology.

2 | MATERIALS AND METHODS

2.1 | Subject

The propositus provided written informed consent for genetic investigation. He is a 45-year-old male with severe HA (both FVIII:C and FVIII:Ag were lower than 1 IU dL⁻¹), is the first hemophiliac in his family, and has no history of inhibitor development. Genetic investigations excluded both intron 1 and 22 inversions

and no detectable *F8* sequence variation was found after direct sequencing of the whole *F8* coding sequence as well as the intron/exon boundaries, the promotor region, and the 3' untranslated region. Multiplex ligation-dependent probes amplification assay (MRC-Holland, Amsterdam, The Netherlands) was performed and found a normal allele copy number for *F8* exons.

2.2 | Long-range PCR

Long-range PCR (LR-PCR) was performed on genomic DNA for the complete region of *F8* by amplification of 28 overlapping fragments using primers previously described by Pezeshkpoor et al⁶ and PrimeSTAR GXL DNA polymerase (Takara, Saint-Germain-en-Laye, France) according to the manufacturer's instructions.

To identify breakpoint locations, LR-PCR was performed using specific primers located in the last known DNA region from either side of the breakpoints and PrimeSTAR GXL DNA polymerase with an elongation time of 15 min. The PCR products were resolved on agarose gels and sequenced.

Essentials

- No *F8* genetic abnormality is detected in about 2% of severe hemophilia A patients.
- Detection of *F8* structural variants remains a challenge.
- We identified a new *F8* rearrangement in a severe hemophilia A patient using nanopore sequencing.
- We highlight the value of single-molecule long-read sequencing technologies in a genomics laboratory.

2.3 | *F8* RNA isolation and Reverse Transcription-PCR assay

The RNA analysis was performed as previously described.⁹ Briefly, *F8* cDNA was subdivided into four regions covering all splice sites: region A (exons 1 to 7), region B (exons 7 to 14), region C (exons 14 to 22), and region D (exons 22 to 26). The reverse transcription reaction was performed using four antisense-specific primers (one for each region). Then, the cDNA was amplified in two rounds of PCR (nested PCR approach).

The PCR products were resolved on agarose gels and sequenced.

2.4 | 3'RACE and long-read sequencing

The sequence located upstream of exon 25 (pseudoexon) in the patient *F8* RNA was identified using a 3' rapid amplification of cDNA ends (3'RACE) approach strategy (3'RACE system kit; Life technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. In the first step, reverse transcription was performed using oligo(dT)-containing adapter primer and then two rounds of PCR were performed using *F8* specific forward primers and a universal reverse primer. The 3'RACE PCR products were analyzed using MinION long-read sequencer (Oxford Nanopore Technologies, Oxford, UK). Library preparation was performed using the SQK-LSK-108 kit for 1D sequencing on a R9.4 flowcell (Oxford Nanopore Technologies) following the manufacturer's instructions. The MinION instrument was piloted through MinKnow v1.13 software. Albacore v2.2 software (Oxford Nanopore Technologies) was used for base calling. Reads were aligned to the GRCh38 human reference genome using BWAMEM with -ont option. Data polishing was performed with Canu using reads aligned to the *F8* locus before reads' realignment on *F8* ENST00000360256.8 cDNA sequence.¹⁰ Visual analysis of reads was done with Integrative Genomics Viewer v2.3, and BLAT (<https://genome.ucsc.edu/cgi-bin/hgBlat?command=start>) was used for soft-clipped reads' alignment.¹¹

2.5 | Cytogenetic microarray analysis

Cytogenetic microarray analysis (CMA) was performed using 3 µg of genomic DNA from the patient and a reference male DNA on an 180 000-oligonucleotide array (AMADID: 022060) following the manufacturer's instructions (Agilent Technologies, Santa Clara, CA, USA). Cytogenomics v3.3 software (Agilent Technologies) was used for data extraction and copy number variant detection.

2.6 | Karyotype/FISH

Standard G-bands by Trypsin using Giemsa (GTG) and Reverse bands by Heating using Giemsa (RHG) banded karyotypes were established from peripheral blood according to standard procedures. Fluorescent in situ hybridization (FISH) was performed on metaphases preparation using RP11-284B18 (Xq11.2), RP13-130F17 (Xq12), and X centromere (Vysis FISH, Abbott Molecular, Chicago, IL) probes.

3 | RESULTS AND DISCUSSION

Long-range PCR was performed to determine the integrity of *F8* by amplifying 28 overlapping fragments. Long-range PCR could not amplify an 8.805-kb region (fragment T2) from *F8* intron 25, indicating a large insertion, deletion, or genomic recombination within this region. In order to narrow down the location of the breakpoint further, this nonamplifiable region was divided into three overlapping fragments and primer pairs were designed to amplify each fragment. Using this approach, the nonamplifiable region containing the breakpoint was reduced to 6.150 kb in size (chrX: 154 071 050-154 077 199, GRCh37/hg19; Figure 1A). This region contains a 6.137-kb LINE sequence (L1PA5).

The *F8* mRNA amplification of region D (exon 22 to 26) failed, while all other segments amplified normally. A new reverse transcription was performed using an antisense primer located in exon 25.

The region spanning from exon 22 to 25 was correctly amplified in the patient (Figure 1B). This implies absence of the normal joining of exon 25-26 in

the *F8* mRNA. The mere presence of *F8* transcript lacking exon 26 suggested that it may be still polyadenylated. Fusion transcript characterization in

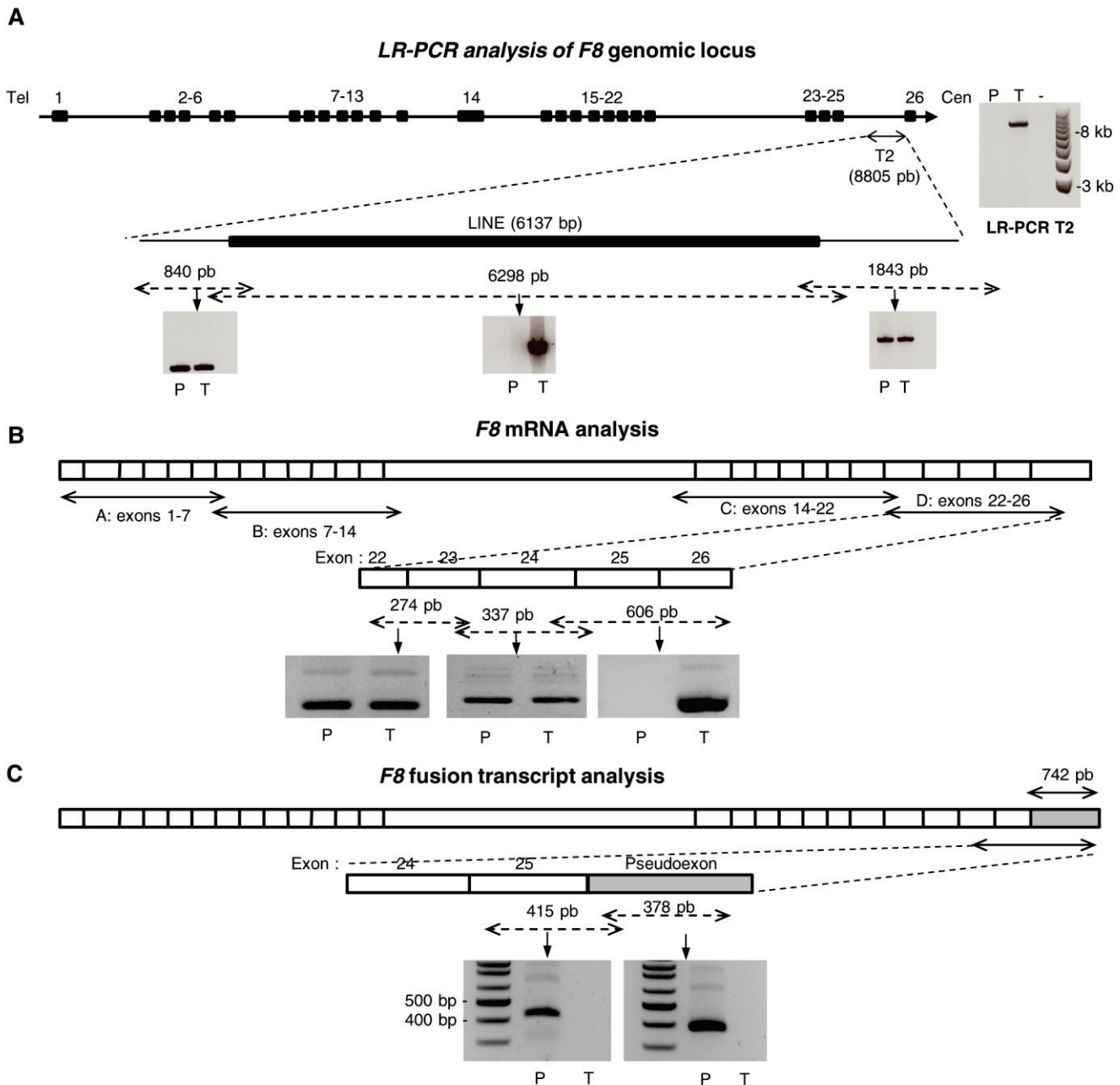


FIGURE 1 Location of the breakpoint in *F8* intron 25 and *F8* mRNA analysis. A, Schematic representation of the *F8* genomic locus is shown in the upper part and the PCR amplification of the region T2 is shown upper right. The detailed localization of the breakpoint using three overlapping PCRs is shown in the lower part. B, The *F8* mRNA analysis failed to amplify the region spanning from exon 22 to 26. The region spanning from exon 22 to 25 was correctly amplified using a reverse primer located in exon 25. This implies absence of the normal joining of exon 25-26 in the *F8* mRNA. C, Confirmation by conventional approach of the presence of 742-bp pseudoexon downstream exon 25 in patient *F8* transcript. DNA, deoxyribonucleic acid; mRNA, messenger ribonucleic acid; P, patient; PCR, polymerase chain reaction; T, control nonhemophilic DNA

case of an unknown partner is theoretically possible using the 3'RACE approach. The 3'RACE initiated from *F8* specific primer was used to amplify the unknown sequence beside exon 25. However, after several experiments using different *F8* specific primers, many nonspecific products were observed for both patient and control preventing the analysis by Sanger sequencing. Single-molecule nanopore sequencing was used to analyze the 3'RACE products in order to identify the unknown sequence beside exon 25 in the transcript of interest. Eighty polished reads aligned to *F8* ENST00000360256.8 transcript. All reads were soft-clipped from the same position after exon 25 and with the same sequence (considering nanopore technology error rate, Figure 2). A 742-bp pseudoexon beside exon 25 in the patient *F8* transcript followed by a poly-A tail was identified. This sequence corresponded to a noncoding region located at the beginning of the long arm of chromosome X (Xq12; chrX: 66 310 352-66 311 093, GRCh37/hg19). The analysis of this sequence found that the pseudoexon was flanked by a canonical AG on its 5' side and contained two putative polyadenylation signals (AATAAA) on its 3' side. The existence of an *F8*/ Xq12 (noncoding sequence) fusion

transcript was confirmed by a conventional RT-PCR approach using an antisense primer located in the pseudoexon (Figure 1C). Due to the undetectable FVIII:Ag in the patient's plasma, the modified FVIII protein encoded by this fusion transcript (p.(Val2301Argfs*32)) was presumably not synthesized by the liver or not secreted into the blood.

In order to amplify the first DNA junction fragment, a LR-PCR approach was performed using a forward primer located in intron 25 (T2F) and a reverse primer located in the pseudoexon (PE_R). A 13-kb PCR product was generated from the patient's DNA and was absent in control DNA. Then, PCRs were performed using several nested primers from each side of the breakpoint in order to reduce the size of the PCR. A 226-bp fragment was obtained using the 4S/2R primer pair. The sequencing of this product found that the *F8* intron 25 was joined to Xq12 noncoding sequence at the breakpoint coordinates chrX: 154 072 766 (*F8*) and chrX: 66 319 203 (Xq12); GRCh37/hg19. These breakpoints showed a 2-bp microhomology (CT). The Xq12 breakpoint is located 8.108 kb upstream to the pseudoexon (Figure 3).

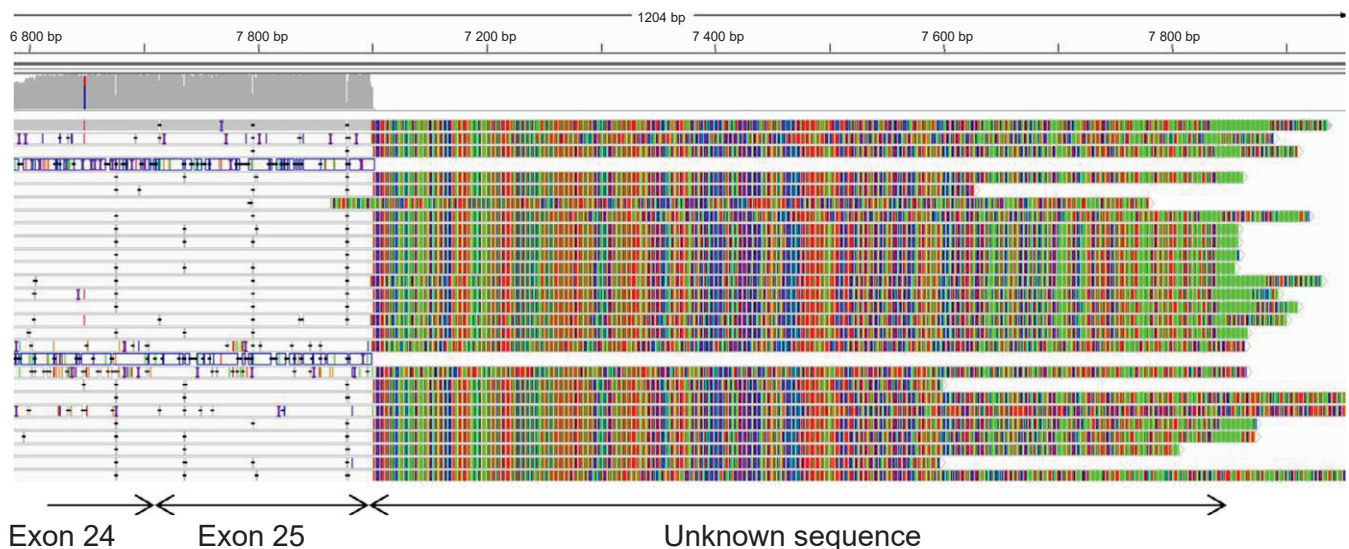


FIGURE 2 Single-molecule nanopore sequencing of the 3' RACE products. The integrative genomics viewer visualization of a subset of reads aligned to *F8* ENST00000360256.8 transcript was shown. All reads were soft-clipped from the same position after exon 25 and with the same sequence (considering nanopore technology error rate) IGV, Integrative Genomics Viewer; 3'RACE, 3' rapid amplification of complementary deoxyribonucleic acid ends

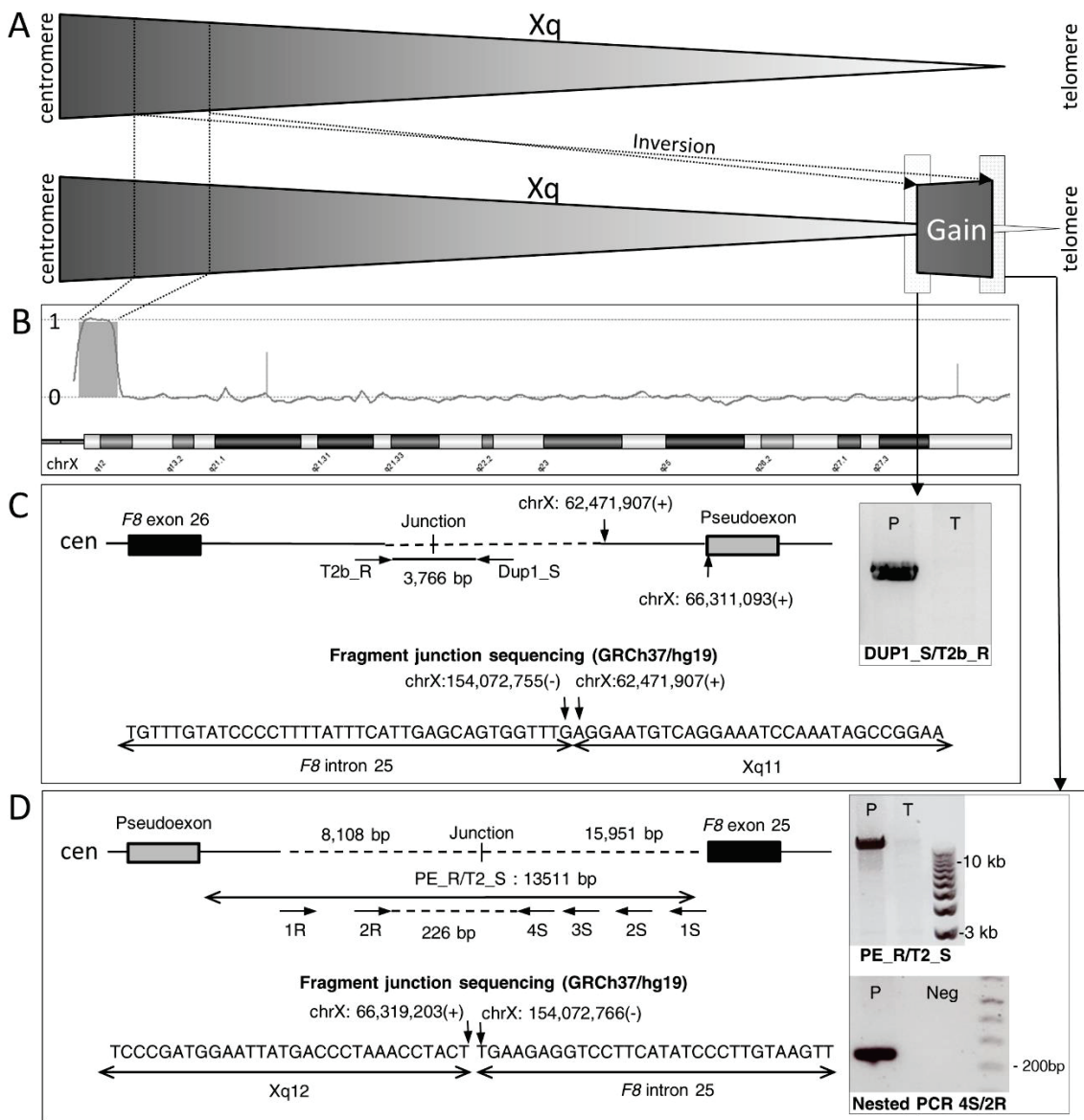


FIGURE 3 Chromosomal rearrangement identified in the patient. A, Schematic representation of a normal and the derivative chromosome X long arm showing the inverted insertion of a Xq11.1q12 duplication within *F8*. B, Cytogenetic microarray analysis confirmed a 3.8-Mb duplication of Xq11.1q12 region in the patient. C, Analysis of the intron 25/Xq11.1 junction. A LR-PCR was performed using a forward primer located in Xq11.1 (DUP_1S) and reverse primer located in *F8* intron 25 (T2b_R). The DUP_1S/T2b_R amplification is shown at the right. The sequencing of this fragment showed that the Xq11.1 region was joined to *F8* intron 25 at the breakpoint coordinates chrX: 62 471 907 and chrX: 154 072 755, GRCh37/hg19. D, Analysis of the Xq12/intron 25 junction. A LR-PCR was performed using a forward primer located in intron 25 (T2F) and a reverse primer located in the pseudoexon (PE_R). The T2F/PE_R PCR amplification is shown upper right. Then nested PCRs were performed using several nested primers from each side of the breakpoint in order to reduce the size of the PCR product. The 4S/2R nested PCR amplification is shown lower left. The sequencing of this product showed that the *F8* intron 25 was joined to Xq12 noncoding sequence at the breakpoint coordinates chrX: 154 072 766 and chrX: 66 319 203; GRCh37/hg19. (+) sense strand sequence; (-) antisense strand sequence; LR-PCR, long-range polymerase chain reaction; Neg, negative control; P, patient; T, control nonhemophilic DNA;

As the Xq12 noncoding sequence was inverted with regard to *F8*, we suspected a paracentric inversion of the q-arm of the X chromosome. Specific primers were designed to amplify the recombinant (INVXS/T2bR, expected size 2001 bp) or the nonrecombined (INVXS/INVXR, expected size 346 bp) Xq12 region. Surprisingly, the 346-bp product was observed in both patient and control DNA. These results invalidated the paracentric inversion hypothesis.

We then hypothesized that a large Xq12 duplicated region could be inserted in the *F8* intron 25. Cytogenetic microarray analysis confirmed a 3.8-Mb duplication of the Xq11.1q12 region (arr[GRCh37]Xq11.1q12(62473920_66315527)x2; Figure 3). This copy number variation contained 13 genes (*SPIN4*, *ARHGEF9*, *AMER1*, *ASB12*, *MTMR8*, *ZC4H2*, *ZC3H12B*, *LAS1L*, *FRMD8P1*, *MSN*, *VSIG4*, *HEPH*, *EDA2R*). The CMA located the start of the duplication in a \approx 68 kb region (chrX: 62 405 278-62 473 920, GRCh37/hg19). As the patient exhibits no other clinical feature except for spontaneous bleeding, the pathogenicity of this copy number variation remained uncertain. Then, the second recombinant region was amplified using a forward primer located in Xq11.1 (DUP_1S) and reverse primer located in *F8* intron 25 (T2b_R). A 3.766-kb PCR product was obtained from the patient's DNA and was absent in control DNA. The sequencing of this fragment showed that the Xq11.1 region was joined to *F8* intron 25 at positions chrX: 62 471 907 and chrX: 154 072 755, GRCh37/ hg19. A 10-bp deletion of *F8* intron 25 was found at the breakpoint (Figure 3). These results corroborated the hypothesis of a large insertion of Xq11.1q12 duplicated sequence in *F8* intron 25 (Figure 3) and were confirmed by FISH analysis (data not shown). Considering the 2-bp homology and the 10-bp deletion identified at breakpoints, this large insertion in *F8* intron 25 could be due to a DNA double strand break repairing mechanism such as nonhomologous end joining or microhomology-mediated breakage induced repair as previously

suggested by Gu et al¹² for other large intrachromosomal insertions.

Single-molecule sequencing techniques have rarely been used as an add-on to karyotype or short-read sequencing to confirm known chromosomal structural variations.¹³ We report herein the first constitutional chromosomal rearrangement detected using a nanopore first approach. This strategy was chosen to resolve nonspecific PCR amplifications while the read quality was convenient for structural variant calling. This is one example of the value of nanopore sequencing in a genomics laboratory.

Because of the large number of repetitive elements such as Alu elements and LINE present in its sequence and its Xq28 location, the *F8* appears to be extremely susceptible to genetic rearrangements.¹⁴ Whole *F8* sequencing approaches previously reported were not able to detect *F8* structural variant.^{15,16} While short-read whole genome sequencing can identify chromosomal structural variations,¹⁷ repeated regions remain a challenge. Indeed, because of ambiguous mapping, variants affecting repeated elements larger than the typical insert size used for short-read paired-end sequencing (1 kb) can be missed. Herein, the 6-kb LINE-1 repeat element in *F8* intron 25 is known to be poorly mappable using 100-bp reads (UCSC genome browser) and the rearrangement might have been incompletely resolved. Whether this could have been prevented using whole-genome long-read sequencing remains to be evaluated.

In summary, we characterized a novel genomic rearrangement in which a 3.8-Mb Xq11.1q12 gain inserted in the *F8* intron 25 led to an aberrant fusion transcript in a patient with severe HA, using comprehensive molecular techniques. This study highlights the value of single-molecule long-read sequencing technologies for molecular diagnosis of HA especially when conventional genetic approaches have failed.

ACKNOWLEDGEMENT

The authors thank Philip Robinson (DRCI, Hospices Civils de Lyon) for help in manuscript preparation.

CONFLICTS OF INTEREST

The authors state that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

N. Chatron, M. Frétygny, D. Sanlaville, C. Vinciguerra, and Y. Jourdy designed the experiments. N. Chatron, C. Schluth-Bolard, M. Frétygny, A. Labalme, G. Vilchez, and Y. Jourdy performed the experiments and analyzed the data. S.-M. Castet enrolled patients in the study, collected the clinical data, and analyzed the data. N. Chatron and Y. Jourdy drafted the manuscript. M. Frétygny, D. Sanlaville, C. Vinciguerra, and C. Négrier helped in revising and editing the manuscript. All the authors approved the manuscript submission.

ORCID

Yohann Jourdy  <https://orcid.org/0000-0001-7408-6474>

REFERENCES

- White GC 2nd, Rosendaal F, Aledort LM; Factor VIII Factor and IX Subcommittee. Definitions in hemophilia: Recommendation of the scientific subcommittee on factor VIII and factor IX of the scientific and standardization committee of the International Society on Thrombosis and Haemostasis. *Thromb Haemost.* 2001;85:560.
- Antonarakis SE, Rossiter JP, Young M, Horst J, de Moerloose P, Sommer SS, et al. Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood.* 1995;86:2206–12.
- Bagnall RD, Waseem N, Green PM, Giannelli F. Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood.* 2002;99:168–174.
- El-Maarri O, Herbiniaux U, Graw J, Schröder J, Terzic A, Watzka M, et al. Analysis of mRNA in hemophilia A patients with undetectable mutations reveals normal splicing in the factor VIII gene. *J Thromb Haemost.* 2005;3:332–9.
- Swystun LL, James PD. Genetic diagnosis in hemophilia and von Willebrand disease. *Blood Rev.* 2017;31:47–56.
- Pezeshkpoor B, Rost S, Oldenburg J, El-Maarri O. Identification of a third rearrangement at Xq28 that causes severe hemophilia A as a result of homologous recombination between inverted repeats. *J Thromb Haemost.* 2012;10:1600–1608.
- Bagnall RD, Waseem NH, Green PM, Colvin B, Lee C, Giannelli F. Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two haemophilia A patients. *Br J Haematol.* 1999;107:766–761.
- Xin Y, Zhou J, Ding Q, Chen C, Wu X, Wang X, et al. A pericentric inversion of chromosome X disrupting F8 and resulting in haemophilia A. *J Clin Pathol.* 2017;70:656–61.
- Jourdy Y, Janin A, Frétygny M, Lienhart A, Négrier C, Bozon D, et al. Recurrent F8 intronic deletion found in mild hemophilia A causes Alu exonization. *Am J Hum Genet.* 2018;102:199–206.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27:722–736.
- Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res.* 2002;12:656–664.
- Gu S, Szafranski P, Akdemir ZC, Yuan B, Cooper ML, Magriñá MA, et al. Mechanisms for complex chromosomal insertions. *PLoS Genet.* 2016;12:e1006446.
- Dutta UR, Rao SN, Pidugu VK, V SV, Bhattacharjee A, Bhowmik AD, et al. Breakpoint mapping of a novel de novo translocation t(X;20)(q11.1;p13) by positional cloning and long read sequencing. *Genomics.* 2018. pii: S0888-7543(18)30299-4
- Lannoy N, Hermans C. Review of molecular mechanisms at dis-tal Xq28 leading to balanced or unbalanced genomic rearrangements and their phenotypic impacts on hemophilia. *Haemophilia.* 2018;24:711–719.
- Pezeshkpoor B, Zimmer N, Marquardt N, Nanda I, Haaf T, Budde U, et al. Deep intronic “mutations” cause hemophilia A: application of next generation sequencing in patients without detectable mutation in F8 cDNA. *J Thromb Haemost.* 2013;11:1679–87.
- Inaba H, Shinozawa K, Amano K, Fukutake K. Identification of deep intronic individual variants in patients with hemophilia A by nextgeneration sequencing of the whole factor VIII gene. *Res Pract Thromb Haemost.* 2017;1:264–74.
- Schluth-Bolard C, Labalme A, Cordier M-P, Till M, Nadeau G, Tevissen H, et al. Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *J Med Genet.* 2013;50:144–50.

How to cite this article: Chatron N, Schluth-Bolard C, Frétygny M, et al. Severe hemophilia A caused by an unbalanced chromosomal rearrangement identified using nanopore sequencing. *J Thromb Haemost.* 2019;17:1097–1103. <https://doi.org/10.1111/jth.14460>

Suite de ce travail

Séquençage Nanopore

En parallèle de la publication précédente nous avons souhaité aller plus loin et évaluer la capacité du séquençage nanopore à identifier un remaniement dès la suspicion de variant de structure. En effet, ici, par chance la présence d'un ARN anormal a permis d'identifier le remaniement et ceci après de nombreuses techniques de PCR pour avancer le long du gène (ADN) et du transcrit (ARN). Dans l'idéal il faudrait pouvoir appliquer cette technique plus économique et plus rapide dès la suspicion de variant de structure dans une pathologie monogénique comme solution à notre question de départ applicable largement.

Nous avons donc utilisé l'ADN total du patient (3µg) pour réaliser un séquençage de génome sur le MinION (Oxford Nanopore Technologies) avec une préparation de librairie selon le protocole du fournisseur utilisant un kit de préparation de librairie LSK-108 pour des lectures dites 1D (dans un seul sens) sur une flowcell R9.4 (Oxford Nanopore Technologies). Le séquençage a duré 48 heures avant alignement de séquences utilisant le programme Minimap2 (Li, 2018) de telle sorte que les alignements de séquences étaient disponibles 3 jours après le début de la préparation de librairie. Au total, 13Gb ont été séquencées pour une profondeur moyenne sur le génome GRCh38 de 4x. Nous avons ensuite utilisé le programme Sniffles pour la détection de variants de structure (Sedlazeck et al., 2018) et filtrer les variants affectant le gène *F8*. Le remaniement a ainsi pu être caractérisée à la base près très rapidement comme il l'avait été précédemment en accumulant tout un panel de techniques différentes (Tableau 1, Figure 31).

Il faut cependant noter que plusieurs éléments ont facilité notre résultat pour ce cas précis :

- L'hémophilie A est une pathologie monogénique dont le diagnostic est biologique avec un accès direct à l'activité de la protéine. Aucun autre gène ne peut être évoqué.
- Le gène est situé sur le chromosome X. Chez un sujet masculin, toutes les séquences obtenues proviennent donc du chromosome remanié.
- Le laboratoire d'hémostase extrait habituellement les ADNs avec une technique manuelle plus favorable à l'obtention de grands fragments.

- Le locus *F8* est particulièrement concerné par les éléments répétés de grande taille de type LINE-1 qui font encourir un risque accru de non-détection de variants de structure par une technique de séquençage *short-read*.

Ces nouveaux résultats ont été présentés au Congrès de l' « European Society of Human Genetics » en juin 2019.

Tableau 1 : Synthèse des résultats obtenus lors de la caractérisation du variant structural du patient hémophile A. L'approche « PCR-Sanger-CGH-array-Nanopore ciblé est celle publiée dans l'article 2. Les résultats de la stratégie alternative « Nanopore WGS » ont été analysés par visualisation des données de séquençage d'une part et en aveugle avec le logiciel Sniffles avec un écart d'un nucléotide maximum.

Breakpoint	PCR-Sanger-arrayCGH-Nanopore ciblé	Nanopore-WGS Visualisation IGV	Nanopore-WGS Sniffles
BP1 (<i>F8</i>)	154,844,480-91	154,844,480-92	154,844,480-92
BP2	63,252,030	63,252,030	63,252,029
BP3	67,099,361	67,099,361	67,099,361

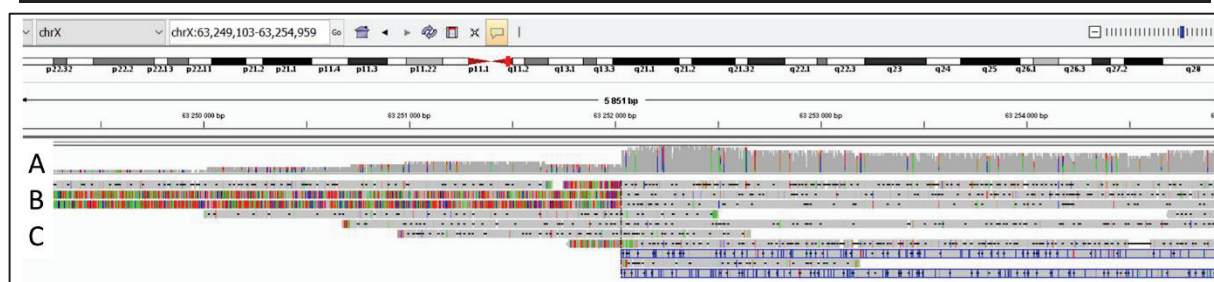


Figure 31 : Visualisation IGV des séquences long-read obtenues par séquençage de génome du patient hémophile A étudié dans l'article 2. Ici au locus du point de cassure en Xq11 on note une augmentation brutale de la profondeur de séquençage (A) signant un gain de copie. Deux types de lectures sont alignées. Les lectures qui réalisent la jonction anormale entre le gain et l'intron 25 du gène *F8* sont visibles identifiables par leur alignement de mauvaise qualité sur leur partie centromérique (gauche) (B). Les lectures qui correspondent à la région Xq11 native s'alignent entièrement correctement (C).

Nous projetons désormais d'appliquer cette stratégie dans diverses pathologies monogéniques dans les cas où un variant de structure est fortement suspecté.

Séquençage linked-read

En collaboration avec 3 autres laboratoires de cytogénétique hospitalier français et le Centre Nationale de Recherche en Génomique Humaine nous avons réalisé une étude pilote sur 13

échantillons connus pour avoir un caryotype anormal (dont les patients 1, 15 et 19 de la publication sur les *chromoanagenesis* (article 3)) séquencés en parallèle en *short-reads* classiques et en *linked-reads* (Uguen et al., *soumis*) (Tableau 2). Pour 8 remaniements sur 13 les points de cassure identifiés en *short-reads* ont pu être confirmées en *linked-reads* avec cependant un important bruit de fond dans les données *linked-reads*. Pour deux échantillons, la technologie *linked-reads* a manqué une partie des points de cassure identifiés en *short-reads*. Pour trois échantillons, le séquençage *short-reads* n'a pas permis d'identifier de point de cassure mais l'un d'entre eux a pu être identifié en *linked-reads*. Ces résultats décevants peuvent être en partie le fait d'une qualité d'ADN sub-optimale et/ou d'une nécessaire phase d'apprentissage de la technologie qui ici n'a pas eu lieu. Au total, dans notre expérience, la technologie ne pourrait se placer qu'en deuxième intention après un séquençage *short-reads* et avec un rendement incertain.

Patient	Indication	Short-read strategy	Linked-read strategy	
			Call (blind)	Candidate
1	46,X,t(X;13)(q22.1;q34)	+ (2)	+ (2)	-
2	46,XY,t(9;13)(p24.2;q21.31)	+ (4)	-	+ (4)
3	45,XX,rob(13;14)(q10;q10)	-	-	-
4 (P19- article 2)	Chromoanagenesis	+ (84)	+ (2)	+ (18)
5	CNV on chromosome X	+ (4)	-	-
6 (P15- article 2)	46,XY,t(1;2)(p13.2;q31.2) Chromoanagenesis	+ (33)	+ (2)	+ (31)
7	46,X,t(X;1)(p12;p36.1)	-	-	+
8	46,XY,t(3;22)(q13-21;p11)	-	-	-
9 (P1- article 2)	46,XX,inv(3)(p13p22),inv(3)(p12q26.3) Chromoanagenesis	+ (23)	+ (2)	+ (21)
10	46,XY,t(6;8;9;13)(q26;p23;p21;q21) CCR	+ (13)	+ (2)	+ (11)
11	Suspected chromomanasynthesis	+ (47)	+ (18)	+ (29)
12	Suspected chromothripsis	+ (2)	-	+ (2)
13	46,XX,t(9;17)(p13;q21)	+ (2)	-	+ (2)

- Tableau 2 : Résumé des résultats obtenues par séquençage short-read et linked-read. + = variant de structure détecté. - = variant de structure non détecté. Le nombre de points de cassure est indiqué entre parenthèses.

L'efficacité de ces approches dans d'autres situations cliniques conditionnera leur place dans l'arsenal d'outils diagnostiques.

II – 2/ Article 2 : Identification of mobile retrocopies during genetic testing: Consequences for routine diagnosis

La deuxième question posée en introduction était :

- **Une fois identifié, par quel(s) mécanisme(s) un remaniement de structure peut-il être responsable d'un phénotype ?**

Cette question a été largement abordée en introduction. Dans ce chapitre nous nous sommes concentrés sur une classe méconnue de variants structuraux évoquée en introduction : les rétrocopies.

Ce qui était connu :

Un ARN messager peut être capté par la machinerie d'un élément L1 et ainsi être rétrotranscrit et inséré à une nouvelle position génomique. Cette classe de rétroéléments est la plus rare avec une mobilisation toutes les 6300 méioses.

De nombreux *pipelines* bioinformatiques dédiés ont été développées pour détecter les éléments Alu et LINE-1 à la fois très nombreux et très homologues entre eux.

Ces rétrocopies sont peu décrites dans la littérature et méconnues du monde de la génétique médicale. L'objectif du travail était de permettre aux généticiens de reconnaître ces événements et de présenter les conséquences possibles en pathologies et les interférences possibles avec les autres tests génétiques de routine.

Ce que ce travail apporte :

Nous décrivons ici le profil typique de ces éléments rétrocopiés obtenu lors d'un test en séquençage haut débit *short-read*. Une rétrocopie apparaît de façon détournée en fonction des différents programmes de détection de variants génomiques utilisés :

- La détection de CNV rapportera un gain de l'ensemble des exons d'un transcrit donné sans affecter les régions introniques. Sur des données de séquençage ciblées, c'est tout

le gène qui apparaîtra dupliqué. Sur des données de séquençage de génome, des duplications isolées de tous les exons peuvent être détectées.

- La détection de variant de structure équilibré mettra en évidence des jonctions inter-exons sautant un ou plusieurs introns en fonction de la taille des exons. Le site d'insertion sera également identifiable.
- La détection de variants nucléotidiques ponctuels appellera de fausses insertions de séquences à toutes les jonctions exon/intron correspondant en fait à l'enchaînement des exons.

L'analyse des séquences dans un logiciel de visualisation est cruciale pour comprendre et associer ces trois types de signaux. C'est la combinaison de ces différents éléments qui doit faire évoquer une rétrocopie. En effet, à l'inverse des autres familles d'éléments mobiles, l'identification de rétrocopies à partir de données *short-read* est tout à fait possible sans développer de pipeline bioinformatique dédié, pour peu que leur signature soit reconnue.

A notre connaissance une seule observation de rétrocopie pathogène a été rapportée chez l'Homme (de Boer et al., 2014). Il s'agissait de l'insertion incomplète d'une rétrocopie du gène *TMFI* dans l'intron 1 du gène *CYBB*. Cet évènement a fait apparaître de nouveaux sites d'épissage permettant l'exonisation de 117 nucléotides entre les exons 1 et 2 du transcrit de *CYBB*. Ce nouvel exon contient un codon stop empêchant la synthèse d'une protéine fonctionnelle et à l'origine de la granulomatose chronique présentée par le patient. Il s'agit là d'un mécanisme proche de ce qui a été décrit pour d'autres éléments mobiles (Alu notamment) mais les rétrocopies ont des mécanismes pathogènes potentiels spécifiques.

L'origine codante du matériel transposé permet, si la situation au site d'insertion est favorable à la rétrocopie d'être transcrite puis traduite. Il peut ainsi y avoir un effet-dose délétère ou l'expression aberrante de ce transcrit dans un type cellulaire différent des tissus où est habituellement exprimé le gène d'origine. Si la rétrocopie est insérée sur le brin opposé au gène d'origine, la synthèse d'un ARN antisens est alors possible et alors responsable d'une haploinsuffisance. Enfin, il peut y avoir formation d'un transcrit de fusion « viable » donnant naissance à une nouvelle protéine aux fonctions difficiles à prédire.

Ce travail a été présenté au congrès de l'Association des Cytogénéticiens de Langue Française à St Malo en septembre 2018. La présentation a permis d'initier une collaboration ayant mené à la publication ci-dessous dans le journal Human Mutation (Impact Factor = 5,359) et à une nouvelle communication orale au congrès de l'Human Genome Variant Society à Göteborg en juin 2019.

BRIEF REPORT

Identification of mobile retrocopies during genetic testing: Consequences for routine diagnosis

Nicolas Chatron^{1,2}  | Kevin Cassinari³ | Olivier Quenez³ | Stéphanie Baert-Desurmont⁴ | Claire Bardel^{5,6} | Marie-Pierre Buisine⁷ | Eduardo Calpena⁸ | Yline Capri⁹ | Jordi Corominas Galbany¹⁰ | Flavie Diguët^{1,2} | Patrick Edery^{1,2} | Bertrand Isidor¹¹ | Audrey Labalme¹  | Cedric Le Caignec^{11,12} | Jonathan Lévy¹³ | François Lecoquierre⁴ | Pierre Lindenbaum^{14,15} | Olivier Pichon¹¹ | Pierre-Antoine Rollat-Farnier^{1,5} | Thomas Simonet^{16,17} | Pascale Saugier-Veber⁴ | Anne-Claude Tabet^{13,18} | Annick Toutain^{19,20} | Andrew O. M. Wilkie⁸ | Gaetan Lesca^{1,2} | Damien Sanlaville^{1,2} | Gaël Nicolas³ | Caroline Schluth-Bolard^{1,2}

¹Genetics Department, Hospices Civils de Lyon, Lyon, France

²GENDEV Team, CRNL, INSERM U1028, CNRS UMR5292, UCBL1, Lyon, France

³Department of Genetics and CNR-MAJ, Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France

⁴Department of Genetics, Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France

⁵Bioinformatics group of the Lyon University Hospital NGS facility, Groupement Hospitalier Est, Lyon, France

⁶Biostatistics and Bioinformatics Department, HCL, Lyon, France

⁷Department of Biochemistry and Molecular Biology, JPA Research Center, Inserm UMR-S 1172, Lille University, Lille University Hospital, Lille, France

⁸Clinical Genetics Group, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK ⁹Genetics Department, Clinical Genetics Unit, Hôpital Universitaire Robert Debré, Paris, France ¹⁰Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands ¹¹Genetics Department, CHU Nantes, Nantes, France

¹²INSERM UMR_S915, Institut du thorax, Nantes University, Nantes, France

¹³Genetics Department, Cytogenetics Unit, Hôpital Universitaire Robert Debré, Paris, France

¹⁴INSERM, UMR_S1087, Institut du thorax, Nantes, France

¹⁵CNRS, UMR 6291, Nantes, France

¹⁶Cellular Biotechnology Center, Hospices Civils de Lyon, Lyon, France

¹⁷Nerve-Muscle Interactions Team, Institut NeuroMyoGène CNRS UMR 5310-INSERM U1217-Université Claude Bernard Lyon 1, Lyon, France

¹⁸Neuroscience Department, Human Genetics and Cognitive Function Unit, Institut Pasteur, Paris, France

¹⁹Genetics Department, Hôpital Bretonneau, CHU, Tours, France

²⁰UMR 1253, iBrain, Tours University, Inserm, Tours, France

Correspondence
Nicolas Chatron, Service de
Génétique, Centre de
Biologie Pathologie Est
2ème étage, Groupement
Hospitalier Est, Hospices
Civils de Lyon, 59 Boulevard
Pinel, 69677 Bron CEDEX,
France.

Email: nicolas.chatron@chu-lyon.fr

Funding information

Wellcome, Grant/Award
Number: WT 100127; UK
Department of Health,
Grant/ Award Number:
R6-388; Health Innovation
Challenge Fund

Abstract

Human retrocopies, that is messenger RNA transcripts benefitting from the long interspersed element 1 machinery for retrotransposition, may have specific consequences for genomic testing. Next generation sequencing (NGS) techniques allow the detection of such mobile elements but they may be misinterpreted as genomic duplications or be totally overlooked. We report eight observations of retrocopies detected during diagnostic NGS analyses of targeted gene panels, exome, or genome sequencing. For seven cases, while an exons-only copy number gain was called, read alignment inspection revealed a depth of coverage shift at every exon-intron junction where indels were also systematically called. Moreover, aberrant chimeric read pairs spanned entire introns or were paired with another locus for terminal exons. The 8th retrocopy was present in the reference genome and thus showed a normal NGS profile. We emphasize the existence of retrocopies and strategies to accurately detect them at a glance during genetic testing and discuss pitfalls for genetic testing.

1 | INTRODUCTION

Since McClintock's (1951) demonstration using corn, genomic mobility has largely been studied and related to evolutionary processes and genetic diseases. While DNA transposons are inactive in humans (Lander et al., 2001), DNA retrotransposons use an RNA step for mobilization. They have been separated into two groups depending on whether they contain a retrovirus-derived sequence (human endogenous retroviruses) or not. RNA non-ERV elements are the most abundant class covering together one-third of the total human reference genome sequence (Lander et al., 2001). Four different types of these can be distinguished (Feschotte & Pritham, 2007). Long interspersed element 1 (LINE-1) is the only autonomous element capable of retrotranscription and insertion; the other three types Alu, SINE-R-VNTR-Alu (SVA), and retrocopies require the LINE-1 machinery for mobility (Dewannieux, Esnault, & Heidmann, 2003; Esnault, Maestre, & Heidmann, 2000; Hancks, Mandal, Cheung, & Kazazian, 2012; Raiz et al., 2012). Retrocopies (also referenced as retrogene, retroCNV, retroduplication, processed pseudogenes, or retrotransposed pseudogenes in the literature) are messenger RNAs that have been reverse-transcribed and inserted in the genome. Studying primate genomes, 127 retrocopies out of 8,000 present in the reference human genome could be demonstrated as

human-specific (Navarro & Galante, 2015; Zhang, Harrison, Liu, & Gerstein, 2003). While LINE-1 mediated events are estimated to occur in 1 of 21 meioses for Alu (Xing et al., 2009), 1 of 95–270 for LINE-1 (Ewing & Kazazian, 2010; Xing et al., 2009) and 1 of 916 for SVA (Xing et al., 2009), retrocopy events are the rarest, estimated at 1 per 6,256 meioses in the literature (Ewing et al., 2013).

The first description of a pathogenic mobile element was the insertion of a LINE-1 sequence within *F8* (Kazazian et al., 1988) responsible for hemophilia A and, since then, more than 130 similar events (Kazazian & Moran, 2017) have been detected in diseases genes associated such as Duchenne muscular dystrophy (Smith et al., 2011), beta-thalassemia trait (Lanikova et al., 2013), and haemophilia B (Mukherjee, Mukhopadhyay, Banerjee, Chandak, & Ray, 2004) among many examples. With the spread of genome sequencing-based

testing, such observations are likely to become more frequent. Beyond structural variant (SV) calling, data interpretation requires expertise in genome biology and the ability to recognize such phenomena. Herein, we present eight observations of retrocopies identified during the genetic interpretation of next generation sequencing (NGS) data with diverse techniques and diverse clinical situations.

We detail potential pitfalls for genetic analysis and discuss their

possible implications in human diseases.

2 | MATERIALS AND METHOD

All observations of retrocopies from Lyon, Rouen, and Nantes University hospitals made since 2016 were included. Patients received gene panel (P1–P3), exome (P4, P8), or genome (P5–P7) sequencing after providing informed written consent. Testing indications and individual methods are summarized in Table 1. All techniques used paired-end sequencing either on an Illumina platform (Illumina Inc., San Diego, CA; Patients 1–4, 6–8) or BGI technology (Shenzhen, China; Patient 5).

All sequencing tests used BWA-MEM v.0.7 (Li & Durbin, 2009) for genomic alignment over the GRCh37 assembly. Single nucleotide variant (SNV) calling was performed using GATK Haplotype Caller v3.8 (P1–P5, P8) while CANOES (Backenroth et al., 2014; P1, P2, P4), DeCovA (Dimassi et al., 2015; P3), and ERDS (Zhu et al., 2012; P5–P7) were used for CNV calling. Finally, SVs were called using either GRIDSS v1.4.1 (Cameron et al., 2017) for targeted sequencing (P1, P4) or BreakDancer v1.4.5 (Chen et al., 2009) for genome data (P5–P7). IGV 2.3 was used for data visualization (Robinson et al., 2011). Aberrant junction fragments were confirmed using classic PCR amplification (P1–P4, P6). For patient 1, QMPFS was performed using custom fluorescent primers, with the protocol previously described (Charbonnier et al., 2000). Moreover, a commercial kit was used for Multiplex Ligation-dependent Probe Amplification (SALSA MLPA probe mix P158-C2 Juvenile Polyposis; MRC-Holland, Amsterdam, The Netherlands). For Patient 5, droplet digital PCR was performed on the QX200 droplet digital PCR platform (Bio-Rad, Hercules, CA) using hydrolysis probes (TaqMan probes) on target (FAM tag) and *HMBS* reference gene (VIC tag). All primers are available upon request.

3 | RESULTS

Patients 1–7 presented a similar profile regarding SNV, copy number (CNV), and SV calling. CNV calling showed an exons-only duplication of genes *SMAD4*, *CCDC88C*, *DKC1*, *MTMR2*, *TYRO3*, *RCBTB1*, and

C20orf27, respectively, for patients 1–7 (Table 1). In each case, a detailed depth of coverage analysis showed an immediate shift at every exon-intron junction for a specific transcript. The structural variant analysis revealed clustered calls within the duplicated gene and apparent translocations linking both extremities of this same gene to another locus. Finally, SNV calling systematically called large indels at every splice acceptor and donor sites of a specific transcript. Alignment visualization revealed that indels were exclusively supported by soft-clipped reads (i.e., reads containing two concatenated DNA sequences mapping to two exon extremities with aberrant exon-exon junctions) paired with reads aligned to the

preceding or the following exon instead of flanking intronic sequence (Figure 1 for patients 3 and 5, Figure S1 for other patients). Patient's 8 *RPL13* retrocopy was not detected during the initial analysis of NGS data. It could only be identified during the complementary DNA (cDNA) study of an intronic SNV identified in the exome. Several unexpected SNVs were observed on the Sanger sequence of this RT-PCR product while absent from exome results.

Further analysis showed that this sequence could perfectly align to an unknown *RPL13* retrocopy present in the reference genome on chromosome 19 (chr19:56216980-56217800 [hg38]; Figure S1). The retrocopy had been preferentially amplified and sequenced after inappropriate DNase treatment.

None of these rearrangements directly disrupted a coding sequence or a topologically associated domain boundary. While six out of eight insertions occurred in another gene, none was likely to produce a fusion transcript regarding transcription strand or translation phase but exonization could not be ruled out. Events detected for patients 2, 3, and 6 were inherited from a healthy parent and they were considered as likely benign. For patient 5, the *TYRO3* retrocopy was initially detected as a candidate de novo event during genome analysis of a sporadic early-onset Alzheimer's disease patient—unaffected parents trio, but detailed data inspection and confirmation techniques showed that both parents were heterozygous carriers of the retrocopy and the proband carried it in a homozygous state (Figure 1). In both parents, the insertion occurred within intron 2 of *ENOX1* on chromosome 13 at the same position. In addition, the presence of two SNVs phased to this retrocopy favored a remote nonfunctional event spread within the general population. Reviewing our in-house exome and genome data, we found this *TYRO3* retrocopy to be a common event among controls, also previously described in the literature (MacDonald, Ziman, Yuen, Feuk, & Scherer, 2014; Schrider et al., 2013), confirming its likely benign nature. The patient 1 *SMAD4* retrotransposition within intron 17 of *SCAI* has also been demonstrated to be a rare likely benign event with a frequency of 0.25% in European populations (Millson et al., 2015; Watson et al., 2017). The patient 8 *RPL13* retrocopy is present in the reference genome and could not even be considered as a variant. Finally, looking at chimeric reads spanning exon-exon junctions in the 1000G project dataset of 2,535 individual genomes, *MTMR2*, and *RCBTB1* retrocopies could also be suspected in 6 and 3 individuals respectively.

4 | DISCUSSION

Through these eight observations, we aim to emphasize the potential consequences of genome mobility on genome biology and genetic testing. We demonstrate that paired-end sequencing standard pipelines easily highlight retrocopies. Their frequency might, however, be underestimated in routine genetic testing, because they can be either overlooked in NGS standard analyses or undetected by other molecular genetics technologies. While LINE-1, Alu, and SVAs insertions require dedicated

bioinformatics pipelines (Kvikstad, Piazza, Taylor, & Lunter, 2018), retrocopies present a recognizable footprint on sequencing results characterized by the association between clustered indel calls at exon-intron junctions, a unique exons-only CNV gain, and aberrant read pairs spanning introns on the one hand and between the parental locus and the insertion site on the other hand.

The clinical consequences regarding pathogenicity mostly depend on the insertion site. Indeed, as for other mobile elements, retrocopies could disrupt a gene and cause haploinsufficiency but considering their coding origin, they could also form a fusion transcript with unpredictable consequences, as exemplified by HIV resistance observed in owl monkeys after the insertion of a *CyclophilinA* cDNA into the *TRIM5* gene (Sayah, Sokolskaja, Berthoux, & Luban, 2004). Retrocopies could also disturb physiological chromatin interactions and have consequences on neighboring genes' expression. In addition, because of their coding capacities, retrocopies could theoretically still be transcribed and translated. Retrocopies have been distinguished in retrogenes and retropseudogenes based on their transcribed residual activity (Kaessmann, Vinckenbosch, & Long, 2009). This requires fortuitous coupling between the insertion site and a favorable regulatory genomic context. An additional copy of the transcript could then be disease-causing either because of a dose effect or through an ectopic expression. Tissue-specific expression of transcribed retrocopies has indeed been demonstrated to be closely related to the insertion site and not to their parental gene (Navarro & Galante, 2015). Finally, if the insertion occurs on the opposite strand compared with the original gene, a small interfering RNA could be produced leading to paradoxical haploinsufficiency (Guo, Zhang, Gerstein, & Zheng, 2009). While several examples of pathogenic retrocopies exist in animals (Parker et al., 2009; Sayah et al., 2004), to our knowledge, a single case of human pathogenic retrocopy has been reported in the literature. de Boer et al. (2014) reported a patient with chronic granulomatous due to abnormal splicing of *CYBB* caused by the partial exonization of a *TMF1* retrocopy. Here, we highlight the fact that retrocopies detected during NGS analyses may also be interpreted as benign, even within gene panels focusing on a few genes with a high *a priori* probability to be disease-causing.

Beyond these clinical aspects, retrocopies are also important to be recognized because of their misleading consequences on different molecular genetics techniques. Looking at SNVs, as seen here using short-read sequencing, aberrant splice site indels are called and will not be confirmed using Sanger sequencing. To address this, using RNA-Seq aligner as an add-on to standard pipelines has been proposed (Watson et al., 2017). Furthermore, SNVs phased to the retrocopy could yield false positive exonic SNV calls on the parent gene especially if PCR primers were selected in the exonic region (large exons). Regarding CNVs, using intronic or exonic-intronic qPCR primers, the apparent duplication would not be confirmed either. On the contrary, retrocopies would be seen as entire

duplications using MLPA but could mask genic deletions and provoke false negative results as has already been described for *SMAD4* (Millson et al., 2015). Because of backbone design and intronic probes, retrocopies are likely to be missed using cytogenetic microarray as has been the case for patients 1 and 2. Moreover, aberrant read pairs spanning introns and at the insertion site are responsible for structural variant calling. These can be misinterpreted as a chromosomal translocation as has already been the case for the *TYRO3* polymorphic retrocopy misinterpreted as a chr 15–chr 13 translocation (Brastianos et al., 2013). Finally, the presence of retrocopies also justifies the use of DNase treatment of RNAs for RT-PCR assays as they would mimic RNA and introduce false results as has been the case for Patient 8 (Menon, Chang, Clair, & Ham, 1991; Mutimer, Deacon, Crowe, & Sonza, 1998). Interpretation of RNA-Seq data should also be cautious for retrocopies present in the reference genome, as reads could be aligned ambiguously between the retrocopy and its parental transcript. Bioinformatics pipelines should take this risk into account.

We propose here a description that can enable the identification of retrocopies at a glance and thus (a) prevent from false interpretation, (b) warn about potential false positives and false negatives (depending on technique) due to such mechanism, and (c) help to figure out potential mechanisms of pathogenicity.

ACKNOWLEDGMENTS

We are thankful to Joris Veltman, Lisenka Vissers, Nathalie Drouot, Christian Gilissen, and Reza Maroofian. Gaël Nicolas acknowledges Fondation Bettencourt-Schueller, Fondation Philippe Chatrier, Fondation Charles Nicolle, and Association Cerveau Progrès. This study was supported by the NIHR Oxford Biomedical Research Centre (AOMW). Patient's 7 whole genome sequencing was funded by independent research commissioned by the Health Innovation Challenge Fund, a parallel funding partnership between Wellcome (WT 100127) and the UK Department of Health (R6-388) (principal investigator: Jenny C Taylor). The views expressed in this publication are those of the authors and not necessarily those of Wellcome or the Department of Health.

CONFLICT OF INTERESTS

The authors declared that there is no conflict of interests.

ORCID

Nicolas Chatron  <http://orcid.org/0000-0003-0538-0981>

Cedric Le Caignec  <http://orcid.org/0000-0002-0598-653X>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

TABLE 1 Patients, tests and results summary. All genomic coordinates are expressed on GRCh37/hg19 human reference genome. ddPCR stands for droplet digital PCR

Patient	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8
Sex	M	M	F	F	F	F	M	M
Age	69 years	6 months	10 years	19 years	55 years	3 weeks	Not known	12 years
Indication for referral	Attenuated adenomatous polyposis	Hydrocephalus	Intellectual disability	Intellectual disability	Early-onset Alzheimer disease	De novo balanced translocation breakpoint characterization	Complex chromosomal rearrangement characterization	Osteochondrodysplasia
Genetic test	11 genes panel sequencing (introns and exons)	Hydrocephalus panel sequencing (exons)	450 Gene panel for intellectual disability (exons)	Exome sequencing	Genome sequencing	Genome sequencing	Genome sequencing	Exome sequencing
Targeted technology	Agilent SureSelect QXT	Agilent SureSelect QXT	SeqCap EZ, Roche	Agilent SureSelect QXT	Not applicable	Not applicable	Not applicable	Agilent SureSelect XT
Sequencing machine	Illumina MiSeq	Illumina MiSeq	NextSeq, 500	Illumina HiSeq	BGISeq-500	NextSeq, 500	HiSeq, 4000	HiSeq
Read length	2 × 150	2 × 150	2 × 75	2 × 150	2 × 50	2 × 101	2 × 151	2 × 150
Alignment program	BWA v0.7.12	BWA v0.7.12	BWA-MEM v0.7.10	BWA v0.7.12	BWA v0.7.12	BWA-MEM v0.7.10	BWA-MEM v0.7.10	BWA v0.7.12
SV caller	GRIDSS v1.4.1	-- GRIDSS v1.4.1			BreakDancer v1.4.5	BreakDancer v1.4.5	BreakDancer v1.4.5	--
CNV caller	CANOES	CANOES	DeCovA	CANOES	ERDS v1.1	ERDS v1.1	ERDS v1.1	--
SNV caller	GATK Haplotype Caller 3.8	GATK Haplotype Caller 3.8	GATK Haplotype Caller 3.8	GATK Haplotype Caller 3.8	GATK Haplotype Caller 3.8	--	--	GATK Haplotype Caller 3.8
Results	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8
Retrocopied gene	<i>SMAD4</i>	<i>CCDC88C</i>	<i>DKC1</i>	<i>MTMR2</i>	<i>TYRO3</i>	<i>RCBTB1</i>	<i>C20orf27</i>	<i>RPL13</i>
Retrocopied transcript	ENS-T000003984-17	ENS-T0000553403	ENS-T000003695-50	ENS-T000003932-23	ENS-T000005590-66	ENS-T00000258646	ENS-T00000399672	ENS-T00000567-815
Copy	Yes (strictly)	Yes (strictly)	Yes (strictly)	Yes (strictly)	Yes (strictly)	Yes (strictly)	Yes (strictly)	.

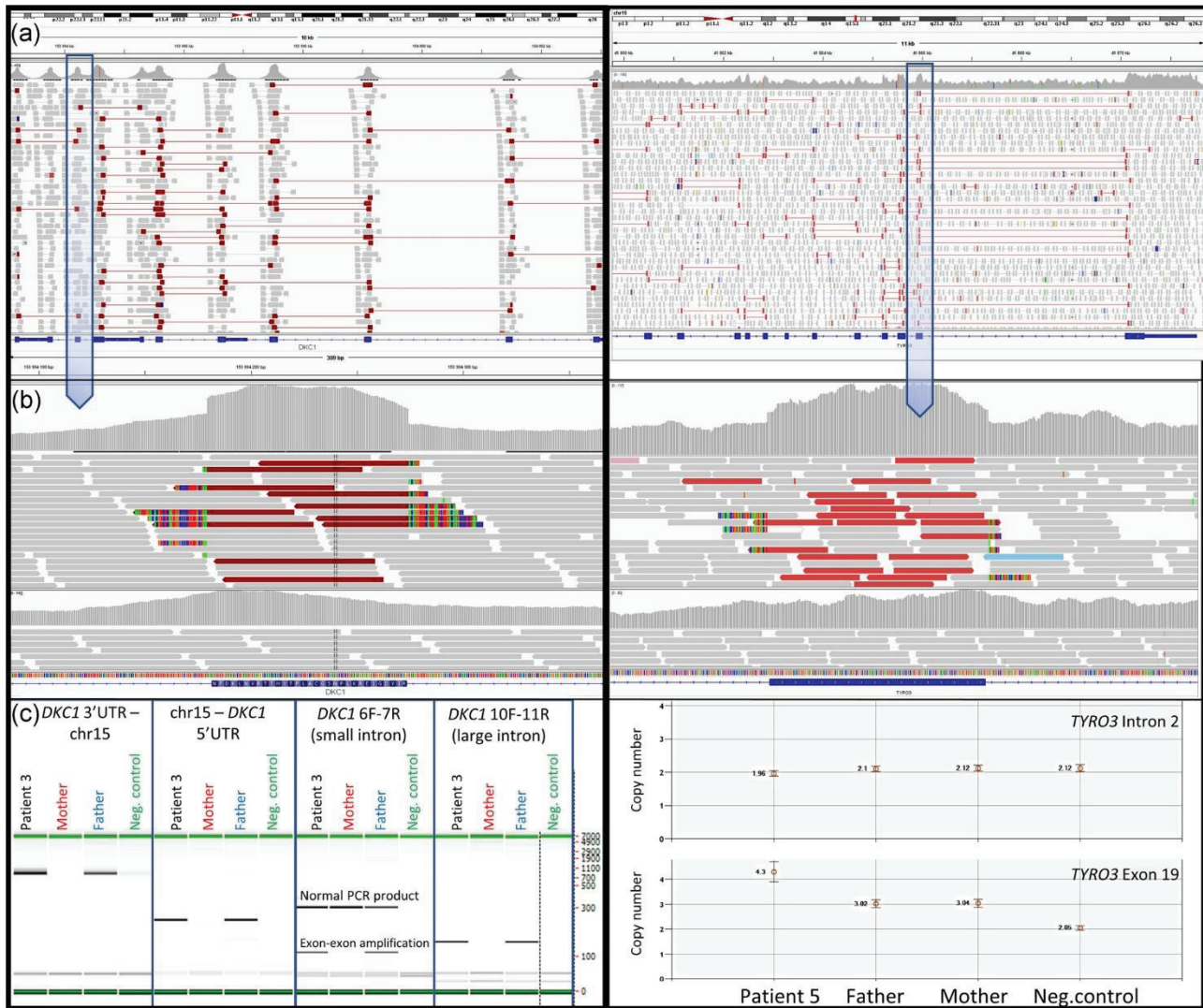
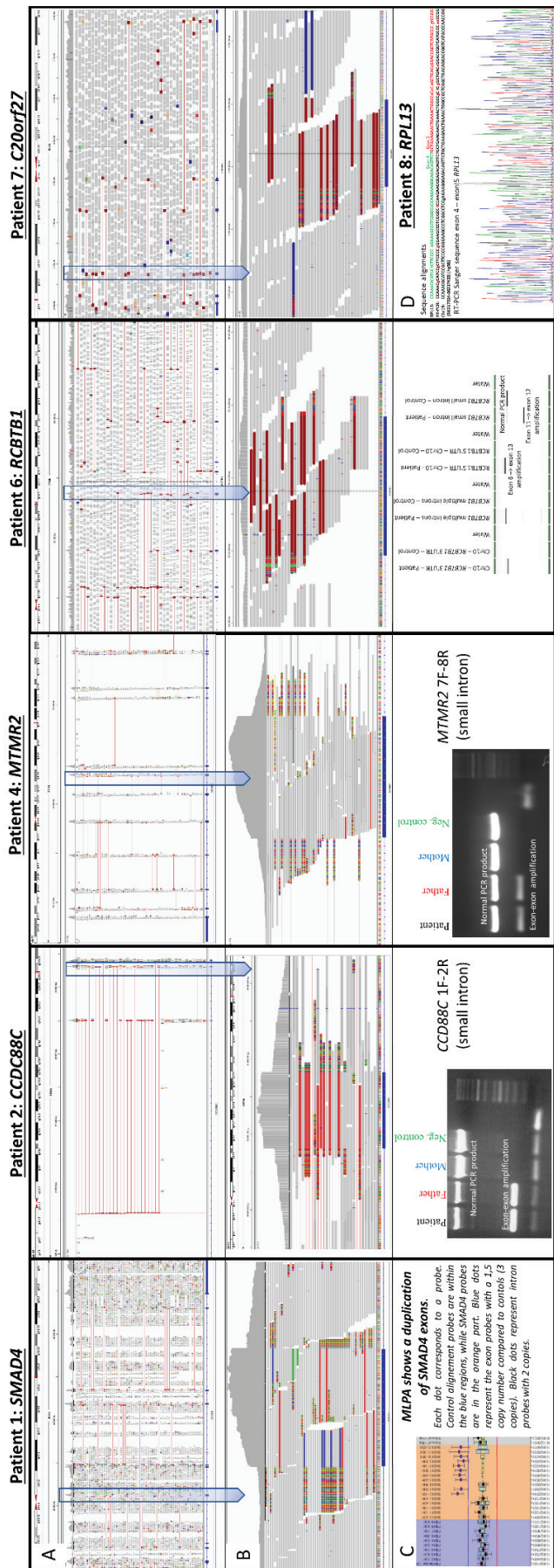


FIGURE 1 Short read sequencing alignment visualization and confirmation of retrocopy. IGV v2.3 visualization at gene scale for patient 3 (left panel, *DKC1* gene analyzed by targeted sequencing) and 5 (right panel, *TYRO3* gene analyzed by genome sequencing) showing aberrant read pairs (colored in red) spanning entire introns (a). When zooming at exon level (b) the depth of the coverage plot reveals an increase in exonic sequence with clear shifts at both exon-intron junctions. Aberrant read pairs are split reads with the nonreference sequence in introns. (c) Confirmation techniques: exon-exon and *DKC1* retrocopy-insertion site junction fragment PCR amplification for the trio of patient 3 and results of ddPCR CNV testing for patient 5 confirming homozygosity of *TYRO3* retrocopy

REFERENCES

- Backenroth, D., Homsy, J., Murillo, L. R., Glessner, J., Lin, E., Brueckner, M., ... Shen, Y. (2014). CANOES: Detecting rare copy number variants from whole-exome sequencing data. *Nucleic Acids Research*, *42*, e97–e97.
- deBoer, M., vanLeeuwen, K., Geissler, J., Weemaes, C. M., vandenBerg, T. K., Kuijpers, T. W., ... Roos, D. (2014). Primary immunodeficiency caused by an exonized retroposed gene copy inserted in the CYBB gene. *Human Mutation*, *35*, 486–496.
- Brastianos, P. K., Horowitz, P. M., Santagata, S., Jones, R. T., McKenna, A., Getz, G., ... Sunkavalli, A. (2013). Genomic sequencing of meningiomas identifies oncogenic SMO and AKT1 mutations. *Nature Genetics*, *45*, 285–289.
- Cameron, D. L., Schröder, J., Penington, J. S., Do, H., Molania, R., Dobrovic, A., ... Papenfuss, A. T. (2017). GRIDSS: Sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Research*, *27*, 2050–2060.
- Charbonnier, F., Raux, G., Wang, Q., Drouot, N., Cordier, F., Limacher, J. M., ... Frebourg, T. (2000). Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. *Cancer Research*, *60*, 2760–2763.
- Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., ... Fulton, R. S. (2009). BreakDancer: An algorithm for high-resolution mapping of genomic structural variation. *Nature Methods*, *6*, 677–681.
- Dewannieux, M., Esnault, C., & Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics*, *35*, 41–48.
- Dimassi, S., Simonet, T., Labalme, A., Boutry-Kryza, N., Campan-Fournier, A., Lamy, R., ... deBellescize, J. (2015). Comparison of two next-generation sequencing kits for diagnosis of epileptic disorders with a user-friendly tool for displaying gene coverage, DeCovA. *Appl Transl Genomics*, *7*, 19–25.
- Esnault, C., Maestre, J., & Heidmann, T. (2000). Human LINE retro-transposons generate processed pseudogenes. *Nature Genetics*, *24*, 363–367.
- Ewing, A. D., Ballinger, T. J., Earl, D., Harris, C. C., Ding, L., Wilson, R. K., ... Haussler, D. (2013). Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. *Genome Biology*, *14*, R22.
- Ewing, A. D., & Kazazian, H. H. (2010). High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Research*, *20*, 1262–1270.
- Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, *41*, 331–368.
- Guo, X., Zhang, Z., Gerstein, M. B., & Zheng, D. (2009). Small RNAs originated from pseudogenes: Cis- or trans-Acting? *PLoS Computational Biology*, *5*, e1000449.
- Hancks, D. C., Mandal, P. K., Cheung, L. E., & Kazazian, H. H. (2012). The minimal active human SVA retrotransposon requires only the 5'-hexamer and Alu-like domains. *Molecular and Cellular Biology*, *32*, 4718–4726.
- Kaessmann, H., Vinckenbosch, N., & Long, M. (2009). RNA-based gene duplication: Mechanistic and evolutionary insights. *Nature Reviews Genetics*, *10*, 19–31.
- Kazazian, H. H., & Moran, J. V. (2017). Mobile DNA in health and disease. *New England Journal of Medicine*, *377*, 361–370.
- Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., & Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*, *332*, 164–166.
- Kvikstad, E. M., Piazza, P., Taylor, J. C., & Lunter, G. (2018). A high throughput screen for active human transposable elements. *BMC Genomics*, *19*, 115.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Gage, D. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860–921.
- Lanikova, L., Kucerova, J., Indrak, K., Divoka, M., Issa, J.-P., Papayanno-poulou, T., ... Divoky, V. (2013). β -Thalassemia due to intronic LINE-1 insertion in the β -globin gene (*HBB*): Molecular mechanisms underlying reduced transcript levels of the β -Globin L1 allele. *Human Mutation*, *34*, 1361–1365.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760.
- MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L., & Scherer, S. W. (2014). The database of genomic variants: A curated collection of structural variation in the human genome. *Nucleic Acids Research*, *42*, D986–D992.
- McClintock, B. (1951). Chromosome organization and genic expression. *Cold Spring Harbor Symposia on Quantitative Biology*, *16*, 13–47.
- Menon, R. S., Chang, Y. F., Clair, J., St, & Ham, R. G. (1991). RT-PCR artifacts from processed pseudogenes. *PCR Methods and Applications*, *1*, 70–71.
- Millson, A., Lewis, T., Pesaran, T., Salvador, D., Gillespie, K., Gau, C.-L., Bayrak-Toydemir, P. (2015). Processed pseudogene confounding deletion/duplication assays for SMAD4. *J Mol Diagnostics*, *17*, 576–582.
- Mukherjee, S., Mukhopadhyay, A., Banerjee, D., Chandak, G. R., & Ray, K. (2004). Molecular pathology of haemophilia B: Identification of five novel mutations including a LINE 1 insertion in Indian patients. *Haemophilia*, *10*, 259–263.
- Mutimer, H., Deacon, N., Crowe, S., & Sonza, S. (1998). Pitfalls of processed pseudogenes in RT-PCR. *Biotechniques*, *24*, 585–588.
- Navarro, F. C. P., & Galante, P. A. F. (2015). A genome-wide landscape of retrocopies in primate genomes. *Genome Biology and Evolution*, *7*, 2265–2275.
- Parker, H. G., VonHoldt, B. M., Quignon, P., Margulies, E. H., Shao, S., Mosher, D. S., ... Acland, G. M. (2009). An expressed Fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science (80-)*, *325*, 995–998.
- Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., ... Schumann, G. (2012). The nonautonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Research*, *40*, 1666–1683.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*, 24–26.
- Sayah, D. M., Sokolskaja, E., Berthoux, L., & Luban, J. (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature*, *430*, 569–573.
- Schrider, D. R., Navarro, F. C. P., Galante, P. A. F., Parmigiani, R. B., Camargo, A. A., Hahn, M. W., & deSouza, S. J. (2013). Gene copy-number polymorphism caused by retrotransposition in humans. *PLoS Genetics*, *9*, e1003242.
- Smith, B. F., Yue, Y., Woods, P. R., Kornegay, J. N., Shin, J.-H., Williams, R. R., & Duan, D. (2011). An intronic LINE-1 element insertion in the dystrophin gene aborts dystrophin expression and results in Duchenne-like muscular dystrophy in the corgi breed. *Lab Invest*, *91*, 216–231.
- Watson, C. M., Camm, N., Crinnion, L. A., Antanaviciute, A., Adlard, J., Markham, A. F., ... Bonthron, D. T. (2017). Characterization and genomic localization of a SMAD4 processed pseudogene. *J Mol Diagnostics*, *19*, 933–940.
- Xing, J., Zhang, Y., Han, K., Salem, A. H., Sen, S. K., Huff, C. D., ... Jorde, L. B. (2009). Mobile elements create structural variation: Analysis of a complete human genome. *Genome Research*, *19*, 1516–1526.
- Zhang, Z., Harrison, P. M., Liu, Y., & Gerstein, M. (2003). Millions of years of evolution preserved: A comprehensive catalog of the processed pseudogenes in the human genome. *Genome Research*, *13*, 2541–2558.
- Zhu, M., Need, A. C., Han, Y., Ge, D., Maia, J. M., Zhu, Q., ... Gumbs, C. (2012). Using ERDS to infer copy-number variants in high-coverage genomes. *American Journal of Human Genetics*, *91*, 408–421.

How to cite this article: Chatron N, Cassinari K, Quenez O, et al. Identification of mobile retrocopies during genetic testing: consequences for routine diagnosis. *Human Mutation*. 2019;1–8. <https://doi.org/10.1002/humu.23845>



Supplemental Figure S1: Short read sequencing alignment visualization and confirmation of retrocopy.

IGV v2.3 visualization of short read alignment for patients 1, 2, 4, 6 and 7 from left to right at gene scale (A) and exon level (B). Aberrant read pairs are split reads with non-reference sequence in introns. (C) Confirmation techniques: MLPA for patient 1 confirming exonic duplication, and PCR amplification for patients 2, 4 and 6 showing unexpected exon-exon amplifications. (D) Patient 8: Sanger sequence of RT-PCR product amplified with *RPL13* exon 4 – exon 5 primers showing better alignment with chromosome 19 than with *RPL13* cDNA sequence.

II – 3/ Article 3 : *Chromoanagenesis* breakpoints are enriched in late-replicating chromatin.

Ce qui était connu :

Les techniques d'analyse chromosomique sur puce à ADN puis de séquençage de génome ont permis de révéler l'existence de remaniements chromosomiques complexes particuliers dénommés *chromoanagenesis*. Le profil de ces remaniements (points de cassure réunis sur un chromosome, équité des fragments inversés et non-inversés, profil du nombre de copies) a permis de formuler l'hypothèse d'une pulvérisation chromosomique en un seul temps suivi d'une réparation plus ou moins fidèle.

Les origines de cette « pulvérisation » sont encore méconnues. L'hypothèse selon laquelle un chromosome serait séquestré au sein d'un micronoyau permet d'expliquer le regroupement des points de cassure sur un chromosome mais le scénario complet n'est pas encore clair et la plupart des résultats obtenus le sont dans des conditions relativement extrêmes. Il s'agit généralement de modèles cellulaires mimant un contexte tumoral pour favoriser les aberrations génomiques (*knockout TP53*) et ayant subi une irradiation pour favoriser l'apparition de micronoyaux.

Bien que fréquents dans certaines tumeurs, ces remaniements sont encore exceptionnels dans des échantillons constitutionnels.

Nous rapportons ici la plus grande série de *chromoanagenesis* constitutionnels (n=20) groupés avec les données de la littérature de remaniements du même type tumoraux ou constitutionnels. L'analyse descriptive de la distribution des points de cassure qui a été menée avait pour but de pouvoir appuyer ou infirmer les hypothèses existantes à partir de données de patients et donc sans préjuger *a priori* des hypothèses existantes.

Ce que ce travail apporte :

Notre analyse confirme le regroupement quasi exclusif des points de cassure sur un seul chromosome dans une majorité de cas. Elle confirme la coexistence de *chromothripsis* (remaniements issus d'une pulvérisation chromosomique avec perte de certains fragments lors de la reconstruction) et *chromoanasythesis* (gains du nombre de copie obtenues par un mécanisme répliatif sans étape de pulvérisation). L'étude des fragments de jonction montre la difficulté de définir le mécanisme de réparation impliquée jonction par jonction et l'utilisation de ce critère pour distinguer *chromothripsis* et *chromoanasythesis*.

Par-delà la caractérisation de ces 20 réarrangements les deux questions biologiques de départ étaient :

- **Les *chromoanagenesis* tumoraux et constitutionnels partagent-ils les mêmes mécanismes originaux ?**
- **Remaniements simples et remaniements complexes sont-ils les deux extrêmes d'un même continuum biologique simplement séparés par un seuil arbitraire de nombre de points de cassure ? Ou bien sont-ils le résultat de mécanismes biologiques distincts ?**

Nous avons groupé nos 20 remaniements avec 20 autres cas de remaniements chromosomiques complexes de la littérature pour étudier la distribution de leurs 1028 points de cassure, celle des *chromoanagenesis* tumoraux (n = 13,310 points de cassure) et celle des remaniements simples (n = 468 points de cassure) et l'avons comparé à une répartition aléatoire simulée *in silico* (hypothèse nulle). Pour les 3 groupes de points de cassure (*chromoanagenesis* constitutionnels, tumoraux et remaniements simples), nous observons un enrichissement des points de cassure dans les régions se répliquant tardivement au cours du cycle cellulaire. La distribution des points de cassure des remaniements complexes apparaît proche d'une répartition aléatoire (hypothèse nulle) s'agissant des autres critères analysés (Topologically-Associated Domains (TADs), Lamina-Associated Domains (LADs), éléments répétés, marquage en bandes G, compartiment A/B, distance aux origines de réplication, sites de fragilité chromosomique). En revanche, pour les remaniements simples, il existe une déplétion significative en points de cassure dans les LADs. Ce dernier résultat infirme l'hypothèse d'un continuum entre remaniements simples et complexes.

La probabilité accrue de cassure dans les régions à réplication tardive appuie l'hypothèse d'une condensation prématurée d'un chromosome dans un micronoyau. Cette séquestration explique le regroupement des cassures sur un seul chromosome et est le lieu d'une réplication retardée. La condensation hâtive entraîne le décrochage des fourches de réplication encore actives dans les régions les plus en retard. Si les fourches rattachent d'autres fragments d'ADN, le chromosome reconstruit portera des cicatrices de ces mécanismes répliatifs. Si les fourches se décrochent complètement et font apparaître de véritables cassures double brin, les fragments seront raboutés le mieux possible.

Nos résultats suggèrent une origine commune des remaniements complexes mitotiques et méiotiques.

Le manuscrit présenté ci-dessous doit encore être relu par l'ensemble des co-auteurs et les données de séquençage générées soumises à une base de données avant qu'il ne soit soumis pour publication à *Genome Research* (IF = 10.101)

Nota bene : Le Tableau *Supplemental Table 1* est constitué de plus de 800,000 lignes et n'ai donc pas présenté ici mais disponible sur demande. Il sera disponible en ligne au moment de la publication.

1 **Title page**

2 Chromoanagenesis breakpoints are enriched in late-replicating chromatin.

3

4 **Authors**

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

14

15 **Authors' affiliations**

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

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

55

56

57 **Corresponding author:**

58 Dr Nicolas CHATRON

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

60 Groupement Hospitalier Est, Hospices Civils de Lyon

61 69677 Bron Cedex, France

62 Tel : +33471129640

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

64

65 **Keywords:**

66 chromoanagenesis, chromothripsis, replication-timing, lamina-associated domains, chromosomal

67 breakpoint

68

69 **Abstract**

70 The rise of pangenomic molecular assays allowed uncovering complex rearrangements named
71 *chromoanagenesis* and hypothesized to result from catastrophic shattering events. We characterized 20
72 new *chromoanagenesis* and report undescribed features. While literature differentiates shattering event
73 repaired through non-homologous end joining (*chromothripsis*) from *chromoanasyntesis* born to
74 aberrant replicative processes, we identify a grey zone of shattered chromosomes repaired through a
75 mixture of mechanisms. At sequence level, three rearrangements involve what we designate as a “hub”,
76 a kilobase locus fragmented and spread in several pieces throughout the rearrangement.

77 To assess the mechanisms at play, we merged our data with those of 20 published constitutional complex
78 chromosomal rearrangements cases. We evaluated if the distribution of their 1028 combined breakpoints
79 was distinctive using bootstrap simulations and found that breakpoints tend to keep away from
80 constrained genes suggesting purifying selection. We then compared their distribution with that of
81 13,310 and 468 breakpoints of cancer and simple rearrangements samples, respectively. Both complex
82 rearrangement groups showed breakpoint enrichment in late replicating regions suggesting similar
83 origins for constitutional and cancer cases. Simple rearrangements breakpoints but not complex ones
84 were depleted from lamina-associated domains (LADs), possibly as a consequence of reduced mobility
85 of DNA ends bound to lamina.

86 Observing breakpoint enrichment in late-replicating chromatin for both constitutional and cancer CCRs.,
87 we provide the first orthogonal demonstration of the premature chromosome condensation hypothesis
88 to explain *chromoanagenesis*.

89 Introduction

90 Since the identification of a supernumerary chromosome 21 in Down syndrome (Lejeune et al., 1959),
91 each technological development (chromosome banding, FISH, cytogenetic microarray, massive parallel
92 sequencing, long-read sequencing) has further exposed the complexity and variability of human
93 chromosomes (Caspersson et al., 1970; Cooper et al., 2011; Sudmant et al., 2015). This “complexity
94 entropy” climaxed with the identification of extremely complex chromosomal rearrangements (CCRs)
95 presenting specific hallmarks in cancer (Stephens et al., 2011) and constitutional samples (Kloosterman
96 et al., 2011a) christened *chromothripsis*. Their chromosomal breakpoints are clustered to a single
97 chromosome, a chromosome arm or even a cytoband and their copy number profile contrary to the
98 classical tumorigenesis scenario mostly oscillates between two states (1 and 2 copies). The equal amount
99 of inverted and non-inverted breakpoint-junctions is consistent with a single catastrophic shattering
100 event (*-thripsis* in Greek) followed by random repairing coupled to fragments loss (Korbel and
101 Campbell, 2013). Identification of different copy number profiles and alternative repairing mechanisms
102 prompted to also define *chromoanasythesis* (Liu et al., 2011) and *chromoplexy* (only in cancer)(Baca
103 et al., 2013), the three being grouped under the umbrella term *chromoanagenesis* (Holland and
104 Cleveland, 2012).

105 There are two main hypotheses for the underlying mechanisms, mostly deriving from cancer cell
106 observation and cellular models. Firstly, a non-spindle bound lagging chromosome is segregated in a
107 micronucleus (Ganem et al., 2009) where DNA replication is deficient (Crasta et al., 2012; Terzoudi et
108 al., 2015) generating two asynchronous compartments, the micronucleus and main nucleus. This
109 replicating and entrapped chromosome undergoes premature chromosome condensation (PCC), which
110 is associated with pulverization of chromatin (Kürten and Obe, 1975; Obe and Beek, 1975). This
111 hypothesis is consistent with the mostly biased parent-of-origin of constitutional *chromoanagenesis* as
112 paternal chromosomes are exposed to more cell divisions (Kong et al., 2012). The second model relies
113 on the dicentric chromosomes that are formed upon excessive telomere shortening and deficient cell
114 death mechanisms (Titen and Golic, 2008). Dicentric chromosomes can create bridges between the
115 cytokinesis poles that can only be resolved through multiple DNA breaks (Maciejowski et al., 2015),
116 sometimes with entrapment in a micronucleus (Pampalona et al., 2016).

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

123 (sequence, genome folding, genome function) and breakage probability in regard to complexity of the
124 rearrangement.

125 Here we take advantage of a cohort of rearrangements to assess possible differences in the distribution
126 of breakpoints in simple rearrangements, constitutional and cancer CCRs to investigate this potential
127 continuum and mechanisms at play.

128

129 **Results**

130 We recruited 14 unbalanced and 6 balanced *chromoanagenesis* cases. Through sequencing, we
131 identified thirteen to hundred breakpoints per rearrangement involving from one to thirteen
132 chromosomes for a total of 678 breakpoints (Table 1). An individual description of each cases is
133 provided as supplementary data. Based on our inferring criteria (see Methods) we classified 13 cases as
134 *chromoanasyntesis* and 4 as *chromothripsis*, while leaving 3 rearrangements without clear
135 denomination, mostly due to a limited number of inferable junctions.

136 This characterization effort uncovered noteworthy features. Individual 20 carries the most complex
137 *chromoanagenesis* described to date with a single derivative chromosome 14 made of sequences
138 originating from 13 different chromosomes. At several chromosome 14 breakpoints, we observe a
139 “juxtapositions of sequences” from chromosome to chromosome before joining chromosome 14 again
140 until the next “juxtaposition”. Two other single-chromosome cases (individuals 3 and 7) also present
141 with copy number gains embedded in junctions but all gains are chained and inserted as a single block.
142 Individual 10’s’s rearrangement also presents a single chain of gains inserted as a block but is unique as
143 gains are not inserted in their chromosome of origin, 5 gains of chromosome 1 being inserted in
144 chromosome 17 close to a pericentromeric inversion (Figure 1). For these latter three rearrangements
145 the number of identified junctions not implicating a copy number gain boundary is very limited and
146 chromosome shattering unlikely fitting with the *chromoanasyntesis* definition.

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

154 Finally, rearrangements of individuals 12, 13 and 15 are a composite of a shattered chromosome and
155 one or two translocations, a situation compatible with a preexisting parental balanced translocation or a
156 two-step process during a single gametogenesis.

157

158 *Genotype/phenotype correlation*

159 Nineteen patients recruited had an abnormal phenotype. The twentieth (Patient 7) was identified only
160 after a positive first trimester screening for trisomy 21 revealing an abnormal chromosome 21 in a fetus.
161 Fifteen patients presented with different degrees of intellectual disability from mild (P1, P18) to severe
162 (P14). When present intellectual disability was associated with different congenital malformations in all
163 patients but P6 and P17. Of the four patients without intellectual disability, two could not be assessed
164 formally as they were identified prenatally (Patient 13) or in the first days of life (Patient 12). We did
165 not observe any correlation between the phenotype severity and the rearrangement complexity regarding
166 either number of breakpoints, number of chromosome involved or genomic imbalance. However, for
167 several patients the phenotype could be mainly explained by the identification of disrupted genes or a
168 position effect. *FOXP1* (Mental retardation with language impairment (MIM 613670)) (Patient 1),
169 *MEF2C* (Mental retardation, autosomal dominant 20 (MIM 613443)) (Patient 5 and 14), *ADNP*
170 (Helsmoortel-van der Aa syndrome (MIM 615873) (Patient 6) , *ZEB2* (Mowat-Wilson syndrome (MIM
171 235730)) and *COL3A1* (vascular Ehler-Danlos syndrome (MIM 130050)) (Patient 11), *YY1* (Gabriele-
172 de Vries syndrome (MIM 617557)) (Patient 20) are driving the phenotypes. For Patients 2, 3, 4, 8, 9,
173 10, 12, 13, 17 and 19 no single gene could be identified but as the rearrangement had occurred *de novo*
174 or was shared with another affected carrier in the family (Patient 10) the genomic imbalance was
175 considered large enough to be causative. The striking absence of phenotype for Patient 7 while she
176 carries a total genomic imbalance of 4,85 Mb on chromosome 21 could be explained by the unique
177 insertion site of the copy number gains located on the short arm of chromosome 21. Placed in a
178 heterochromatin context the copy number gains could be silenced through a protective position effect.

179

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

181 We reached nucleotide resolution in both forward and reverse directions for 498/678 junctions allowing
182 inferring the repair mechanism through analysis of junction sequence (Supplementary Table 2). Using
183 our stringent criteria, 164/498 junctions could not be inferred. Non-homologous end-joining (NHEJ)
184 was inferred for 144 junctions (29%), while 121 junctions (24%) fitted with our criteria for
185 Microhomology-mediated breakage induced repair (MMBIR) / Fork Stalling and Template Switching
186 (FoSTeS). Three of these 121 replicative junctions could also be considered as Microhomology

187 Mediated End Joining (MMEJ) as criteria entirely overlap. For 69/498 junctions (14%) we could not
188 distinguish between NHEJ and MMBIR/FoSTeS. Only two rearrangements (individuals 10 and 14)
189 could be fully inferred with a single repairing mechanism (NHEJ or ambiguous), whereas all others
190 presented different levels of ascertained mixed signatures.

191 At sequence resolution, we describe unknown features that we named “rearrangement hubs”. They
192 consist of short genomic intervals (<1 kb) contacted by more than four junctions each coming from
193 different loci and using a few dozens of nucleotides of the hub before restarting another rearrangement
194 loop. We identified 5 of these in individuals 9 (n=2), 12 (n=2) and 19 (n=1). The most complex “hub”
195 presents with 12 junctions within a 211 nucleotides span (individual 12; Figure 2).

196

197 *Purifying selection affects breakpoint distribution of constitutional rearrangements*

198 We then investigated CCR breakpoint distribution to gain insights on potential biological mechanisms.
199 To increase biological relevance and statistical power we combined the 678 breakpoints of our 20
200 constitutional CCRs with the 350 breakpoints of 20 published cases (Yang et al., 2015; Redin et al.,
201 2017). To investigate if the mechanisms involved in constitutional and cancer rearrangements are similar
202 we compared the distribution of these 1028 breakpoints to that of 13,310 breakpoints of cancer CCRs
203 deposited in ChromothripsisDB (Yang et al., 2015). We also compared their distribution with that of
204 468 breakpoints of 234 simple rearrangements described in (Redin et al., 2017) and (Schluth-Bolard et
205 al., 2019) to gauge the possible continuum from simple to complex rearrangements. In parallel we
206 assessed if the breakpoint distribution of these three different groups of rearrangements were enriched
207 within diverse genomic features, such as constrained genes, repeat elements, G-band staining,
208 chromosomal fragile sites, A/B compartments (i.e. open/expression-active and closed/expression-
209 inactive chromatin), lamina (LADs) and topologically associated domains (TADs), replication origins
210 and replication timing. Direction of the distribution (enrichment/depletion) and statistical significance
211 were gauged by both comparing to a random distribution, in which each breakpoint could occur
212 anywhere in the genome, and a “constrained” distribution that better reflects the clustering of
213 breakpoints on a single chromosome observed in *chromoanagenesis* (see Methods). All annotations are
214 presented in Supplementary Table 1.

215 As our samples and the published cases were from liveborn individuals (made exception of our 13
216 individual) we first assessed purifying selection. Indeed, beyond genes affected by copy number variants
217 we surmise that accumulation of dysregulated genes (directly disrupted or through improper chromatin
218 folding (Robson et al., 2019) would reduce fitness. Consistent with this hypothesis, we observed a
219 depletion of both gene-disrupting and (Relative Risk Ratio (RRR) = 0.816, $p = 0.0015$) (Table 2) and
220 TAD-disrupting breakpoints in constitutional CCRs (RRR = 0.870, $p = 0.0456$). The gene-disrupting

221 breakpoints of simple rearrangements tend to affect haplosensitive genes (higher loss-of-function
222 intolerance (pLI); RRR= 2.256, $p = 9.39 \times 10^{-7}$), probably mirroring the ascertainment bias in recruiting
223 these affected individuals. On the contrary we observed a nominally significant depletion of high pLI
224 genes in CCRs (RRR= 0.752 ; $p = 0.0219$), suggesting that their pathologies have an oligogenic basis
225 with multiple breakpoint contributing and/or that affecting multiple high pLI genes is incompatible with
226 life. We also observed a significant enrichment in genes and TADs in cancer rearrangements (RRR =
227 1.07 and 1.113; $p = 1.256 \times 10^{-4}$ and 2.195×10^{-9} respectively) consistent with a distribution driven by oncogenic
228 forces rather than purifying selection.

229

230 *Chromosomal breakpoints are enriched in late-replicating chromatin*

231 Aside from purifying selection consequences, we identified several genomic features affecting
232 breakpoint distribution. Both groups of CCRs had their breakpoints depleted in open chromatin
233 compartment and enriched in repeat elements, G-stained cytobands, LADs and late replicating regions
234 of the chromatin. Constitutional CCRs breakpoints were depleted in fragile sites in both simulations
235 (only free simulation for cancer CCRs) (Table 2). However, we also confirmed an important and
236 complex interplay between these variables with LADs, for example, being known to be late-replicating
237 gene poor regions (Supplementary Figure 1). To resolve all covariables and confounders (including
238 purifying selection elements) that influence the distinction between simple and CCRs (constitutional
239 and cancer) we built a multinomial logistic regression model. We reached the best Akaike information
240 criterion using as covariables disruption of a repeated element, a gene, a lamina associated domain, a
241 topologically associated domain, chromatin compartment, G-band staining, replication timing status and
242 distance to replication origin (Table 3). In this model, if all these elements contribute to explain
243 breakpoint distribution, replication timing appears as the unique significant element in the three groups,
244 indeed chromosomal breakpoints are enriched in late-replicating chromatin (RRR = 3.629, $p = 4.203 \times 10^{-12}$
245 for constitutional CCRs; RRR = 1.431, $p = 7.194 \times 10^{-14}$ for cancer CCRs; RRR = 2.149, $p = 0.0050$ for
246 simple rearrangements)(Table 3).

247 For constitutional *chromoanagenesis*, the contribution of repeat element disruption and distance from
248 replication origin (log-transformed) seem modest with coefficients being between 0.9 and 1.1. Beyond
249 replication timing, depletion in genes (RRR = 0.58, $p = 0.0003$) and enrichment in open chromatin
250 regions (RRR = 1.49, $p = 0.0159$) appear as important elements. For cancer rearrangements, none of the
251 analyzed covariables but repeat element disruption had a coefficient out of the [0.9-1.1] interval
252 suggesting breakpoint distribution to be close to random once adjusted for replication timing.

253

254 *Simple rearrangements' breakpoints are depleted in lamina associated domains*

255 Finally, while late-replication is known to be a characteristic of LADs, simple rearrangements'
256 breakpoints appeared depleted within LADs (RRR= 0.456 ; $p = 0.0076$). Conversely, complex
257 rearrangements' breakpoints were independent of LADs as their distribution is comparable to breakpoint
258 simulation (RRR= 0.987 and 1.041 (NS) for cancer and constitutional cases respectively).

259 In summary, preference towards non-peripheral, i.e. non lamina-associated, chromatin appears as the
260 unique criterion distinguishing the distribution of simple rearrangement breakpoints' from that of
261 complex cases that are naive to this feature.

262

263 **Discussion**

264 Using short-read genome sequencing we were able to characterize extremely complex chromosomal
265 rearrangements. With a total of 20 cases, this is the most important cohort of constitutional CCRs to
266 date offering new insights on the origins and mechanisms of such rearrangements *in vivo*. While
267 individual 20 carries the most complex constitutional *chromoanagenesis* reported to date with a total of
268 100 breakpoints, short-read sequencing stays blind to some challenging regions (short arm of acrocentric
269 chromosomes for individuals 12, 13 and 16). Paired-end short-read sequencing only identifies 91% of
270 breakpoints visible on the karyotype (Redin et al., 2017). In such complex cases, even when the
271 rearrangement seems resolved, the presence of an extra loop of junctions cannot be ruled out.
272 Considering our downstream analysis of breakpoint distribution, as we excluded from our simulations
273 genomic regions where structural variants are difficult to call using short reads, this shall not bias our
274 interpretation of genomic features surrounding chromosomal breakpoints.

275 *Grey zone between chromothripsis and chromoanasyntesis*

276 Our stringent criteria used to infer repairing mechanism limited the number of inferred junctions and
277 our ability to classify 3 rearrangements in a subgroup of *chromoanagenesis*. However, the observations
278 made here tend to question these subgroups for constitutional CCRs. Indeed, within complexity different
279 profiles emerged independently from subgroup classification.

280 Literature definitions and criteria used herein do not disentangle situations where gains are inserted in a
281 single block (individuals 3, 7 and 10) from replicative repair of a high number of breakpoints in a single
282 chromosome "using" a variety of other chromosomes. The latter situation appears as an a grey zone
283 between *chromothripsis* and *chromoanasyntesis*. In *chromoanasyntesis*, the complexity rather than
284 resulting from chromosome shattering is the consequence of a DNA break, or a simple replication fork
285 stalling resolved "using" other sequences of the genome through a replicative process
286 (MMBIR/FoSTeS) (Liu et al., 2011; Holland and Cleveland, 2012). Noteworthy, the fact that none of

287 the eight junctions resolved for individual 10 could be inferred as replicative questions our ability to
288 infer a repairing mechanism. In the grey zone, a shattering event seems likely but junction sequences
289 present distinct signs from what should be observed with classical non-homologous end joining in
290 *chromothripsis* (Kloosterman et al., 2011a; Chiang et al., 2012).

291 *Chromothripsis* has been used extensively to designate all kinds of *chromoanagenesis* (Pellestor, 2014;
292 Fukami et al., 2017; Redin et al., 2017; Collins et al., 2019) so that it is challenging to know whether
293 such grey zone could have been suspected before. We emphasize that this simplification towards
294 *chromothripsis* is incorrect with a limited number of constitutional cases being clearly *chromothripsis*
295 in our experience.

296 With repairing “hubs” we describe a new feature of complex rearrangements where a single locus is
297 “used” several times within the derivative chromosome. If handling of short sequences within CCR had
298 been reported before (Slamova et al., 2018), these “hubs” had never been described before while we
299 found them in 3 out of 20 cases. No common feature, but G-band positive staining, between the 5 hubs
300 we identified appear. It will be of particular interest to collect additional observations and then
301 investigate the genomic features shared by these loci. As Slamova *et al.* pointed before the fact that
302 extremely short sequences can be handled, and here even spread throughout a chromosomal
303 rearrangement, challenges the 50 bp definition of a structural variant (Sudmant et al., 2015; Slamova et
304 al., 2018).

305 While there are no CCR in the 1000 genomes project database (n=2504) (Sudmant et al., 2015), a single
306 one was identified in the gnomAD dataset (n=14,891) (Collins et al., 2019). Interestingly, of the 13
307 rearrangements we included based on their abnormal array-CGH result, they all presented additional
308 breakpoints and complexity. The coincidence of three non-recurrent non-polymorphic CNVs identified
309 during array-CGH on a single chromosome is a good indication for an unbalanced CCR and shall prone
310 to genome sequencing for detailed characterization.

311 *Late replication timins is a risk factor for chromosomal damage*

312 The observed enrichment of gene and TAD disruption in cancer rearrangements in our univariate
313 analysis is in line with the recent retrospective analysis of 75,710 cancer karyotypes where enrichment
314 in gene-rich region was observed (Mitelman et al., 2019). However, once adjusted for covariables in our
315 logistic regression model, these signals disappear. Mitelman *et al.* explained that these enrichments seem
316 to be a common feature of both simple and complex cancer rearrangements and reflect the importance
317 of structural variants as oncogenic factors rather than genomic noise but did not adjust on covariables
318 (Mitelman et al., 2019). Our results tend to balance interpretation towards “genomic noise” for complex
319 cases.

320 Interestingly, apart from variables affected by purifying selection we could not distinguish constitutional
321 *chromoanagenesis* from cancer cases. This suggests that the shattering scenario is comparable for
322 constitutional and tumoral cases. However, while the influence of each variable goes in the same
323 direction for both cancer and constitutional cases, the observed driving forces are still of different
324 magnitude. We hypothesize that this could result from additional complexity delineating possible
325 subgroups. Drier *et al.* have previously shown that cancer rearrangements' breakpoint, without detailing
326 complexity of each rearrangement, has a bimodal distribution, some cancers having breakpoints in early-
327 replicating, highly expressed, GC-rich chromatin while others had their breakpoints preferentially
328 positioned in late-replicating, low transcribed, GC-poor chromatin (Drier et al., 2013). Focusing on
329 CCRs, breakpoints of sarcoma samples are enriched in early-replicating chromatin (Anderson et al.,
330 2018) oppositely to our finding pooling all cancers together.

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

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

350 In our analysis, late replication timing appears as a key risk factor for *chromoanagenesis* and
351 chromosomal breakage more broadly as it is for other types of genomic variation. It was first associated
352 with important variations in SNP-density across evolution and cancer development with 2-fold and 6-
353 fold enrichment for transitions and transversions mutations respectively (Koren et al., 2012), but profiles
354 are similar in germline and cancer cells (Liu et al., 2013). One possible explanation is that error-prone

355 repairing mechanisms are more active in late replication (Pol. Theta) while early-replicating gene-dense
356 chromatin benefits from high fidelity transcription-coupled nucleotide excision repair (Waters and
357 Walker, 2006). Interestingly, polymerase Theta had previously been considered to explain the features
358 observed in a *chromoanagenesis* case (Masset et al., 2016). CNV distribution was also reported to be
359 influenced by replication timing, gains and losses being enriched in early and late regions respectively
360 in cancer cells, with breakpoints being part of the same replicating domain. However, to date, there is
361 no robust explanation for such findings (De and Michor, 2011). Finally, if partners of balanced
362 rearrangements preferentially replicate at the same time (Ryba et al., 2010) preference towards late
363 replicating regions had never been reported before.

364 *Breakpoint enrichment in late replicating chromatin as an orthogonal demonstration of the PCC*
365 *hypothesis*

366 Based on our observations, our scenario for *chromoanagenesis* is that after one (or a few)
367 chromosome(s) got trapped in a micronucleus it can only replicate slower compared to the main nucleus.
368 While entering the next cell division, replication is not completed in the micronucleus and the entrapped
369 material has to start condensation prematurely. This generates a stress on replication forks still present
370 in the physiologically late replicating regions of this delayed chromosome. The forks would either
371 collapse and provoke DNA breakage or stall and start FoSTeS cycles. This would explain the co-
372 occurrence of different repairing mechanisms observed in a same rearrangement. The rearrangement
373 puzzle will be reassembled after reintegration within the main nucleus, opening the possibility for
374 FoSTeS loops to use other chromosomes.

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

384 Overall, we highlight the importance of precise characterization of CCRs as we were able to identify (i)
385 rearrangement “hubs” that had not been described previously and a grey zone between *chromothripsis*
386 and *chromoanagenesis* challenging actual definitions. By extensive annotation of the largest cohort of
387 constitutional CCRs and the use of publicly available data, we identify replication timing as a new
388 element driving chromosomal breakpoint probability. Thus, our *in vivo* dataset implies premature

389 condensation of genetic material segregated in a micronucleus. This new hallmark shall be used to
390 validate the relevance of cellular models and improve knowledge of complex chromosomal
391 rearrangements. Finally, we identify a statistically significant depletion of simple rearrangements'
392 breakpoints within LADs. This strongly suggests that *chromoanagenesis* and simple rearrangements are
393 not part of a continuous spectrum and opens new perspectives towards chromosomal rearrangement
394 understanding.

395

396 **Material and Methods**

397 *Patients*

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 included after
401 genome sequencing results fitting this criterion while this patient was not tested through cytogenetic
402 microarray. Second, we included all balanced rearrangements characterized in our laboratory as part of
403 our routine activity having more than 10 breakpoints.

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

408

409 *Genome sequencing*

410 All cases, but case 10, were genome sequenced using our validated approach (Schluth-Bolard et al.,
411 2019). Briefly, after PCR-free library preparation (Nano, Illumina Inc., San Diego, CA, USA), 2x100
412 bp paired-end sequencing was performed on a 300 cycles High Output FlowCell for NextSeq500
413 instrument (Illumina, Inc.). Additionally, samples 19, 2, 3, 6, 9, 10 and 19 were 2x151 bp paired-end
414 sequenced on a HiSeq X5 instrument or a HiSeq4000 instrument (Illumina Inc.) for the latter.
415 Breakpoints were detected using BreakDancer (Chen et al., 2009) and ERDS was used for CNV calling
416 (Zhu et al., 2012). For all calls, supporting alignment data was systematically inspected using the IGV
417 v2.4 software with soft-clipped reads set as visible (Robinson et al., 2011). At each breakpoint, split-
418 reads were extracted manually to align the soft-clipped sequence using BLAT and obtain fragment
419 junction sequence directly (Kent, 2002). For all junctions, the process was repeated in both directions
420 of the junction.

421

422 *Inferring mechanisms*

423 Repairing mechanisms were inferred following the same criteria as in (Schluth-Bolard et al., 2019). The
424 same mechanism had to be consider for both strands of a junction to infer it. Based on existing literature,
425 we established a decisional algorithm to infer the *chromoanagenesis* subgroup of each rearrangement
426 as a whole (Liu et al., 2011; Holland and Cleveland, 2012; Baca et al., 2013; Korbel and Campbell,
427 2013). First, rearrangements presenting several copy number gains are classified as *chromoanasythesis*
428 as the concept of *chromothripsis* does not foresee genomic gains. Second, the rearrangement is
429 considered as *chromothripsis* or *chromoanasythesis* depending on the dominant repairing mechanism
430 (NHEJ for *chromothripsis* ; MMBIR/FoSTeS or clear FoSTeS events with several shifts for
431 *chromoanasythesis*) with a minimum of 8 inferred junctions to qualify the rearrangement.

432

433 *Breakpoint annotation*

434 Considering extreme complexity, all breakpoints distant from less than a kilobase were considered as
435 one for further analyses. We annotated each breakpoint with publicly available datasets using Svagga
436 (<https://gitlab.inria.fr/NGS/svagga>). RefSeq genes coordinates were extracted from the UCSC Genome
437 Browser; pLi scores were extracted from the gnomAD dataset (Lek et al., 2016). Annotation was also
438 made across lamina associated domains (LADs) (Guelen et al., 2008); topologically associated domains
439 (TADs) (Rao et al., 2014) ; repeated elements (RepeatMasker); chromosomal cytoband and Giemsa
440 staining (Furey and Haussler, 2003); chromosomal common fragile sites (Fungtammasan et al., 2012);
441 replication origins (Massip et al., 2019) ; A/B compartments (Fortin and Hansen, 2015) and replication
442 timing (Dixon et al., 2018). For the latter, only constitutive early and late domains were considered.

443

444 *External cases*

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

454 simulation data, both distributions were either not significant or in the same direction considering
455 individual genomic features (Supplementary Table 3).

456 Constitutional simple rearrangements published in (Schluth-Bolard et al., 2019) and (Redin et al., 2017)
457 were pooled in a third group above (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 of
467 the reference genome cannot or poorly be analyzed for structural variants, we used the “DangerTrack”
468 bed file (Dolgalev et al., 2017) to exclude all breakpoints simulated in these regions and be comparable
469 with the observed dataset. A single breakpoint of our in-house dataset overlaps the DangerTrack.
470 Overall, 399,346 and 399,915 were simulated and used for analyses for the “free” and “constrained”
471 models respectively.

472

473 *Statistical analysis*

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

478

479 **Data access**

480 All raw and processed sequencing data generated in this study will be submitted to the NCBI Gene
481 Expression Omnibus (GEO;).

482

483 **Acknowledgments**

484 The authors would like to thank Isabelle Rouvet and Emilie Chopin for lymphoblastoid cell line support
485 (Cellular Biotechnology Center, Hospices civils de Lyon, France). NC acknowledges « Fondation
486 Thérèse et René Planiol », Réseau Régional de Rééducation et de Réadaptation Pédiatrique en Rhône
487 Alpes and IDEX Lyon and Lyon University Hospital. GG is recipient of a Pro-Women Scholarship from
488 the Faculty of Biology and Medicine, University of Lausanne. This work was supported by grants from
489 the Fondation Maladies Rares (WGS20140103) to DS, from the French Ministry of Health (DGOS) and
490 the French National Agency for Research (ANR) (PRTS 2013 : PRTSN1300001N) to CSB, the Swiss
491 National Science Foundation (31003A_182632) and Horizon2020 Twinning project ePerMed (692145)
492 to AR.

493 **Disclosure declaration**

494 The authors have no conflict of interest to declare.

495 **References**

- 496 Ader F, Heide S, Marzin P, Afenjar A, Diguët F, Chantot Bastaraud S, Rollat-Farnier P-A, Sanlaville
497 D, Portnoï M-F, Siffroi J-P, Schluth-Bolard C. 2019. A 14q distal chromoanagenesis elucidated by
498 whole genome sequencing. *Eur J Med Genet* 103776.
- 499 Anderson ND, Borja R de, Young MD, Fuligni F, Rosic A, Roberts ND, Hajjar S, Layeghifard M,
500 Novokmet A, Kowalski PE, Anaka M, Davidson S, et al. 2018. Rearrangement bursts generate canonical
501 gene fusions in bone and soft tissue tumors. *Science* 361:.
- 502 Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N,
503 MacDonald TY, Ghandi M, Allen E Van, Kryukov G V, et al. 2013. Punctuated evolution of prostate
504 cancer genomes. *Cell* 153:666–77.
- 505 Caspersson T, Zech L, Johansson C. 1970. Differential binding of alkylating fluorochromes in human
506 chromosomes. *Exp Cell Res* 60:315–9.
- 507 Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang
508 Q, Locke DP, Shi X, Fulton RS, et al. 2009. BreakDancer: an algorithm for high-resolution mapping of
509 genomic structural variation. *Nat Methods* 6:677–81.
- 510 Chiang C, Jacobsen JC, Ernst C, Hanscom C, Heilbut A, Blumenthal I, Mills RE, Kirby A, Lindgren
511 AM, Rudiger SR, McLaughlan CJ, Bawden CS, et al. 2012. Complex reorganization and predominant
512 non-homologous repair following chromosomal breakage in karyotypically balanced germline
513 rearrangements and transgenic integration. *Nat Genet* 44:390–7, S1.

514 Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Khera A V., Francioli LC, Gauthier LD, Wang
515 H, Watts NA, Solomonson M, O'Donnell-Luria A, et al. 2019. An open resource of structural variation
516 for medical and population genetics. *bioRxiv* 578674.

517 Cooper GM, Coe BP, Girirajan S, Rosenfeld J a, Vu TH, Baker C, Williams C, Stalker H, Hamid R,
518 Hannig V, Abdel-Hamid H, Bader P, et al. 2011. A copy number variation morbidity map of
519 developmental delay. *Nat Genet* 43:838–46.

520 Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova E V, Pan Y, Nezi L, Protopopov A,
521 Chowdhury D, Pellman D. 2012. DNA breaks and chromosome pulverization from errors in mitosis.
522 *Nature* 482:53–8.

523 De S, Michor F. 2011. DNA replication timing and long-range DNA interactions predict mutational
524 landscapes of cancer genomes. *Nat Biotechnol* 29:1103–1108.

525 Dixon JR, Xu J, Dileep V, Zhan Y, Song F, Le VT, Yardımcı GG, Chakraborty A, Bann D V., Wang
526 Y, Clark R, Zhang L, et al. 2018. Integrative detection and analysis of structural variation in cancer
527 genomes. *Nat Genet* 50:1388–1398.

528 Dolgalev I, Sedlazeck F, Busby B. 2017. DangerTrack: A scoring system to detect difficult-to-assess
529 regions. *F1000Research* 6:443.

530 Drier Y, Lawrence MS, Carter SL, Stewart C, Gabriel SB, Lander ES, Meyerson M, Beroukhim R, Getz
531 G. 2013. Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA
532 breakage and rearrangement-induced hypermutability. *Genome Res* 23:228–35.

533 Fortin J-P, Hansen KD. 2015. Reconstructing A/B compartments as revealed by Hi-C using long-range
534 correlations in epigenetic data. *Genome Biol* 16:180.

535 Fukami M, Shima H, Suzuki E, Ogata T, Matsubara K, Kamimaki T. 2017. Catastrophic cellular events
536 leading to complex chromosomal rearrangements in the germline. *Clin Genet* 91:653–660.

537 Fungtammasan A, Walsh E, Chiaromonte F, Eckert KA, Makova KD. 2012. A genome-wide analysis
538 of common fragile sites: What features determine chromosomal instability in the human genome?
539 *Genome Res* 22:993–1005.

540 Furey TS, Haussler D. 2003. Integration of the cytogenetic map with the draft human genome sequence.
541 *Hum Mol Genet* 12:1037–1044.

542 Ganem NJ, Godinho SA, Pellman D. 2009. A mechanism linking extra centrosomes to chromosomal
543 instability. *Nature* 460:278–282.

544 Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, Klein A de, Wessels L,
545 Laat W de, Steensel B van. 2008. Domain organization of human chromosomes revealed by mapping
546 of nuclear lamina interactions. *Nature* 453:948–951.

547 Hatch EM, Fischer AH, Deerinck TJ, Hetzer MW. 2013. Catastrophic Nuclear Envelope Collapse in
548 Cancer Cell Micronuclei. *Cell* 154:47–60.

549 Holland AJ, Cleveland DW. 2012. Chromoanagenesis and cancer: mechanisms and consequences of
550 localized, complex chromosomal rearrangements. *Nat Med* 18:1630–8.

551 Hsu TC. 1975. A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics*
552 79 Suppl:137–50.

553 Kent WJ. 2002. BLAT---The BLAST-Like Alignment Tool. *Genome Res* 12:656–664.

554 Kloosterman WP, Guryev V, Roosmalen M van, Duran KJ, Bruijn E de, Bakker SCM, Letteboer T,
555 Nesselrooij B van, Hochstenbach R, Poot M, Cuppen E. 2011. Chromothripsis as a mechanism driving
556 complex de novo structural rearrangements in the germline. *Hum Mol Genet* 20:1916–24.

557 Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson
558 A, Jonasdottir A, Jonasdottir A, Wong WSW, Sigurdsson G, et al. 2012. Rate of de novo mutations and
559 the importance of father's age to disease risk. *Nature* 488:471–475.

560 Korbel JO, Campbell PJ. 2013. Criteria for inference of chromothripsis in cancer genomes. *Cell*
561 152:1226–36.

562 Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, McCarroll SA. 2012. Differential
563 Relationship of DNA Replication Timing to Different Forms of Human Mutation and Variation. *Am J*
564 *Hum Genet* 91:1033–1040.

565 Kürten S, Obe G. 1975. Premature chromosome condensation in the bone marrow of Chinese hamsters
566 after application of bleomycin in vivo. *Mutat Res Mol Mech Mutagen* 27:285–294.

567 Lejeune J, Gautier M, Turpin R. 1959. Etude des chromosomes somatiques de 9 enfants mongoliens. *C*
568 *R Hebd Seances Acad Sci* 248:1721–2.

569 Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware
570 JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, et al. 2016. Analysis of protein-coding genetic
571 variation in 60,706 humans. *Nature* 536:285–291.

572 Lemaître C, Grabarz A, Tsouroula K, Andronov L, Furst A, Pankotai T, Heyer V, Rogier M, Attwood
573 KM, Kessler P, Dellaire G, Klaholz B, et al. 2014. Nuclear position dictates DNA repair pathway choice.
574 *Genes Dev* 28:2450–2463.

575 Liu L, De S, Michor F. 2013. DNA replication timing and higher-order nuclear organization determine
576 single-nucleotide substitution patterns in cancer genomes. *Nat Commun* 4:1502.

577 Liu P, Erez A, Nagamani SCS, Dhar SU, Kołodziejaska KE, Dharmadhikari A V, Cooper ML,
578 Wiszniewska J, Zhang F, Withers MA, Bacino CA, Campos-Acevedo LD, et al. 2011. Chromosome
579 catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell*
580 146:889–903.

581 Maciejowski J, Li Y, Bosco N, Campbell PJ, de Lange T. 2015. Chromothripsis and Kataegis Induced
582 by Telomere Crisis. *Cell* 163:1641–1654.

583 Masset H, Hestand MS, Esch H Van, Kleinfinger P, Plaisancié J, Afenjar A, Mollignier R, Schluth-
584 Bolard C, Sanlaville D, Vermeesch JR. 2016. A Distinct Class of Chromoanagenesis Events
585 Characterized by Focal Copy Number Gains. *Hum Mutat*.

586 Massip F, Laurent M, Brossas C, Fernández-Justel JM, Gómez M, Prioleau M-N, Duret L, Picard F.
587 2019. Evolution of replication origins in vertebrate genomes: rapid turnover despite selective
588 constraints. *Nucleic Acids Res* 47:5114–5125.

589 Mitelman F, Johansson B, Mertens F, Schyman T, Mandahl N. 2019. Cancer chromosome breakpoints
590 cluster in gene-rich genomic regions. *Genes, Chromosom Cancer* 58:149–154.

591 Obe G, Beek B. 1975. The human leucocyte test system. VII. Further investigations concerning
592 micronucleus-derived premature chromosome condensation. *Humangenetik* 30:143–54.

593 Pampalona J, Roscioli E, Silkworth WT, Bowden B, Genescà A, Tusell L, Cimini D. 2016. Chromosome
594 Bridges Maintain Kinetochore-Microtubule Attachment throughout Mitosis and Rarely Break during
595 Anaphase. *PLoS One* 11:e0147420.

596 Pellestor F. 2014. Chromothripsis: how does such a catastrophic event impact human reproduction?
597 *Hum Reprod* 29:388–393.

598 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol
599 I, Omer AD, Lander ES, Aiden EL. 2014. A 3D Map of the Human Genome at Kilobase Resolution
600 Reveals Principles of Chromatin Looping. *Cell* 159:1665–1680.

601 Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, Hanscom C, Pillalamarri V, Seabra
602 CM, Abbott M-A, Abdul-Rahman OA, Aberg E, et al. 2017. The genomic landscape of balanced
603 cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet* 49:36–45.

604 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011.
605 Integrative genomics viewer. *Nat Biotechnol* 29:24–6.

606 Robson MI, Ringel AR, Mundlos S. 2019. Regulatory Landscaping: How Enhancer-Promoter
607 Communication Is Sculpted in 3D. *Mol Cell* 74:1110–1122.

608 Roukos V, Misteli T. 2014. The biogenesis of chromosome translocations. *Nat Cell Biol* 16:293–300.

609 Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM. 2010.
610 Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and
611 distinguish closely related cell types. *Genome Res* 20:761–770.

612 Schluth-Bolard C, Diguët F, Chatron N, Rollat-Farnier P-A, Bardel C, Afenjar A, Amblard F, Amiel J,
613 Blesson S, Callier P, Capri Y, Collignon P, et al. 2019. Whole genome paired-end sequencing elucidates
614 functional and phenotypic consequences of balanced chromosomal rearrangement in patients with
615 developmental disorders. *J Med Genet* jmedgenet-2018-105778.

616 Slamova Z, Nazaryan-Petersen L, Mehrjouy MM, Drabova J, Hancarova M, Marikova T, Novotna D,
617 Vlckova M, Vlckova Z, Bak M, Zemanova Z, Tommerup N, et al. 2018. Very short DNA segments can
618 be detected and handled by the repair machinery during germline chromothriptic chromosome
619 reassembly. *Hum Mutat* 39:709–716.

620 Soutoglou E, Dorn JF, Sengupta K, Jasin M, Nussenzweig A, Ried T, Danuser G, Misteli T. 2007.
621 Positional stability of single double-strand breaks in mammalian cells. *Nat Cell Biol* 9:675–682.

622 Stankiewicz P, Lupski JR. 2002. Genome architecture, rearrangements and genomic disorders. *Trends*
623 *Genet* 18:74–82.

624 Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D,
625 Stebbings LA, McLaren S, Lin M-L, et al. 2011. Massive Genomic Rearrangement Acquired in a Single
626 Catastrophic Event during Cancer Development. *Cell* 144:27–40.

627 Storchová Z, Kloosterman WP. 2016. The genomic characteristics and cellular origin of chromothripsis.
628 *Curr Opin Cell Biol* 40:106–113.

629 Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G,
630 Hsi-Yang Fritz M, Konkel MK, Malhotra A, et al. 2015. An integrated map of structural variation in
631 2,504 human genomes. *Nature* 526:75–81.

632 Terzoudi GI, Karakosta M, Pantelias A, Hatzi VI, Karachristou I, Pantelias G. 2015. Stress induced by
633 premature chromatin condensation triggers chromosome shattering and chromothripsis at DNA sites
634 still replicating in micronuclei or multinucleate cells when primary nuclei enter mitosis. *Mutat Res*
635 *Genet Toxicol Environ Mutagen* 793:185–98.

636 Titen SWA, Golic KG. 2008. Telomere loss provokes multiple pathways to apoptosis and produces
637 genomic instability in *Drosophila melanogaster*. *Genetics* 180:1821–32.

638 Waters LS, Walker GC. 2006. The critical mutagenic translesion DNA polymerase Rev1 is highly
639 expressed during G(2)/M phase rather than S phase. *Proc Natl Acad Sci U S A* 103:8971–6.

640 Yang J, Deng G, Cai H. 2015. ChromothripsisDB: a curated database of chromothripsis. *Bioinformatics*.

641 Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q, Heinzen EL, Cirulli ET, Pelak K, He M, Ruzzo EK,
642 Gumbs C, et al. 2012. Using ERDS to infer copy-number variants in high-coverage genomes. *Am J*
643 *Hum Genet* 91:408–21.

Individual	Nb of chromosomes	Nb of breakpoints	Main chromosome	Copy Number Variant			Nb of resolved junctions		Nb of inferred junctions			Inferred mechanism			Inferred mechanism for rearrangement
				Gain in Mb (n gain)	Loss in Mb (n loss)				NHEJ	MIMBIR (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	Chromoanasyntesis				
3	1	18	6 (100%)	1,43 (7)	6,56 (1)	12	9	1	5	3	Chromoanasyntesis				
4	1	18	21 (100%)	5,30 (5)	11,73 (4)	9	6	5	1	0	Chromoanasyntesis				
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	Chromoanasyntesis				
7	1	25	21 (100%)	4,54 (11)	0,31 (1)	14	14	3	8	3	Chromoanasyntesis				
8	1	62	7 (100%)	0,76 (15)	19,68 (3)	27	27	13	7	7	Chromoanasyntesis				
9	1	68	11 (100%)	8,74 (14)	4,40 (1)	53	44	11	23	10	Chromoanasyntesis				
10	2	13	1 (76,9%)	0,83 (5)	0	8	8	8	0	0	Chromoanasyntesis				
11	2	18	2 (94%)	0	2,97 (2)	15	9	3	6 (1)	0	Chromoanasyntesis				
12	2	46	3 (98%)	0	9,00 (5)	41	29	12	9	8	Chromoanasyntesis				
13	2	57	13 (98%)	35,91 (15)	15,24 (8)	32	31	5	24	2	Chromoanasyntesis				
14	3	18	1 (72%)	0	0	17	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)	70	28	8	12 (2)	8	Chromoanasyntesis				
20	13	100	14 (70%)	0	4,23 (15)	79	46	32	5	9	Chromoanasyntesis				
Total		678				498	334	144	121	69					

644

645 **Table 1:** Rearrangements' characteristics summary. We only inferred the junctions for which a same repairing mechanism was considered at both ends. Full details can be
646 found in Supplementary data.

Risk relative ratios	Gene disruption	pLi	TAD disruption	Repeated element	G band positive staining	Fragile site	Open chromatin compartment	Distance to Replication Origin (log-transformed)	LAD disruption	Late replication timing
Constrained simulation										
Cancer CCRS	1,070	0,920	1,113	1,242	1,152	1,001	0,904	0,953	1,288	1,517
p	1,256E-04	8,243E-03	2,195E-09	0,000E+00	3,109E-15	9,562E-01	4,231E-07	0,000E+00	0,000E+00	0,000E+00
Constit CCRs	0,816	0,752	0,870	1,257	1,389	0,767	0,866	1,000	1,578	4,491
p	1,514E-03	2,194E-02	4,558E-02	2,471E-04	4,827E-07	6,533E-03	4,547E-02	9,743E-01	3,553E-15	0,000E+00
Simple rearrangements	1,013	2,256	1,105	1,184	0,972	1,119	0,732	0,956	0,959	1,303
p	8,916E-01	9,388E-07	2,878E-01	6,766E-02	7,625E-01	3,739E-01	3,639E-03	4,135E-03	6,604E-01	1,347E-01
Free simulation										
Cancer CCRS	1,057	0,929	1,119	1,234	1,146	0,928	0,898	0,951	1,242	1,400
p	1,869E-03	1,964E-02	3,059E-10	0,000E+00	3,175E-14	2,796E-03	7,024E-08	0,000E+00	0,000E+00	0,000E+00
Constit CCRs	0,806	0,760	0,876	1,249	1,380	0,711	0,860	0,999	1,520	4,146
p	7,451E-04	2,763E-02	3,417E-02	3,752E-04	7,872E-07	4,657E-04	3,647E-02	9,110E-01	2,384E-11	0,000E+00
Simple rearrangements	1,000	2,277	1,112	1,176	0,967	1,036	0,727	0,954	0,924	1,203
p	9,996E-01	6,887E-07	2,593E-01	7,905E-02	7,144E-01	7,790E-01	2,978E-03	2,947E-03	4,161E-01	2,969E-01

Table 2: Univariate analysis of breakpoint distribution against two simulation models. Bold characters indicate statistically significant results ($p < 0.01$).

	Gene disruption	TAD disruption	Repeated element	G band positive staining	Open chromatin compartment	Distance to Replication Origin (log-transformed)	LAD disruption	Late replication timing
Constrained simulation								
Cancer CCRS	0,998	0,933	1,126	1,064	1,075	0,988	0,987	1,431
p	9,485E-01	8,174E-02	5,723E-04	9,386E-02	7,807E-02	7,488E-02	7,730E-01	7,194E-14
Constit CCRs	0,584	0,856	0,983	0,836	1,543	1,042	0,909	3,629
p	2,677E-04	3,158E-01	9,018E-01	2,378E-01	7,989E-03	2,146E-01	5,640E-01	4,203E-12
Simple rearrangements	1,050	1,386	1,279	0,743	0,665	0,982	0,466	2,149
p	8,148E-01	1,838E-01	2,222E-01	1,752E-01	7,201E-02	6,499E-01	7,587E-03	5,036E-03
Free simulation								
Cancer CCRS	0,998	0,950	1,135	1,103	1,009	0,982	0,960	1,305
p	9,555E-01	1,951E-01	2,298E-04	7,788E-03	8,235E-01	8,482E-03	3,525E-01	3,181E-08
Constit CCRs	0,586	0,879	0,992	0,870	1,487	1,037	0,906	3,337
p	2,826E-04	4,045E-01	9,563E-01	3,586E-01	1,592E-02	2,737E-01	5,468E-01	1,047E-10
Simple rearrangements	1,055	1,421	1,285	0,768	0,618	0,976	0,454	1,932
p	7,943E-01	1,536E-01	2,145E-01	2,302E-01	3,616E-02	5,448E-01	5,500E-03	1,642E-02

651

652 **Table 3:** Multivariate analysis of breakpoint distribution across different types of chromosomal rearrangements. The logistic regression model presented is the
653 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
654 group respectively. Bold characters indicate statistically significant results ($p < 0.01$) with late replication-timing being the only significant variable for both
655 cancer and constitutional CCRs. Simple rearrangements diverge from simulation with a depletion in LADs.

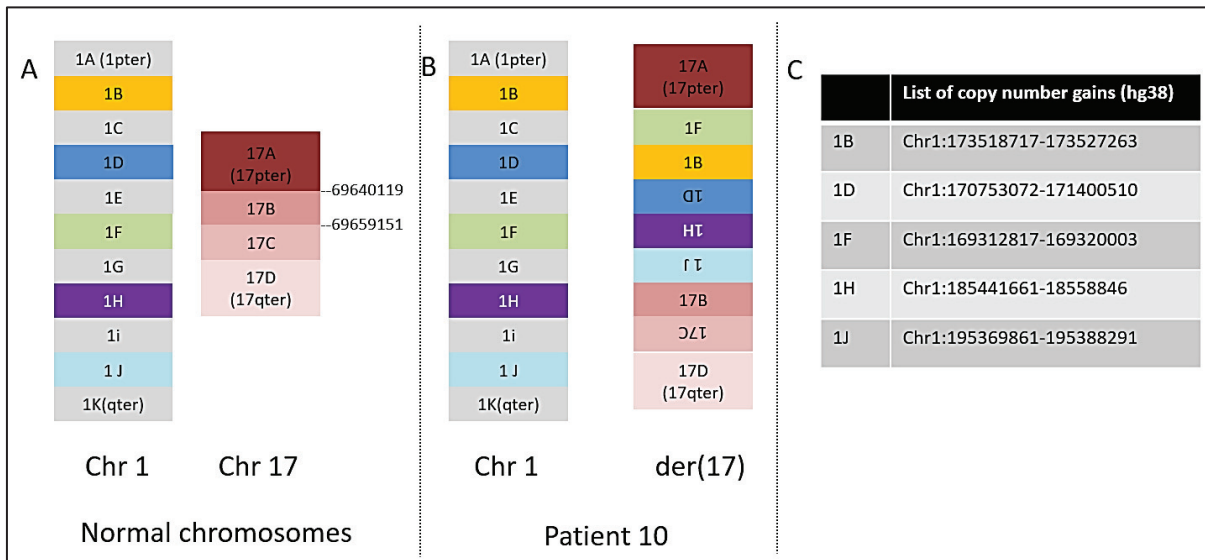


Figure 1: Schematic reconstruction of Patient’s 10 rearrangement with 5 chromosome 1 gains being inserted together in chromosome 17 nearby a pericentromeric inversion.

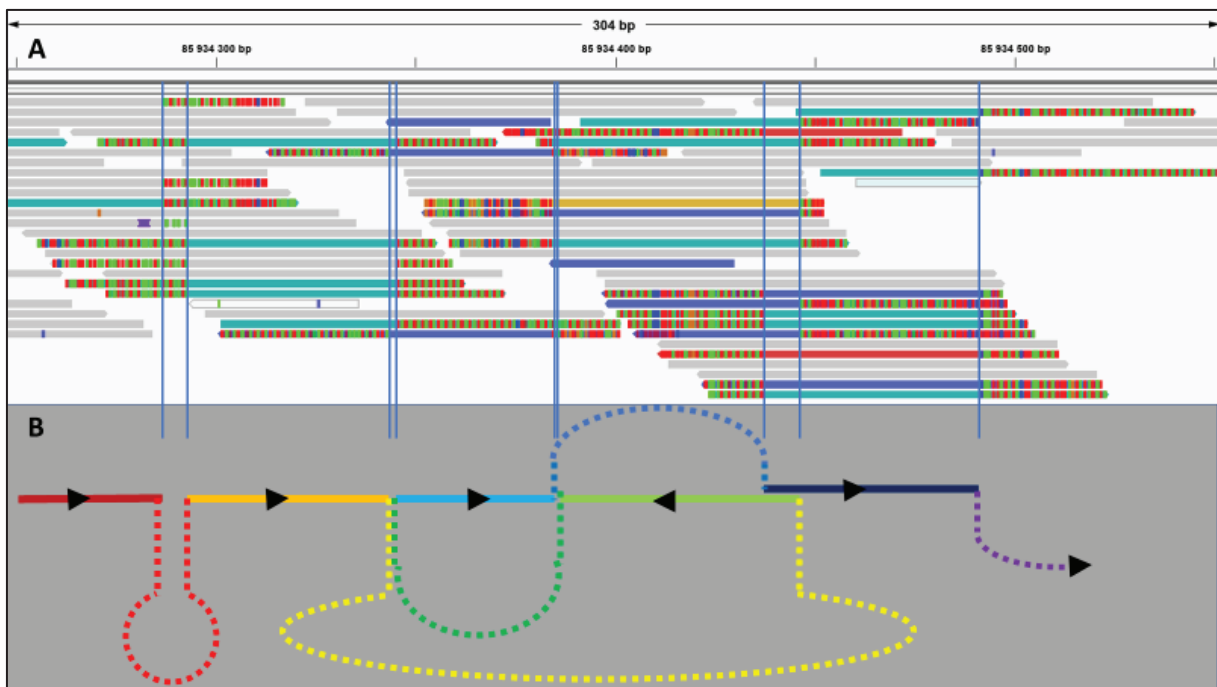
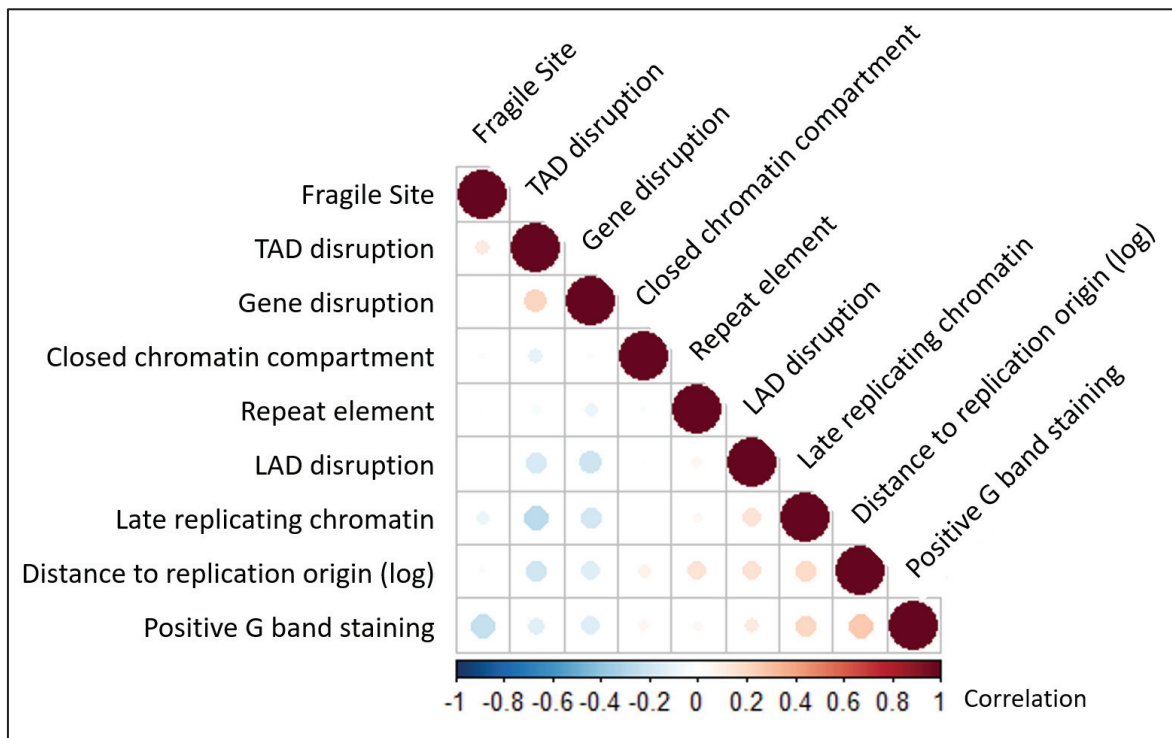


Figure 2: Read alignment with the IGV browser (v2.4) of the largest “hub” observed in our cohort (Patient 12) consisting of 12 breakpoints in 211 nucleotides (A). Reads paired with a mate aligned to a different locus are coloured. Partial schematic reconstruction of the rearrangement with plain lines being sequences of the “hub” and dashed-lines depicting other fragments of the rearrangement (B).



Supplementary Figure 1: Correlation matrix of the analyzed variables based on our dataset of constitutional CCRs.

Supplementary Table 1: List of annotated breakpoints detected in our cohort of complex chromosomal rearrangements (CCRs), in external cases of CCRs (cancer and constitutional), simple rearrangements and simulations. For replication timing, CE, CL and S stand for constitutive early, constitutive late and switching respectively. Only the two first ones were used for statistical analyses.

Supplementary File 1: Individual description of the 20 individuals included in this study.. Split-read sequences are color-coded with sequence aligned on chromosome A locus in black; sequence aligned on the other side of the junction in red; inserted sequence in green and sequence aligned to other loci in blue. Homologous nucleotides between breakpoints of the same junction are in bold characters.

Relative risk ratios	Gene disruption	pLi	TAD disruption	Repeated element	G band positive staining	Fragile site	Open chromatin compartment	Distance to Replication Origin (log-transformed)	LAD disruption	Late replication timing
Constrained simulation										
Our cohort	0.802	0.923	0.953	1.291	1.214	0.855	0.829	0.979	1.489	3.325
p	1.257E-04	8.235E-03	2.207E-09	0.000E+00	3.109E-15	9.585E-01	4.312E-07	0.000E+00	0.000E+00	0.000E+00
Literature	0.864	0.485	0.760	1.275	1.849	0.630	0.998	1.041	1.945	8.022
p	4.431E-03	5.881E-01	5.307E-01	7.291E-04	1.291E-02	1.675E-01	2.803E-02	1.095E-01	1.756E-07	0.000E+00
Free simulation										
Our cohort	0.791	0.930	0.959	1.285	1.208	0.792	0.824	0.977	1.436	3.081
p	2.610E-03	6.242E-01	5.819E-01	9.241E-04	1.559E-02	4.042E-02	2.322E-02	8.959E-02	2.097E-06	0.000E+00
Literature	0.853	0.491	0.764	1.267	1.839	0.584	0.991	1.042	1.875	7.364
p	1.452E-01	1.704E-03	1.197E-02	2.702E-02	1.384E-07	2.922E-03	9.474E-01	4.857E-02	4.155E-09	4.219E-15

Supplementary Table 3: Univariate analysis of breakpoint distribution in our cohort and reported cases compared to two distinct simulations.

Supplemental File 1

Phenotypic information, cytogenetic results, circos plot and fragment junctions identified through genome sequencing are presented individually.

Each junction is presented on two lines, one from each partner of the junction as presented below:

Ref. Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homology	Insertion	Inferred mechanism
hg19	21	41874669	+	1	21	45120234	-	CCAGATAATCCACAATAATTTCTTATGTCAAAGGACAGCTGTATAGACAGCTGGAACGGTGCC AGGAGAAATCCCAAGTGTGGATTACAGGTG/TGATGCTTCTTGGAGTCTGTGGCCCCAC GGTGCCGGCTC	0	57	MMBIR/FOSTeS
hg19	21	45120234	+	0	21	41874669	-	GAGCCGGCACCGTGGGGCCACAGCAGACTCCAAGAAGACATCA/CCCTGTAAATCCCAGCAGCTTT GGGAGTCTCTGGCAGCCGTTCCAGCTGTATACACAGCTGTCCTTGACATAAGAAAATTTATGT GGATTATCTGG	0	57	MMBIR/FOSTeS

Reference genome

ID of junction (pairs)

Partner A

Partner B

Nb of homologous nucleotides

Nb of inserted nucleotides

Split read sequence : Beginning of splitted sequence is signaled by a slash. The black sequence is partner A, red sequence is partner B, inserted sequence is green. Homologous nucleotides between partners are bold characters. Other colours signal other fragments close to the

Inferred mechanism for split read. Can vary because of CAN. Both directions have to be concordant for the junction to be inferred.

Patient 1

Phenotype: Mild intellectual disability, speech delay, cleft palate, single palmar crease, strabismus.

Karyotype: 46,XX,der(3)inv(3)(p13p22)inv(3)(p12q26.3)

Array-CGH : normal

Circos plot:



Comments : *FOXP1* disruption contributing to the neurodevelopmental phenotype. Rearrangement transmitted to her daughter, with the same phenotype. This patient has been previously published as IL/1901 in (Schluth-Bolard et al., 2019).

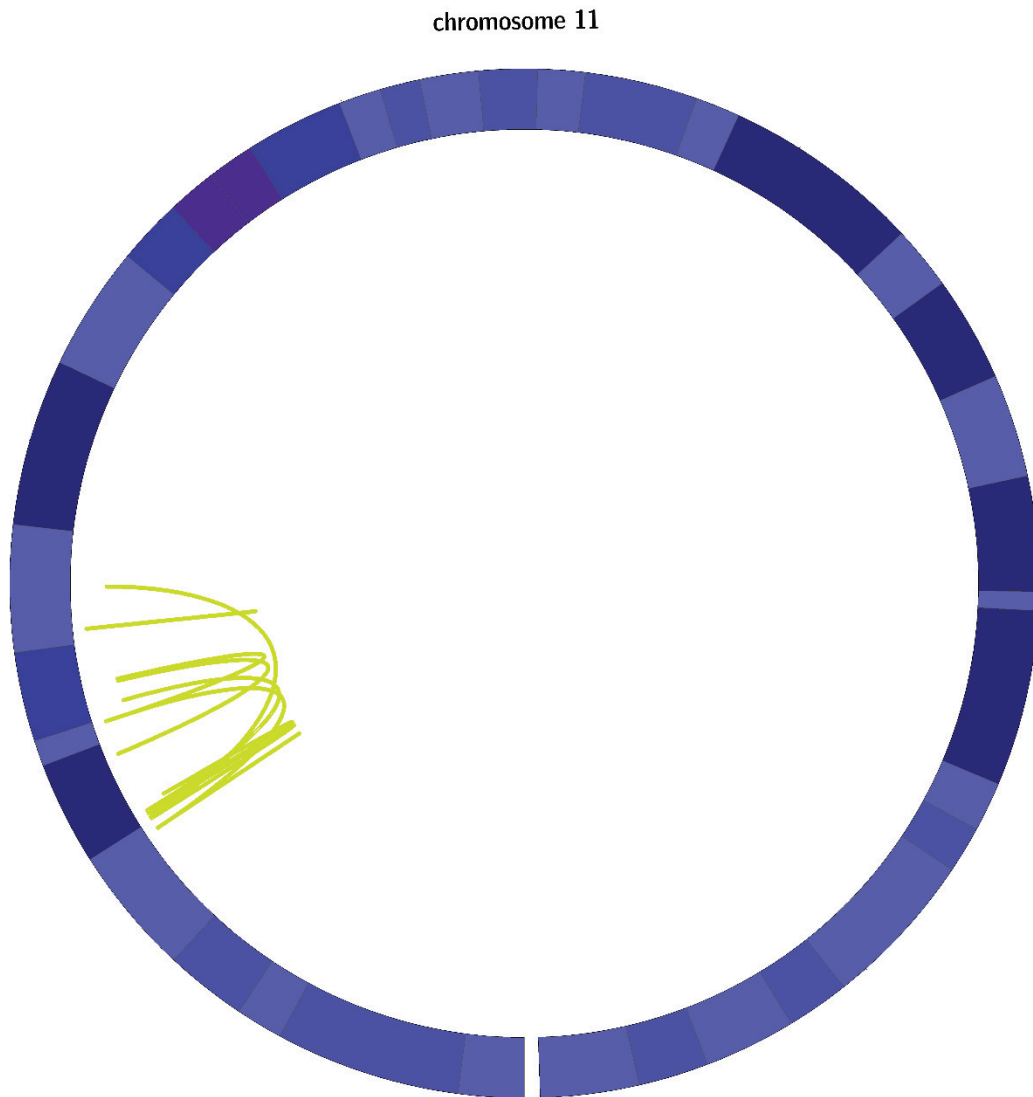
Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insert -ion	Inferred mechanism
hg19	12	3	75494494	+	-3730	3	75498415	-	CCCCGGGTGGCCTTGGCAATAGTCAACTCAATAGCACACTGGTGGGCATGCTGATGGAAAGCTGCTTCTGCCCCAG CCCTGTTTTA/ACCCACCATCACTTCTGCTCTGGGTAGGTCAGTCAAGCTCTGTAGCTGATCAAGATGCTCTG	5	0	MMEIMMBIR/FoSTeS
hg19	12	3	75498420	+	-3	3	75494489	-	GTCTCTCTCTACAGACATCTGATCAGCTACAGAGCTTGAAGCTACCCAGAGGCGAATGATA TGGTGGTTAAA A/CAGGGCTGGGGCAGAAGCAGCTTCCATCAGACATGCCACCAGTGTGCTATGTGAGTTGACTATTGCCA	5	0	MMBIR/FoSTeS
hg19	13	3	75498424	-	-3	3	163583575	-	TCCCCACCCAGGGAGATCAGTAGATAGTCTCCAGACCAATTTCCAAATAGTGTAGACAAAATAA/TGCCCTGGGCGGAG ATTTGCCACTGCACCTCCAGCTGGGTGACAGAGTGACTCCATCTCAAAAAGAAAAAAGTTGCTCT	0	0	NHEJ
hg19	13	3	163583575	-	1	3	75498424	-	CCCCAGGCTGGAGTGCAGGGGGCAAGATCTCGGCCAGGGCA/TTATTTTGTCTACACTAGTGGAAATGGTCTGGAGACTA TCTACTATGATCTCCCTGGGTAGGGAGACCTCTTTTGTAGTTTGTCTAAAAATCTTTCTCTGGCAATGAGT	0	0	NHEJ

Patient 2

Phenotype: Aniridia, Growth advance, hyperlaxity.

Karyotype: 46,XY.ish dup(11)(p13p13)(RP11-133E13++,RP1-74J1++)

Circos plot:



Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19	1	11	21227974	-		11	27050476	+	AATGGAAATTTACCAAACTTTCTAAAATGAGTAACTGCTGTT/ATGTCACATTC AGAGGTTCTGGTGTACAAAATTTGGGGTGGTACTGTTCAACCA	0	3	NHEJ
hg19	1	11	27050476	+		11	21227974	-	ACAGACATGCAGCCTAAATGGAAATTTACCAAACTTTCTAAAATGAGTAACTGCTGTT G/CTTATGTCACATTCAGAGGTTCTGGTGTATACAAAATTTGGGGT	0	3	NHEJ
hg19	2	11	21788748	-		11	22234413	+	ATATACCAATTTGATTTTTACTCTTGATTTTTGGAGCTCTTTATATATATTA/I TGCATCTCCAAATCATACTTACTAGAAATGAGAAATTCCAA	3	0	NHEJMMBIR/FoSTeS
hg19	2	11	22234416	+		11	21788751	-	AATATACCAATTTGATTTTTACTCTTGATTTTTGGAGCTCTTTATATATATTA/I TG/CATCTCCAAATCATACTTACTAGAAACGAGAAATTC	3	0	NHEJMMBIR/FoSTeS
hg19	3	11	21863699	+					TATCTAGGGAGACTACAAAGATAGTTAG/ATATTTATATTTCTATATTTAA CACTAAATATATAGTGTAAATATAGTATGATGATATATAGAAGAT	NA	NA	NA
hg19	3	11	33760158	+					CTAGAAATCAAGTTGACCTTTCCACTCCCTCACATCT/TCATGTGACAGGATTT CACATTCAGAAATCTAAATTTATAGATACTCTATATATCATG	NA	NA	NA
hg19	4	11	22053229	+		11	22528869	+	CAAAGGCTCAATGTGAGCCAGAGTTAAACACCAAGACCTTATCAGGCTTTTGGCT /TGAATCCCAACTTTGGAGGCTAAGGTGGCAGATTCCTTGAG	2	0	NHEJMMBIR/FoSTeS
hg19	4	11	22528871	+		11	22053227	+	CTGGCTCAAGCAATCTGCCACTTAGCTCCCAAAGTGTGGATACAAAG/C AAAGACCTGATAAGGCTTGTTAACTCTGGCTCACAATTGAGC	2	0	NHEJMMBIR/FoSTeS
hg19	5	11	22234372	-		11	27801985	+	ATACCAATTTGGATTTGGTATAGAAATCCACTTAAACCGTCTACAAAATATAC CATTTGA/TTTTTACTCTTGTATTTTTGGAGCTTTTATATATATTTTGGCAT	0	53	MMBIR/FoSTeS
hg19	5	11	27801985	+		11	22234372	-	AAAGTCCCAAGTTCTTTCTATACCAAAATTTG/GATTTGGTATAGAAATCCACTT AAACCGTCTACAAAATATACCAATTTGATTTTTACTCTTGTATTTTTGGAGCT	0	53	MMBIR/FoSTeS
hg19	7	11	25349369	-		11	28843199	-	CTTTATATATTTTATTTGTCATCTTCCAAATCATACTTACTTAGAAACG GGTCTGTAAGATGCAAGCCCTCTGGAGCTTAAAGTTCA/CTTTGCAATTTTG GAGGAAACAGAAATGCAAAATGCAAGTTCAGCAATTTGTTTAAAGAAAGGATT	1	0	NHEJ
hg19	7	11	28843198	-		11	25349370	-	AAACAATTTGCTGACTTTGCAITTTCTTTTCCCAAAATTTGACAAG/GTGAACCTT TAGCTCCAAGAGGCTTGGCATCTTTACAGACCATAGCATTTAGCAC	1	0	NHEJ
hg19	8	11	27048510	-		11	28954963	+	TAGTAGAGATGAGGTTTCGT/CATGTTGGCCAGGCTTGTCTCGAACTCTCTGACCT CATGATTCACCCACTCTTTTAAAGGTTCTACTATTCCTTAGTATAT	35	0	MMBIR/FoSTeS
hg19	8	11	28954998	+		11	27048545	-	TTTGTATTTTAGTAGAGATGAGGTTTCGT/CATGTTGGCCAGGCTTGTCTCTGATC TCTGACCTC/ATGATTCACCCACTCTTTTAAAGGTTCTACTATGC	35	0	MMBIR/FoSTeS
hg19	9	11	31689509	-		11	31694900	-	ATGACCCACTGTACTGCAGCCCTGGGTGACAGAGCAAGACACTGT/CATCTGCAT CATTTCTAAAATAAAATTTGCTCCCAATTTTCCCAAGCATTTATTT	1	0	NHEJ
hg19	9	11	31694899	-		11	31689510	-	GCTTGGGAAAAATGGGAGCAATTTATTTAGAAAATGATGCAGAT/GACAGT GTCTTGTCTGTCAACCCAGGCTGCAGTACAGTGGTGTGATCATAGCTC	1	0	NHEJ
hg19	10	11	21228602						NA	NA	NA	NA
hg19	10	11	21231380						NA	NA	NA	NA

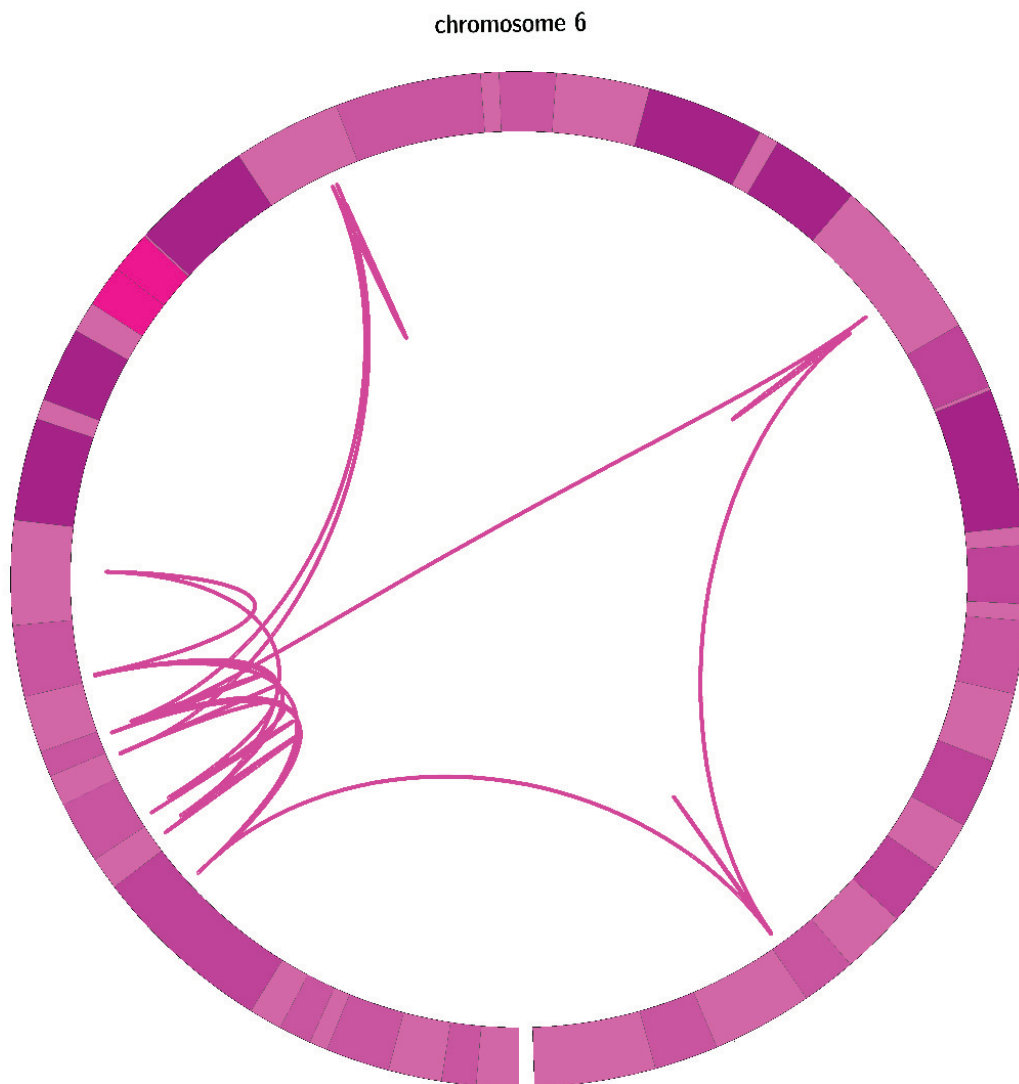
Patient 3

Phenotype: Moderate intellectual disability, growth delay, ophtalmologic problems, unexplained frequent bleeding

Karyotype:

46,XX,ins(6)(p24q26q22).ish ins(6)(wcp6+).arr[GRCh37] (26033642_26276650)x3,
6p21.32(33140010_33359008)x3, 6p21.2p21.1(36942747_43490809)x1,
6q21(111254270_111451467)x3

Circos plot:



Ref.	Junct-ion ID	Chr . A	Chr . A Position	chr. A strand	CN A	Chr . B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insert-ion	Inferred mechanism
hg19	11	6	33138619	+		6	22711182		AGCCTCATCCCAGTGTCTCTGTAAAAAGACACTGTGCTTTTACACAGAGTCTTGCATGGGTCCTGGGGGCCCATTTGAA/AGTGCAGGGCGGTGGTGG	0	10	MIMIR/FoSTeS
hg19	12	6	22711173					NA	NA	NA	NA	NA
hg19	12	6	155062702	+		6	22711173	+	AGCTGGAATTACAGGATGTACCACACACCCG/CCCTGCACTTTCAATGGGCCCCCAAGGACCATGCAAGACACTGTGTAAAAAGCACAGTGTCTTTTAC	NA	NA	NA

Patient 4

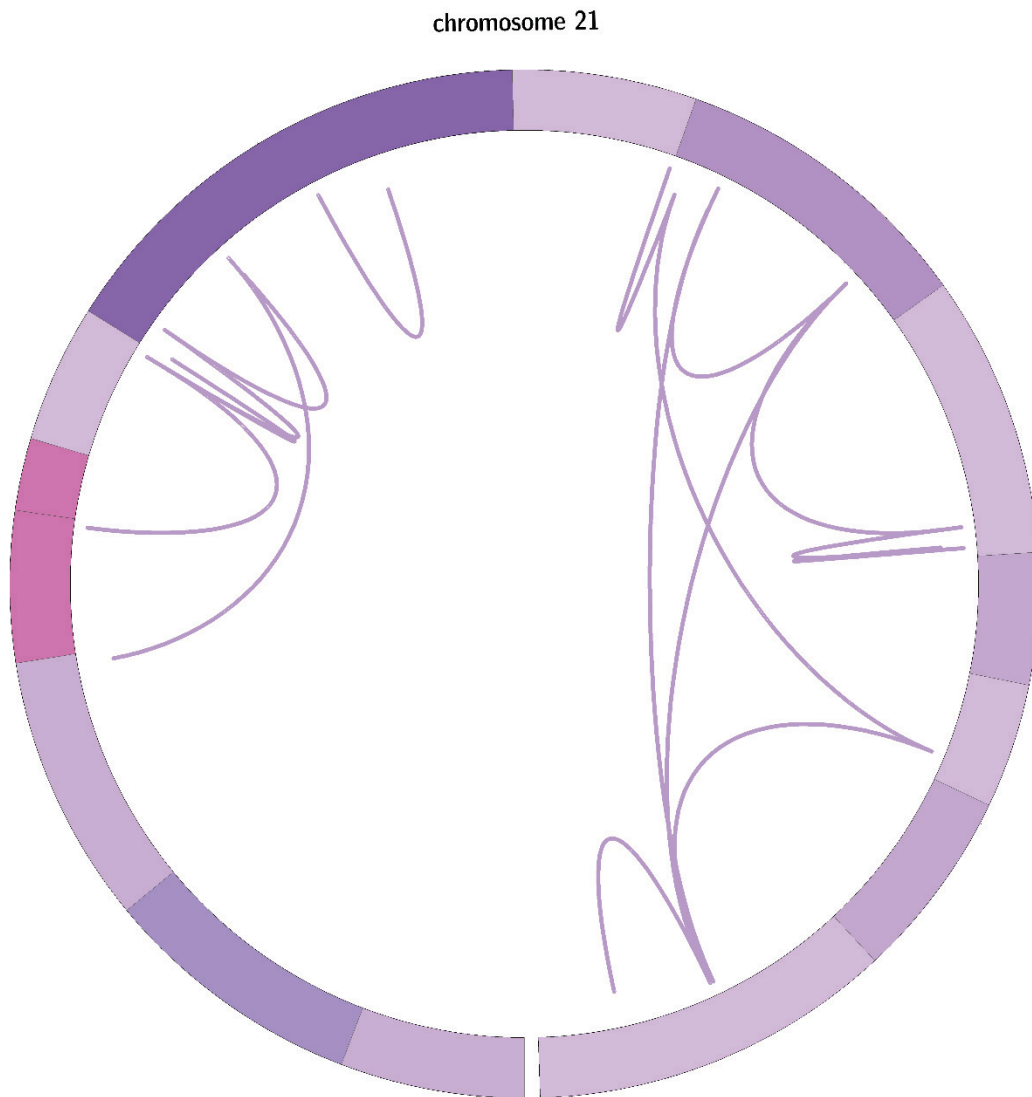
Phenotype: Intellectual disability, speech delay, facial dysmorphism

Karyotype: 46,XY,r(21)dn

Array-CGH : arr[GRCh37]

21q11.2q21.1(14795268_17628820)x1,21q21.1(17639720_17833313)x3,21q21.1(18223919_10849109)x3,21q21.1(21812036_22994075)x1,21q21.3(28195991_28358156)x3,21q21.3q22.11(29085595_31832089)x3,21q22.12(36658424_36982577)x4,21q22.2q22.3(40655668_48129895)x1

Circos plot:



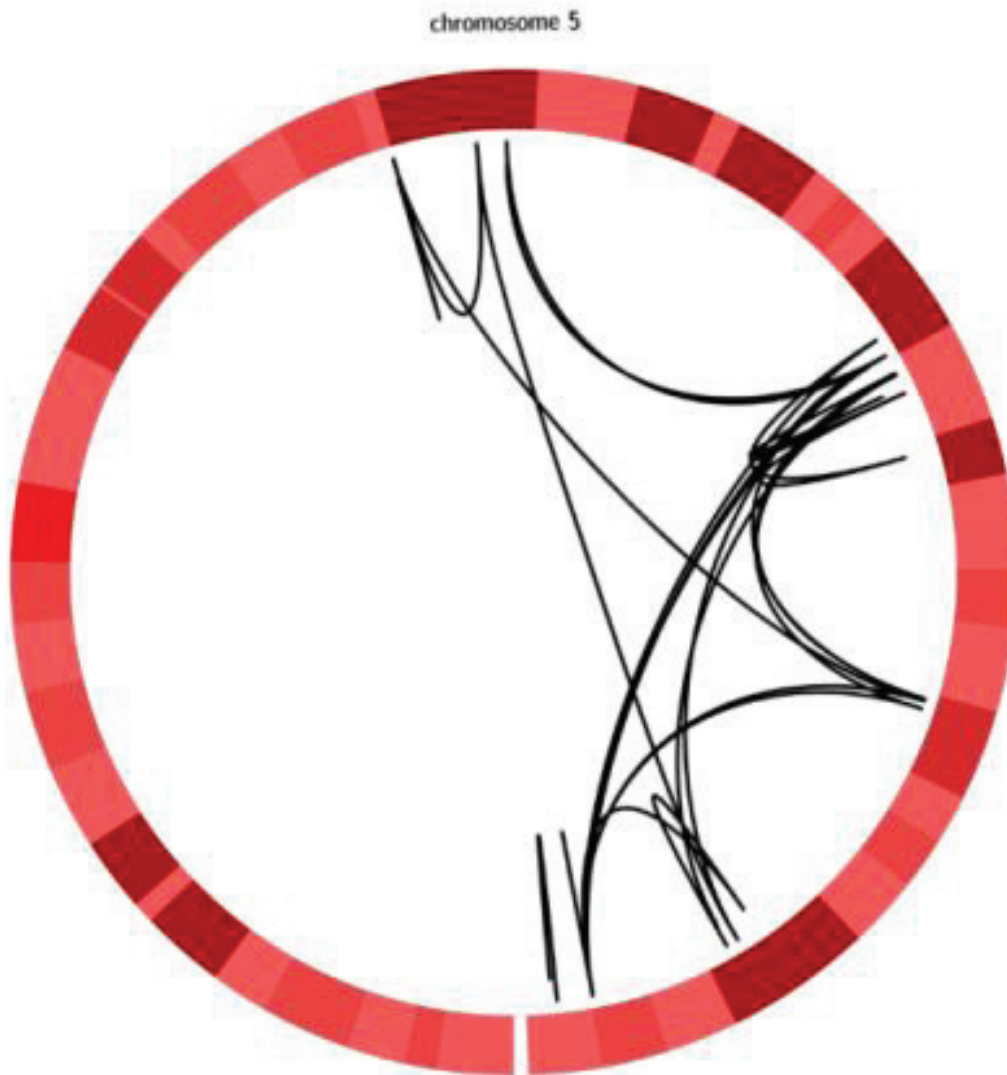
Ref.	Junction ID	Chr .A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insert- ion	Inferred mechanism
hg38	1	21	16261001	-		21	16462841	-	TTTTAAGTCCACCTTTTAAAAACATGCTATTCATTAAAGCATGCTATTTGT/TTGAAT AAAATTGTTTCAATAATGTGGAGTTATGTGCTGCGCTGTTAAGAC	0	0	NHEJ
hg38	1	21	16462841	+		21	16261001	+	CTTGTGAATATTGACTAAATCAATTTATTTAAAGCCACCTTTAAAAACATGCT ATT/CATTAAAGCATGCTATTTGTTTGAATAAAAATGTTTTTCATAATG	0	0	NHEJ
hg38	2	21	10713076	+					NA	NA	NA	NA
hg38	2	21	18512397	+		21	10713073	-	CCAGTTACTTACTGACTCCTCCCTCAATCACTCTGCTGCCACGAACTTCAITTA A/GTCAAGTGGAAATGGAGTGGAGTGGAAATGGAAATGGAAATGGAAAT	0	NA	NA
hg38	3	21	16257817	-					GGGCATAAGAAAAAGTCTTTAC/TTTTTGTCAAAAAAGTATTTTCAGATATGAATA GGTAAATTTTCAAGCTGCTCTTTTGGAAAGTGTATATTTG	0	0	NHEJ
hg38	3	21	16828187	-		21	16257817	-	GAATTTACCTATTCAGAAATCTTTCTGCACAAAAAGTAAAGA/CTTTTC TTATGCCCTTAAATCCATACTCAAAGTTCATTTACCACCTGTACA	0	0	NHEJ
hg38	4	21	20435335	+		21	21622221	-	CAGGAGATTGAGACCCTCGCTAAATGGTAAATGCCATGTCTACTAAAA/ TAGTACCTCAAAAAATGAAATCAATGCGAATCTTATCAATTTTATCT	0	0	NHEJ
hg38	4	21	21622221	-		21	20435335	+	CATCTGGCTAAAATGGTGAATGCCATGCTACTAAAA/ATGACCTCATAAAA ATGAAATCGGATACTTATTTATATCTAGCTTATTTTCACTT	0	0	NHEJ
hg38	5	21	30487845	+		21	35278549	-	GCCTCTGATGCTGGGATTAACAGGTGCTGCCACCACTCAGCTAAATTTTA TATTTTATAGAGAC/TTGGTGTCTCCCAAGATGCAGACTGAAAT	0	1	NHEJ
hg38	5	21	35278549	-		21	30487845	+	AGGTGCTGCCACCCTCTCAGCTAAATTTTATATTTTATAGTAGAGACT/TTGGTGC TTCCCAAGATGCAGACTGAAATGACCGGTACATGACACCATCTT	0	1	NHEJ
hg38	6	21	30489408	-		21	45014811	+	AGGCAGACAGGACTAGT/CTGCTCAACAATCTATCCATATTTAAATTTATCA CAATATTTTGTCTTATGTTTTCTTTTCCAAATCTGTGTCCTC	0	12	MMBIR/FoSTes
hg38	6	21	45014811	+		21	30489408	-	CAACAAGAGGCCAGGAAATTTCCAGCAGGAGGACAGACAGGACTAGTCTGC CTAACAAA/CTATCCATAATTTAAATTTTATCACAAATTTTGTCTCT	0	12	MMBIR/FoSTes
hg38	7	21	27687858	-					NA	NA	NA	NA
hg38	7	21	44956614	-		21	27687844	-	ATTATTTTCCAAAAGCGAATAAGTAAAGACTTT/GAATAATTGAATTTCAAGA ATTTTGTGCTGTGTTCTCTCCTGAGCTCATGCAGATCTTTGTGGA	0	27	MMBIR/FoSTes
hg38	8	21	27012531	+					CGGCATGGTGACACATGCTCTAGTCCCGCTACTTTGGAGGCTGAGATGAGA GGATCGCTGAGCCTGGAGTG/CTTTTTTTTTTAAAGGTGGAATA	0	1	NHEJ
hg38	8	21	39279649	+		21	27012535	+	GCCATCACCTCAGCCTGGCAACAGAGTGAGACTCCATCTCAAAAAA /CGCACTCCAGGCTCAGGGATCTCTC	0	1	NHEJ
hg38	9	21	35602836	+					GAATCCAGGCTGACGGAGGAGGAGGAGGAGGTTAG	NA	NA	NA
hg38	9	21	35646169	+		21	35602834	+	CACAGATAGATGCAGATTGCTTGTACCAGGCTTAGAGAAAAGGGAAAGG AGAAAGTACTGCT/AACCTTCCCCTCTCTCCCTCTCTGACGCTGATT	0	0	NHEJ
hg38	unresolved	21	26768682	-						NA	NA	NA

Patient 5

Phenotype: IUGR, post-natal growth retardation, microcephaly, intellectual disability, sleeping and feedings difficulties, epilepsy (hyperthermic seizures), choreic and dystonic abnormal movements, stabismus

Karyotype: 46,XY,ins(5)(q15q23.3q34)dn

Circos plot:



Comments : A likely position effect on *MEF2C* was considered as likely pathogenic. This patient has been previously published as EB/0401 in (Schluth-Bolard et al., 2019).

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homology	Insertion	Inferred mechanism
hg19	10	5	167429301	+	7	5	121094014	-	TGAGTTAGGTCAATTTGTGGGAGATTCGTGCAACATGTTGGCACTCCCTTAGG/TAG GACTGGCATCACCTGAGAGTAAGGATGATCTAAATAACATT	0	0	NHEJ
hg19	11	5	121175463	-	4	5	176446831	+	AACTAGCTGGTATCGTGGATTCCTGTAATCCAGCTACTCGGGAGGCTG/AG GGTTTTCTTGAGTTCAACTAAACAATTAATTAACAACACTAG	0	3	NHEJ
hg19	11	5	176446831	+	1	5	121175463	-	ACTGGGAGG/CTAGCTGAGGCGAG	0	3	NHEJ
hg19	12	5	121175466	+	4	5	165945124	-	AGCTGCAGAGGATAAATCCACTCTAATTTTTTATCCCTTCACTAGTACTACCAAGGG/ GACTTGTGAAGAGGGTGCCTCAITTACTCAGCCACAGCTCT	2	0	NHEJMMBIR/FoSTes
hg19	12	5	165945122	-	-4	5	121175464	+	AAATTTTTTATCCCTCAGTGACTACCAAG/GGGACTTGTGAAGGGTGCCTCATTT ACTCAGCCACAGCTCAACTCTCTGGGAGGGAGCA	2	0	NHEJMMBIR/FoSTes
hg19	13	5	122429965	+	-4	5	145663114	+	AAAGGAGGCGAGTGGTGTGACTTCCCTAATCCCTCCACCAGGGGGAAAAACG CAGGGCTGTTGGGTAAC/TTTTTTTTTTTTGGACAGGGTCTTGCTGATCC	3	0	NHEJMMBIR/FoSTes
hg19	13	5	145663117	+	1	5	122429962	+	GTGAGTGTGATCTCCCACTGCACATCCAGCTGGGAATCAGAGCAAGACCCTGTCTC AAAAAAGAAAAAAG/TTACCCAAACAGGACCTGCGTTTTTC	3	0	NHEJMMBIR/FoSTes
hg19	14	5	122429970	-	-4	5	123494994	+	GCTTTAAAAAGTATAGAAATAGTT/CCCAAGTGGCACACAGGCTTTTTGTGCAAC CGAGAAATCCACTGGCCCTCTGCCCTGGCCCTCCT	4	0	MMBIR/FoSTes
hg19	14	5	123494998	+	-5	5	122429974	-	CATTTAGCCAGATAAAATAGGTGTATTTGTCTGTTTTAAAAAGTATAGAAATAGTT CCCA/GTGGCACACAGGCTTTTTGTGCAACCGAGGAAC	4	0	MMBIR/FoSTes
hg19	15	5	122565264	-	1	5	166779229	+	AAATGACAAAGTCTACTGGAGCTGAATAGCAGCTCTTTCCCAAAGG/GGTATTCCC ACAGAAAAGTGGAAATTAATTTCTCAGCTCCCTACAGC	1	0	NHEJ
hg19	15	5	166779230	+	1	5	122565265	-	CCACCTTAAATGACAGAACTACTGGAGCTGAATAGCAGCTGTTTTCCCAAAGG G/GCTTATCCACAGAAAGTGGAAACTTACTTTCTCAGC	1	0	NHEJ
hg19	16	5	122565264	+	1	5	144980453	+	TGGAGGCTACCAAGTGAAGTCTCAGTGGAGGCAAGGAGTCTGCGAGGAGG/ AATAAGCCATCACACCAGCTGGCTTTTTTTTGATATTTTA	1	0	NHEJ
hg19	16	5	144980454	+	-13	5	122565263	+	AAAGAAACAAAATAAATATCAAAAATAAGCCAGGCTGGTGTGATGGCTTATC/T CCTCTGCAGAGACTCCCTGGCTCCACTGAGACCTCAGCTTGG	1	0	NA
hg19	17	5	123495004	-	-5	5	123913723	-	AAATGATCTGCCGCTCGGCTCCAAAAGTIIAGATIIA GTTTTAGCCAGGATTTACTAGATCAGAAAGCCAGATCGCACAGTCTTCAGT	7	8	MMBIR/FoSTes
hg19	17	5	123913723	-	-3	5	123495004	-	ATCTAGGTAATCCTGGCTGAACTGGGTAACCTTAACATCTAACATCT/AACATTTGG GAGCCGAGCGGGCAGATCACATTGAGTCTAGTTCGAGATCAGCTGGCCAAACAT AGTGAACACCTGTC	7	8	MMBIR/FoSTes
hg19	18	5	123913719	+	-3	5	128098596	-	CATAATAAACAGACTGGCTGGCAATGGTTTCATGCCTGTAATCC/GGGATATAC CTTACGGTATAGTAGAGCAGACCTTACTTGGCAGAAAAAT	0	1	NHEJ
hg19	18	5	128098596	-	-6	5	123913719	+	CAGCAGTTGGCTGGCACAATGGTTTCATGGCTTAATCCG/GGATATACCTTACCGGTA GATAGCAGACCTTACTTGGCAGAAAAATTTCTTCTGGT	0	1	NHEJ
hg19	19	5	145115094	-	2	5	176446831	-	TGGCACCATTCTAGTTCACTGCAACTCCACTCTGGTTCAAGCAATTTCTCGCTCA GC/GATAATTCATGCTGGAGAAAAAGAAAAAGAAAGAAAGTAAAGACATTAATCT TTGGTATTTAGCATTACATGACATTA	0	0	NHEJ
hg19	19	5	176446831	-	1	5	145115094	-	ATATC/GCTGAGCAGGAGAAATTTGCTTGAACCCAGGAGTGGAGTTGCAGTGAACTG AGATGGTGCCACTCCAGCTGGCAACAGACTCCCGTC	0	0	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	20	5	145663117	-	1	5	167436338	+	TTGATCAGTTGGTTAGAGGCCAGGTTGAAAATTATGATAAAATCT/GGTACATATCTATATTT TGGTAGGGCTATTTGATACATCAGTTTGGTATGTCCCATG	9	0	MMBIR/FoSTes
hg19	20	5	167436347	+	10	5	145663126	-	AAACTTGCCTTTTATGATCAGTTGGTTAGAGGCCAGGTTTGAATTTATGATAAAATCT GGTACA TAT/CTATA TTGGTAGGGCTATTTT GATACATCAGTTTGGT	9	0	MMBIR/FoSTes
hg19	21	5	164462058	-	3	5	167429296	-	AAAGGTAATATGCCCTTCTTTTATAAAGACAGAGTCCCTAGG/GTGGGCAGCACTGGGAG GTTATTGACTTGCACAATGTTAGATCCTGGAAGTCGCCTCC	1	0	NHEJ
hg19	21	5	167429295	-	7	5	164462059	-	CATTGTGCAAGTCATAATAACCTCCCACTGCTGCCCA/CCCTAGGGA CTGTCTTTATAA AAGAAGGCATATTACCTTGTCTTTTGCCTTTTGCAGTCATG	1	0	NHEJ
hg19	22	5	164462060	+	3	5	166779231	-	TCATCTTTCTGCTCTATTCTTTTCTGGGATCGTATCTAAAGTCCAGAGT G/ACCTGCAC TCCTTCTATTTTCCACAGGCCCTTCTTTATTTG	1	0	NHEJ
hg19	22	5	166779230	-	1	5	164462059	+	CCTGGATCGTATCTAAAGTCCAGAGT/GA CTGCACTTTTCTATTTTCCACAGGCC CCACTGTTCTTATTGCCTTCTACAATGGCTGAGTTAGCT	1	0	NHEJ
hg19	23	5	167436338	-	10	5	176428600	-	TCAAGATTGTTAAACTGTTCTGGGCCCATGCCCTTCTATA/TC CTACATATTCATTTCTAT GAGCAGGTTTAAACTCTAATGCACTGCTGCTGTATATCTGC	1	0	NA
hg19	23	5	176428599	-	-2	5	167436339	-	ACAGGCCAGATGCATTAGAGTTTAAACTGCTCATAGAAATGAATGTAGC/AT ATAGAA AGGCATGGGCCCCAGAACAGTTTAAACAATCTTGACAAAGAAAT	1	0	NHEJ

Patient 6

Phenotype: Intellectual disability

Karyotype: NA

Comments: *ADNP* disruption is responsible for Helsmoortel-van der Aa syndrome (MIM 615873)

Circos plot:



Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	1	20	2387743	+		20	14620345	+	AGGAGAA TTGCTTGAACCTGGAGGCGAGGTTGACGTAGCCAGGATCGCACAC <u>TG/TGTGCTAAAGGACTGTGAACCTGCAAGAAATAGAGAAGCCTGT</u>	3	0	NHEJMMBIR/FoSTes
hg19	1	20	14620348	+		20	2387440	+	AAGTCTACGGCTTCTAAATCTCTTTCAGTTCACAGTCTTATAGCCAA <u>CAG/TGTG</u> CGATCTGGCTACTGCAACCTCCGCTCCAGGTTCAAAGCAAT	3	0	NHEJMMBIR/FoSTes
hg19	2	20	3539218	-		20	14094767	-	<u>CTTTACATAAATAATGGGCTGTAAATCAGACTAATAGAGAATTTGT/AAITCAAC</u> ACTTTTAAACAGATCTCATATGATAGAAAGATGCTGGAAAT	2	0	NHEJMMBIR/FoSTes
hg19	2	20	14094765	-		20	3539220	-	<u>TCCAGACATCTTTCTATGATAGGATCTGTTAATAAAGTGTGA/TTACAAAT</u> CTCTAAATAGTCTGATTTACAGCCATATTTTATGTAAGAATA	2	0	NHEJMMBIR/FoSTes
hg19	3	20	1682397	-		20	4140782	+	<u>AGGACTGAGGCTGGCTAGTTTGTCTGCTGCTCAAG/CCTAATC</u> CAAAATTAATAATTTCCGTGAGTGCAGTGCACACCACTTTGTGGAA	7	0	MMBIR/FoSTes
hg19	3	20	4140789	+		20	1682404	-	AGATCAGGGTGAAGGACTGAGGCTGGCTGGCTAGTTAGTGGCTGGCAA <u>GCTAATC</u> <u>/TTCCACTGAAAAGCAAAAATAAAAATTTCCGTGAGTGCAGTGCAC</u>	7	0	MMBIR/FoSTes
hg19	4	20	5624009	-		20	5631544	+	<u>CTGGTTCAAGTGAATCTCGTGCTTCAAGCTTCAAGT/AGCTGGGATTACAGGTGC</u> CTGCCACCTGGCTGGCTAAATTTTGCATTTAATAAAGACGAAAT	18	0	MMBIR/FoSTes
hg19	4	20	5631562	+		20	5624027	-	<u>CCTCTGGTTCAAGTGAATCTCGTGGTTCAGCTCCTCAAGT/AGCTGGGATTACAGG</u> <u>TG/CTGCCACCTGGCTGGCTAAATTTTGCATTTAATAAAGACG</u>	18	0	MMBIR/FoSTes
hg19	5	20	30631654	-		20	31119657	+	<u>TGGACATTTGTGTAGGAATCTCTCTATCTGGCTCTGACG/CTAATGAT</u> ATTAAGCACCTTCCATGTCTATTGGCCATTTGTATAGCTT	1	0	NHEJ
hg19	5	20	31119658	+		20	30631655	-	CTGAGTCTGACATTTGTGTAGGAATCTCTCTATCTGGCTCTGACG/ <u>TAATGATTAAGCACCTTCCATGTCTATTGGCCATTTGTA</u>	1	0	NHEJ
hg19	6	20	32577536	-		20	34092328	+	<u>CAGGATCTCGATACAGGAGGATGTCAGCAGCTGCAGAGGCACT/TGSGCTCA</u> ATGNCAATCAATCTTCAATATATATGCAAGGCTGGCTCT	2	0	NHEJMMBIR/FoSTes
hg19	6	20	34092330	+		20	32577538	-	GGACAGATCTCCGATACAGGAGGATGTCAGCAGCTGCAGAGGCACT <u>TG/G</u> <u>CTCCAATGACATCAATTAATCTGCAATATATCTGCAAGGCTGGC</u>	2	0	NHEJMMBIR/FoSTes
hg19	7	20	48898612	-		20	49514845	+	<u>AATGTTTACACCCAGAACTAGCAGTCTGGCAACTGCTGGCAAC/ACGTTG</u> CCTGCCCTCAAAAACCCAGTTCCTCCCAAGCTCTCCCAAGC	0	21	MMBIR/FoSTes
hg19	7	20	49514845	+		20	48898612	-	TGAATAACGATCTTGAATTTGTTTACACCACCCAGAACTAGCAGT <u>TCTGGCAACTG</u> <u>CGTTGCCAACCGTTGCTGGCCCTCAAAAACCCAGGTTCCCCC</u>	0	21	MMBIR/FoSTes
hg19	8	20	49514838			20	55871319		NA	NA	NA	NA
hg19	8	20	55871319	+		20	49514838		ACACCTAACAGGGACCCCAAGTATTTCTGGTATTCTGAA <u>CTAGCAGTTAAT</u> <u>CAACATCACAGCCAGTCCGCCCTGCGCTTCTGCGTACCAACAG</u>	0	0	NHEJ
hg19	9	20	49514869			20	60927048		NA	NA	NA	NA
hg19	9	20	60927048	+	3	20	49514865		<u>AGGGAAAGGGGGAGGAGAGAGAAGAAAGAGAGAGAGGGAGGAGCGACT</u> CACCCGGGGTGAACCGTGGTCCGCGCAGCCTTCCCAATGAGATGGCCAGC AGCGTGTGTACCGAGAAAGCGAGGCGGAC <u>CTGGCTGTGTGTAATGAC</u> <u>TGCTAGTTCAAAATACCAGAAATCACT</u>	0	0	NHEJ
hg19	10	20	55339301	-		20	61027614	+	<u>TGGCTTAGGAAGCCCCAGCCCTCTCAGTCTTCTCTCAGCCTTCG/CTGAGTA</u> GCTGAGACTACCGATGCACAAACCTGACCAATATTTTACT	3	0	NHEJMMBIR/FoSTes
hg19	10	20	61027617	+		20	55339304	-	CCCCCTGGCTTAGAAGCCCCAGCCCTCTCAGTCTTCTCTCAGCCTTCG <u>CTG/</u> <u>AGTAGCTGAGACTACCGATGCACAAACCTGCCCCAGATAATTTT</u>	3	0	NHEJMMBIR/FoSTes
hg19	11	20	59953108	-		20	59968630	+	<u>GGTCTTTATGGAGCAGGATCTCAGTAGCCCACTTGGCACAAGCT/AGACCAGC</u> CTAGACAACATGGCAAGACCCCATCTCTATAAAAATACAAATA	0	1	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	11	20	59968630	+		20	59953108	-	GGGTCTTTATGGAGCAGGATCCTCAGTAGCCACATTCGCCACAGCC/AGACC AGCCTAGACAACATGGCAAGACCCCATCTCTAAAAACTACAAA	0	1	NHEJ
hg19	12	20	60926956	-		20	61027051	+	AGGAAAAGGGAGAGAGAAAGAGAGAGGGAGGGGA/GCAGC TCACCGCGGTGACCGTGGGTCCCGCAGCCCTTCCCCATGAGATGGCCAG CAGCGTTGGTACGCAGGAAGCGCAGGGGACCTGGCTGTGATGTTGAATTGAC TGCTAGTTCAGAAATACGAAATATCACTT	1	0	NHEJ
hg19	12	20	61027052	+		20	60926957		GGGAGAGAGGGGGAAAGGGGGAGGAGAGAAAGAAAGAGGGAGAG GAGGGAG/CAGCTACCCGGGTGACCGTGGGGTCCCGCAGCCCTCC	1	0	NHEJ
hg19	13	20	60927046	-	3	20	61803263	-	ACATGGTGTACATGCCTGTAGAAGTTGCCAACACATAGACTGCTA TGCCATCTCGCGTTGACCAAGGCT/GACGTTGGTGGCCTTGGTG	0	15	MMBIR/FoTeS
hg19	13	20	61803263	-		20	60927046	-	NA	NA	NA	NA
hg19	14	20	61803283			20	62236493	+	NA	NA	NA	NA
hg19	14	20	62236493	+		20	61803283	-	GGTGTACATGCCTGTAGAAGTTG/CCAACACATAGACTGCTATGTGTGCC ATCTCGCGTTGACCACCAAGGCTGACGTTGGTGGCCTTGGTGAAC	0	18	MMBIR/FoTeS

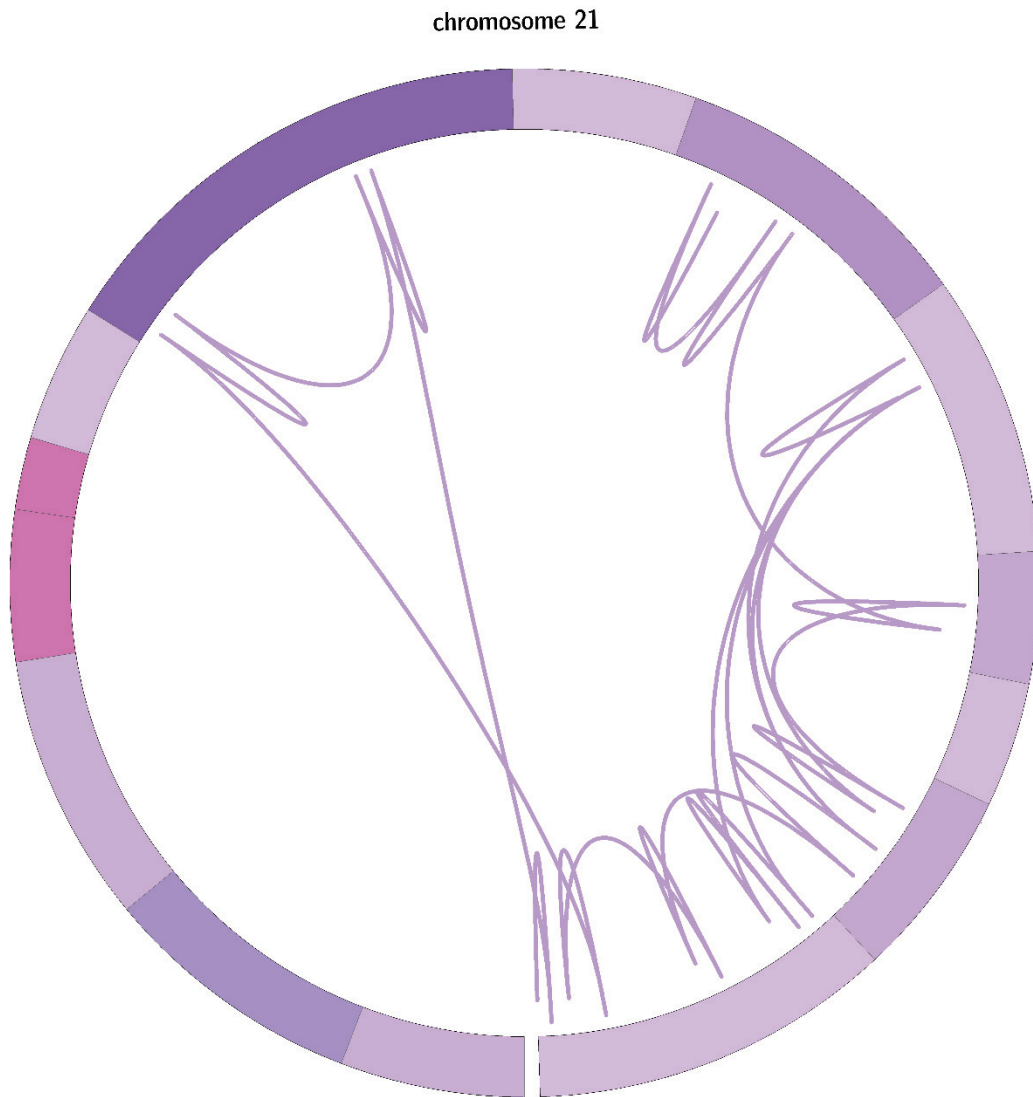
Patient 7

Phenotype: No phenotype. Karyotype was performed in her fetus after positive first trimester screening. It showed an abnormal chromosome 21 so that the maternal karyotype was performed.

Karyotype: 46,XX,der(21)

Comments: All duplications are chained and inserted on the short arm of chromosome 21.

Circos plot:



Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19	12	21	41267524	-		21	32173603	-	AGATTGTTATTATACACTGAACAATTTGCTTAATAGA/GGCTTTGAGCAAATTCCTCTGTGCTCAGG ATATAATTGGTAAAGTGAGCTGATGAAGACACTA	1	0	NHEJ
hg19	13	21	32729160	+		21	41267516		TCATGCTCCTCTAAAGAAATGTAGATTCCTCTGGGGCAG/CATAATGGATAATACACAGCTCT GCCTGCTCCTCTCCCAAGACCATATACATGGAATAGATATACAGGGATAATTTGCTATTAG CCCCATAGCATGCCTGGCTAAATTTT	0	21	MMBIR/FoTeS
hg19	13	21	41267516	+		21	32729160	+	AAAAATTAGCAGGCAATGCTAAATGGGCTAATAGACAAATATCCCTGATATATCTATTCCATGTA TATGTCCTTTGGGAAAGAGCGAGACAGCAGA/GCTGTGTAATATCCAAITAGCTGCGGCCAGATG GAATCTACATTTCTAAGGAGGACATGA	0	21	MMBIR/FoTeS
hg19	14	21	32729204	+		21	43550726		ATGTTTGCAAATGTGGGGCTCAGAATTCACCAAAATTAATAGGACTCGATAATGTTGTAATAACA GATAAACTCGAGCCGTTCCAGGTGTGGGGCAGTGCTCAGGCTGTAATCTTAGCACTTTGGGA TACTGAGACAGG/TGGCCGTTTACTCACATGGCCACACTTAGCTGCCAGAGCTGGCTCTGGGTGG AAAGCAGGGCGCAGCCAGGAGGACAGTAGGG	0	13	MMBIR/FoTeS
hg19	14	21	43550726			21	32729204		ATGTTTGCAAATGTGGGGCTCAGAATTCACCAAAATTAATAGGACTCGATAATGTTGTAATAACA GATAAACTCGAGCCGTTCCAGGTGTGGGGCAGTGCTCAGGCTGTAATCTTAGCACTTTGGGA TACTGAGACAGG/TGGCCGTTTACTCACATGGCCACACTTAGCTGCCAGAGCTGGCTCTGGGTGG AAAGCAGGGCGCAGCCAGGAGGACAGTAGGG	0	13	MMBIR/FoTeS
hg19	15	21	36666550	-		21	40409718	-	TTTTCCAGCGTGTACTCCTGCTCCCTAAAGCATTTCAAAAGAAAGG/CAGAAAGACTTGT CTGAAAAACGCTCTGATAAGCCATGGTAAAGAAACA ACGTTTTCCAGACAAGCTCTTCTGCTTTCTTT/TGAAATGCTTTAGGAGCATGGAGTACATACGC TGGGAAGAAAGATCCCCAGCAAAAGCAAGGAG	0	9	NHEIMMBIR/Fo TeS
hg19	15	21	40409718	-		21	36666550	-	ACGTTTTCCAGACAAGCTCTTCTGCTTTCTTT/TGAAATGCTTTAGGAGCATGGAGTACATACGC TGGGAAGAAAGATCCCCAGCAAAAGCAAGGAG	0	9	NHEIMMBIR/Fo TeS
hg19	16	21	43189415			21	43550704		ATGTTTGCAAATGTGGGGCTCAGAATTCACCAAAATTAATAGGACTCGATAATGTTGTAATAACA GATAAACTCGAGCCGTTCCAGGTGTGGGGCAGTGCTCAGGCTGTAATCTTAGCACTTTGGGA TACTGAGACAGG/TGGCCGTTTACTCACATGGCCACACTTAGCTGCCAGAGCTGGCTCTGGGTGG AAAGCAGGGCGCAGCCAGGAGGACAGTAGGG	0	8	NHEIMMBIR/Fo TeS
hg19	16	21	43550704			21	43189415		ATGTTTGCAAATGTGGGGCTCAGAATTCACCAAAATTAATAGGACTCGATAATGTTGTAATAACA GATAAACTCGAGCCGTTCCAGGTGTGGGGCAGTGCTCAGGCTGTAATCTTAGCACTTTGGGA TACTGAGACAGG/TGGCCGTTTACTCACATGGCCACACTTAGCTGCCAG/AGCTGGCTCTGGGTGG AAAGCAGGGCGCAGCCAGGAGGACAGTAGGG	0	8	NHEIMMBIR/Fo TeS
hg19	17	21	44796146	-		21	47544777	+	ACTAAGCCACCCAGCTTTGGCCCAAGAGTCTCA/GCCTCCCAACTCGTCCCCATGCCAAAGGCC TCAGGTGCTGGGAGAGTGAACCGGGGTGAGCCCA	5	0	MMBIR/FoTeS
hg19	17	21	47544782	+		21	44796151	-	GCAGTGGCTGGGAGACTCAGCCACCAGCCTTGGCCCAAGTCTCAGGCTG/CCTCAACTCG TCCCCCATGCCAAAGGCCCTCCAGGTGCTGGGAGAGT	5	0	MMBIR/FoTeS
hg19	18	21	27870932	+					TATGAGCCATGCTAACATGAAACTGAAACATGAACAATAACAATAA/GTGGACATTTCTCAGGCTTTGCAGC CTATGGAAGAAAAACTTAATACCTTCCACATAAAATC	NA	NA	NA

Patient 8

Phenotype: Hypotonia, intellectual disability, testicular ectopia

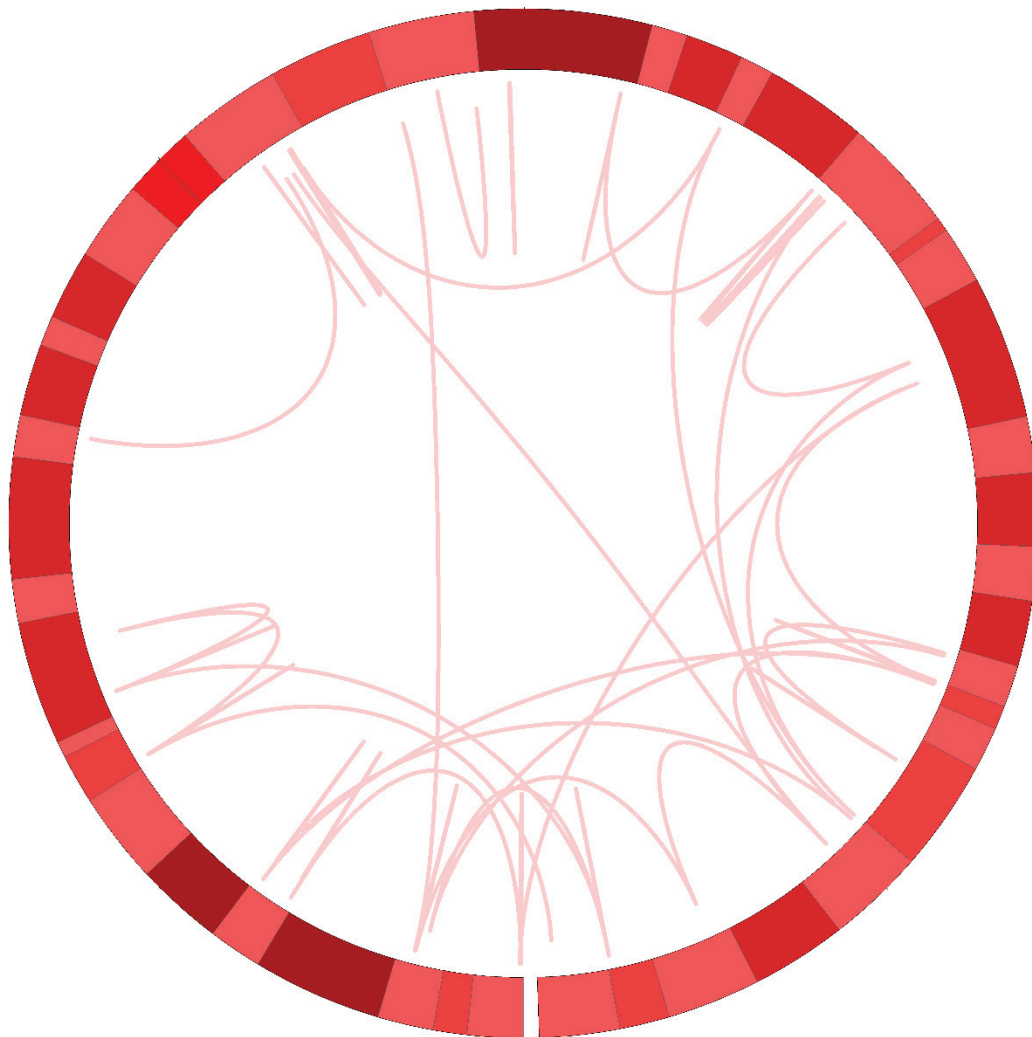
Karyotype: 46,XY,der(7)[23]/46,XY[2]

Array-CGH : arr[GRCh37]

7q11.21(65806576_65842999)x3,7q11.23(75062013_77095846)x1,7q22.1(98769647_99035705)x3,7q22.1q31.1(100798980_109868414)x1,7q34q36.1(140654844_149074160)x1

Circos plot:

chromosome 7



Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	1	7	140101	-	7	14152919	-	-	TGTCACATGAAATAGCCCATGGCATCTTGGAAACA TGGT/CCCATCCCTCCGC AGTCTCAGGCATGGCCTGCACGGGTCTCTCGCAGGTGCCCCACAC	0	4	NHEJ
hg19	1	7	14152919	-	7	140101	-	-	GACCCGTGACGGCCATCCCTGAGACGTGGGAGGGATGGGACCA/TGTTCC AAGATGCCATGGCTATTCATGTGGACAAATCCCTTCAGTGAAGCAGGG	0	4	NHEJ
hg19	3	7	5740782	-	7	154947874	-	-	AGACAAATATGGACATGGAGAA TGGCTCGCATGAGTACAGG/A/GGCC TTCTGAACTTGAATAACGTGGACTAAAGGAGAGCTGCTATTTCTAT	0	1	NHEJ
hg19	3	7	154947874	-	7	5740782	-	-	CAAGTTCAGAGAGGGGCT/CCTGTACTCAGATGCAGACAAATCTCCAAATG TCCATAATGTCTATTTTATTCCTGTGTGACACGTGAAAACACAGCAATG	0	1	NHEJ
hg19	4	7	6301744	-	7	149076999	-	-	TAAAGCAAGCTTCAAATGGAGTAGGTTGACAGGCCAAGT/TGGAATATAAGT CAGAAGAATACAGAGTGAACGATGGAGGCAAGGAAAGGGATAAAGGACA	0	0	NHEJ
hg19	4	7	149076998	-					NA	0	0	NHEJ
hg19	5	7	6344384	+	7	72521079	-	-	TGGAGTGCAGTGGCGGATCTGTCTCACTGCAACCTCCGGCTCTGGGCTC AAGCGA/CATGAAATACCTCTGACTGCTTCAGAGGCTCTGGCTCGTC	0	33	MMBIR/FoStes
hg19	5	7	72521079	-	7	6344384	+	+	GGCTCTGGGCTCAAGCGACATGAATATACCTCTGACTCTGACTTTCAGAGGCTCT /GGCTCTGGCTCTCCTCCATGTCAITGGCCAGATAGCTGAGGACAGA	0	33	MMBIR/FoStes
hg19	6	7	16072393	-	7	129437618	+	+	ATTTTTCTTTTCTCTAAATTTCTTTGAAAACATATGTGGATA/AATGTGTGA AAATCATATACATACATATATGGTTTTATCTTTTGTGGCATTT	0	5	NHEJMMBIR/FoStes
hg19	6	7	129437618	+	7	16072393	-	-	GGGATTTATTTCTTTCTCTAAATTTCTTTGAAAACATATGTGGATAA ATGTGAAAATTCATACATACATATATGGTTTTATCTTTTGTGGCATTT	0	5	NHEJMMBIR/FoStes
hg19	7	7	25938688	+	7	158213791	-	-	GTTCAAGGCAATTTGGACTGAGGATCCCGTACCCCATACAGAAAG/GAGG ACTTATATGGCAAGGCACTTATTTGCTACTGTTACCTAAATTTCC	0	7	NHEJMMBIR/FoStes
hg19	7	7	158213791	+	7	25938688	-	-	CTGCTCAAGTATCTGAGTTCAAGGTCATTTCTGGACTGAGGATCCCGTACCC CAT/ACAGAAAGGAGGACTTATTTGCAAGGCACTCTGTCTGTTGCTAC	0	7	NHEJMMBIR/FoStes
hg19	8	7	25964124	+	7	33365390	-	-	ATCCTTTAATCTTTGCAATCAAGCTCCTCAACTCTGTACACCATGACACTTTCT AAAGAATGTCTCAACTTAGAAGTTGAGAGAAT/AATTTCTGGATTA	0	0	NHEJ
hg19	8	7	33365390	-	7	25964124	+	+	ACTTTCTACAGAAATGCTCAACTTAGAAGTTGAGAGAAT/AATTTCTGGATTA ACTAGGAGAGACTTGCCTTTCTCCCTTACCTTTCCCAACAAAA	0	0	NHEJ
hg19	9	7	30039119	-	7	33364800	-	-	TGAAAATATACAGTCAGAGGTATATTT/CATCTCGGCTCCCAAAGTCTGGG ATTACAGGCGTGAGCCACCGTCCCGGCTCCAGTTTGGTCTTTTAT	0	6	NHEJMMBIR/FoStes
hg19	9	7	33364800	-					GGCTCACGCTGTAAATCCAGCACTTTGGAGGCCGAGATGATATA/CCTC TGACTGTATATTTCAAATAGCTGTCTCAAGCTTACTAAATTTCTTCTA	0	6	NHEJMMBIR/FoStes
hg19	10	7	30058065	+	7	154967350	+	+	AATGGCGAATTAAGGAGAGGAGCCTTTCTCTTTCTTTACCCAAAGGC TCAGA/AGGAATATACCCCATAAATACATCTCTCTTTCACTTTAG	0	17	MMBIR/FoStes
hg19	10	7	154967350	+	7	30058062	+	+	TGGTTAAACTCTCGATGTGGTTCTACAAGTGAAGAGAGGAGGATGT/TTA TGGGGTATATCTTCGAGCCTTTGGTAAAAGAGAAAAGAGAGG	0	17	MMBIR/FoStes
hg19	11	7	44897419	-	7	64631736	-	-	CTTATAAATAGCCCGGAATGGTCCCATGTCTTCTGGTCCAGCTACTCG AGAGGCTGA/CGGCGCTGGCCCTAGTAGTTAAACAATATTATACCCAGTT	0	0	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	11	7	64631836	-					TTTTTGTGATACAGGGGCTCACTATCTAGTGCATGCTT/GTCTTTAACTCCTGGCTCAAGTGATCCTTTGCCTTGGCCCTCTAAAGTGTGAGATTG	0	0	NHEJ
hg19	12	7	63945339	-	7	63945339	+		TGTAACCTTAGTAATCTTTGGGAAA/TAAAAATTTTAGATTATGTGAACTTAATTTTTAATGTACATTTTTAACATTTAAGACTATTGTGCATTC	9	0	MMBIR/FoSTeS
hg19	12	7	63945348	+	7	63933073	-		AAAAAAAACCTGTAATGCCCTTTAGTTGATGAACCTTTAGTAATCTTTGGGAATAAAAATTT/TTAGATTATGTGAAAGTAAATTTTTAAITGTACA	9	0	MMBIR/FoSTeS
hg19	13	7	65148197	+	7	140647123	-		TTTGTCTTTAGTACTAAGTACAAAACCTAATAGGTAGTAACTGGACCCCAATTATACATGTGCAGTGTGACGATCAG/CTCACCTGGTATACTTA	0	9	NHEJMMBIR/FoSTeS
hg19	13	7	140647123	+	7	65148197	+		CATTTATAAAATAAAGTCTGATATACTGAAATTTAAAGTATCA/CCAGGTGAGCTGCTCAGCATCTGCACATGTATAATTTGGTCCAGACTAGCT	0	9	NHEJMMBIR/FoSTeS
hg19	14	7	65660718	-	7	65669431	+		GGGGTTTCACTATGTTGGCCAGGCTGGTCTCGAACTCTG/GGGTGGCAAGAAGCTGGTAGCCCTTAGCTCTTTTGTGTTTTTTTGTAGACA	0	0	NHEJ
hg19	14	7	65669431	+	7	65660718	-		ATTTTAGCAGGGACGGGTTTCACTATGTTGCCAGGCTGGTCTCGAACTCTG/GGGTGGCAAGAAGCTGGTAGCCCTTAGCTCTTTTGTGTTTTTTTGTAGACA	0	0	NHEJ
hg19	15	7	65851114	+	7	91707373	-		CCTCCACCTCAGCCCTCCTAGTGTGAGACTACAGGTGCAATGTCATCAT/CAGCTGGCTCACTGCAACTCCACTCTAGGTTCAAGTGAATCTCC	0	3	NHEJ
hg19	15	7	91707376	-	7	65851114	+		AGCTGAGACTACAGGTGCATGTCATCAG/CCTCGGCTCACTGCAACTCCACCTCCTAGGTTCAAGTGAATTTCTCTGCTCAGCTCCAGTGTAGCTG	0	3	NHEJ
hg19	16	7	79132794	-					TATGTT/ACTTTTGAATCCCAAAAACATGACTACTAATAGCCTACTGTTGACTAGAAATTTACAATAACATAACAGTCATTTACACACATATTTATAT	6	0	MMBIR/FoSTeS
hg19	16	7	79174614	+	7	79132800	-		CATTCTGGATTGCCTAGAGGCAATGTTCCATTGATCTATATGTTACTTTT/GACTCCCCAAAACATGACTACTAATAGCCTACTGTTGACTAGAAAT	6	0	MMBIR/FoSTeS
hg19	17	7	85631766	+	7	98120963	-		TATCTGAGTTATTACTCTGTATCTCAGTGAGTTCTTTAAAATCATTATAGTGAATC/ATCCCCCTGCTCCCATCTGTATCGGTGAGCTGGGGAAACAG	0	0	NHEJ
hg19	17	7	98120962	-	7	85631765	+		TGTATCTCAGTGAGTTCTTTAAAATCAATTATTGTGAAT/CATCCCCCTGCTCATCTGTCTGGGTGAGCTGGGAAACAGGATGTTCCACCCAGGGAT	0	0	NHEJ
hg19	18	7	98682717	-	7	98686430	+		TGCCTGTAATCAGGGCTCATGCCTG/TAATCCAGCACTTTGGAGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGTTGAGACCAGCCTGGCCAACAT	0	0	NHEJ
hg19	18	7	98686481	+					NA	0	0	NHEJ
hg19	19	7	98769397	-	7	138585186	-		CTATACTTTTAAATTTATAGAAGCTTTTAAAATGGAAT/AAGCACAGGCAGGTGTGAGATCCTGAAAACAGGGACACTGACCCAGGAAAAGTGAAGTGGAG	0	0	NHEJ
hg19	19	7	138585183	-	7	98769400	-		TTTTCTGGTCCAGTGCCTTTTGGGAAACAGGATGTTCCACACTGCCTGTG/CCTTATCCATTTTTAAAAGTTCTAATAAATTTAAAAGTATAGTTTCTTCTTC	0	0	NHEJ
hg19	20	7	98150722	+					NA	0	0	NHEJ
hg19	20	7	99047260	+	7	98150720	+		TGGGTTGAGGGCCCTTTCCAGGCACTGGAGGACAATGGTGAAGAGGTCCTTGGCCCTGTGAAGCCT/GAGTCTCATAGAACTCTTTGGGAGGGACAGGT	0	0	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	21	7	111248208	+	7	129280668	+		TGGAGGTGTCCTAGCTATGGTGGCTAAGCAGCGGAGCTTGTGGAGTACAGGAGCTACAGGAGAC/TAGTTGATGTTCTCATTTCCAAATGTGACCA	6	0	MMBIR/FoSTeS
hg19	21	7	129280674	+	7	111248202	+		AACTACTGGGTCTTCTGGTCCAAATGGAAATGAGAACTCAAACTA GTCTCC/TGTAGCCTCTACTCCCAAGCTGCCTGTAGCCACCA	6	0	MMBIR/FoSTeS
hg19	22	7	127550156	-	7	134312504	-		AAAATGGTACAGCCACTGGGAAAGAGTTGATTTCT/ATGGAAGCATAA GAAAAAATTAATCTCTTTGGTAAAGAGAGTTTTAAAAAGCCGACG	0	0	NHEJ
hg19	22	7	134312506	-	7	127550154	-		TAATTTTTCTTATGCTTCCATA/GAAATCAAACTCTTCCACAGCTGGCTG TACCATTTAAATGCCACCAGTAGTACAAAAGACTCCAATTTCTC	0	0	NHEJ
hg19	23	7	5749892	+					ACCTCATCTCTACTAAAAATATAAAAAATCAGTCAGGCTGGTGGCGCATACC TGTAGCCAGCTACTGGGAGGCTG/GGCCAGCCTTGGATTCTGCACA	0	6	NHEJMMBIR/FoSTeS
hg19	23	7	127703816	+	7	5749892	+		ATAAGTAATGGAGTGGTGCATAATAATCACATCTGTGAGAATCCAAGG/C TGGCCAGCCTCCAAAGTAGCTGGGATCACAGGTATGGCCACCAGGCC	0	6	NHEJMMBIR/FoSTeS
hg19	24	7	129431277	-	7	140648305	+		CATTCAATCCAAATATAATTTCTGGTGGACACGATGGCACATCGT/A TTTGTGATCTCTGTTTTAATCATGAAAATCCAGATATGGCCGGT	0	7	NHEJMMBIR/FoSTeS
hg19	24	7	140648307	+	7	129431277			AGCTTAGTCCATTCAATCCAAATATAATTTCTGGTGGACACGATGGCA CA/ACATCGTATTTGTTGATCTCTGTTTTAATCATGAAAATCCAGA	0	7	NHEJMMBIR/FoSTeS
hg19	25	7	91711406	+	7	138326362	-		CTCTCCCTCTGAGCAGCTGGACTACAGTGTGCACCACAACA/GGTGCCA AGAGCAAGCCACTACACTCCAGCCTGGATGACAAGAGCGAAACTCCATC	0	3	NHEJ
hg19	25	7	138326362	-	7	91711406	+		TCTCCTCTGAGCAGCTGGACTACAGTGTGCACCACAACAAGGT/GCCAA GATCATGCCACTACCTCCAGCCTGGATGACAAGAGCGAAACTCATCT	0	3	NHEJ
hg19	26	7	109880984	-					GGACATATGGAGTTTAGGGTTATTTAGAGAA/ACCTGGAGGACATTA TGCTATGAAATAAGCCAGACATAGAAAAACAATACTGCATAATCTC	0	34	MMBIR/FoSTeS
hg19	26	7	138329547	+	7	109880984	-		ATCTTGGCAACGCGCTCCTGAGTGGCCAGGGCAGCCCGAG/GACATATGTG GAGTTTAGGGTTATTTAGAGAA/AACTGGAGGACATTTGCTATATGA	0	34	MMBIR/FoSTeS
hg19	27	7	16098549	+					AAAAAACCAATACAAAAAATAAATATTTAATTTTGAACCTTTAAAAAA AGGCCATTTTGTCCCTGCTGCTGGATAAA/AAAAAAAAAAAAAAAAAT	2	0	NHEJMMBIR/FoSTeS
hg19	27	7	138586089	+	7	16098547	+		TCTGTATAACATTTGGATTACATGGCATACTACTTTTGTAGCCGCTTGA TTTTATTTTTTTTTTTTTTTT/TATCCACAGCAGCGGACAAAAAT	2	0	NHEJMMBIR/FoSTeS
hg19	28	7	111245610	-	7	211575	+		CGGGCAGAGTGGCCAGGCTCCAGAGCCCTCCCTAGAAATAAAGAAG AGAAATAA/AGATGCTGGAGCATGTTTATACAGCCTGAAGAACCAT	0	19	MMBIR/FoSTeS
hg19	28	7	211575	+	7	111245610	-		CGGGCAGAGTGGCCAGGCTCCAGAGCCCTCCCTAG/AAATAAAGAA GAGAAATAAGATGCTGGAGCCATGTTTATACAGCCTGAAGAACCAT	0	19	MMBIR/FoSTeS
hg19		7	14203549	+					NA	NA	NA	NA
hg19		7	33365463	+					AGGGAGAGACTTCTTCTTCCCTTACCTTTTCCCAACAAAAAGAGTCTC TCTCT/AGTAATCCAGAGAAATTTCTGGATCTCAACCAAGTCCACCA	0	0	NHEJ
hg19		7	65799184	-					AGAAAGATATAGGAGGCA/CCCACCTCCCCTCTGCAGAGAGATAGATTCT CCTTAGGGAGGAAGAAAAACTAGAGAAAAATAAGACACTAGGGGGAGAGT	0	0	NHEJ

Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19		7	74942225	-					NA	NA	NA	NA
hg19		7	77097025	-					TTATCAGCTTAAGGAGATTTGGGCTGAGACAATGTC/AAAAAGAGGAAGAAT CTCTGGGCCCCCACCCTTACAAGGCCACAGAACAGCCAGCCTAGGGT	0	0	NHEJ
hg19		7	85620677	-					NA	NA	NA	NA
hg19		7	100786967	+					NA	NA	NA	NA
hg19		7	127558121	+					CTTAATCGGCCATAGATTTTTCTTTTTTTTTTTTTTTT/TATTTTACATTTTACAAA AAGTAATAATTTTTACCCCGCAAGACAAGAATGGGAAAAACAA	0	0	NHEJ
hg19		7	127701041	-					GAGGCAGAGGTTGCAGTGAGCCGAGATCATGCCATTGCACACATTT/CTTA GCACAAGAAATGGAGGTGGAGCGAGGGAATGAGGCCACCTGGCCC	0	5	NHEJMBIR/FoSTeS
hg19		7	129239163	-					CAACAAATACATGAATACATAAACATGAATACATGAATACATGAATACATAA A/CATGAATACATAAAAAACAAGGCAAGACAAAACCCCAAAAACG	0	0	NHEJ
hg19		7	134352090	+					TCAAGTTACATGATCATATGCCCCATTTTGTGGTTAATGCCTGTTGTCCGAA T/ACAAACAATACATGAATACATAAACATGAATACATGAATACATGAAT	0	0	NHEJ
hg19		7	140647194	-					GACTTTGTAGCCC/AGGCTGGAGTGCAGTGGTGAATCTCAGCTCACTGCAA CCTCCGCTCCCGGGTTTAAAGTATTCTTATGCCTCAGCTCCCGAGTA	0	0	NHEJ
hg19		7	158210296	-					NA	NA	NA	NA

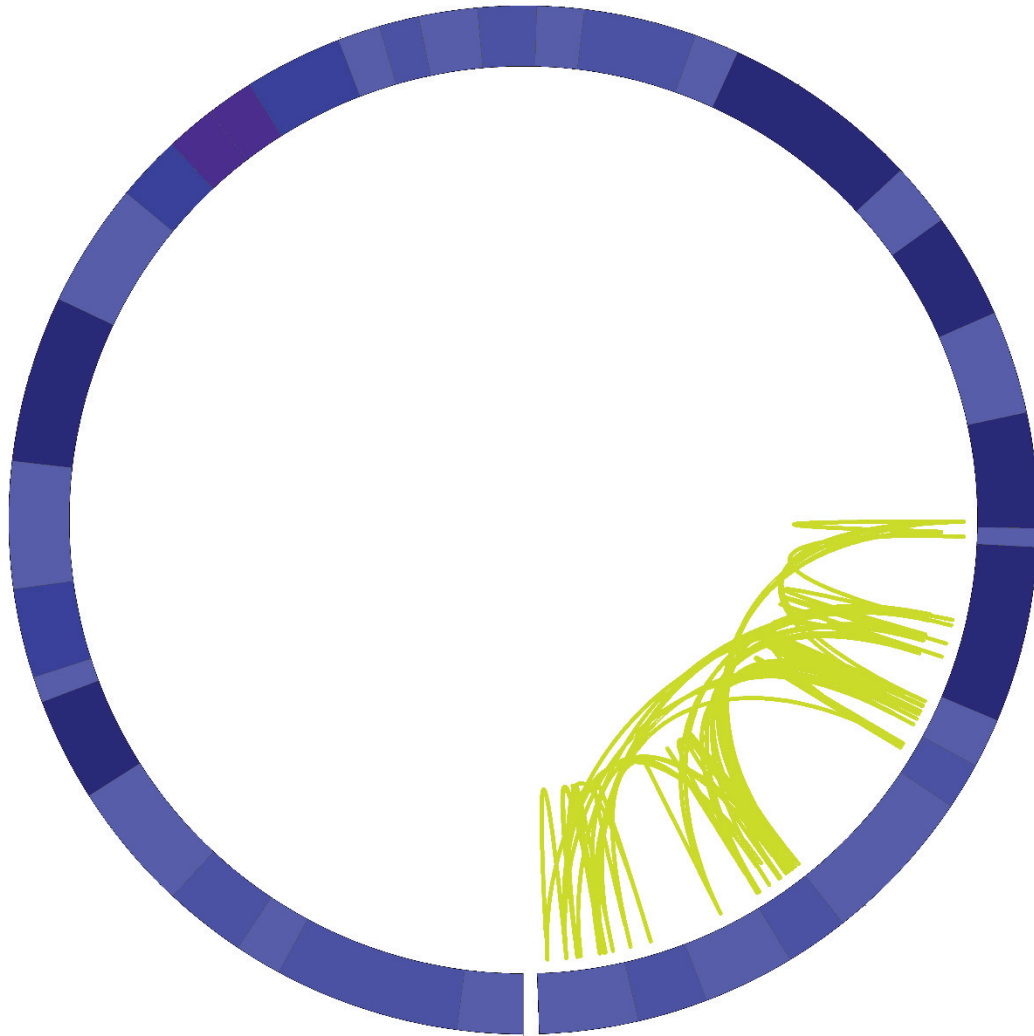
Patient 9

Phenotype: Intellectual disability, facial dysmorphism and achromic spot

Karyotype: NA

Circos plot:

chromosome 11



Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insertion	Inferred mechanism
hg19	1	11	101839543	-		11	102356609	+	AGATGAAATCCCATGTGAAAGGGTGTTCCTCTATACCAGAGAAAC/AC TTAGTGTGTTACTAATCAATCAATATATTTACCCTAGAAATAG	0	0	NHEJ
hg19	1	11	102356609	+	3	11	101839543	+	AGCAGGAGATGAAATCCCATGTGAAAGGGTGTTCCTATACCAGAAAGGA AAC/ACTTAGTGTGCTTACTAATGAATCAATATATTTACCCTAG	0	0	NHEJ
hg19	2	11	101839715	+	7	11	121534596	+	GATAGTTAGTGACATGTTTAGGATTTAGAAAGGGGATACCTGGAAA ATTTAAACA/GGTATGTCAACAACAAGGAAATGGCATGAACAGGAAGGGA GT	0	7	NHEJMMBIR/FoSTes
hg19	2	11	121534596	+		11	101839715		TACCTTGCCATTCCTGACTCCCTCCTGTTTCATGCCATTCCTTGTGT TGTTG/ACATACCTGTTAAATTTCCAGGTATCCCTCTAAAATCCT	0	7	NHEJMMBIR/FoSTes
hg19	3	11	102356606	-	3	11	121530833	+	GTTTTGAAATAAAAAGTCATGACCA/AAACGTTATTCCTATTCATGAGGAA GGAAAGTTGAAGCAAATGCAGATAGATGTTCTAAGGTACAGAAAAAAT	3	0	NHEJMMBIR/FoSTes
hg19	3	11	121530836	+		11	102356609	-	CTTTCTGGAGATGAGTTTCGAAATAAAAAGTCATGACCAAA/CGTTATT CTTATCATTGAGGAAGGGAAGTTGAAGCAATTTGCAGATAGATGTTCTAA	3	0	NHEJMMBIR/FoSTes
hg19	4	11	102356646	+		11	121699204	-	TGACCAAAACGTTATTCCTATCATGAGGAAGGGAAGTTGAAGCAA/TTG CAGATAGATGTTCTAAGGTACAAGAAAAATTCACCTGAGGAAGGGGCT G	3	0	NHEJMMBIR/FoSTes
hg19	4	11	121699204	-	4	11	102356643	+	TCTTATCATTGAGGAAGGGAAGTTGAAGC/AAAATTGCAGATAGATGTTCTA AGTACAAGAAAAATTCACCTGAGGAAGGGGCTGCTAGTTCCCTCCAC	3	0	NHEJMMBIR/FoSTes
hg19	5	11	102569574	+		11	110908152	-	ATTTATCCCGAGGACTCCCTCAAAGTTATTCCTTCATGAACTCCCTTT ATC/AGTCCATAGGGAATTAATCTCCCTCTGTAAGGTTAAAAGGGC	5	0	MMBIR/FoSTes
hg19	5	11	110908147	-		11	102569569	+	CTCCCTCTAAAAGTTATTCCTCATGAACTCCCT/TTATCAGTTCCATAAG GAGTAATTAATCCCTCTGTAGGTTAAAAGGGCTATGATGATCAAC	5	0	MMBIR/FoSTes
hg19	6	11	106682089	-	152	11	132850739		ATTAAGTGACACGGGACTGTATGTGCTGTGAAATAAGTCAACCGTAC/ACTA TTGGATTGGACTAAGTTAACTAAGCACACCTGGAGAAAAAGCCAGTT GCAATTTACTTAGAAAAGAAATAGGCTAAAAGTTGACCAGTTGGGACTGTGAT CAGTTTCATCTATGTTCCACTACTACATGGAATGTGCTCAATACATGTA CTCGGTGACACAGGAAGTGAAGGGGCTGCAGTTGGGTTATTTTC	2	0	MMBIR/FoSTes
hg19	6	11	132850737	-		11	106682091	-	TTTTTCCAGGTGTAGTTAACTTAGTCCAAATCCAAATA/GTGTACG GTGACTATTTACAGCACATACAGTCCCGTGCATTAATGACAGGGA	2	0	NHEJMMBIR/FoSTes
hg19	7	11	106682242	+	152	11	131898574	-	ATTAAGTGACACGGGACTGTATGTGCTGTGAAATAAGTCAACCGTACACTAT TTGGATTGGACTAAGTTAACTAAGCACACCTGGAGAAAAAGCCAGTTG CAAATTTACTTAGAAAAGAAATAGGCTAAAAGTTGACCAGTTGGGACTGTGATC AGTTTCATCTATGTTCCACTACTACATGGAATGTGCTCAATACATGTA/CACC TCGGTGACACAGGAAGTGAAGGGGCTGCAGTTGGGTTATTTTC	0	6	MMBIR/FoSTes
hg19	7	11	131898574	-		11	106682242	+	TCATCTATGTTCCACTACTACATGGAATGTGCTCAATACATGTAACCTC/G GTGACACAGGAAGTGAAGGGGCTGCAGTTGGGTTATTTCCCTCCCT	0	6	NHEJMMBIR/FoSTes
hg19	8	11	106698462	-	11	11	125647992	+	TCACTACAGGACCTTATCTTGCCTAGTGAAGGGGCAATAACTTCAC TGTGGGTAGT/TTTGTAGTTTTTCATAACAGGAAAGCTGTTACCAAGAG	0	34	MMBIR/FoSTes

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insertion	Inferred mechanism
hg19	15	11	106969739	-	-	11	131700488	-	AGCAATCTAAGCAATAATAGAACAAAAGCTGGAGGCACTCACACTACTGT/G CAAAGTAAAAACAAAATCACATCAAAAAGTGGGCTAAGGACATGAATA G	0	3	NHEJ
hg19	15	11	131700488	-	-	11	106969739	-	TTCATGCTTAGCCACATTTTGATGATTTTTGTTTTATCTTTTCACACA/G TAGTGTAGTGCCTCCAGCTTTGTTCTATTATTGCTTAGGATTGCTTTG	0	3	NHEJ
hg19	16	11	107005789	-	128	11	131898744	-	CAAAATGGCTGCCCTTGCACTCCCTCTGGATGATGAGGGGA/TTTCT ATTTATCAAAATCCTATCTGATGGTAAACCTTCAGCTTGTCTCAA	3	0	MMBIR/FoSteS
hg19	16	11	131898747	+	1	11	107005792	-	TGTCATATTTCAAAATGGCTGCCCTTGCACTCCCTCTGGATGATGAGG GGATTT/CTATTTATCAAAATCCTATCTGATGGTAAACCTTCAGTCTC	3	0	NHEJMMBIR/FoSteS
hg19	17	11	107005918	+	128	11	131701543	-	CTCTATACTAATGTAATCCCAAGTTTGGCTCCTCTTGC AAAATGGATC CCATA/ACTAGGTAGTATGCTTACGTAGCAAAATAGCTTTCTTGAAAA	0	13	MMBIR/FoSteS
hg19	17	11	131701543	-	-	11	107005918	+	AATCCCAAGTTTGGCTCTTGGCAAAATGGATCCCACTACGCTAGTT AT/GCTTACGTAGCAAAATAGCTTTCTTGAAAACTCCTCAATAATCT	0	13	MMBIR/FoSteS
hg19	18	11	107745155	+	-	11	134504777	+	GCTGAGTAGTGGGATACATGCACCAACACCCGGCTAATTTGTATTT TAGAA/TACTACTTAGTGAGAGCTTGGTAAATATTGGCTGAGTTAATGA	0	3	NHEJ
hg19	18	11	134504777	+	-	11	107745155	+	GAGCCTACTTCTTCAATTAACCTAGCCAAATTTACCAAGTCTCTACTAAGT A/GTATTTAAAAATACAAAATTAGCCGGTGTGGTGCATGTAATC	0	3	NHEJ
hg19	19	11	107889088	-	-	11	111715709	+	GCTGTATTTCTACTGACCACTATATCAGTATAATATGAAAGGAACTCTTT TAA/AAAGATTAATAATACTCAGAAAAGCTGGGGTGTGGCGCATGC	0	27	MMBIR/FoSteS
hg19	19	11	111715709	+	-	11	107889088	-	AAATAGTAGTGTATTTCTACTGACCACTATATC/AGTATAATATGAA GGGAATCTTTTAAAGATTAATAATACTCAGAAAAGCTGGGGTGTAT	0	27	MMBIR/FoSteS
hg19	21	11	108087607	+	-	11	123496319	-	CATGCCCAACTTGGCAACTTGGTCTCATTAGATTAGATTAGATTGATGT ATTCTGAGCTGGG/TCTGACTTTTCCATTGACTAAGCTCTGGCCCTGGC	11	0	MMBIR/FoSteS
hg19	21	11	123496308	-	-	11	108087604	+	AACTTTGGCAACTTGGTCTCATTAGATTAGATTAGATTGATGT/CTGAG CCTGGGTCCTGACTTTTCCATTGACTAAGCTCTGGCCCTGGC	11	0	MMBIR/FoSteS
hg19	22	11	108576353	-	-	11	122874737	-	TTTTCCATAGAAAAAAGTTTTCAAACTCTTACCCTTACT/TTTCAATGCC ACCCACAAAATTTCTCAGTAAAAAGTTTATTTGCTTCTCACATTTT	4	0	MMBIR/FoSteS
hg19	22	11	122874733	-	-	11	108576357	-	AGAAGCAATAAACTTTTACTGAGAACTTTGTGGTGGTGGGCA/TGAA AGTAAATGGTAGGAGTTTGAACCTTTTCTATGAAAAATAAAAGT	4	0	MMBIR/FoSteS
hg19	23	11	108622263	+	-	11	113485664	-	AGTCAAGTGGGAGGGTCCATAAATGTATACCCTAATCCAAATGTAT/C AGGGAATGAGCATTCAAAATTTTGTGTACAGTTTGGAAACCCGAGGGT	5	0	MMBIR/FoSteS
hg19	23	11	113485659	-	-	11	108622257	-	TGTATCAGGGAATGAGCATTCAAAATTTTGTCTGT	5	0	MMBIR/FoSteS
hg19	24	11	108773624	+	-	11	123474515	+	ATGCTGTAGTCCCAGCTTACAGGAGTTTGTACTGACGATGACATAA TCCTTCCATTTGTAAGTAGAATGTAAAAAATTTGCACCTTTACATAAA AATG/TGCTGAAAATGTAATCAGACCTTAGCAATATCTTGCCTTCCAACTTT	1	0	NHEJ
hg19	24	11	123474516	+	-	11	108773623	+	GGGAAAATTTGGCAAGATAATTTGTAAGGCTGATTTAACATTTCCAGCAG /ATTTTTAATGTAAGGTTGCAATTTTTTACATTTACTTTACAAAATGGA	1	0	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insertion	Inferred mechanism
hg19	25	11	110909142	-		11	111185673		CTTCTGATTTAAGCTAGGAAGGTTGCTTAAACATCCATGTAATGGATGTAC ACATATCTCTGTTGCTCCGGATG/ATAAGATTAGTCTTCTTATCATC	0	51	MMBIR/FoSTeS
hg19	25	11	111185673	-		11	110909142	-	CAGTGATAGAAGAGCAATCTTATCATCCGGGAGCAACAGAGAATAT GTGTACATCCATTACATGGATTAAAG/CAACCTTCTAGCTTAAATCAG	0	51	MMBIR/FoSTeS
hg19	27	11	110909427	+		11	131911306	+	TGCAACTCTGTTACCTTGAGAAGCTAGGTTTACAACACA/TACTGTGTGT GTACCTTTAAACATTTCCATCCACATTTGCTTCTCAGATACACTT	0	31	MMBIR/FoSTeS
hg19	27	11	131911295	+		11	110909427	+	GCCAAGTATCTGAAGACAAGAAATGTGGAA/GTGAAGAATGTTAAAG GTACACAACAGTATGTGTTGTAGAACCCTAGCTTCTCAAGGTAACACAGAGT	0	31	MMBIR/FoSTeS
hg19	29	11	110919847	+		11	121534512	-	TTTCAGAAACATCTATGTCTATCCAAAGGGTGTACCCAGTGTCTCAACA CACCAG/AGCATGGCTAGAACAGAGGGTGTGGAGAGGCTTGTGGAGCC	1	0	NHEJ
hg19	29	11	121534513	+	9	11	110919846	+	ATCACCTCTAACCTTGGCTCAGCAAGCCCTCCACACCCTCTGTCTAGC CATGCTG/TGGTGTGTGAGGACACTGGGTAGCACCTTGGAAATAGAC	1	0	NHEJ
hg19	30	11	110920767	-		11	121445845	+	TAGAGATCTCTGGTCCAACTCTTTAACTTCTACTTTCTAAAG/TGGTCATG CATCATGTAAGGTTGAGACTACGTTCTGACAAATGCATCCAGAGCGGAT	2	0	NHEJMMBIR/FoSTeS
hg19	30	11	121445847	+		11	110920769	-	AATGATAGAGATTTCTGGTCCAATCTTTAACTTCTACTTTCTAAAGTG/G TCATGCATCATGTAAGGTTGAGACTACGTTCTGACAAATGCATCCAGAG	2	0	NHEJMMBIR/FoSTeS
hg19	31	11	111095573	+		11	111445161	+	AAATAGTAGACTTCTGTACATGAACAAATTAACACTAGGTGAAAAAGAAAT CAAGAAGCAATCCG/TTCTCTATAATTTGTATATGTTCTTAAATTTCCA	1	0	NHEJ
hg19	31	11	111445162	+		11	111095572	+	ATTTAAATTTTATGTTGAAATTTAAGACATATACAAAATATAGAGAA G/GGATTGCCTTCTGATTTCTTTTACCTAGTTAATTTGTCATGTAC	1	0	NHEJ
hg19	32	11	111100829	-		11	121699207	+	CTTCAAAGGAAATCTAACACGTTCTTCTATGCGAAGGAAAATGTTAA/ AGCCAATGCTTTCTCCCATTTCTCAGATGAGAACTGAGGCACAGAG	0	5	NHEJMMBIR/FoSTeS
hg19	32	11	121699207	+	4	11	111100829	-	TTTTCTCAAAGGAAATCTAACACGTTCTTCTATGCGAAGGAAAAT/GT TAAAGCCAATGCTTTCTCCCATTTCTCAGATGAGGAACTGAGGCAC	0	5	NHEJMMBIR/FoSTeS
hg19	33	11	111105186	-		11	121781513	+	CATATTAACAAAACTAAGAAGTGTGAGCAATGTATGTGTGT/GAGGTGG CCAAGAAAGAGAACATAGAGCAAGGGGCTGGCGGAACATTAGAAAACTC AC	0	0	NHEJ
hg19	33	11	121781513	+		11	111105186	-	TTTAACTTTAGATCATATTAACAAAACTAGGAAGTGTAGCAATGTATGT GTGT/GAGGTGCGCAAGAAAGAGAACATAGAGCAAGGGGCTGGCGGAAC	0	0	NHEJ
hg19	34	11	121106933	-		11	111143258	+	ATAAAGAACCCAGACATAGTTGTCACTAAAAATGAATACCTGATGTTAA CTGATGACTGTGGAGGC/AAATAAGTTTTCTGTTTTGGCTGGGTTATA	0	32	MMBIR/FoSTeS
hg19	34	11	111143258	+		11	121106933	-	AGAGTACATAAAGAACCCAGACATAGTTGTCACTAAAAATGA/ATAACT GATGTTAACTGATGACTGTGGAGGCAGATAAATTTTTCTGTTGGCTGG	0	32	MMBIR/FoSTeS
hg19	36	11	111579295	+		11	121534505	+	AAAAATAGTCTATTTTTTCTATACATTTAAACCACCAAT/ACATGTTGC CATGCTGAGTTGCCCTTATTTAAACAAGCCCTTGTCCCATTTCTATCTCT	0	7	NHEJMMBIR/FoSTeS

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insertion	Inferred mechanism
hg19	36	11	121534505	-	9	11	111579295	+	ACATTTAACCACTAATAACATGTT/GCCATGCTGAGTTGCCTTATTTAAACAA GCCTACTTGCCCATCTACTCTGATCCCTTCTGTCATGCCATTC	0	7	NHEJMMBIR/FoSTeS
hg19	37	11	111628252	-		11	133561215	-	TTTATCTTTTCTCTTGAATGATGATCCACTTTGAAT/CATATTT GTTAAAAAATTAATCTCTGCTGGTGCAGTGGCTCACACTGTAAT	0	0	NHEJ
hg19	37	11	133561215	-		11	111628252	-	TGAGCCATGCACAGACAGAAAGTATTTTAAACAATATG/ATTTCA AAGTGGATCAAGACTACATCTAAGAGAAAAAGTAAAAAATTTTCAGA	0	0	NHEJ
hg19	38	11	111865245	-		11	122229823	-	TCTTGACTGTGTTGGAATAAGCCCAAGACCACTGACAC/ATATATGT ATATGTGTATATGATATGATATATATATATATATGATATGATATGAT	0	32	MMBIR/FoSTeS
hg19	38	11	122229823	-		11	111865245	-	ACATATACATATACACACATATACATATATGTCAGTTGGCTTGGGGCTT ATCCAAACAT/AACAGTCAAGATGTTATTTCAACACATTTCTTCCTTG	0	32	MMBIR/FoSTeS
hg19	39	11	112154915	+		11	121563369	-	CGATCTGGCTCACCGCAACCTCCGCTCCAGGTTCAAGGATTTCTCTGC CTCAGCTCCCJ/GTCCAGTGTATGTCTCTAATGTGCCCTGTTCAAT	1	0	NHEJ
hg19	39	11	121563368	-		11	112154914	+	ACCTCGCCTCCAGGTTCAAGGATTTCTGCTCAGCTCCC/IgTCCA GTGATGTGCTCTAATGTGCCCTGTTTATGTTGTCACCCACCCCGA	1	0	NHEJ
hg19	40	11	108773565	-		11	123581795	+	ATGCTGTAGTCCAGCTTACAGGAGTTGTTACTGCAGCATGAC/ATA ATCCCTTCCATTTGTTAAGTAGAATGTAAAAAATTTGCACCTTTACATTA	6	0	MMBIR/FoSTeS
hg19	40	11	112179538	+		11	108773565	-	ATGCTGTAGTCCAGCTTACAGGAGTTGTTACTGCAGCATGAC/ATA TC/CTTTCCATTTGTTAAGTAGAATGTAAAAAATTTGCACCTTTACATTA	6	0	MMBIR/FoSTeS
hg19	41	11	112179560	-		11	123581795	+	NA	NA	NA	NA
hg19	41	11	123581795	+		11	112179560	-	AATTAGCTAAGTGTGGTGGCACATGCTGTAGTCCAGCTATTCAGGAGGT T/TGTTACTGCGCATGACATAAATCTTCCATTTGTTAAGTAGAATGTAA	0	0	NHEJ
hg19	42	11	113242173	-		11	113485694	+	CCATCTAATCCAAIGTATCAGGAAATGAGCATTCAAAAATTTGCTGTACA GG/TTTGGAAACCCGGGTACAGAGGGGCTGACTGTATCTGCTTTGTTGGA	0	5	NHEJMMBIR/FoSTeS
hg19	42	11	113485694	+		11	113242173		TGTATCAGGGAATGAGCATTCAAAAAATTTTGCTGT	0	5	NHEJMMBIR/FoSTeS
hg19	43	11	113285974	+		11	131250089	-	GTCATGGACCTGCCTCCCTCTCTGGAGACTCAGATTGCTCCA/GTACCCT TCCCTTGAGTCTTTTTTTTTTTTTTTTTTTTATAGATGGGTTTCCCTCCT	0	0	NHEJ
hg19	43	11	131250087						NA	NA	NA	NA
hg19	44	11	121223869	-		11	121699243	+	GAGGAAGGGAAGTTGAAGCAAAATTCAGATAGATGTTCTAAGGTACAAGA AAAAATTCAG/CCTGAGGAAGGGGTGCTAGTTCCCTCCAGTCTATTCT	0	1	NHEJ
hg19	44	11	121699243	+		11	121223869	-	TCTTATCATTTGAGGAAGGAAAGTTGAAGCAAAATTCAGATAGATGTTCTAA GGTACAAGAAAAAATTCAG/CCTGAGGAAGGGGCTGCTAGTTCCCTCCAC	0	1	NHEJ
hg19	45	11	121272048	+		11	121512213	-	GCAATTTGATACATATTTTGTGGTATATGTTTTCATACCTCTTGAGTAT ATGCCAAGGAAT/TGCAATTCAGAAAAACATTTGTTACTGGCTGCCTCT	8	0	MMBIR/FoSTeS
hg19	45	11	121512205	-		11	121272040	+	ATAATTTTGTGGGTATATGTTTTCATACCTCTTGAGTATGCT/CAAGGA ATTGCAATTCAGAAAAACATTTGTTACTTGGCTGCCTCTGCTCTTATTGTA	8	0	MMBIR/FoSTeS

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	55	11	125647883	-	11	11	125647893		AGTCACTCCCAATATATCAAGAAATTTGTTTATGCATAAACCAGTAAATGAA ATCCAGCATTTGGAATATATTTTACATTTACATTAACACATTTATTTCAAAAAATG TTGAAATATGAAG/ACATTAITTTCAATCGCTCCAGATACAAAAAGAGGTG CTTAGAGGATGCAGTCTTGAGAAATATTTCTCTGGGGCTGGGGCTTCACTAC AGGACCTTATTTCTTGCCTAGTAGTGAAGGGGCAATAACTTCACTGTGGGT GTAGTTTTGTAGTTTTTCATAACAGGAGAGCTTTACCAAGAGAA	11	19	MMBIR/FoSTeS
hg19	55	11	125647893	+	11	11	125647883		AGTAAATGAAATTTCCAGCATTTGTAATATATTTTTACATTTCAATTAACACATTA TTTCA/AAAATGTTGAAATATGAAGACATTTTTCAGTCGCTGCCAGATA	11	19	MMBIR/FoSTeS
hg19	56	11	129344020	-		11	133022623	+	TCTCTAATTTGATGGCTGGGACTCCTAGGCCCTCATTTTCTCTC/AGTGA AATGGCTCAGTAAATTCCTAAGAAATATAAAATTCAAAATGAAAAGG	0	0	NHEJ
hg19	56	11	133022623	+		11	129344020	-	TGTAATTAATTTCTCTAATTTGATGGCTGGGACTACCTAGGCCCTCATTTT CCTC/AGTGAATGGCTCTCAGTAAATTTCTAAGAAATATAAAATTCAAAATGAAAAGG	0	0	NHEJ
hg19	57	11	131709405	+		11	132999532	+	TCCACCTGAGGAAAGCAGCCAGGGCAGAGGCCCTCAGGCCATGCTATCCTG CTCAGTGTG/AGACACACACCCAGTCTTCTTTACATGCTGATACT	6	0	MMBIR/FoSTeS
hg19	57	11	132999538	+		11	131709399		GTTTTAAAAAGTATCAGGCCATGTAAGAAAGCAAGGACTGGGTGGTGTCT CACACT/GAGCAGGATAGCATGGCCTGAGGGCTTCTGCCCTGGCTGCTTCC T	6	0	MMBIR/FoSTeS
hg19	58	11	131626353	+					ACCTTGTCTGGGAGGCTGAGCAGGTCAATTGCTTGAACCTGGAGACGG AG/CGAGACCCCTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA	0	0	NHEJ
hg19	58	11	131795300	+					NA	NA	NA	NA
hg19		11	101839709	-	7	11	111185562	-	ATTTGTTTTCTGTGGTGTGTAATAATAAATCGGTTATTTCTTATGTT AACATAAGATCCTTTT/TTAACACAGATTTTTTAATGTTAATTTT	0	40	MMBIR/FoSTeS
hg19		11	121534310	-	4	11	111185590	-	TGTTAACATAAGAAATAACCGATTTTATTTACAAACCGACACCCAGAAAATA CAAAATACACACAATGTTATAAGTTG/TTTTCCATCTCCAGCATGAAAAC AGGAAA	0	23	MMBIR/FoSTeS

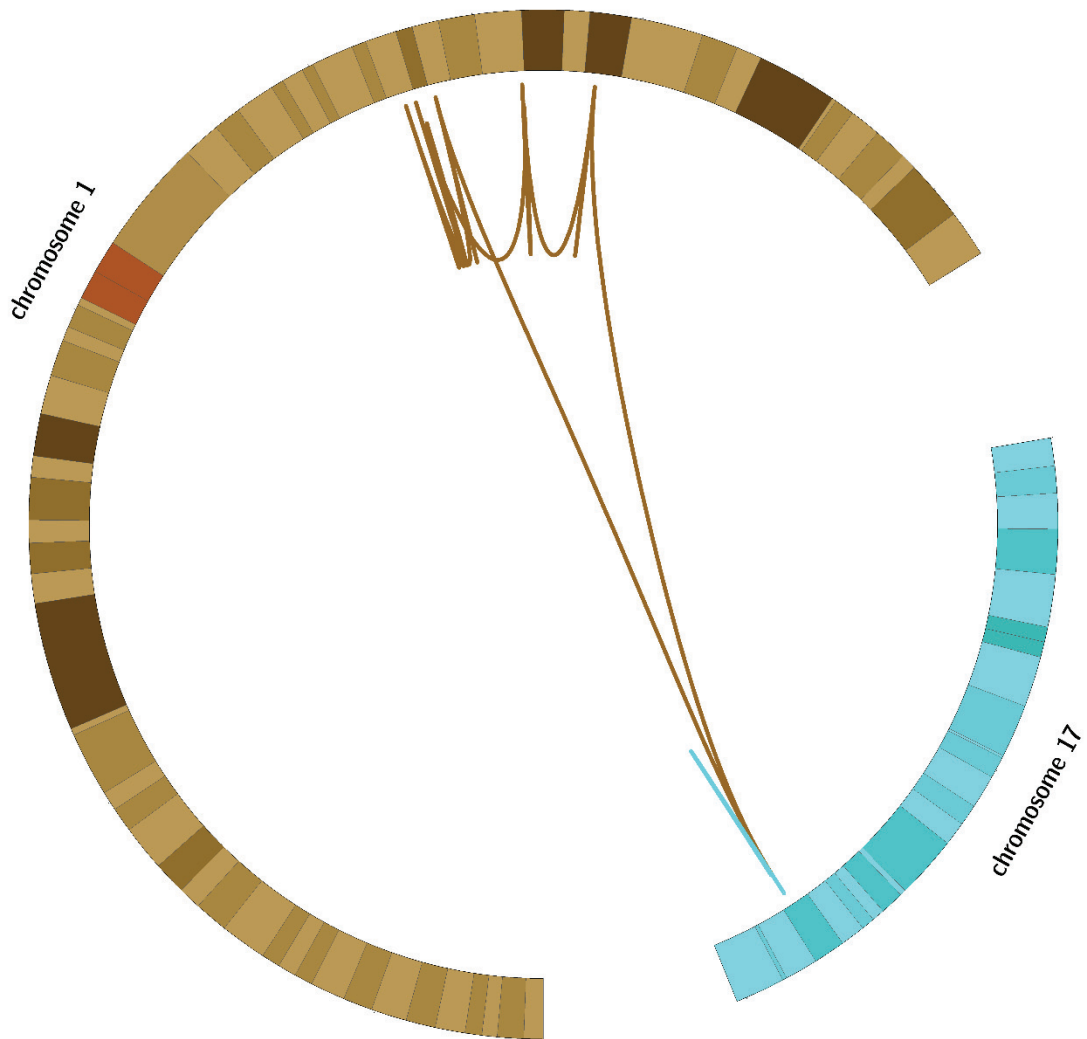
Patient 10

Phenotype: Maxillary prognathism

Karyotype: Not performed

Comments: Inherited from affected parent

Circos plot:



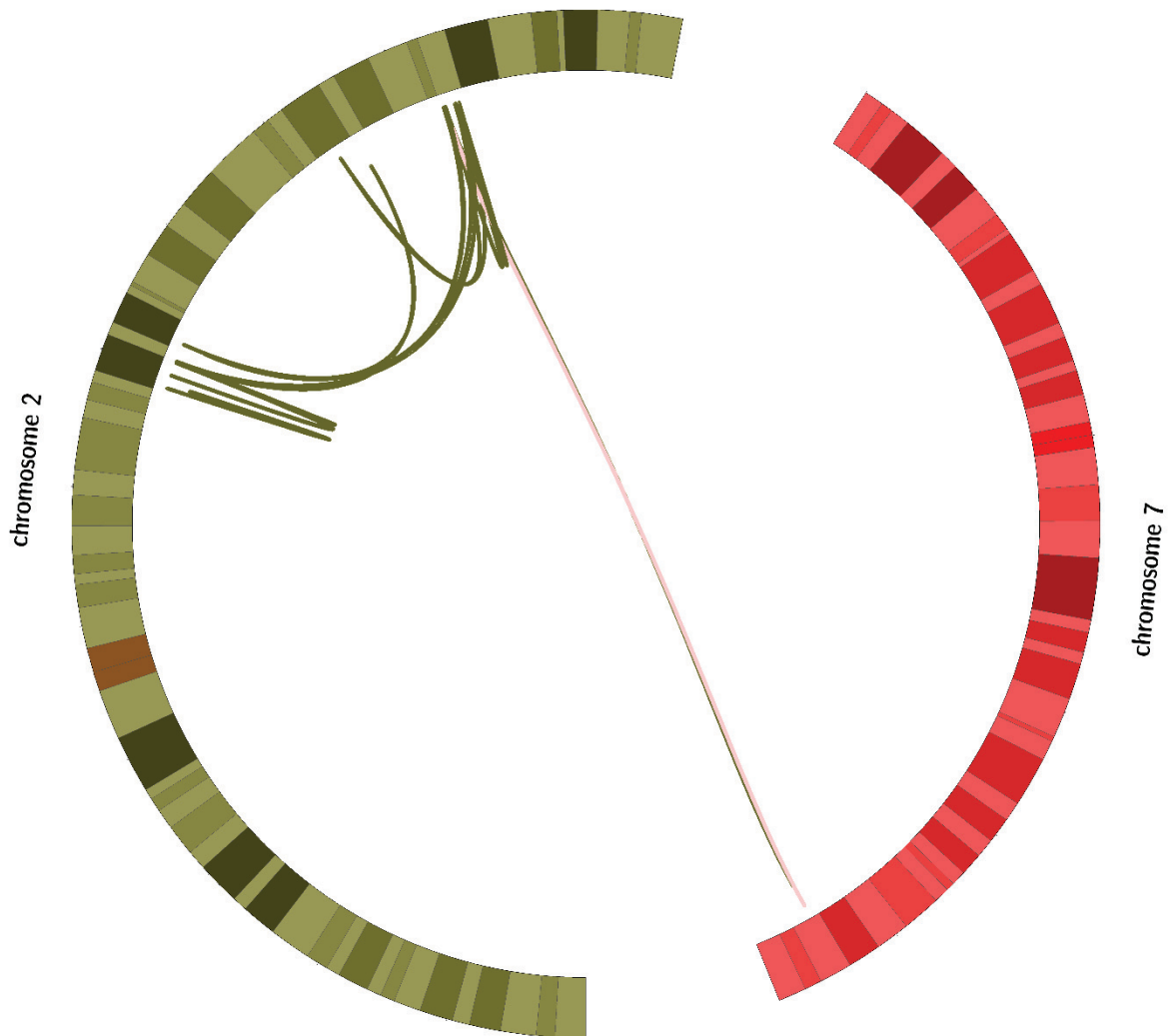
Patient 11

Phenotype: Intellectual disability, facial dysmorphism, cardiac malformation, urogenital malformation

Karyotype: 46,XY,ins(7;2)(q31;q23.31)dn

Comments: Disruption of *ZEB2* (Mowat Wilson syndrome) and *COL3A1* (Ehler-Danlos syndrome)

Circos plot:



Ref.	Junct -ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19	10	2	205573995	-	-21	2	186183969	+	TAAACAAAAGGGGGCAGAGGAAAAATG/TAAACATTTTCCGCATGATGC AGAACGTACTGCAGTCTTTCTGTGAGAGAAAAACAGTACCAAGAGAGAAA A	1	0	NA
hg19	11	2	186183974	-	-3	2	207963107	+	TCTTCCCTTCAAGATGAACAATACACCAAGT/GAATCTGCGGTTTATATT AGATAACAGACAGACAGATGGGTAGGGAAACAATGAGATATGCATTTGT	7	0	MMBIR/FoSTeS
hg19	11	2	207963114	+	-6	2	186183981	-	TCTTCCCTTCAAGATGACAACATACACCAAGT/GAATCTG/	7	0	MMBIR/FoSTeS
hg19	12	2	205295565	-	-87	2	207277526	+	/CTGCAAAACAGTTTGACTTCTCTTTTCTTATTTGAATA	0	1	NA
hg19	12	2	207277526	+	-22	2	205295565	-	GTCTATTCAGAAAATCCACTTCCTCTAGTCTTGGAGAGTGTATGT GT/ACTGCAAAACAGTTTGACTTCTCTTTTCTTATTTGAATACCCTTTA	0	1	NA
hg19	13	2	205595820	+		7	149331068	+	GTAAGATATTTTGGAGGGTAGTGGTTGATTACAATGTGATGCTTTATAATC CCTGACCA/ACCCCAAGTCCCTCTCTCTACATTAGAATCCCTT	0	1	NHEJ
hg19	13	7	149331068	+		2	205595820	+	GCAATTCAGGGGAAATTAATGTAGGAGAGGGGACTCTGGGG T/TGGTCAGGGATTAAAGCATCACATTGTAAACACCACCTCCCTCCAA	0	1	NHEJ
hg19	14	2	205607886	-		7	149332019	-	GCTGGCATTGCTCTAAATTCCTCTCTTCCAGTCTCCAGGC/AAATGT TGGTAAATGAATGAAGTGGCTTTTAAAGATTTTGGAGAGACT	0	0	NHEJ
hg19	14	7	149332019	-		2	205607886	-	TTAACAAAGCCAATCTTCAATTTAACCAACATTT/GCCTGGGAGGATCT GAAGAGGAGGAGGAAATAGGACAATGCCAGCTGCAGGAGGCAGAGAA G	0	0	NHEJ
hg19	15	2	207344333	-	-21	2	207963124	-	GAGTGATAACATTCCTGCTAAATGGAAAAGAAAATGTTCTCAAAATG/AGTGT GATATTTGATTTGAATAGAGGTTCTGTTGGATAACAGACACACACTT	3	0	MMBIR/FoSTeS
hg19	15	2	207963121	-	-6	2	207344336	-	TATCCAAACAGAACCTCTTCAATACAAATATCC/ACTCATTGTAGAAC ATTTCTTTCCATTTGGCAGCAATGTTATCCCTCTGTAAAGTGAAGTTCT	3	0	NHEIMMBIR/FoSTeS

Patient 12

Phenotype: Hypotonia, unilateral cryptorchidism, pharyngomalacia

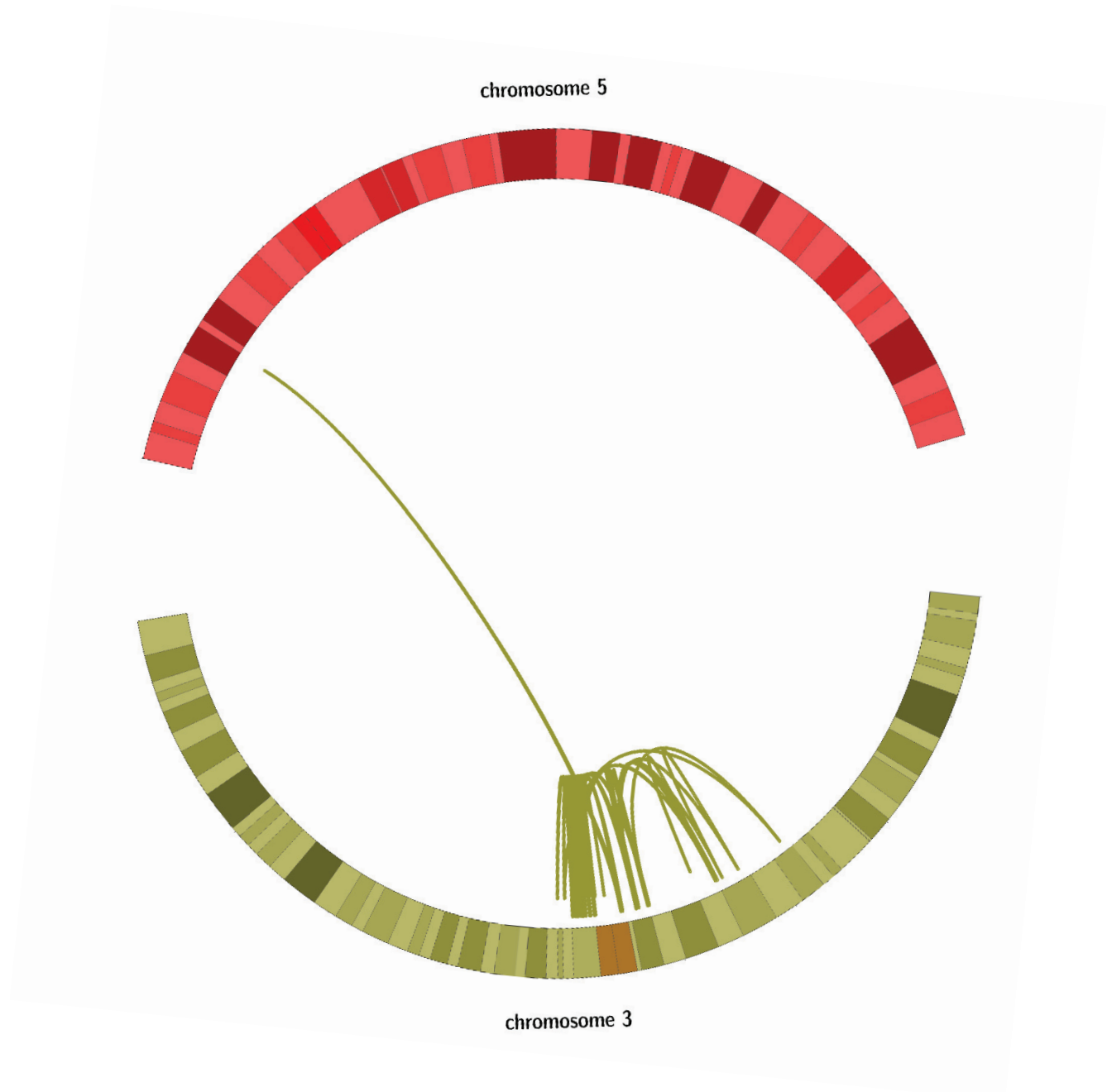
Karyotype: 46,XY,t(3;13;5)(p11.1;p12;p14)dn

Array-CGH : arr[GRCh37]

3p14.1(65238298_68667113)x1,3p13(70127345_73724765)x1,3p12.1(83784489_85467284)x1,3q11.2(97180779_97270083)x1

Comments : Breakpoint on the short arm of chromosome 13 not identified.

Circos plot:



Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insertion	Inferred mechanism
hg38	1	3	68622008	-		3	85960244	-	GTGATGCTATGGGTATGGCATGCCATCTTTCTAAGAGGAAGGATATA/TATGAGATTTAAATAAAGTCTATGAACGTCTCCAGATGCCACTGTCTTCCCA	0	0	NHEJ
hg38	1	3	85960242	-		3	68622010	-	GAACAGATGGCATCTGGAAGCAGCTTCTAGGACTTTTAAACTTCA/TATATATCTTCTCTTAGAAAGATGCCATACCATAGACATCACTTACT	0	0	NHEJ
hg38	2	3	83423395	+	-9	3	83435480		TGCTGAAACTCCCAAGAAATTTACGAATTAATAAACTGAGTTGGT/TGGATCACTTAGAAGGGCAGCACTGCTTCTCGAGCTCTGGGCTCT	0	5	NHEJMMBIR/FoStEs
hg38	2	3	83435480	+	2	3	83423395	+	TATTTAAAGCCAGGAGCTGCAGGAAGACAGTGTCTCGCTCTTCTAAGTG/ATCCAACTCAGTTTTTAAATTCGTAATAATTTCTGGGAAGTT	0	5	NHEJMMBIR/FoStEs
hg38	3	3	83423405	-	-9	3	85968714	-	ATATACTGGTAGAAGACTGTACATTTGGTAAAGCTACAGCC/AAAGGAAATAGGTTGGTAGCAGAGTAGTTAAAGATCCACAAAGGACAA	0	1	NHEJ
hg38	3	3	85968714	-	-2	3	83423405	-	GTGGATTTAACTACTGCTACCAACCTATTTCTCCTTG/GGCTGTAGCTTTACCAAAATGTACAGAGCTCTACCAAGTATATGTGTAAATTG	0	1	NHEJ
hg38	4	3	69655521	-	3	3	85968709	+	GGCACTGGTCTCTATATTTGTTATGTGGAATAG/GCTCCTTGGAAAGCCTATTATCTGAAGGCCCTAACATCTCAGAAAAATTCACCTGAAAGC	2	0	NHEJMMBIR/FoStEs
hg38	4	3	85968711	+	-2	3	69655523	-	AAAGAGAAGGATAGAATCTGTGGACTCGTCCCTCTATATTTGTTATGTGGAATAAGGC/TCCCTTGGAAAGCCCTATTAATCTGAAAGGCCCTAACATCT	2	0	NHEJMMBIR/FoStEs
hg38	5	3	69655523	+	3	3	85863974	-	GCCTTTGTTAATGAAGCACAGATCATGTTTGGCAATGTGATAATCTGCT/AGGGCTGGTATGGTTATGAAACCTGAAAGAGCACACCCACAGAGTACTGAG	0	0	NHEJ
hg38	5	3	85863974	-	1	3	69655523	+	TTAATGAAGCACAGATCATGTTGGCAATGTGATAATCTGCT/AGGGCTGCGTGTATGGTTATGAAACCTGAAAGAGCACACCCACAGAGTACTGAGAA	0	0	NHEJ
hg38	6	3	70057568	+	-142	3	85863967	+	TAGAACTTTTGAATAACAGAAATGAACAGAAAAATTTCCCTTGGAAAACTGTAGCT/TCCATATCCACACATGCTACTGTGACCCACCAA	0	0	NA
hg38	6	3	85863974	+	1	3	70057568	+	CTGGAGACGTGGGAGTTGGTGGTCAAGCAAGTACATGTGGAAATATGGAA/AGCTACAGTTTTTCCAAAGGGAAATAATTTCTGTTCATGTTCTGTT	0	0	NHEJ
hg38	7	3	70057711	-	-142	3	95981617	+	CTACACTTCTGGCTCCAGTAACTACCAAACTA/TAGTGAGAAAT/TTGGA	0	10	MIMBIR/FoStEs
hg38	7	3	95981617	+	-9378	3	70057711	-	TTTACTTAAATAAACCCTCTTTTATCCTTCTCCCTACACTTCCCTGCGCTCAGTAACTACCAAACTA/TAGTGAGAAATTTGGAGAAATTTGCTTTTAA	0	10	MIMBIR/FoStEs
hg38	8	3	95596124	-		3	95990996		CTTTTGAGAAAGTGTCTGTTCATATCCTTTGCCTACTGTGTGGGGT/TCCAGGCTGGAGGGGAGCTGATGATTCCTTCACATCTAGAACCTCCATGTTT	1	0	NA

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insertion	Inferred mechanism
hg38	8	3	95990995	-	-9378	3	95596125	-	TCTAGAAATGTGAAGGAATCATCAGCTCCCTCCAGCCTGG/AACCCACACA ACAAAGTAGGCAAGGATATGAACAGCCACTTCTCAAAA GAAAA CATTTA TG	1	0	NA
hg38	9	3	95558386	+		3	91620277	+	CACATAATGCAGTAGGATGTTACAATGACTTTGACGTCACTAGGCGAAA GGGATG/AGAAATCTCTGTAGCATTATGAAAAATCCGGTTTCCAAC G	0	NA	NA
hg38	10	3	83435479	-	2	3	96057406	-	CAAAATTTGGGAGAACAACATTTAGTACCAAAAAATGA/TGGCATGTATGA AGAGACATCTGATTTAGCAGCTGACAGCTCACAGGAAACTTGACCCA GA	3	0	NHEJMMBIR/FoSTeS
hg38	10	3	96057403	-	-3	3	83435482	-	TCTTCATACATG/CCA TCAATTTTTGGTACTAAATTTGTTCTCCAAATTT GTATGTTGAAGCCCTAACCCCATGTACTGTATTTGGAGACAGGCTCT	3	0	NHEJMMBIR/FoSTeS
hg38	11	3	96057399	+	-3	3	97457781	+	AATCTCAGCCATATTTCCAGTAGATTACTTTTCCTTAAATAATATTA TATGAGTCATCTTTTAAAAATATT/AAGCTGCTTCTAACTCTTC	0	0	NHEJ
hg38	11	3	97457784	+		3	96057396	+	TAAACCCAGCCAGTCCACAAAGTCTGTATGAAGAAATTAAGAAGCAGC TTAAT/AAATTTTAAAGAAAGATGACACATATAATATTTAAGGAAAA GT	0	0	NA
hg38	12	3	97556630	-		3	98083513	-	TTTTGTTTAAACCAAGGAAAA/GATTATGGGAACACTACAAITCAAGATG AGATTTGGTGGGACAGACCCAAACAATATCAGTGTCTTACTATAC AG	0	0	NA
hg38	12	3	98083513	-	5	3	97556630	-	GCTCTGTCCCCACCCAAATCTCATCTTGAATAGTGTCCCATATC/TTTT CCTTTGGTTTTAAACAAAAACAATTTGAAATAAAACACATATATCT	0	0	NHEJ
hg38	13	3	98083517	+	5	3	10096381	+	CTTTATCTACAGGGAATCCTTTGGAAGTGAAGGGAATATCTTCTTT CCTTTTTC/AGTGAATCAGAGCCTCCCTAGGAGAGAGACCAGCTCTGAG C	0	0	NHEJ
hg38	13	3	10096381	+		3	98083517	+	GTAGGGTGTCTGCTCAGAGCTGGTCTCTCTCTAGGGAGGCTCTGATTC ACT/GAAAAAGGAAAGGAAAGATATCCCTTACACTTCCAAAGGATCCC T	0	0	NHEJ
hg38	14	3	97053398	+	1	3	10096381	-	ACTTGTGCTGACTGAGATAGAACCAATTAGAT/CCCATGAA GCCATTCA CTAAGCAGCCCTGATAGTCTCAGTTAGTTATCTCTCACTCATCCTTCAGA 6	0	9	NHEJMMBIR/FoSTeS
hg38	14	3	10096381	-	1	3	97053398		GTACATGGGAAACATGTGTGTTAGTACTAGATGAAACAATAGATCCAATG AA/TCCATTCACTAACCAAGCCTGAAAGTCTAAGTTAGTTAACACTCACT C	0	8	NHEJMMBIR/FoSTeS
hg38	15	3	70060749	+		3	97053398	-	CAATGCACAGGACAGTCCCCACAGCAAAGAAATGATCGCCCAACATGT CAATA/TTGACCCAGAAATGGAAAGAAATTAAGGAAAGTATAGGAAATTC CTC	0	1	NA
hg38	15	3	97053398	-	1	3	70060749	+	CAAAAGATGATCTGGCCCAACATGTCAAAT/TGACCAGAAATGGAAAG AATTAAGGAAAGATAGGAAATTCCTCTCTGACCACAGTAGACACTTAA A	0	1	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insertion	Inferred mechanism
hg38	21	3	85433268	-	1	17_GL00020 5v2_r andom	44006	+	CCTGAGCTTCCAGCGTGATCTTGAAGCTGAAAAGTGAACCTCCAT/CA TAAATTGTTTTCCAGACATCTGTCTATAGTGATAATTCATATGTATATA	0	0	NHEJ
hg38	21	17_GL00020 5v2_r andom	44006	+	1	3	85433268	-	TGGAATGTGGGAAGTCCCTGAGCTTCCAGCCGTGATCTTGAAGCTGAAA AGTGAAACCTCCAT/CATAAAATGTTTTCCAGACATCTGTCTATAGTGAT A	0	0	NHEJ
hg38	22	3	85433266	+	1	3	97041863	+	TCCTTTGAAGTCCAAATTTAGCCAAATGCGTGGGAATATCAATGACCT/C TTGCTGTTCCGATGATAGTGAGTGAGTTCTCAGGATATCCGGTTGTTTAA	0	0	NHEJ
hg38	22	3	97041863	+	2	3	85433266	+	GTGAAGGAGAGGGCTACACATTTAAACAACCGGATCTCCTGAGAAC TCACTACTATCATGCGAAACAGCAAG/AGGTCAATTTGATATCCCTAGGC AT	0	0	NHEJ
hg38	23	3	96046690	+	4	3	97041863	-	ATTGCATGTACAAAAAGAAACAAAAAGCTTTTAAAGTAAAAATTTAAA AGTCACATAAAGATA/GGGGAAAAATCTCCCGAGTGATCTAATCACCTC C	1	0	NHEJ
hg38	23	3	97041862	-	2	3	96046688	+	TAAAGTAAAAATTTAAAGTACATAAAAGAT/AGGGGAAAATCTCCCC AGTGATCTAATCACCTCCCTTAGCCCTCTCCAACACTACAAATTACA	1	0	NHEJ
hg38	24	3	88557356	-	200	3	96046687	-	AAGTCTCAGTATGGA/GAATGTTCTTCCATTTCTGTATCCTCGTTTATT TATTTGAGAAGGATTCGAGTTCTCCTTGAAGAGGTCTACGCATCCC	0	3	NA
hg38	24	3	96046687	-	4	3	88557356	-	ACTGCTCAAAGAAATAAAAGAGGATACACAGAAATGGAAGAACATTC C/ATAGTGAGAACTATGAAGGCAATAAGAACAGTCACCTAAAAATAACTGT GA	0	3	NHEJ
hg38	25	3	88557155	+	200	5	22906970	+	CTGTTTCAGTACCAGTACATGCTGTTTTGGTTACTGTAGCCTTGTAGTAT AGTTGAAAGTCATGATGAT/CTTGACCTCCACCTTATACACCA	2	0	MMBIR/FoSTes
hg38	25	5	22906972	+	1	3	85557153	+	GATGTGACGATTCACCTAAGTATCTGAAATAAATAAGGGTGTGGTGT ATAAGGTGGGAGGTGCAAGAA/CACTACCTGACTTCAAACTATACTA CA	2	0	NHEJMMBIR/FoSTes
hg38	26	3	88891616	-	3	5	22906974	-	TGCTTGTATGTTGCTCAGATGTTCCATTTGCAACTTCTCCAAA/CAGAT GGATTCACAGCCGAATTCACCAGAGGTACAAGGAAGAGGCTGGT GC	2	0	NHEJMMBIR/FoSTes
hg38	26	5	22906972	-	1	3	88891618	-	TGAATCCATC/TGTTTTGGAGGAAGTTGCAATGGAACATCTGAGCAACAT AACAGCATTATCTCATAGTAAGCTCCTTTAAGATGCTACTTGAAGTGC	2	0	NHEJMMBIR/FoSTes
hg38	27	3	88891618	+	3	3	95417215	-	NA	NA	NA	NA

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insertion	Inferred mechanism
hg38	27	3	95417218	-	0	3	88891615		TTACCAACAAAACAGTCCAGGAC/CAGTTGGTAGGTTATCAGGGAAATGAGTTCTGATGAAAGGATAAATGTTGCCTCTTACTACTACAGACTTC	3	NA	NA
hg38	28	3	93944945	+	0	3	95417217	+	AAAAATTAGATAAATAAAGAAGATTGCAAGACTAATAAAGAAGAAAAAGAG/CCTTTTTCATGGTATAACAGCATTAAACATTCATGAAGTCTCTG	0	6	NHEJMMBIR/FoSTes
hg38	28	3	95417217	+	0	3	93944945	+	TGGGATTTTAAATAGCAGAGACTTCATGAATGATTAATGCTGTTAATAACATGA/AAAGGGCTCTTTTCTTTTATAGCTTTCGCAATCTCTTTA	0	6	NHEJMMBIR/FoSTes
hg38	29	3	93944946	-	0	3	94971547	-	AAGATGCTGGTGAAAATGGAGAT/AAAGATCAAAATAGACGCAATAAAAATGATAAAGGGGATAACACCAGATCCACAGAAATACAAACTACCAACA	1	0	NHEJ
hg38	29	3	94971546	-	7	3	93944947	-	TGTTATCCCTTTATCATTTTTATTGCGTCTATTTGATTTCT/TATCTCCATTTTACCAGACATCTGTAGTTATTTCTAATACTTATGTTCTTTCACTC	1	0	NHEJ
hg38	30	3	94660544	-	2	3	94971536	+	TACTATTAACCTTTTCCACTCTTTTTTAATATCCACC/CTAGCTAGCCAAGTATTGTTTCCACATACTCTAAAAATAGTTAAACCACATAAGA	2	0	NHEJMMBIR/FoSTes
hg38	30	3	94971538	+	7	3	94660546	-	ATGAATGTGGCTTCCAAGGCTTTGAGGATATTTATACTATTTAACTCTTTTCCACTCTTTTTTAATATCCACCCT/AGCTAGCCAGGTGATTTATGTTT	2	0	NHEJMMBIR/FoSTes
hg38	31	3	93729798	+	0	3	94660545	+	TAGTCTCTCAGGACTGCTTTATGAATCTGGTGCTCTGTTGGGTGCATATATATTAGGATAGTAGCTCTCTTTGTTGAATGATC/AGAAAATC	0	0	NHEJ
hg38	31	3	94660545	+	2	3	93729798	+	CTAATTTCAAGTCTTTTCTGATTAGCCATATTAAGACACCTAAAAATAAAGGATTTCT/GATCAATTCACAAGAGAGCTAACTATCTTAAATATATA	0	0	NHEJ
hg38	32	3	93470460						NA	NA	NA	NA
hg38	32	3	93729799	-	0				AACTGCTTTGGATACCTTTTGATATTTTTGTACAGTATAGAATATATA/CCTTTACTATTATAATAGCCCTTCTTTGCTCTTTTGTATCTTTGTTGTTT	NA	NA	NA
hg38	33	3	96175170	-		3	96359339		ATAAAGTAAAAATAAAGCATAAGT/GAGGCAGTCAAAGCCCTAGGCAATGATAGGGGACGGTCCCCAGTCAAAGTAAAGTCTA	1	0	NHEJ
hg38	33	3	96359340	+		3	96175171	-	TCTAAAAATAAAGTTGAAAAAGTAAAAATAAAGTAAAAATAAAGCATAAAGT/AGGCAGTCAAAGCCCTAGGCAGATAGGGACGGTCCCCAGTGAAT	1	0	NHEJ
hg38	34	3	85934286	+					AGCCATAGGCCTCAGTTTGACCAGAACAGTCTGGTTCACAACTGTCATCTTCATAATTAATA/ATTATAAAATAATAATAATTTTTCTTTTAAAAA	0	>50	MIMBIR/FoSTes
hg38	34	3	96175100	+					AAGAGATATGACTTCTGTTTTCCAAAAACAGCCTCCCTGTCATCATGACATAGTTAA/TATATACTACTATTTATATAAAGTATATATAATA	0	>50	MIMBIR/FoSTes
hg38	35	3	85934294	-					ATCTCATAATTAATAAATAATAAATAATAATA/TTTTCTTTTAAATAATGCTCTGGAGATGAAATTAATGTTTATCTATATATAATA	NA	NA	NA
hg38	36	3	85934274	-		3	85961075	-	GATATATAATAATAATAATAATAAAGAAATAATAATATTTCTTTATATATAATAATAATAAATAAT/ATATAAATAATTTATATTACTTATATATCCAAA	0	0	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg38	36	3	85961075	-	62	3	85934274	-	GTATATATAATAATAATAATAATAATAAAGAAATTAATAATATATTTT ATATATAATAAATAT/ATATAAATATTTACTTATATCCAAA	0	0	NA
hg38	37	3	83384016	+	368	3	85934379	+	AAAAAGTTTAGGCTCCATAGCTGGAAACTTAAACACCAATTTTC /CAATCAGCACATCTAGGAATCAGTTTGGACTAATAATATTATA	0	0	NA
hg38	38	3	85934287	+		3	85934344		GTATATATAATAATAATAATAATAAAGAAATTAATAATATTTT ATAT/ATATAAATAATAATAATAATAATAATAATAATCCAAA	0	>50	MMBIR/FoStes
hg38	38	3	85934344	+		3	85934287		AAAAAGTTTAGGCTCCATAGCTGGAAACTTAAACACCAATTTTC CAATCAGCACATCTAGGAATCAGTTTGGACTAATAATATTATA	0	>50	MMBIR/FoStes
hg38	39	3	83384385	-	368	3	83404203	-	AGGTTCTCGAGGGGAGAGAGCCAGGAGGAGAGAGAC/ATGGG AGTAAATTAATCAGGGGGTGGTTACCTCCATGCTGTTCTCGTAGATG GAGT	2	0	MMBIR/FoStes
hg38	39	3	83404201	-	453	3	83384387	-	ATGGAGGTAACCAACCCCGTGAATCAATACCTCCC/ATGCTGCTCCC TGCTGGCTCTCTCCCTCAGGAACCTGCTCAAACCTCAAAGCAGGAT	2	0	MMBIR/FoStes
hg38	40	3	83403747	+	453	3	96672726	-	GCTCCTAGATGCCTCTCAGCCATCCAAACCATGGATACCGACCCAGG GTCCTACTCTATGGAGCAGGCA/CCTTTTTGCTTTTGAATAATA	1	0	NA
hg38	40	3	96672725	-	2	3	83403746	+	TATGGAGCAGG/ACCTTTTTGCTTTTTGACTTAGAATAAATCTGAACTG GTGAGGTGCTCACAATGAGGTTCTCTAAAAGTTACTTTTCTACTTT	1	0	NHEJ
hg38	41	3	96672726	+	2	3	99535332	-	TAAATACCTTTTTTAACAGAAATAGCCCATACTTAAAGATTTTGAGTTT GTAAGTAC/GAATTTCTGGAACATAATGTGTATGCATAATGGGAGGAG	0	0	NHEJ
hg38	41	3	99535332	-	16	3	96672726	+	AGCCCCACTTTAAGATTTTGTAGTTTGAAGCTAC/GAATTTCTGGAACA TAAATGTATGCATAATGGAGGAGATCAGGCGGAAATGCAGGCTGG GG	0	0	NA
hg38	42	3	55754537	+		3	83435578		GTACAAAGGCACGTAGGCATGTTGGAAAGAAATTAATTCATATTTATA /TTGAAATTTAAAAAAGAGGAGTGGGGACTAAGAAATCTGAGTCCCT GG	0	0	NHEJ
hg38	42	3	83435578	-	10	3	55754537		GTACAAAGGCACGTAGGCATGTTGGAAAGAAATTAATTCATATTTATA /TTGAAATTTAAAAAAGAGGAGTGGGGACTAAGAAATCTGAGTCCCT GG	0	0	NA
hg38	43	3	55754556	-		3	99535311	+	GTACAAAGGCACGTAGGCATGTTGGAAAG/AAATTAATTCATATTTAT ATTGAAATTTAAAAAAGAGGAGTGGGGACTAAGAAATCTGAGTCCCT GG	4	0	MMBIR/FoStes
hg38	43	3	99535315	+	16	3	55754552	-	GTACAAAGGCACGTAGGCATGTTGGAAAGAAAT/TAATTCATATTTAT ATTGAAATTTAAAAAAGAGGAGTGGGGACTAAGAAATCTGAGTCCCT GG	4	0	MMBIR/FoStes
hg38	44	3	83435587	+	10	3	85934387	-	TTGACCCAGACAGAAACAGAATGTGAAATAACAGGACTCATCCTTG AATTA/TTAATAAAGTATGTTAAATCTGATATCAATTTTTTCTCTA	2	0	MMBIR/FoStes
hg38	44	3	85934385	-		3	83435585	+	GTGAAATAACAGGAGCTCATCCTTGAATTT/AAATAAAGATGTTT AAATCTGATATCAATTTTTTCTCTATCAGTAAAAAAGTATATTT	2	0	NHEJMMBIR/FoStes

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -log ₁₀	Insertion	Inferred mechanism
hg38	45	3	85934384	+		3	83384007	-	TATATATATAAATATATATAAATATATATAGTCAATAAAGTCAATCAATGATTCCATA GATGCTGATAATGGAAA/TATGGTGGTTTTAAGTTCCAGCTATGG	0	41	MIMBIR/FoSTes
hg38	45	3	85934344	-	2	3	85934343	+	TATATATATAAATATATAAATATATAAATATATA/TAGTCAATA ACTGATCCATAGATGCTGATAATGGAAAATATGGTGAGTTAAGT	41	0	MIMBIR/FoSTes
hg38	46	3	65243091	+					TTAAGATAACTGTGTTAAATAAATACTAGGAAAATGATCCAAATGCT GT/ATATAATATATATATAAATATATAAATATATAAATATATA	NA	NA	NA
hg38	46	3	85934446	+	9				AAAGTATGTTAAATCTGATAATCAATTTTTCTCATCAGTAAAAAA GTAT/ATAATTTTTATAAATATATAAATAAATATATAAATTTTTCTC	NA	NA	NA
hg38	47	3	95973473	+	-2	3	96276232	-	GGTGTCTCAGATGGAAATGAGAAATCATGTTGGGAGTAGAGCAGAGA TGACTCTTGTTATGTTTTAGCAAAGAACTGG/ATCAGTTGTAATAAGCC ATG	0	0	NHEJ
hg38	47	3	96276232	-		3		-	NA	0	0	NHEJ
hg38	48	3	95973474	-	-2	3			AGCTTCTGCACAGCAAAAAGAACTACCATCAGAGTGAACGCAACC/T GACATTTTGCCCTCTTAGCAATTTGTGGAACCTTTGAACTTGAGAGAGA TG	0	>50	MIMBIR/FoSTes
hg38	48	3	94851098	+	9	3			AAAAAGACTTCTGCACAGCAAAAAGAACTACCAGAGTGAACAGGC AACCTTCAAAAT/CGGACAAAATTTAGCACCCCTCACAACTCACAAAAG ACT	0	>50	MIMBIR/FoSTes
hg38	49	3	94851090	-	9	3	96213736	+	CAATAATCCTCTGGCTAAGCCCAAGTTTTGGTG/CTACAAAATGGGAG AAAAATTTCAACAACCTACTCATCTCACAAGGGCTAATATCCAGAACTAC	2	0	NHEJMMBIR/FoSTes
hg38	49	3	96213738	+		3	94851092	-	TTTTATTGTACCACCTTTTCACTGTTGAATTTCTCCCTGGGCAAGCCAAAT AATCCTCTGGGCTAAGCCCAAGTTTTGGTGCT/ACAAAATGGGAGAAAA	2	0	NHEJMMBIR/FoSTes
hg38	unresolved	3	85934345	+	2				TTAAAAATGCTCCTGCTGGAGATGAAATTAATATGTTTATCCTA/TATATA TATATATATAAATATATAAATATAAATAAATAAATAAATAAATAAATA	NA	NA	NA
hg38	unresolved	3	85934438	-	9				TTTTCTTATATAAATATAAATATAAATATAAATATAAATATAAATATAA AATATAAATATAA/AAAAAGTATGATGAATGTATCATATTTTTTCAGAT TAATTTTTCTCCTAAGCAATTAATACCTTAATATATTTAGTATATAATA TATATATAAATATAAATATAAATATAAATATAAATATAAATATAAATATAA	NA	NA	NA
hg38	unresolved	3	85934497	+					ATGTTATCATATTTTTTCAGATTAATTTTTCTCCTAAG/CAATTAATAC TTAATATAATTTAGTATATAAATATAAATATAAATATAAATATAAATATAA	NA	NA	NA

Patient 13

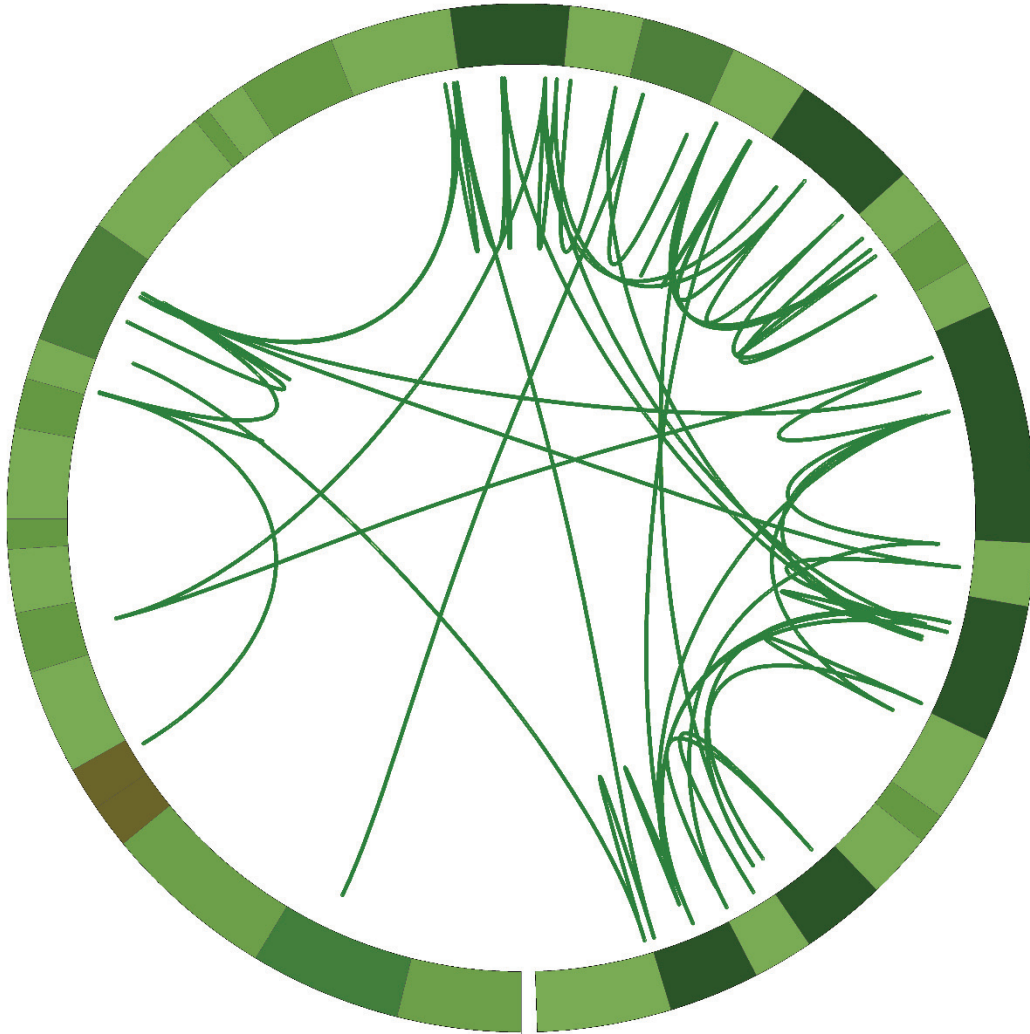
Phenotype: 3,37 mm cystic hygroma at 15 weeks of gestation

Karyotype: Amniotic fluid : 45,XX,rob(13;15)(q10;q10)[20] with abnormal structure of chromosome 13 on the derivative chromosome

Comments: Breakpoint on short arm of chromosome 15 not identified.

Circos plot:

chromosome 13



Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	15	13	5885957	+	2	13	24453512	+	AGAAAATCTTGTCTATACAGGAATGCTCAAAACAACCTTTACTGG/CAAGGTAACCTGT AAAAAtatATAAGATGATGATTTCTAAATAAAAATTCCTTTGGAAATTCACCTGCTCTTAATAT TTTAAATC	0	11	MMBIR/FoSTeS
hg19	16	13	2445353			13	79929560		AGAAAATCTTGTCTATACAGGAATGCTCAAAACAACCTTTACTGG/CAAGGTAACCTGT AAAAAtatATAAGATGATGATTTCTAAATAAAAATTCCTTTGGAAATTCACCTGCTCTTAATAT TTTAAATC	0	4	NHEJ
hg19	16	13	7992956	+		13	24453536	+	GATTTAAATAATTAAGAACAGTGAATTTCCAAGGAA/ATTTATATTAGAAAAATCATCTA ATATAAATTAACAGTACCTTCCAGTAAAGTTGTTTTGA	0	4	NHEJ
hg19	17	13	5885957	-	2	13	69950856	-	CCATGGTGGTTTGGCTGACCCATCAACCTCATCTACATA/GGTGAGTCAAAAACAAG AAATAAACAATGTCCATCAATGGTATTGATAAACACAT	3	0	NHEJMMBIR/FoSTeS
hg19	17	13	6995085	-		13	58859579		CAATAACCAATTTGATGGACATTTGTTTTTCTTTGACTC/ACCTAAATGTAGATGA TAGGTTGATGGTACAGCAACCCATGGCAGCTGTATA	3	0	NHEJMMBIR/FoSTeS
hg19	19	13	5934865	+		13	70728648	+	ATACTTTATAGCAATTTACTGAAATATAAAATATTT/ATATTTTATAAATATTTATAAAT ATTTATAAAATTTAAAAAATTTTATA.AAAATTTATAAAATTTTATAtttATTTTT ATATAGTATAGTACTACTACTGATGCACCACCTTTTGGTGTGGTACTGAGGGCGGAA ATTTAA	0	114	MMBIR/FoSTeS
hg19	19	13	7072864	-		13	59348651	+	NA	0	114	MMBIR/FoSTeS
hg19	22	13	6785488	-		13	70731432	+	CTGATAGGGTGCATTTTGG/TTGAATACACATTTCTGGATGTTTTTTCATATTTAG ACATGTTTTGAAAATCAAGATATGATATTTTTCAAAATA	0	11	MMBIR/FoSTeS
hg19	22	13	7073143	+		13	67854880	-	TTAAAAAGAAATGTTGTCTCAAGAAACAGTTTTGACAACTGATAGG/GGTGCAATTT TGTTGAATACCAATTTGGATTTTTTCTCATTTAG	0	11	MMBIR/FoSTeS
hg19	23	13	7472402	-		13	75055605	+	GGTATACTAT/TTCCATCTGCTGGGACTCTGCCCTAGATATGCTGGTTGCATTT CCTCACCTTCTGTTCTGCTCACCAGCTAGTGAGGGCTT	0	10	MMBIR/FoSTeS
hg19	23	13	7505560	+		13	74724020	-	CCATCCCTCAGAAAAAAGACTGACAAATACCACCTATTAGTAGT/GGTATACTATTTTC CATCTGCTCGGGACTCTGCCCTAGATATGCTGGTTG	0	10	MMBIR/FoSTeS
hg19	24	13	7287557	-		13	76467126	+	GGAAAACTTAGTTTTATAGCACAAAGTCTACCCCTGCCAAGGCATGCTGCCA AG/GCATAGGATCTGGAAATTCATCTGAACTTTGTGATACTG	0	27	MMBIR/FoSTeS
hg19	24	13	7646712	+		13	72875571	-	GTCAATGGTCTAATAGGAAAACTGAGTTTTATAGACACAAGACTCTA/TACCCCTG TCCAAGGCATGCTGCCAAGGCATGCTGGAAATTCATA	0	27	MMBIR/FoSTeS
hg19	25	13	5715364	+		13	92227571	+	GTTTTAAATGTAGCTGCAAGGACTGTTTTTCAGATCTGGTAGTGAATTTACATTGAC TTGGAAA/CCAAGTCAATGATCCCAATAAATGCAATTTT	0	18	MMBIR/FoSTeS
hg19	25	13	9222757	+		13	57153641	+	AGTATTTTTGTACAGCATATTTTTAAcaAAAAATGCAATTTTGTGGAAATCAATTGAC TTGGTTTCCAAGTCAATGAAATTTCACTACCAGTACTGTA	0	18	MMBIR/FoSTeS
hg19	26	13	8229351	+	21	13	92227548	-	AGTATTTTTGTACAGCATA/TTTTTAAcaAAAAATGCAATTTTGTGGAAATCAATTGAC TTGGTTTCCAAGTCAATGAAATTTCACTACCAGTACTGTA	0	2	NA
hg19	26	13	9222754	-		13	82293513	+	NA	NA	NA	NA
hg19	27	13	8229349	-	21	13	95602658		GAGTTTCTGCAACTGTATATAAAGTGTATAAagtaACACAGTACTTTTAACTGTG TACAAA/AAATTTTTTTGTACAGCAATAATTTTCCAAA	0	51	MMBIR/FoSTeS
hg19	27	13	9560265	+		13	95602658		GAGTTTCTGCAACTGTATATAAAGTGTATAAagtaACACAGTACTTTTAACTGTG GTACAAAAGTATTTTTGTACAGCAATAATTTTCCAAA	0	51	MMBIR/FoSTeS

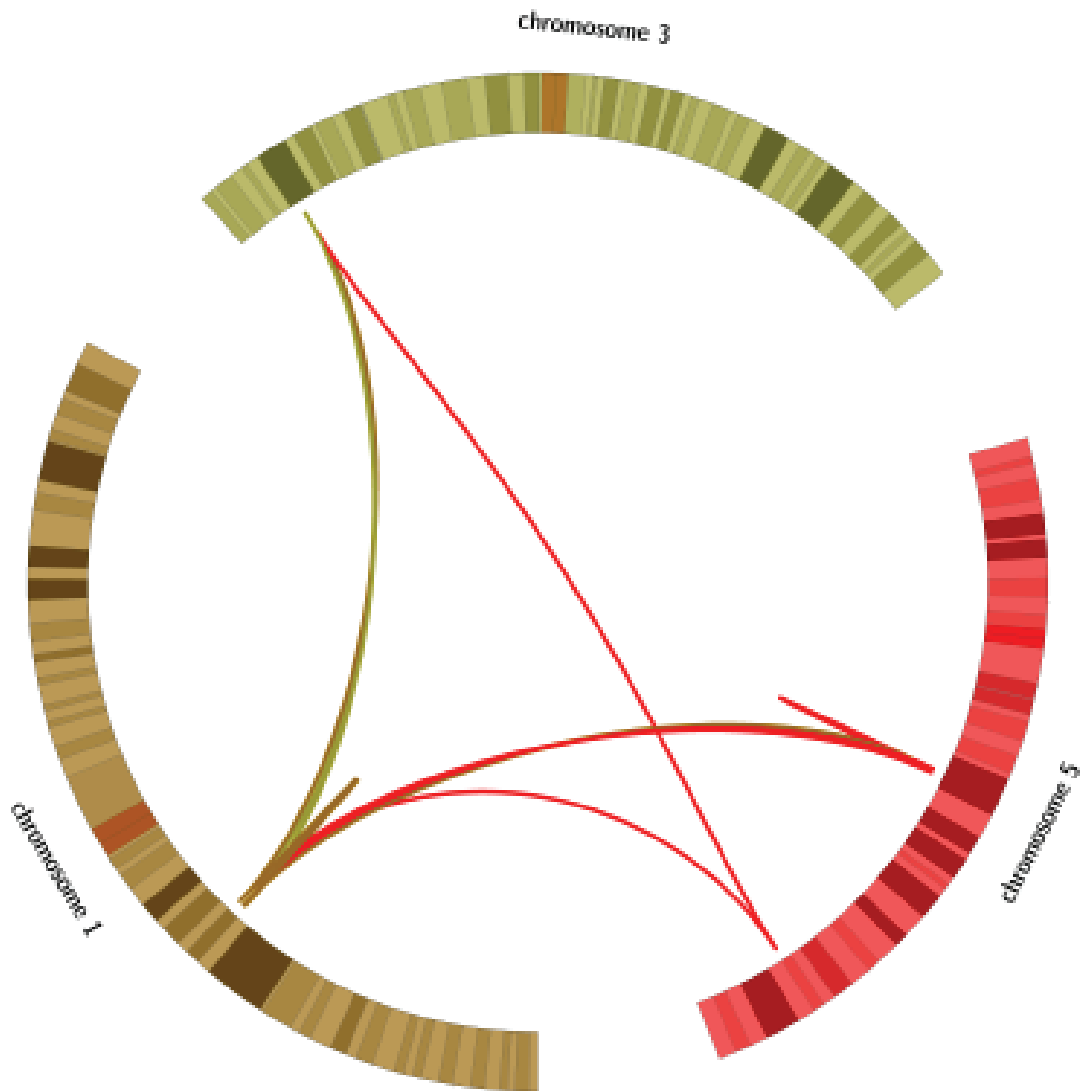
Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19	41	13	1049147 36	+	4	13	91220038	-	GAAGTGACCTAAGCTGAAATAATGGCAAGCTTCCGGAGATATTTTTGAAGGCCAAAAATA ACACATGTCT/ATCCACACTGA/CACCACTCTGGGTTCAAGCAATTC	0	13	MMBIR/FoSTeS
hg19	43	13	1025042 76	-		13	10551890 1	+	TGCCTGACTCCATGGGAAACCACTCTCTGA/GAAGTAGCCTCTATGGTCCACACAGC TTGTAGGTGGTGGAAATCAGAGGCTCAGTCAGTCAAGCCGTGC	0	4	NHEJ
hg19	43	13	1055189 01	+		13	10250427 6	-	GGTGTACAGAAAGTGTCTGGCCTTGGCTGCCTACTCATGGGAAATCCATCTCC/TT GAAGAGTAGCCTCTATGGTCAACAGCTTGTAGGTGGTGGGA	0	4	NHEJ
hg19	44	13	8794849	-		13	10681066 3	+	GATGTTATAGTCATAGCCACTCTGGACCAATTTTGG/TAACCCACACAGTGTAAAGGG AAATATCCAACTTCACTAGCAACTAGCTTTTAGTCTCACCT	0	3	NHEJ
hg19	44	13	1068106 63	+		13	87948491	-	GGATTTAGGGAACAGAGAGCCCGCAACTGCGAGGATGTATTAGTCATAGCC ACTCTGGACAGCAATTT/TTGGTAAACCCACACAGTGTAAAGGGAAA	0	3	NHEJ
hg19	45	13	3590775 4	-		13	11051282 3	-	GGTCAGTGTAGAGGGAGGAGTGTAGAGGGAGCCTAAGGCTGGCCTTGAGGGCAGA TTGACATG/TCAATATTAGGCCTTATTCATCTTTCTTAATAAT	0	45	MMBIR/FoSTeS
hg19	45	13	1105128 23	-		13	35907754	-	AAATTAGAAAAGAAATGAAGGCCATAATTTGAATGTCAATCTGCCTCAAGGCCAAGCCT TAGGCTCCCTTAACACT/GCCTCCCTTAACACTGACCCCTC	0	45	MMBIR/FoSTeS
hg19	46	13	5515929 6	+		13	11010586 3	+	CATAGGAAATGTAGTCTCTTTCCCTTATAAATACCAGTCTGGGTAGTATCTTTAC AGCA/CTAAAGTACCCGCACATCCATAGGATGGGCCAAC	0	21	MMBIR/FoSTeS
hg19	46	13	1101058 63	+		13	55159296	+	CTCACCAACAGCAAATGAGACCATCTGTGGCCACATCCT/ATGGATGTGCCGGTGA CTTTAGTGTGTAAAGATACTACCAAGACTGGGTAATTTAT	0	21	MMBIR/FoSTeS
hg19	47	13	6630484 8	-		13	75065069	-	ATTGTAATGTCTGGAGAGTTGATGGGTGCATCAACTCAATATAAAAAAGA/TTAAA ATATGGTACAGTCAAAATGATTTAAGTACTACTCTGTCTGTC	NA	>50	MMBIR/FoSTeS
hg19	47	13	7506506 9	-		13	66304848	-	ATCTTTTAAATATGAGAGTTGATGCACCCATCAAC/TCCTCCAGACATTACCAATGTCCC CTTTGTAGCAAAAATTCACATTTGAGAAATCACTGTTCTAG	NA	>50	MMBIR/FoSTeS
hg19	Unreso lved	13	7506516 9	+					CAAAAATGCACATTTGAGAAATCACTGTTCTAGAGCCACAGCACATTGGGTATGCTGT C/CAAGGCCATGCTTGGCTCCCAAAGTGCCGAGATTGCA	NA	NA	NA
hg19	Unreso lved	13	8224397 5	-					CACTTTTTTATAAGTGTGAGCATGGCTCACACTTTACTGTTGAGGGATGTATACAC AACTTG/ACAGCATGACATGCTCTGGAAAGTTAGTTACCCG	NA	NA	NA

Patient 14

Phenotype: Severe intellectual disability (walking at 9 years old), absent speech, autistic spectrum disorder, epilepsy, facial dysmorphism, constipation

Karyotype: 46,XX,t(1;5;3)(p21;q?34;p24)dn

Circos plot:



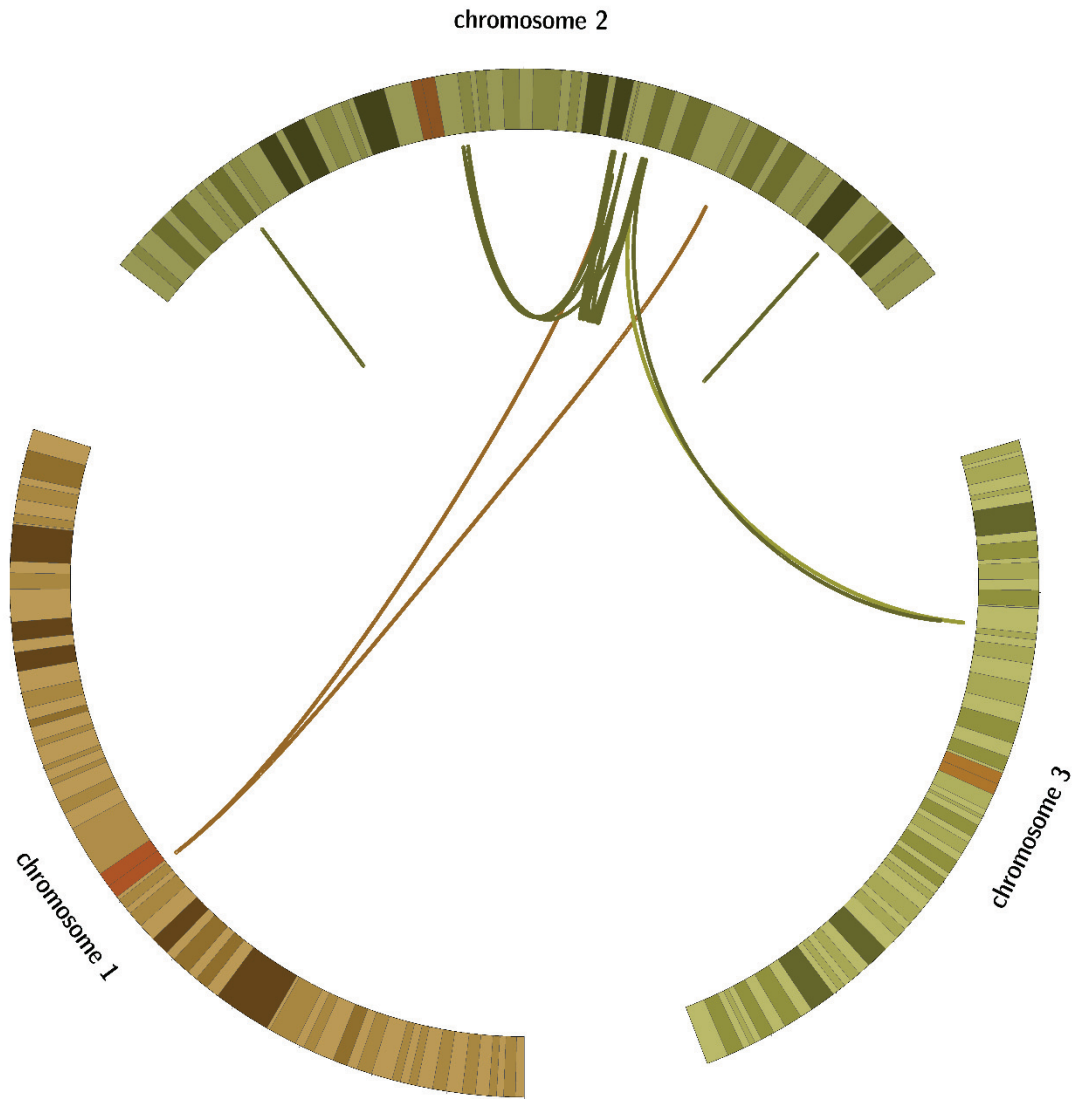
Comments : The phenotype can be explained by *MEF2C* disruption. This patient has been previously published as MD/2203 in (Schluth-Bolard et al., 2019).

Patient 15

Phenotype: Intellectual disability, speech delay, hypotonia, anxiety

Karyotype: 46,XY,t(1;2)(p13.2;q31.2)dn

Circos plot:



Comments : This patient has been previously published as NM/0201 in (Schluth-Bolard et al., 2019).

Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo -logy	Insert- ion	Inferred mechanism
hg19	10	2	150753372	-	0	2	147559888	+	ATCAATAGAAAGTTGGAAATTCATCTCTT/CCTTCAAAGTGATGCTGAAATGAT GCAAGTGCCTCCCAATGAGGGGCATAAGCTTGAGCAAGGAAAGCTTCTTTTCAG TTTTGGGAGAG	0	0	NHEJ
hg19	11	2	147559889	-	0	2	157274750	-	GTGATCCGCCCACTGGGCTC/TAACATCAGTTTTGCCCGTAGGACGACCTG AAGAAAATCTTGAATGGAAAACCTAATGCTTCAACAATGCTTAAGTTGTAT	0	0	NHEJ
hg19	11	2	157274489	-	0	2	147559889	-	CTGCCTACGGGCAAACTGATGTTA/GAGGCGAGGTGGGGGATCACGAG GTCAGGAGTTCGAGACCAGCTGGCAATATGTTGAAAACCCCATCTCTACTAA AAATACAAAAAATAGCTGGG	0	0	NHEJ
hg19	12	2	150803303	+		2	157274215	+	NA	NA	NA	NA
hg19	12	2	157274489	+	0	2	150803421	+	AGAAAAATGTACATTAGGCTGGTGGCCATGGCTCAGCCCTAATCTGAATA CTTTGGG/GGCTTTCCCTTGGGGTAAACCGACCTTCTCTGGCTG	NA	0	NA
hg19	13	2	156072831	-	1	2	157372067	-	TTTTCTCATGATCTAAGAGATAAATCTTTCTCTGCC/TTACCAGCTCCATTTCC AGAAAGATAATCTGGCAGCAATTTGAGGAGAAATAAGAAATAA	2	0	NHEJMMBIR/FoSTes
hg19	13	2	157372065	-	4	2	156072864	-	TGCCAGAATTAATCTTTGAAATGGAGCTGT/AAAGGCAGAAAGAAATATTCT CTTAGAATCATGAGGAAAAGAAATCTTTAATCTAGAAATGATCTTTATAGCCT GTTATAGTTGCCAAAATATGAG	2	0	NHEJMMBIR/FoSTes
hg19	14	2	156265463	+	-29	2	157269449	+	AAATTTGATAGGATTTTAAAATTTCTAATGCTGAGTAATGAGCCAT/TGCTCAT CACTCCAAAGTGTGGAAATGGAGGCATGAGCCACTCACCTGGCT	0	9	MMBIR/FoSTes
hg19	14	2	157269856	+	-9	2	156265463	+	GTGATCTGAAAGTAAATAACAGCCAGGTGAGTGGCTCATGCCTGCAATCC AGCACTTTGGGAG/TGATGAGCAATGGCTCACTCAGACATTAGAAAT	0	9	NHEJMMBIR/FoSTes
hg19	15	2	156265493	-	-29	2	157270282	+	TGCCAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCCACC/CATGTAACCAT TTTAAACATTAACATGGTTATATCTTAAACACAGAAATAACCAA	0	0	NA
hg19	15	2	157269866	-	-9	2	156265493	-	AGAAATAACCATGTTAAATGGTTAAATGGTTAGCATG/GGTGGATCACCTGA GGTCAGGAGTTTGAGACCAGCTGGCAACATGGCGAAATCCCTCT	0	0	NHEJ
hg19	16	2	156459486	+	-58	2	157275114	-	TTTTAGTAGAGATGGAATTTTGGCCATGTTGGCCAGTTGGTCTGAAATCTCTGA CCTCAGGTGA/TGAAAGGTTCACTGGAAATCACTGACATAAAGC	0	2	NA
hg19	16	2	157275114	-	-2	2	156459486	+	ATGTTGGCCAGGTTGGTCTCGAAATCCTGACCTCAGGTGA/TGAAAGTTTCCAC TGAAATCACTTACATAAAGCAGATAATCGGATAAAGGCAATTCAGA	0	2	NHEJ
hg19	17	2	157372068	+	4	3	49454932	-	TTTGGAAATTTGTTGAAATATAGATTTTATTTGATTTCTACATGGAGAGCCAAAG GGAAACCTTTGTTACACAGG/GGCAAAACCTCTGAAAGGACAGCCC	3	0	NHEJMMBIR/FoSTes
hg19	17	3	49454929	-	0	2	157372065	+	CATGGAGAACCAAGGAAACCTTTGTTACACA/AGGGGCAAACTCCTGGAA GGGACAGCCCACTGAGCCAGCTTCACTTGAAGGTATAGTAGTTGT	3	0	NHEJMMBIR/FoSTes
hg19	Unreso lved	2	156072831	+	1	2	156072831	+	GTGGAGATAAACATAAAGGAGATTTGGGATACACGTTACTTTAAAT/G CTGGAAAACCTTGCTAGTCAATTTGTAGAAAAGCTGAAAGCTGGATCCCTTC	0	0	NHEJ
hg19	Unreso lved	3	49454928	+	0	2	156427982	-	CCTCACCTTAACTGCCACCCAGTAGTCTGCTCAGCTTCTTTGACTAACCC CCTTGT/TACCACTTTAATAACAGGAAATCTTTCCCAATCTCTTG	0	2	NHEJ

Patient 16

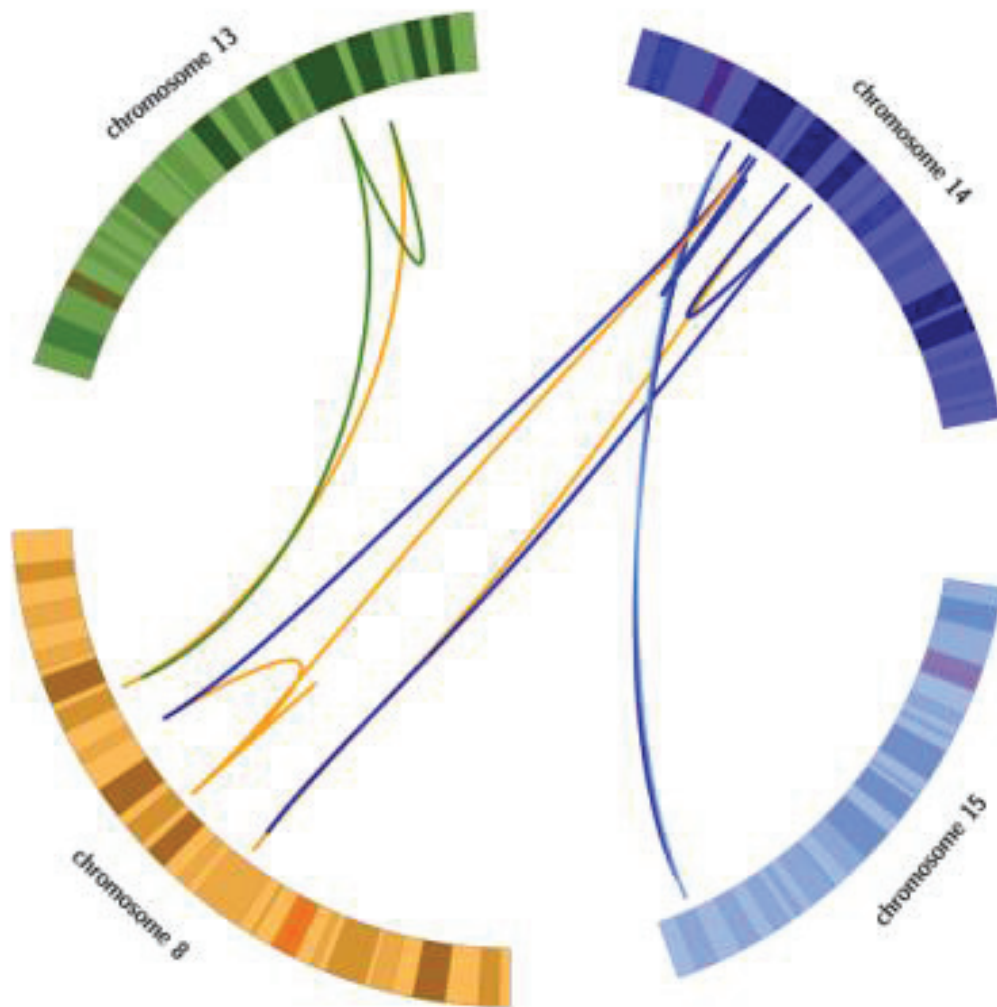
Phenotype: Macrocephaly, intellectual disability, psychomotor delay, speech delay, left club-foot, café-au-lait spot

Karyotype:

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)(q21;q24),der(22)t(13;22)(q31.1;p11.2)dn

Comments: Breakpoint on the short arm of chromosome 22 not identified.

Circos plot:



Comments : This patient has been previously published as KT/1403 in (Schluth-Bolard et al., 2019).

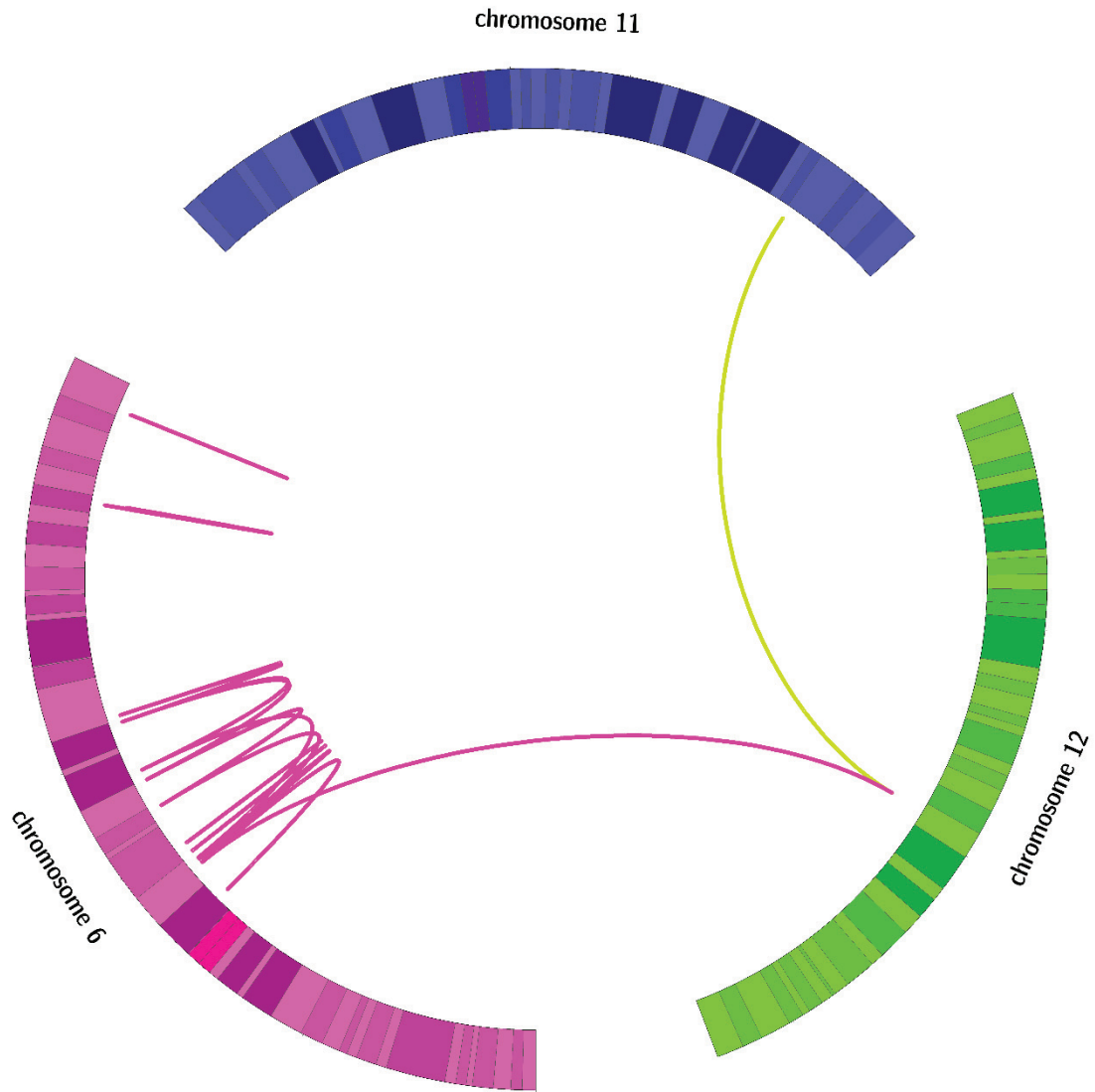
Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homology	Insertion	Inferred mechanism
hg19	11	15	94885846	+	-7	14	25904167	-	TTATATACACACACAATAATATTTTTATCCCATATATACACACACACATATATATTTAT/CTTGGCAGTATGGCATTTCATGATGTTGATTCCTCCATGATGTTGAACCTCCTG/C	0	3	NHEJ
hg19	12	14	25904165	+	-1	15	94885854	-	AAATATATATTTGTCTCCATATAACGATGTTTATCTCCA	0	0	NHEJ
hg19	12	15	94885854	-	-7	14	25904165	+	ACATGATTTGAACCTACTG/CAATATATATTTGTCTCCATATACGCATATTTATCTCCATATAAACACATATAATCTCCATATATACACAAA	0	0	NHEJ
hg19	13	14	32709589	+	-12	14	33757312	+	CCTCCCCAGCCTTGTGCTGCCCTTGCAGTTTTGATCTCAGACTGCTGTGCTGGCAATGAGTCC/GGACTAACATCCATTTCCATAGAAAGTTTGGCA	0	0	NA
hg19	13	14	33757312	+	2	14	32709589	+	AAATTAATTTGCAATCCTTCCAAATTTTGCATTTGCAAACTTTCTATGAAATGGATGTTA	0	0	NHEJ
hg19	14	14	32709602	-	-12	14	33757311	-	GTCC/GGAGTCTCACTCAITGGCCAGCACAGCTCTGAGAT	0	0	NA
hg19	14	14	33757311	-	2	14	32709602	-	TACAGATCTAACCTTTCTAAATCAGTTTCTCTCTGG/CCCTCCGAGCCACGTGCCGGATAAATCTCTGGTGGCGTTTTTTAAGCCCGTTGGAA	0	0	NHEJ
hg19	14	14	33757311	-	2	14	32709602	-	TTATATCCCGACGTGGCTCGAGGG/CCAGAAGAGGAACTGATTTAGAAAGGTTAGATATCTGTACAGTTTGGCATGGGTCGATGTGGAGGGCCAGGT	0	0	NHEJ

Patient 17

Phenotype: Intellectual disability

Karyotype: NA

Circos plot:



Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insert-ion	Inferred mechanism
hg19	1	1	74385012	-	240	11	115009441	+	ACAGCCTCGCAAACTGTGGGGTCTACTCTCCAGG/TGCTATGCATCGCTTCTTCCAC TTGTACCTCGATGCATGCTATTACAGAGATCTTGGATTTT	3	0	MMBIR/FoSteS
hg19	1	11	115009444	+		1	74385012	+	AAATCTTCCAGAGATCTATTAGATGTGCACCCCCAGCCATACCTTTAGCACACCCTC TGCAACTGTGGGTCTACTCTCAGGTGG/TATGCGAGCCTTTCTTACTGTACT	3	0	NHEJMMBIR/FoSteS
hg19	2	1	74385251	+	8085 756	6	66299496	+	CAAAAGGAGTGAATGAATAAATCAAAATCTCAGTAGGAAACTATGTTTAAAG T/GATAAGTCCAATTTGAAGTTTATATATGGCTGTTTGAAGAA	0	NA	NA
hg19	2	6	66299496	+		1	74385251	+	CTTATCTGAATCTGTCTTACTCTCAGGTAGCAAAATTTTTTCAAACGCCAATATATAA ACTTCAATGGACTATC/ACTTTCAATAACATAGTTTCC	0	NA	NA
hg19	3	6	73482766	-		12	77394524	-	TATAATATATCTGAATGTGGAGCAGGCTTGAAGTAGAGT/GCTATATTGAAT ATATCAATAAACAAGGATCTAGAAGTCAATACCAATTTG	0	0	NHEJ
hg19	3	12	77394552	-	-131	6	73482766	-	TCCTTTGTTTATGATATATCAATATAGG/ACTCTACTTCAAGGCTCTGTCTCCACA TTCAGAATAATATATCAGAAGTTCAATGCGTTACGG	0	0	NA
hg19	4	6	74677278	-		6	94902790	-	AGTTCCACCTGTAATCTCAGA/GGTTTTGGCTTCTATGATCTTCTGTCTGTCTTTGTG GGCACAGATCAATAAATAATTAACCTCACCAGGTGACCTAATATCCCTTTCAA	0	0	NHEJ
hg19	4	6	94902790	-	0	6	74677278	-	GAGATCATAGAAAGCCAAACC/GCTGAGATTACAGTGGAACTACTGCACCAGGC CTAGTTATTTGTAGAGATAAGGAGAAAGTTAGTGAACCTG	0	0	NHEJ
hg19	5	6	76432084	+	1696	6	76433781	-	TAATCTTTTTGTAATTTAGTAGAGACAGGTTTACAGTGTACTCAGGATGGTCTC GATCCTGACTCGGATCCGCCACCTCGGCTCCCAAGTCTGAGATTACA/AGTG TGGCCACTGCGT	0	0	NA
hg19	5	6	76433781	-	1696	6	76432053	+	CGATCTGACTCGGATCCGCCACCTCGGCC/TCCCAAAGTGTGAGATTACAAG TGTGGCCACTGCTCCAGCAGGAAATTTATTTTAAAGCAAATTTGTGTATAAAGC ATGTGATACCAACAAAGAAACAGC	0	0	NA
hg19	6	6	78354190	+	3044	6	78357239	-	CCTGGAGTGAATCATCAGAAAGTCTCACCTATATTAGGTTTTCTGACCAGATCAAG CCACTAGA TTTTCAGTGTGTCGAAGGAGTGTGGGGTTGAAGGGTTGAAGGGA GGG/TGCGGTGGAGGCTTACTACCTACCAAGCTGAATG	4	0	MMBIR/FoSteS
hg19	6	6	78357235	-	3044	6	78354186	+	GGGGTTGAAAGGTTGAAGGG/AGGGTGGGGTGGGGCTTAGCTTACCACCTACA AGCTGAATGTCCTTGGATGAA TTAATCAATCGGTTTCTTATTG	4	0	MMBIR/FoSteS
hg19	7	6	73996720	+		6		+	GCCCTGACTAGGATCCTGAAAGATGACGCAATCCTACTACCAGGATGGCTCCAGAGG CATGGAATAACAGT/CCTTTACCA TTAGTAAATGGCTTCTT	0	0	NHEJ
hg19	7	6	87104720	+		6	73996699	+	TCCTAGTCTGGATAAACAGTCTTAAACCAACAAAGATCAAAAGAGACAAAGAGG CCATTACATAATGGTAAAGG/ACTGTTTCTTCCACTGCTTGG	0	0	NHEJ
hg19	8	6	92806165	-		6	104960498	+	CTAAGAGCCTCA TTTTAAAGGA/GACTTAGCAAAAGTATCTCCAGTCTGCAGGCT CCATTTATGGTAAGTACCTATATGGATGATACCA TTTTAGTCTTTTATATGATTTTCT ATGG	0	1	NHEJ
hg19	8	6	104960498	+	1	6	92806166	-	TAAACAATTTAAAAAAAATTTAGTGCCTACTAGGTGTACAAGAATTTTTAAAAACT AAGAGCCTCCATTTTAAAGGA/GACTTATGCAAAATCTTCCAGTCTGCGAG	0	1	NHEJ
hg19	9	6	94902789	+	0	6	104960498	-	GCCAAAGCTGTTTTCAAACCTTCTACTCAAGCAG TTTGCTCTTGGCTCTGGAAGT /ACCTAATAATGAAGGAGTTATTTCCCACTGTTGTGACA	0	0	NHEJ
hg19	9	6	104960498	-	1355 892	6	94902789	+	AGTTTGTCTCTGCGCTCTGGAAGT/ACCTAATAATGAAGGAGTTATTTCCCACTG TTGTGACAAATGCAAGAAACAGAGTTTGGGAATTTCTGGT	0	0	NA
hg19	10	6	106316389	+	-441	6	106316835	-	AAGTCTCTGTGGAAAGGCTGCAGCCTTTTGTGAC/AGTCACATTTGGTGGCAAGTAGTAC TTTTGTCTCATGGACTAATG	4	0	MMBIR/FoSteS

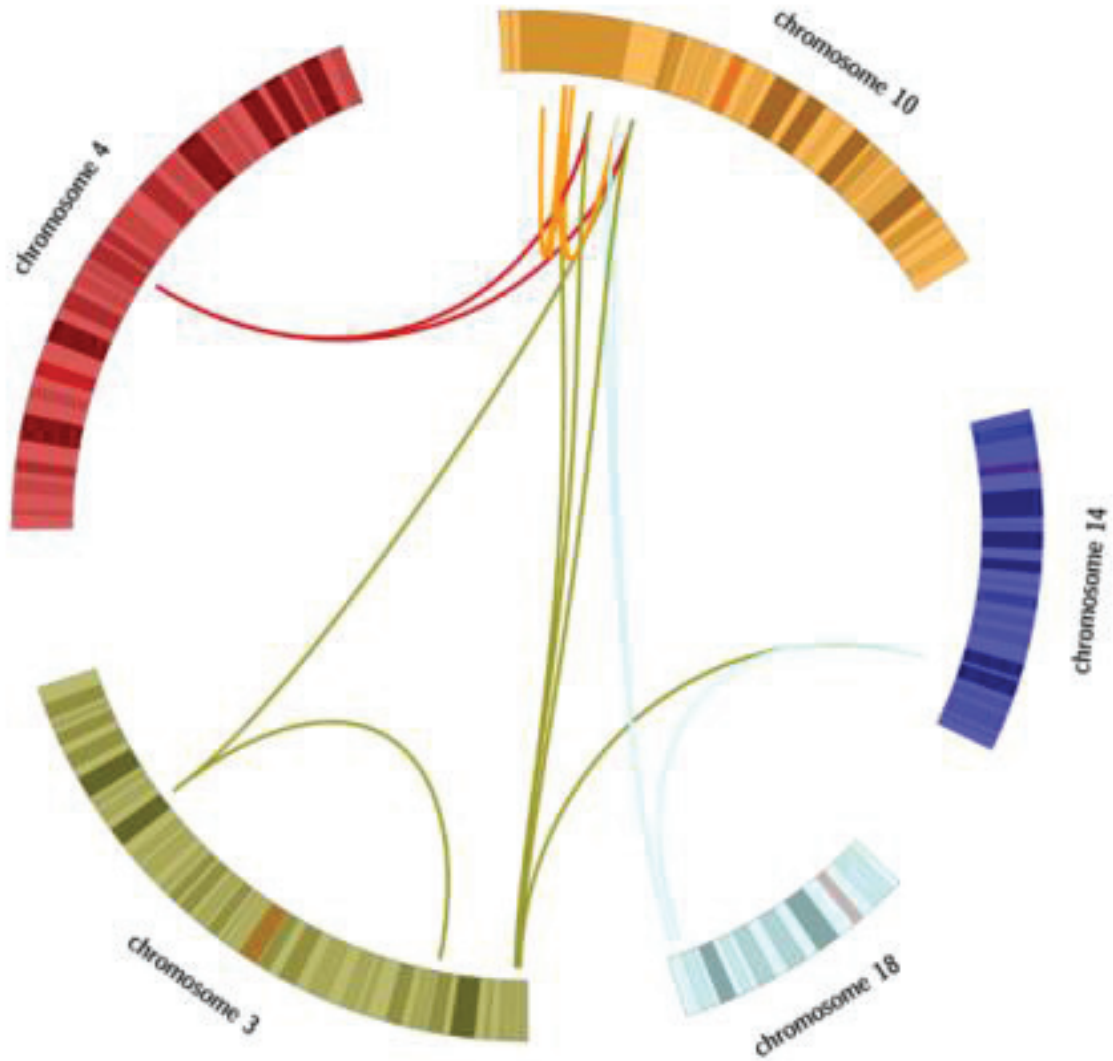
Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homology	Insertion	Inferred mechanism
hg19	10	6	106316831	-	-441	6	106316385	+	AAC TCA CCCAAGTCTCTGTGGAAAGGCTGCAGCCCTTTGT/GTACAGTCACTTGGTGGC AAGTAGTACTTTGGTTCATGGACTAATGGGTGTTGGAGGAG	4	0	MMBIR/FoSTes
hg19	11	6	147513318	-	2471 68	6	147760485	+	TCTC ACT TAATTTGCTTTAGAAATACCTCTTT/JACAAITCAAAGTGAGATTTGGGTGGGGGC ACGCCAAACCATATCAGAAATTACTCCTAGTTTATGAGAA	0	1	NA
hg19	11	6	147760485	+	2471 68	6	147513318	-	ATTTGCATAATGACTTGCCTTTTGTCCCTTTGTTTCTCCTAATTTGCTTTAGAAATACC TCTTT/CACAAITCAAAGTGAGATTTGGGTGGGGGCACAG	0	1	NA
hg19	13	6	165768715	-	11	6	165769625	-	GGTGA CGG TGATTTGAGTCA TGGGGACAGATATAACCAACATA /TATGTTATATAATG AAATGCATAAATTTGAAGGATATTAGTGAGTATGAAAATATAT	0	11	MMBIR/FoSTes
hg19	13	6	165769625	-	194	6	165768715	-	AATATCCTCAAAATTTATGCATTTCA TTTATATAACATAATGTTGTTAT /ATCTGTCCCCA TGACTCAATCACCGCTCACAGCCCTCCCTTGACACATGCAGATTAC	0	11	MMBIR/FoSTes
hg19	14	6	165768725	+	-899	6	165769818	+	TGCAGATTTAGATCATACTTTTATAC TTTAGTTTATATA /AATGCAAAATGC ATATAAGGTAAACCAAGGTTGTATGTACCA TTATGCA	0	14	MMEJMMBIR/FoSTes
hg19	14	6	165769818	+	194	6	165768738	+	GATTA CAATCCGAGATGAGATTTAGTGGGGACACACACGCCAAACAAATATCACTTGCCAA CAACTGATAATGTTGGTCTTTGTAATTTGGCCATCTAGTGGTGCATAATGGTACAT ACACCACCTGGTTTACCTTAT/ATGCA TTTGCA TTTATATAACATAACCTAAAGTATAA AAAAAG	0	14	MMBIR/FoSTes
hg19	15	11	115009122	-	1	12	77394719	+	ATTTAA ACCCAGGGCTTAAC /AGCTTTGTGGCTGAAACAAATATCCTTCACAGCTTG CTCCCAAGTCCAATAAAGCCAGAGAAATGGTAGAGAGATTTAAT	1	0	NHEJ
hg19	15	12	77394420	+	-131	11	115009123	-	TAGGCTCAGACCTATGAGAAACCTTACTGAAAGTCAACACAGCTAGATGCTAGTTACAAA ATTTAAACCCAGGGCTTAACA/GCTTTGTGGCTGAAACAAA	1	0	NA

Patient 18

Phenotype: Mild intellectual disability, behavioral troubles, facial dysmorphism

Karyotype: 46,XY,t(3;4;10;14;18)(p23;q24;p13;q32;q23)dn

Circos plot:



Comments : This patient has been previously published as TS/1101 in (Schluth-Bolard et al., 2019).

Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19	11	10	10335498	-		3	4235296	+	ATGGTAATAAAAAATAAAATAAACAGCTATAAAAT/CAATCTTTTATATCCAAAGTT TAAATGACTTATGCTCAATAAAAAATGATTTCAGCATATTGTA	0	3	NHEJ
hg19	12	10	10607349	+	2	10	12437412	-	ATGTAACAAGTGAAGTCAAAATAAAGTCAGTTCTGAAATGCCAGCCTGTGATGACGGC TCTATAAACCCACAGGGATAGA/ATCTCAGCCACCACTCAC	2	0	NHEJMMBIR/FoSTeS
hg19	12	10	12437410	-	4	10	10607347	+	GCCTGTGATGAGGGCTCTATAAACCCACAGGGATA/GAATCTCAGCCACCAGTCACT CATTAGCCCAAAAAACATCACTGTGGAGTTTCCAAAGATTG	2	0	NHEJMMBIR/FoSTeS
hg19	13	10	12474192	-	-1	18	75078221	+	GGCTTTAAACAAAAGCCATGTCTATGACAAAACCAAGTGTGGGGTGTGT/CCACAGAG GTGGATTGAACTTGCCCGTGGTGGTGGGGAGACAGCCCAA	0	11	MMBIR/FoSTeS
hg19	13	18	75078221	+	-10				TTGTTTCTGTGTCTTACGTGTTTGGATGGCTTAAACAAAAGCCATGTCTATGA CAAAACAAGTG/GTGGGGTGTGTCCACAGAGGTGGATTGAC	0	11	MMEJMMBIR/FoSTeS
hg19	14	10	12981920	-	-24	18	75078232	-	TGGAGTGTGAGCTAGCTGAGCTC/AAATTCAGAGTTGTTAGCTATGGGAATAGTTCAA GGTCGGAGGCCCTTAAAAATAGCATTGACCTGGTCAAGATGG	0	2	NA
hg19	14	18	75078232	-	-10	10	12981918	-	CTCCGACCTTTGAACTATCCCATAGCTAAACAATCTTGAAATTTGA/GCTCAGCTAGCTC ACACTCATTCTGACCTCATCTAGTGTCCCTCCTCCTGC	0	2	NA
hg19	15	14	86863996	-	6				GCATGCTGA/AAATAATGTTGGGTAATTAATTTGCTCTCATTTTAAACAGATAAAGAAATT GAGGCACAGTGAGGATAAGCAGCTTGATTAATTTATATGGCT	0	0	NHEJ
hg19	15	18	71840451	+	6	14	86863996	-	AGGAGTGAGGATGGGACGAGGAGGGTCCCGCTGGCCACAGCTCCTGCACTGCTGA/ AAATATGTGGGGTATATTTTGTCTCATTTTAAACAGATAAAGA	0	0	NHEJ
hg19	16	10	9559342	+	4	10	10335462	+	GAGATTTCCATGGAAAATAAACAAGGAAAGCGAGCATCCCCACCCCATGACTG GGA/CAGAACCGTCACTGGAGAAAACCTACCACAGATTTTATGT	NA	0	NA
hg19	16	10	10335464	+					NA	NA	0	NA
hg19	Unreso lved	3	146852560	+	-27				TCAAAAGTCCCCAGTCCAAAGTCTCAACTGAGAGAAAAGAAAATCCCTCCACCTATGAG CTTTGAAAATTAGAAAGCAAGTTAGTTACTTCCAGA/AGGTTG	0	0	NA
hg19	Unreso lved	10	10315593	+					NA	NA	NA	NA
hg19	Unreso lved	10	10334274	+					NA	NA	NA	NA
hg19	Unreso lved	10	12981895	+	-24				NA	NA	NA	NA

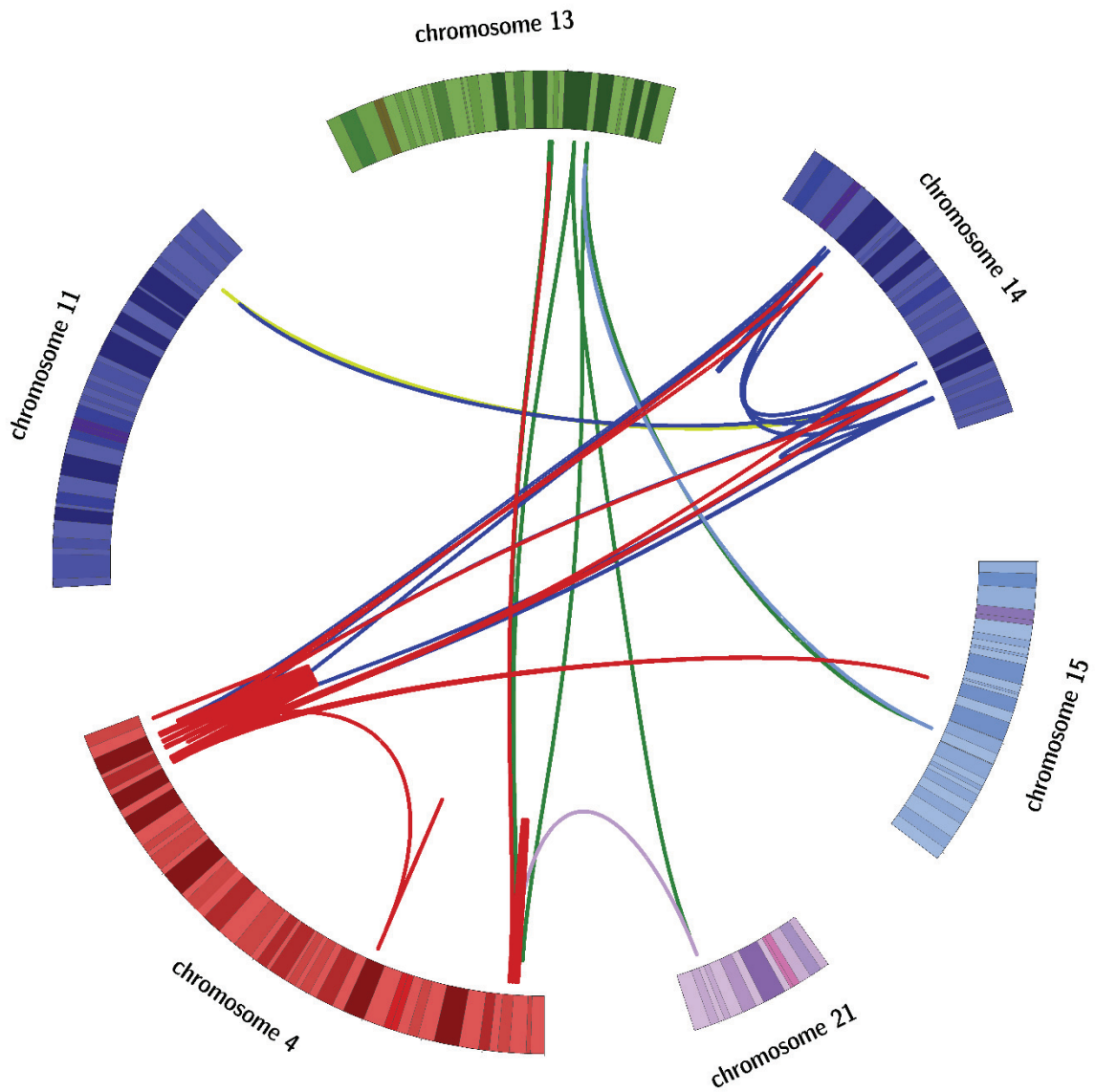
Patient 19

Phenotype: Intellectual disability, facial dysmorphism, bones malformations, valvopathy

Karyotype: NA

Array-CGH: arr[NCBI36] 4q33q34.1(171958564_174625926)x1,
4q34.3q35.1(182539074_183620310)x1, 14q11.2q12(22439503_23819413)x1

Circos plot:



Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19	15	4	13315831	+	- 2880	4	13318715	-	TGAGCAGTTGCAACAGAGTGTATGACCCCAACCCCTAAAATAATTTATCATCTAGATCT TTATAAAAACTTTTGTCAACCCCTAATCAACCCACTGCTT/GATAAATAGTATGATGAGA CCAATTTATAAACCCCAATGGCCATTTGTATAT	3	0	MMBIR/FoStes
hg19	15	4	13318712	-	- 2880	4	13315828	-	TTATCATCTAGATCTTTATAAAAACTTTTGTCAACCCCTAATCAACCCACTGCTTGGAT AATTAGTATGATGACCCATTTTATAAACCCATTTGGCCATTTGTATATCTCTTTTGATAA ATATCTGTTTCCAGTTCATTTGCCCTTTTAAAT	3	0	MMBIR/FoStes
hg19	16	4	64297784	+	-3	4	64424471	-	TGCTAAAATTAAGAACAGCAAGTTAGGGATCATTTAAAGCAATGCAATGGAAAGATGAT AGTGACTTTGGAGCAGATAAATGCACGTGATATAGGGTGGC/CTGCAATCCCAATCT GGATGTATATCCACAAGAAAAATTAATCAATCTAC	1	0	NHEJ
hg19	16	4	64424470	-	1	4	64297783	+	AATGGAAAGATGATAGTACTGGAGCAGATAAATGCACGTGATTTAGGGTGGT/CC TGCAATCCCACTACTGGATGTATATCCACAAGAAAAATTAATCATCTACTGAAAAGACA CATGCACCTGTATGTTCAAGCAGCACTAGTCA	1	0	NHEJ
hg19	17	4	64297788	-	-3	14	86805476	+	TCTACCATAGTTACAAAACATAGATAACTATGTTACAAGCTGGGTATTTAAAGGGTA CAGTCA/TAGGACTGTATTTGAAAGCTACAGCTTCCACCAAAATTTGGTGAAAAATTTGA ATTGAAAATGTTTAGCATGAGAAAAATAAACA	0	61	MMBIR/FoStes
hg19	18	14	86805476	+	- 4987	4	64297788	-	ACTATGAAATAATCATATACTTTGTTAAAATTTTCAAAACGTCTA/CCATAGATTACAA AACATAGATAACTATGACAGCTGTATTTAAACGGTACAGTCAAGCACTGTAT TTTAAAGCTACAGCTTTCCACCAATCTGGTGG	0	61	MMEIMMBIR/FoStes
hg19	19	4	64424470	+	-	4	176777027	+	CATGTTTGACATCTGCTGGTAGTAATGTAAGTATTTCAACCACTGTGAAAGGAGTTG GAGATTGATCTAAAAAACCTTAGAAGCTACAGCTTCCACCAAAATTTGGTGAAAAATTTGA AAATATATATGTTATATTAACCTATAATATGTA	2	0	NHEIMMBIR/FoStes
hg19	19	4	176777029	+	5	4	64424468	+	AATCTGATGATACACATGGAATATTTGTAATAATGTACATATATAGTTAAATATAC ACATATATTTCTTCTCTCAACTGAGAGG/TTGAATGGCAGTTCTAAGTTTTTTTA GATCAATCTCCAACCTCTACTGCTTTTCTGCTCA	2	0	NHEIMMBIR/FoStes
hg19	20	4	170082530	+	-410	14	25081148	-	ATTTTCATTTAAAAAATAAGTATGATAGTGGCCAGCTTTACATTTCAATAATTTTT GTTCCAAATCATATACAGTT/AGCATCCCTAGGGATGGGACTTTCTGACCACTGG ACTGCATTTGGCTTTACTGCTTTTCTGCTCA	0	5	MMBIR/FoStes
hg19	20	14	25081148	-	1364 95	4	170082530	+	GTGGCCAGCTTTACATTTCAAAATAATTTTGTTCCAAATCATACATACAGGTTAGCAT /CCTAGGGATGGACTTTCTGACCACTGGACTGCATTTGGCTGTTACTGCTTTTCTGGG TCTAGCTCCAGTGGCTACCAGTCCAGC	0	5	MMBIR/FoStes
hg19	21	4	170082941	-	8607	14	86810465	-	CTCACTGGTAGAGAAAAATTCATTAACAAAATTCATTTGTTACTACTATAATGGTTAT TA/TTCCAGAAAGAACTGAGGCACAAAGAAATCATAGTAATAGAAAGTACCACACAG AGTAAGAAAGTAAAGAAATGGGTGGAAGAAACAAT	1	0	NA
hg19	21	14	86810464	-	-120	4	170082942	-	GTAATTTTACTATGATTTCTTTGCTCAGTTTCTTCTGGAA/ATAATAACCAATTA TAGTAACAAGTATTTGAAATTTGCTAATGAAATTTCTCTACCACTGAGACTACTTTT CACGTGTTTTCATGATGATGATGTTGCCCT	1	0	NA
hg19	22	4	170091559	-	-11	14	93184543	+	ACATGGTGAATTCCTCTACTAAAAATACAAAAATAG/CTAACGAAAAATTTAAG ATGATAAATTAATAATCGAAATCTTTTGGTCATGGCAACAGACTAAATTTACTATTTT AAAACTCTGCCAATATATTAACATATGCTT	2	0	MMBIR/FoStes
hg19	22	14	93184775	+	- 5642 62	4	170091561	-	CCAGCATTTGGAGGCCAAAGCAGCCGGATCACTTGAAGTCCAGGTTCAAGACCA GCCTGGCCCAATGGTGAATCTCTCTACTAAAAATACAAAAATTAGCT/AACGAA AAAAATTAAGATGAGAAAAATTAATAATCGGAAATCT	2	0	MMBIR/FoStes

Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19	23	4	170094872	-	2	4	176278079	-	GTCCTGCCCCG CAGTCAITGGGGTGA CTGAGAAATAAATAAAGATTTTAAATCC CAGA CTAAAAATATCTCAAGCCATCA /TGGAAACCACTATCTAACCATTTACAAGTACCTTTT TTTTAAATAGTACTTTTAAAGAGGGGGT	2	0	NHEJMMBIR/FoStEs
hg19	23	4	176278077	-	-2	4	170094874	-	CTCCTTTAAAAGTAGCTTTAAA AAAAAGGTAGCTTTGTAATGGTTAGATAGTGG TTTT/CATGATGGCTTGAGATAAATTTAGTCTGGATTA AAAATCTTTATTTATCTCAA TTACCCAATAACTGTGGGGAAGAAACTAAA	2	0	NHEJMMBIR/FoStEs
hg19	24	4	170094873	+	2	14	25222785	-	GGAGGGTTACATGTTAGCTCTGGAGAAA CAATAAAAAGCTTAATACTTCCCTTAA TTTGAGAGGGTTAGCTTACTGAGCTTGAAATATGIG /AA TAAAGCTGTA AA TACTTG CTG GGAAGCCATAGCCCTCCCTGAGTGTG GACTGTGTAATAGGCTTATTTCTTGTAACCT	2	0	NHEJMMBIR/FoStEs
hg19	24	14	25222783	-	3949 268	4	170094871	+	AAGCTGAAATACCTTCCCTAA TTTGAGAGGGTTAGCTTAGAGCTTTGTA AATATG /TGAATAAGCTGTA AATACTTGTGGAAAGCCATAGCCGTGACTTCCCTGAGTGTG GACTGTGTAATAGGCTTATTTCTTGTAACCT	2	0	MMBIR/FoStEs
hg19	25	4	170097391	+	- 2288	4	178598918	+	AGAACTGTGAGAAATAAATCTCTGTTGTTTACTCCCACTCAGGTAATTTGTCCTAGC AGCCAAATCCCTAAGACATTTGGTGAAGGAATAAAC /GAGGTGGGAGGATCATCTG AGCCTGGAGGCCAGGATGCAGG GAGCGGGGAG	0	0	NA
hg19	25	4	178598918	+	-7	4	170096959	+	TGAAACAGTCTCTCTGCCAGGCTGGAGTGCAGTGGACCTATCACAGCTCACT GCAGCTGGCTCCAGGCTAGATGAT CCTCCACTC /GTTTATCTGTTTACC ACT GTCTTAGGGCTCTTGGCTGTAGGACATACA	0	0	NHEJ
hg19	26	4	170099680	-	- 2288	14	93749042	+	CCTGGTTTAAAGCAGAAAAC TCAACTCTCAACTGAGCTGAGTTGAAAAGCCAGGA TATT /TGGACAGGCTCTGAAACACAGTACTTTGGTTCTGACATACTGTGGGGCAA GAGGAGTCTGAAGGTGATGAAATGCAATGCACT	0	0	NA
hg19	26	14	93749041	+	4,7E +07	4	170099680	-	TTTAAGACAAAACCTTCAACCTGTTCAACTCTGCGTTTGA AAAGCCAGGATAT /ATG GACAGGGCTCTGGAACCCAGTACTGATTTCTGACATACTGTGGGGGCAAGACGG GCTGCAGGTGATGAAAGCCAA GCGCACTGCCCCAA	0	1	NA
hg19	27	4	170148746	-	-63	4	171658061	+	TTGATATCTGAAAACAACATTTGGCAGGGTAG /TGCATAGATAGCAACATAGTGGAT TTAAGGTTGTACATGTACATAGTGGATGTGTACACTTATGTACATTTATACACTCCCA TGTAGGTGTATAAAAATAATTCATTTTACTTTG	NA	99	MMEJMMBIR/FoStEs
hg19	27	4	171658061	+	-129	4	170148746	-	AATAGATTGAGTCAAAAAGTAAAATGAAATTTT /TTATACACTACATGGGAGGTGTA TAAATGTACATAAGTACACATCCACTATACATGTACAACTTTAAAATCCACTATGT GTCTATCTATGCACTACCCCTGCAAAATGTTG	0	99	MMEJMMBIR/FoStEs
hg19	29	4	170148773	-	-	4	181134419	-	TCTGAGATGGAGAGCCCTGTG CAGCTGCCGTA TTCGATACT /ATCTATGCATATTT ATTTTACTACTACCATGATCAATTA TTGTAGGACTCAAGGTAGGTACAGCTT GTACAGAGGTAGAAAAGACTGTAGCAGTTGTAG	0	5	NHEJMMBIR/FoStEs
hg19	29	4	181134419	-	-	4	170148773	-	CAAAGCTGACTACTTGTAGCTCAAA TAAITGTAGATCAGTATGGTAGTATAGTAAA ATAAAATAATGCATAGATAGTAT /TCGAATACGGCAGCTGCACAGGCTCTCCCATCTCA GATTGTGAAATGATTAATTAATTTTCACTTAG	0	5	NHEJMMBIR/FoStEs
hg19	30	14	29172050	+	-	14	78766145	-	GGATTCAGGATTTCTACCTAAAATCCAAAATAAACA CCGCAGGAAATTAATGTA GCCAATAAAA /GGTGTAACTCAGCTCACCTCTAGAAAAGAG:AGAAAATCTACAAT ACTCTCCCTGTGTGCCCTTCTCTACCAGTGGCCTG	0	17	MMBIR/FoStEs
hg19	30	14	78766145			14	29172050	NA	NA	0	17	MMBIR/FoStEs

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	68	15	64381865	-	6	13	88536370	+	GAGAAAAGAGTGTGTGCAAAAGGGGAAAGGACCTCATAAAAACCATCACATC/AGAG GTTTGAGACAAGTACATGAGACAACACTGTGCTGACGCTGGTAACATCATCAGCT GCTAGACTCAGACTAAAAAGCAAAACCAATGATACCN	0	0	NHEJ
hg19	69	13	88536415	-	-44	15	64381872	+	AGTATGAGATTTGGGTCTGACCATCAAGTGTCTTTACAGCCAGCTACCCGGCAATAGCA AGTCTCAAGAGGTAT/GGCCCCATATTACAGTTACTTCCACTGTGCCCTCCACAA CAGTGGGATTATGGGAACCTCAATTCAGAT	0	0	NA
hg19	69	15	64381870	+		13	88536413	-	GTTGCTGAAGCAGGCATAAGTAAAGGTTTITAGAGTATGAGATTTGGGTCTGA CCATCAAGTGTCTTTACAGCCAGCTACCGGCAATAGCAAGTCTCAAAGGT/ATTC CCCCATTATTCAGTACTCCCACTGTGCCCTCC	0	0	NHEJ
hg19	70	14	23314244	+	0	14	93184781	-	AAATGGCAGCCTCTTTGGGCAACGCTGGTGGATGTGACCGGTTCAITCCAGAC CTGGTCAATTTGGGCTCTGCCCCAGACAGCGCTCCACCCAGCAGCCAGGC/GC TGGCCTGGCAGTGCACCGCTGCAGTCCCACTTC	0	8	NHEJMMBIR/FoStEs
hg19	70	14	93184773	-	3	14	23314239	+	GAGTGGTGTGTCGCCGCGCATCCAGGCTGTGGTGGTGGCCCGCCCGCCCGC ACTGGCTCCACCCACAGG/CAGGCTGGGATGGCAGTGCACACCTGCAGTCC AAGCTTCTGGGCTGAGTGGGATCACTTG	0	0	NHEJ
hg19	71	14	23356721	+		14	25081130	+	GGTGGCAGTGTCCACTTTGCCCTTAAAGGATTTTAAAGGCAAGGGCTGGAAGT GCTGTTCTGAGATGGATCTTGAC/CATCAAGGATCACCCCATGGGATAATAGAA TCTGAACGACGAGGACTTGAGTTCAGACCTTCCCA	0	0	NHEJ
hg19	71	14	25081130	+	-17	14	23356721	+	TCAGTGAAGGCTGGAAACAAAGTCTGCTTTCAGACTATATCCCAATGGTG ATCCTTTGATG/GTCAAGAGAAATCCATCTCAGGAACAGCACTCCAGGCCCCCTTGCCTTAA AGAACTC	0	0	NA
hg19	72	14	24752552	-		14	86688561	+	CTAGTGGCTGTCTGAGGTTAAGAAATAAAATTTTACAATGCCCTTCTACTATAAG GCTCCAGTATACA/TGGAGGCTGAGCGGGCAGACTCTGAGGCCAGGATTCAC GACCGCTTGGCAACATGCCAAACTCTGTCTCT	0	11	MMBIR/FoStEs
hg19	72	14	86688561	+		14	24752854	-	TAGTCTCCATCAATTTTCTACTCTAGTGTGGCTTCTGAGGTTAAGAAATAAAAT TCTTACAATGCCCTTCTACTATAAGGCT/CCCAGTATACATGGAGGCTGAGGGGGG AGATCACTTGGCCCGAGGTTCAAGACCAGCT	NA	NA	NA
hg19	73	14	86578562	-	1100 00	14	93749041	-	TAAAGGAAAAATGACAGAAAGGAAAGAACTCTGCAITGTACAAA/TAATATCC TAGGACACTAAAGTAAATATATTTCCAGTAAAAAATAAAATTTGCTCTCTAAAC TGGTGTCTTTCAGATAACCAATATCTTAAAGT	3	0	MMBIR/FoStEs
hg19	73	14	93749038	-	4	14	86578565	-	GTTTAGGAGACAAATTTTATTTTGTCTTACTGAAATAATTAACCTTAGTGTCTAG GAAT/ATAATTTGACAATGCAAGGTTCTCTCTCTCTGTCATTTTCCCTTTAGTCA GTAATGCCCTCCCAAGTCTATCTCTCT	3	0	NHEJMMBIR/FoStEs
hg19	74	4	170091547	+	-11	14	86688615	-	CCATGCGTGCACCTGAGACTTTTGTAAAGCATTTGTTGATGATGATAGAAAGG TAAACAGATCATGTAG/GATGTACATGGGTGAAGGTTA/CTCATCTTCCAACTTGT TCCAGGATAAGCTACAAAAATAATTTAATTTA	0	22	MMEJMMBIR/FoStEs
hg19	74	14	86688615	-	-53	4	170091547	+	GTTAAGCAATTTGTGTACTGCA TAAGAAAGGTAACAGATCGTAGGATGTACATG GGGTGAAGGTA/CTCATCTTCCAACTTGTCCAGGATAAGCTACAAAAATAATTA TTTAGTAAATTTAATTTAGTGGAAAGTAATAA	0	22	MMEJMMBIR/FoStEs
hg19	75	4	179104413	+		14	25217696	-	TCCATTGGGCTTGAAGAGGGGATACTTGAGACTATCTTTAAATTAATAATTAAT ACATACACTCTTTTGGTATGCTATTTGGTATATTTCAAATTTCACTCAGTTTCTTA A/TATCACTCCAGATATTAGACCCAGCTTAT	0	1	NHEJ

Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19	75	14	25217696	+		4	179104413		TACATTCAATCAAAA CAATATGAATACTA CTGAATGATATTTCT ATACATGTTTTTC ATTGAGTCTTACACA AAAATAAGCTGGTCT AATATCTGGAGTGAT /TTAGAAA ACT GAGTGA AAATTTGAAATATA CAAAATAAGCATAA	0	1	NHEJ
hg19	Unreso lved	4	170534792	+		4	170148772	+	NA	NA	NA	NA
hg19	Unreso lved	4	172089614	+	1,1E +08	14	86571432	-	ACAAAGTTTGCAGAT CTACTGAAAGCTAAA AGTAAGCAGAAAAGT TGCACTGAAAA CCAACATTGCATGA ACA/CAATGTGTACT GTGT TTTAAATTCCTCTG GGGTGAGGTACTC ATAGCTTGACCAA GCTATGTACAAGCTG TGCCT	0	16	MMBIR/FoSTeS
hg19	Unreso lved	4	176195769	-	6660	4	172089274	+	GACAAAGATACTTAA ACAAAGGACAGTTT GGCCAAATTAATTC ATGACTTGGACA/J GGAGAGCATGTCACA AATACATGCTAAAAT CCATGATCTCAAAC TCTGACAA CATGAATCTCCAGAT GATCTTGTGGATCT GATATT	0	17	MMEJMMBIR/FoSTeS
hg19	Unreso lved	4	178588129	+					TCTCTGTCCCAACAA AAAA CCACAGAGGCAAGT CCCGT TTTAAAGGGGGTGC GGTC/CTTTAAACCT TTAAATACACATCCT GTGTACATTTTTTAT ACACCTACATGGGG GTACCTCACCCAGGG TGAGGTACAGTAGAA TACTTGTACTGATCT ACACGTTATCGA TCCACA	0	0	NHEJ
hg19	Unreso lved	4	179104341	-	4251	14	25222425	+	TCCATTGGGGCTTGA AGAGGGGATA TTCTGAGACTCAT TTTTTAAAT/TA AATTTAA TACATACATACACT TTTTTCTGTTATGCT ATTTTGTATATTTCA AAATTTTCACTCAGT TTCTT AATATCACTCCAGA TATTAGACCCAGCTT AT	0	4	NA
hg19	Unreso lved	4	181134375	+		14	86571413	+	CCAATTACCTGTCCT TGTAAAGTATCTTT GTCAAGGTAGGCA/ CAGCTTGTACATAG CT TGGTACAAAGCTAT GAGTACTCACCCCA GAGGAATTAACACAG ATACACATTGTGT CATGCAATGTTGGT TTTTTAGTGCAACT TTTTCTG	0	35	MMBIR/FoSTeS
hg19	Unreso lved	4	181134426	+					NA	NA	NA	NA
hg19	Unreso lved	4	181549440	+	2463				NA	NA	NA	NA
hg19	Unreso lved	14	25217642	+					NA	NA	NA	NA

Patient 20

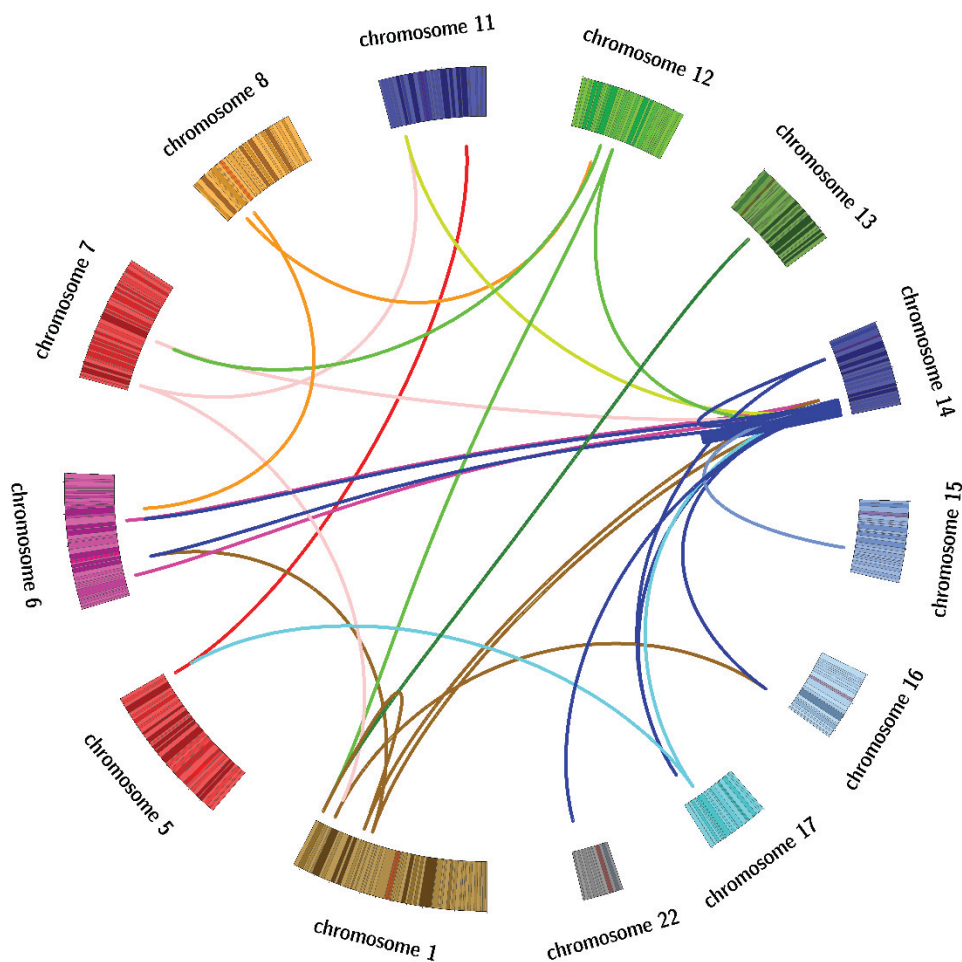
Phenotype: growth retardation, hypotonia, microcephaly, dysmorphic features, coloboma, and hypoplastic corpus callosum

Karyotype:46,XX,der(14)(q31q32.3)dn .ish der(14)(wcp14+,tel14q+)

Comments:arr[hg19]

14q31.1(83088298_83225103)x1 ;14q31.1(83241901_83329898)x1,14q31.2(84163810_84376455)x1 ,14q31.2(84379113_84429967)x0,14q31.2(84431545_84589783)x1,14q31.2q31.3(84730585_85038612)x1,14q31.3(85985712_86019586)x1,14q31.3(88328980_88809657)x1,14q32.11(91425021_9165551)x1,14q32.11q32.12(91766712_91923002)x1,14q32.12(93989679_94156172)x1,14q32.13(95024931_95083116)x1,14q32.2(96853454_96946830)x1,14q32.2(98308170_98679536)x1,14q32.2(98892998_100452660)x1

Circos plot:



Chromosome 14 is shattered with small regions of other chromosomes making several chains inserted in different chromosome 14 breakpoints. The affected chromosome 14 is paternal while the rearrangement occurred *de novo*. This observation has just been published (Ader et al., 2019).

Ref.	Junct- ion ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo- logy	Insert- ion	Inferred mechanism
hg19	1	1	168607753	-		6	48320595		TTTGGCATT TTTTCACACAGCTTTATATAAAGAAATAAAACGAATGAATGAATCAACCCGAGTACAGGG ACAG /GGGAGTTGACAATGATGTTCCAAAGCCAATCAAAGT AC ATTTTCAACAGCTTTATATAAAGAAATAAAAC/	0	23	MMBIR/FoTeS
hg19	1	6	48320595	+						0	23	MMBIR/FoTeS
hg19	2	1	168607787	+		14	95253675	-	ATTATTTCAACAGCTTTATATAAAGAAATAAAACGAATGAATGAATCAACCCGAGTACAGGG GAGTTGACAATGATGTTCCAAAGCCAATCAAAGT ACTTTGCA	0	0	NHEJ
hg19	2	14	95253675	-	0	1	168607787	+	CCCGAGTACAGGGAGTTGACAATGATGTTCCAAAGCCAATCAAAGT ACTTTGCAAACT CCCTCAGGGACACAATATTGGCCATCATGCAACTGCCCT	0	0	NHEJ
hg19	3	1	227162125	-		16	87703254		TCAACTGT CCCGCCTCCACCCACAGCCCA GGC GGCTTTAGAAATGGATCTGAGC AGGGAGAGGTGTGAGCAGCTGGTCTATGCAAAAGACTCTCT	3	0	NHEJMMBIR/FoTeS
hg19	3	16	87703251	-		1	227162128		TGACCCACCTGT CCTCACACTCCTCCCTGCTAGATCCACA TTCTAAGAGC /GCCCTGGCG TGTGGGTGGAGCCGGACAGTGTGAGGGCCGCCTGGTCA	3	0	NHEJMMBIR/FoTeS
hg19	4	1	227162166	+		7	11546481		CCGGCTCCACCCACACGGCCTCTTAGAATGGATCTGAGCAGGGAGAGGT GTGAGG ACAGGTGGTCTGCAAAAGACTCTTAGGAAGATAG	1	0	NHEJ
hg19	4	7	11546482	+		1	227162165		GTTCTGTTATGGCCAGACAATCTATCTCTAGAGAGTCTTTGCATGACCCACCTG TG/CTCACACCTCCTCTCCAGTCCACA TTCTAAGAGC	1	0	NHEJ
hg19	6	3	1095133	-		14	87537527		CTCAGGGTTTCTGGGAGC /GCACACGCCAGCTAAAACTGCTTTTGATGATCACG AGGTCAAGAGATCGAGACAAAATAGTAGTATGTTCTAG	0	0	NHEJ
hg19	6	14	87537527	+	1	3	1095133	+	TGTCTCGATCTCTGGTGTGATCCATAAAAGCAGTTTTAGCTGGGGTGGTGC / GCTTCCAGAAACATGAGCATCTCCCAAAGTCA CCAAAGAT	0	0	NHEJ
hg19	7	3	171614184	+		6	25324760	-	AATAGTGCATATATTGCCAACTAAATACACATGAGGAAGAACTA ACAACATT CTTAGGATTGCTGGCCAGGCTAAGCA CTTGA AAAAAGTTGTT	1	0	NHEJ
hg19	7	6	25324761	+		3	171614183		GTCTGCTTCTAGTGTCTCCAAACACTTTCAAGT GCTTAAAGCCTGGCACAGCAATCCT AAGAA GTGT /AGTTCTTCCCTCATGTTAA TTTAGT	1	0	NHEJ
hg19	8	4	133103888	-		14	83238668	+	AGACTTTCTCAACT TTTGATGTGGCATTAGT/TTTTAA TTACTAATGGCATTGGACA CATATTTCAAATCAATATCTTTACTTAAACAATGTAA	0	0	NHEJ
hg19	8	14	83238668	+		4	133103888	-	CTTAGTCTATAGTTGTGAAGTGGTTGTTAGCTTGAGATCTTTCTAACTATTTGAT GTGGCATTAGT/TTTTAA TTACTAATGGCATTGGACA	0	0	NHEJ
hg19	9	4	133104052	+		8	22739493	+	TCTAGAAGTCAATTTATTTATAGTTTAA AAATATATCATCTT GAAATTA GT/ACTGCC ATCCCTGGCTCAATGGGGCAGATGAGGAAGGTCATGCAG	0	0	NHEJ
hg19	9	8	22739493	-		4	133104052	-	TAAAAATATATCACTTGAAATTA GT/ACTGCCATCCCTGGCTCAATGGGGCAGATG AGGAAGG TTCTG TCATGCAGCTCTGGCTCCAAAGACT TGA CTGC	0	0	NHEJ
hg19	10	4	174563182	-		6	144168815		ATGCAATGGAATCCACCGAGTGGAAATGGAATAAA ACCGCCCGGT CGTATAAGACT AACTTTA /ACTTCTATGTGCTATTATGGTGCATCCTAGAT	1	0	NHEJ
hg19	10	6	144168816	+					ACCGCCCGGTCTGATAAGACTAACTTT TA	1	0	NHEJ
hg19	11	6	25324725	-		14	86688858		GTCTGTCTTAGTGTCTCCAAACA CTTT CAAG /TGCTTAAAGCCTGGCACAGCAATCC TAA GAATGTTAGTTCTCTCATGTTAAATTAGT	1	0	NHEJ
hg19	11	14	86688857	-	2	6	25324726		CTAAATTAACACATGAGGAAGAACT AAACA CTTT TAGGATTGCTGTGCCAGGCTTA AGC/ACTT GAAAGTGTGGAGGACACTA GAGACAGACA	1	0	NHEJ
hg19	12	6	105684218	-		14	76423225		/A GAA GAAT TCAGATTCACCTGGTGCAGT	1	0	NHEJ
hg19	12	14	76423226	+		6	105684219		AAATTA TTCCCTTTCTATATTGAAGCCTCAAGTAAGGACGAGTGA /GAA GAATTC AGAGTTCCACCTGGGTGCAGTCC TTCC AAAAAGCTGAGTCT	1	0	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	13	4	174563336	+		6	121471513		CACATTATCCCATTTAAAGATGAGGAAATATATTGACAAATATGTCTGAAAACT/C CTTGAAGCCAGGAGTTGGACACAGCTTGGCCAACTGGT	0	0	NHEJ
hg19	13	6	121792659	-		4	174563336		ATGAGGAAATTTATGCAAAATGTCTGAAAACCT/CCTGAAAGCCAGGAGTTTGGC ACCAGCTTGGCCAACTGGTGAACCCCTCTCTCAAAA	0	0	NHEJ
hg19	14	7	11546441	-		11	17722395		GCTCTTCAAGGAGTACCTTGGCGTCTTTAAGTCTCTGTTATGGCCAGAC/A TCTATCTCTAGAGAGCTTTGGATGACCACTGCTCCTC	0	0	NHEJ
hg19	14	11	17722395	+		7	11546441		GCTCTTCAAGGAGTACCTTGGCGTCTTTAAGTCTCTGTTATGGCCAGAC/A TCTATCTCTAGAGAGCTTTGGATGACCACTGCTCCTC	0	0	NHEJ
hg19	15	7	79520512	+		14	88326161		GGCATAGGAGACCTTTGAGATAATGTAATGTTTG/ ATCAATATTAGCTGGATAATGAAATGTGGACATATACACCTGGAAC/GGCATAG GAGACCTTTGAGATAATGTAATGTTTGCGAGAAGCTGGTTG	NA	NA	NA
hg19	15	14	88326161	+	- 4865 44	7	79520512		ATCAATATTAGCTGGATAATGAAATGTGGACATATACACCTGGAAC/GGCATAG GAGACCTTTGAGATAATGTAATGTTTGCGAGAAGCTGGTTG	NA	NA	NA
hg19	16	8	22739530	+		12	41730792		TAAAAATATATCATCTTGAATTTATGTACTGCCATCCTGGCTCAATGGGGCAGATGA GGAAGGG/TCATGCAAGCTCTGGTCCAAGAGTCTGAACTGC	3	0	NHEJMMBIR/FoStEs
hg19	16	12	41730795	+		8	22735527		AATTTGGGGCAGTTTCAGACTCTGGAGCCAGAGGCTGCATGACC/TTCCTCATCTGCC CCATTGAGCCAGGGATGGCAGTACATAATTTCAAGATGATAT	3	0	NHEJMMBIR/FoStEs
hg19	17	11	17722368	-		14	86439890		GCTCTTCAAGGAGTACCTTGGCGTCTTTAAGTCTCTGTTATGGCCAGACAACT TCTATCTCTAGAGAGCTTTGGATGACCACTGCTCCTC	0	0	NHEJ
hg19	17	14	86439890	+	1	11	17722368	-	TGCTCTTCAATTTCAACCTTGGTGAATCTGGCAATTTGTCTTGGGATGCTCTTCTCA AGGAGTACCTTTGGCGT/CTTTAAGTCTCTGTTATGGCCAGACAACT	0	0	NHEJ
hg19	18	7	79520548	+					GGCATAGGAGACCTTTGAGATAATGTAATGTTTG/ GCACATATACCACTGGAAAGGGATAGGAGACCTTTGAGATAATGTAATGTTTG/IGG CAGAACTGTTGAGCCAGATTACCAGTGTGTGGTATCAGC	2	0	NHEJMMBIR/FoStEs
hg19	18	12	47730627	-		7	79520546		GCACATATACCACTGGAAAGGGATAGGAGACCTTTGAGATAATGTAATGTTTG/IGG CAGAACTGTTGAGCCAGATTACCAGTGTGTGGTATCAGC	2	0	NHEJMMBIR/FoStEs
hg19	19	12	65825988	-		14	96431001		GCAAGGCTAAGGGTCAAGACTGCTTTGAGGAGAGGGGT/AATGCAAAAATTA TATTCATATATGCAAAAATGTAAGTCTTGGATGTAGCA	1	0	NHEJ
hg19	19	14	96431000	-	1				AAGTATTTGCA/TACCCCTGCTGTAAGCAGTCTGACCCCTAGCCTGCCAAG CAAGGTGCTCCTACCTGGAAACAGTACCCACTCCTGCAG	1	0	NHEJ
hg19	20	1	246568908	-					GATTAAGAATGGACTCTGAGGCCAGGTGCAGT GATTCTGGCACAGGACTGTTTTGTGGCAAGGAAAGATGAATG/ATTAAGAATGGAC TCTGAGCCAGGTGCAGTCTTACCAGTCTTAGGTATTC	1	0	NHEJ
hg19	20	12	65826166	+		1	246568909	+	GATTCTGGCACAGGACTGTTTTGTGGCAAGGAAAGATGAATG/ATTAAGAATGGAC TCTGAGCCAGGTGCAGTCTTACCAGTCTTAGGTATTC	1	0	NHEJ
hg19	21	1	181300811	+		14	76423082	+	AAAATTAATGATTAAGAATGGACTCTGAGCCAGGTGCAGTCTTACAAGTCTT CAAGGTATCTTAATCC/GTATCTAGCTGCTTGGTTCATTA	0	0	NHEJ
hg19	21	14	76423082	-		1	181300811		AAGAATGGACTCTGAGCCAGGTGCAGTCTTACCAGTCTTAGGTATTCCTTAATC C/GTATCTAGCTGCTTGGTTCATTAAGTGAACACAGGCC	0	0	NHEJ
hg19	22	14	21043051	-		14	84627318		TTACAGCAGTAAATAAATGGAATGAAAGTCTAGAGTAAAAAGGT/AGCTTGAG CCTTAGGATAGAACATGTGAGTGAAGGCTGATGAGAC	0	0	NHEJ
hg19	22	14	84627318	-	0	14	21043051	+	TCTCCATGACGCTTCTCATCTCACATGTTCTATCCATAAGGCTCAAGCT/ACCTTTTAC TCTAGACACTTTCATATCCAGTTTAAATAGCTGCTGAAA	0	0	NHEJ
hg19	23	14	21043206	+		17	67466261		CTTGATTGGAATGAAGGATGTAATGATGATCTGGGTGCTGTGAGGG/CCTTGA GCCCGAGGATGAGGTTTGCAGTGGCCCTCATCACCACTTG	0	0	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	23	17	67466261	-		14	21043206		GATTGGATTGAAGGATGTAATAATGATTAAGTCTGGGTGTCTGTGAGGG/CCTTGAGC CCAGGGTAGAGTTTGCAGTGAGCCTCATCACACCTTGGCAC	0	0	NHEJ
hg19	24	14	82711713	+	1769 4	14	83225750	-	TACACATGTTTTCACTGCCAGCAAACTATACATGACATAAAGGAAAGGC/ GATTCA TATAAA TCCAATTCAAGTGAAGTGAATAACAGTTAGAAAAGTT	0	0	NA
hg19	24	14	83225750	-		14	82711713	+	CAGCAAA ACTATACAT GACATAAAGGAGGC /GATTATATAAAATCCATTTCAAGTGA AGTGTAAACAGTTAGAAAAGTTTAAATGGAAAGTGTGAGTAAA	0	0	NHEJ
hg19	25	14	82729406	-	1769 4	14	91413843	+	ATCTCTCCA ACTAAAT CTGCACTCATGCTCTGTGA / AACCTT ACATTTAAAAAGTACT TCTGTGAGGACATCTGCCAGCAGCTGCCTATCTACCTGA	3	0	MMBIR/FoSTes
hg19	25	14	91413846	+		14	82729409	-	TTCGAGGGCAAGCAACTCTCCAACATAAATCTGCACCTCATGCTGTGAA AAAC/CTT ACATTTAAA AGTACT CTGTGAGGACATCTGCCAGCAGC	3	0	NHEIMMBIR/FoSTes
hg19	26	14	82731076	-	2	14	83336492	-	ACATTTATTT ATTAATAAT/ACCCAAATAGAAATCAACCCCTTAGAGGTGGTAACTT AGCACTGATATGAAATTTACACAAGCTAGTTGTTTAT	0	0	NHEJ
hg19	26	14	83336492	-		14	82731076	-	TAAAGTTACC ACTTAAAGGT TTGATTTTAAATTTGGGGT /ATTTATAAAATAAATAATG TATTATGGTCTTTTCACTACTATTGAACCTCTGTGGTAG	0	0	NHEJ
hg19	27	14	82731077	+	2	14	93970535	+	GCACCTATTCAGGAATACACACTCACAGTGGCTAAGCAGCATATCTGGAA GGCAC /TTGCCA GGTGGATTTAAATTTATTT TAAGCTACAGTA	2	0	NHEIMMBIR/FoSTes
hg19	27	14	93970537	+	1892 69	14	83731075	+	TTTTAAITACTGAGCTAATAATAAATTTTAAATCCACTGTG CAAGT/GCTTCCAATG ATATGCTG TTAGCCACTGTGAAGTGTATTTCTCTGTAT	2	0	MMBIR/FoSTes
hg19	28	14	82796695	+	1	14	86023056	-	ATATCAAAAAGCATGGAAATGTCATTTCTGTAGAAATCTGAGCTTCTGGGT CA/ GTTAATACA TAGAAATG CTTGGCACATGGTATATACACAA	1	0	NHEJ
hg19	28	14	86023055	-		14	82796694	+	ATGTCATCTT GTAGAATA CTGAGCTCTTGGGTC/AGTTA TACATAGAATAGTGC TTGGCAGATGGTATATACAAATAAGTATTAACCTGTTGCTA	1	0	NHEJ
hg19	29	14	82796695	-	1	14	84799118	-	GAGTAAAA ATTTCTGTGTGAT/ACTTCGATACAGAGAAGCAGCTTTGATTTTCAGA CCAGCTTCAGTGTTCAGATACAGGATTTTGCACCTACAAT	1	0	NHEJ
hg19	29	14	84799117	-	7119 1	14	82796696	-	AATCAAAGCTCTTCTGTATCGAAG/TA TACACAGAAAATTTTTACTCTATCATAAA TTTTAAACATATGTTTGTATTTCTCAGTATCTCCTTTG	1	0	NA
hg19	30	14	83355463	+	0	14	84824288	+	AGTAGAGCGGTTCCACATGTTGGCAGGCTGGTCTCAAACTCTGGCCTCAAGT/ A CTGTATATGTACCA CTTTCTTTATTT TTTGTCTGTG	0	0	NHEJ
hg19	30	14	84824288	+	2149 05	14	83355463	+	AGTGTCCATCACAGACAAATAAATAAGAAAATGTGGTACATATACACAGT/ ACTTGA GGCCAGG ATTTGAGACCAGCT GGCCAA CATGGT GAAACTC	0	0	NA
hg19	31	14	83355464	-	0	14	85056193	-	TCTTCAAAA GGAGAAA AGACTTAACCCTGTGAC /GATCCGCTGCCTGGCCTCCAAA GTGCTGGGATACAGGCTAGCCACTGAACCTGGCTGCCT	0	0	NHEJ
hg19	31	14	85056193	+	0				AAAATATCTTTTCTAAAATCTCAAAGGAGAAAAGACTTAACCTGTGAC/ GATC CGCTGCCTGCCTCC AAA AGTGTGGATACAGGCATGAG	0	0	NHEJ
hg19	32	14	83376934	+	0	14	84123928	-	TAAAGTAAGCACTGCTGATTTGCTCATGTTATGCAAAAAGTTCGAGGATG/ GTTT TACTCATTGTACTCC CTCTCTGAGCATTT CCCTTAATG	0	0	NHEJ
hg19	32	14	84123928	-	0	14	83376934	+	AGCACTGCTGATTTGCCTCATGGTATGCCAAAAGTTCCAGGATG /GTTTACTTTCAT TGTACTCCCTTCTGAGCATTTTCCCTTAATTTGCACCTGAG	0	0	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	33	14	83376935	-	0	14	89839245	-	GCCTTGGCCTAAGCGCTGGAAATTACAGGTTGAGCC/AGATTCAAAGGGTTAAGGGTAAAGGCAAAAAGAGGAGGAGGAGTGAATAATAGGCAATT	1	0	NHEJ
hg19	33	14	89839244	-	2	14	83376936	-	CAACTGCCTCTCTCTTTTCCCTGTCTACCCCTAACCCCTTGAATC/TGGCTCACACCTGTAATCCAGCGCTTAGGCAAGGCGAAGGATCGCTTGA	1	0	NHEJ
hg19	34	14	83625230	+	1	14	83702210	-	AAAAATTCCTAGAAATAATCTGTGGATTACCAAGACTTCAGGGCATAAGGG/GCC	0	0	NHEJ
hg19	34	14	83702210	-	2603	14	83625230	+	TAAACAACAGTGGACTCATAATTTAACTAAGCTCAGTCCCTTTTGGATTTACCAAGACTTCAGGGCATAAGGG/GCCTAACACAGTGGACTCATAATTTAACTAACCTAGTCCCTTTAGCACATGTTGAGCAATAATTTGCC	0	0	NA
hg19	35	14	83625230	-	1	14	98304843	+	AGCCCCAGCTGGCCCTGCAACCTCTCTTAGATGGTTCTGTAAAGAC/GCAAATAACAA	0	0	NHEJ
hg19	35	14	98304843	+	3749	14	83625230	-	AAGCCAAATGTTGGGGTGGGTGAGTAGGGGAGGTATAAGCAGTTGCAGAAATGCAAAGCTCACCTGTGAGCCCGCTGGCCCTGCAGACCTTCCTT	0	0	NA
hg19	36	14	83676170	+	2603	14	89176197	-	AGATGGTTCCTGTAAGAC/GCAAAATAACAAAAGCCAAATTTGTGCTGATATTTCTAACGGAAACTAAGTTCACCTGCCTATTTTATAATAACAATAATTTAAAGCTAATTTACATGAAT/TACACGCAACAATAATGGAAA	0	0	NA
hg19	36	14	89176197	-	0	14	83676170	+	TCTGCCTATTTTATAAACAATAATTTAAGCTAATTTACATGAAAT/TACACGCAACAATAATGGAAACAGCAGCTGAATCAGTCCCGCTTATAAATTT	0	0	NHEJ
hg19	37	14	84123927	+	0	14	84645180	-	CTCAGGTGATTGATATCTACCAGGGTGGCTGGTTAATACTACTGACTTAACCT/AAGTCAACACATCTGTCTACACTCAAGGGGAGAAGTACA	0	0	NHEJ
hg19	37	14	84645180	-	0	14	84123927	+	GTTAATCTCAITGACTTAAC/AAGTCACACATTTCTACACTCAAGGGGAGAAGA	0	0	NHEJ
hg19	38	14	84594911	-	4316	14	84802113	-	TGACACAAGAAATGTAAGGCTAGGATTTGACCATCAAAATAGTTCAAGTGTCTCTCTGTTGCATACCTTTAAT/TCAATTTCTGAGACAGGAT	1	0	NA
hg19	38	14	84802112	-	47	14	84594912	-	AAATGAAAGTCTTAAAAAAGGAGGATATTTATTTCTTTTTAAGGACTTCATCTCTGTCTCAGAATG/AATTTAAAAGTATGCAACACAGAG	1	0	NHEJ
hg19	39	14	84627317	+	0	14	85914881	-	AGAAGAAAACACTGAAACCTCATCTAAGTTGTTGACTTATATGTTGATCTTCTCAAGTAATAAAATGGCATTGATGAAATTTAAAATCTGAAGGATA	1	0	NHEJ
hg19	39	14	85914880	-	1442	14	84627317	-	GC/CTTTAACTGAAGCTGGTGTGCTTGTATCCAAAAGTTAGCATTGATGAAATTTAAAATCTGAAGGATAG/CTTTTAACTAAGCTGGTGTGCT	1	0	NA
hg19	40	14	84645179	+	0	14	91758296	+	TGATCCAAAAGTTAGGCAATCAGAAGACACAGGGTCTAATAATTTATCTATAAATCCACTCACATTTGCCATATCTATTGATAAAGAAGC/GCTCCGA	0	0	NHEJ
hg19	40	14	91758296	+	1609	14	84645179	+	ATGCACTGTGGGGAGCCCGTGACTCCCTAAGTGCAAAATGCTTCTGCTGCATTAGGAGTACAGGGGCTGCCCCACAGTGCATTGGAGC/GCTTCTTATCAATAGATAGCAAAATGCAAAATGAGTGGATATTTATGATA	0	0	NA
hg19	41	14	84701599	+	55	14	91400188	-	TCAATAATTTTTATTTTTTTTTTTTTTTGATTTGTCAGTGTGTCAGTGTGTTTCT/CTCTAACCCCTCACTTCCATAACCCACCTTAAAGGGCCTTAATTTG	0	0	NA
hg19	41	14	91400188	-	2229	14	91400188	-	CTAACCCCTCACTTCCATAACCCACCTTAAAGGGCCTTAATTTGCTGATCCCTTACCTAGAAATCTCTCTCTGAGATATCT	0	0	NHEJ

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	42	14	84703829	-	-	14	89799705	+	AGGAGAACTCCTGAACCTCAGAGGGGCAAGGTTGGCAGT/ATTGTAAACAGAGCAAAAGC CCTCTAAATCTCAACGTCTCAGGCAGCAGGTCCTTGGTATCC	0	0	NA
hg19	42	14	89799705	+	0	14	84703829	-	AAGCACCTGTAATCCAGCTACTCGGAGGCTGAGCGAGGAACTCACTTGAACCTCAG AGGCAGAGGTTGCAGT/ATTGTAACAGAGCAAAAGCCCTCTA	0	0	NHEJ
hg19	43	14	84727925	+	-	14	86096148	+	GTCTTTTCTGTGATTATAAAAACATAATGGAATACTCTCATCTCTGTTACAGATGTA TAAAGTAGATC/GTTCCTTTATTTCCATAAATGATAAT	0	0	NA
hg19	43	14	86096148	+	-	14	84727925	+	TTTTCATGATGGAGCTATCAACTCTCAGTGGATACATATATATATATTTATGGCAA TAAAGGGAAC/GATCTACTCTTATACATCTGGAACAGAAT	0	0	NA
hg19	44	6	105684250	+	8				AGAAAGAAATTCAGAGTTCACCTGGGTGCAGT/	0	0	NHEJ
hg19	44	14	85056194	-	0	6	105684250		AGCCTCAAGTAAGGACGAGTGAAGAAATTCAGAGGTCACACCTGGGTGCAGT/CCT TTCCAAAAGCTGAGTCTATATGACTGTCTATTAGATAAGGCA	0	0	NHEJ
hg19	45	14	85900452	+	-	17	67466429		CATGATGCCCTGAGCAGAAAAGTGGATCACCTATTAGTCTGGTCTGTCCAGGGGAAA AAAAGTCTACTAGTTTTTAAT/ATTCCACATCCCCTCCAGC	0	0	NA
hg19	45	17	67466429	+	7	14	85900452		GAAACAATGTAAGTGTAGACAACCTTTAGATAAATTGGATTGGCTGGAGGGAT GTGGAAT/ATTAATAAAGTAGTAGACTTTTTTCCCCTGACA	0	0	NHEJ
hg19	46	14	86109927	-	-	14	92709155	-	CCACAGTTAAGCCAGCCTTGTGA/AGGTTTTCACATATTTGCCAGGCTGCTCTCAA ACTCCTGACCTCAGGTGATCCACCCGCCTCAGTTCCTCCAAAT	3	0	MMBIR/FoTeS
hg19	46	14	92709152	-	1				CTGGCAAATATGGTGA/CCCTTTCACAAAGCTGCTGGCTTAACTGTGGAGTCCCTTCC CACCTCCACTGAAATAGACATAGTTCCTTCCCTTCTCATTTCCAC	3	0	NHEJMMBIR/FoTeS
hg19	47	14	86439890	-	1	14	87537525	+	CACCTCCTATGCCCTCTCTACCACCCACCTATGAGTCAAGACAC/TGTTCTCTGT TATTTCTTGAATTTGAAATTTGGCTGCTTGTGCTAGCTTGG	2	0	NHEJMMBIR/FoTeS
hg19	47	14	87537527	+	1	14	86439892	-	CACCTCCTATGCCCTCTACTACCACCCACCAATGAGAACATAAGACACTG/TCTCTG TATTTCTTGAATTTGAAATTTGGCTGCTTGTGCTAGCTTGG	2	0	NHEJMMBIR/FoTeS
hg19	48	14	86688858	+	2	14	89165217	-	TCTCAAAAATACCTTCAGTGCAGTGGGCAAACTCTGTTAGAGAAAATCAAGGTATG GGGAG/CTTAGGAACTTGACTCATGTACCAATCAACAAC	2	0	NHEJMMBIR/FoTeS
hg19	48	14	89165215	-	0	14	86688856	+	TGGGACAAACTCTGTTAGAGAAATCAAGGTATGGGG/ACTTACGAAACATTGACTCAT GTACCAATCACACAACCTCTACGAGATGGGAAACAGATAACAT	2	0	NHEJMMBIR/FoTeS
hg19	49	14	88150158	+	-6	14	90080465	+	AAATAGCACCATGAGTCCCAAGGATATAACACATGAGTAAATAAAGTA/AA ATGATAACACACAGATGATTTGGCATATCATCAATAATTC	4	0	MMBIR/FoTeS
hg19	49	14	90080469	+	4	14	88150154	+	ATAATGTAATATATTTGATGATGCAAAATATCATCTGTGTGTTATCATTTACT/TATT TATTTACTGATGTTTATCTTGGGACTCATGGTGT	4	0	MMBIR/FoTeS
hg19	50	14	88150165	-	-6	14	89176195	+	CATATACTACAACAATATTTTGGCAGTGAAGGCAAAAGGCATT/AAAGCCCTTCTT ACAGCCAATTAATAATAACAACAACAACAACAACA	5	0	MMBIR/FoTeS
hg19	50	14	89176196	+	0	14	88150170	-	CAGCATATCTACAACAATAATTTGCAGAGTGAAGGCAAAAGGCATT/AAAGC/CCTT TCTACAGCACAATAATAATAACAACAACAACAACAACA	5	0	MMBIR/FoTeS
hg19	51	14	88812706	-	-	14	105291088	-	AAACCATGCAACCTCCACAGCTACGATCC/AGTAAATCGTTTATTGTTGAGGTCTCAT AAAAAAGGCAGTGGGAGGTGACTCCTTGGTGTGCTTCCCT	1	0	NA

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	61	14	94289074	+	2	14	98679828	-	TGTTTTGTGCATACACCTAAGAATTGAATTGCTGGTCA TGTGGAATTCGTAATATG/T GTGGAATGTTAAATGTTAATTAATGATGATTAATTAATA	0	0	NHEJ
hg19	61	14	98679826	-	3749 82	14	94289072	+	AAGAAATGAAATGCTGGGTCATGTGGAATTCCTGTAATA/TGTGTGGAATGTTAAATGTT ATTAATGTGATTAATTAATAATTAATAATTAATAATGTTAATAAG	0	0	NA
hg19	62	14	95026062	+	5870 2	14	100741565	+	TGATTAATGCTCTCAATTTGACAGTGAATAAAGTGGAGTTTCAGAGAGGTGGAGT/ACTT TTCAGGTTATCCCTTTAATTTAAAGACTAACTACATTTGAAA	3	0	MMBIR/FoTeS
hg19	62	14	100741568	+	3	14	95026059	+	TTTCAAATGTAGTTAGTCTTTAAATAAAGGATAACTGAAAAGTACT/CCACCTCTCT GAACCTCAGTTTTTTCAGGTGCAATGAGAAAGCAATAATCA	3	0	NHEJMMBIR/FoTeS
hg19	63	14	95084765	-	5870 2	16	87703427	+	GTCACCTGACAGTGTGCGGCTGCAGGACCCAGTGT/ACTGAATCCAGGGCAAA TCTTAGCTGAATTCATGCTGTAGAGGTGACATGAGCTA	0	0	NA
hg19	63	16	87703427	+		14	95084795	-	ACGAAGACAGGAAGTCCAAGACAGTCACTGCTGAGTGGGCTGCAGGCA GCCAGTGT/ACTGAATCCAGGCAAACTTAGCTGAATAATCAT	0	0	NHEJ
hg19	64	14	95253674	+	0	14	97741197	-	CCTCTAGACCATTCTGGTTCCAGCAGTGTTCATCATGGTTCCTTAGGAGATAAG ACAGC/GCATGGCAGCGTCAGGCGGTTCGATGTTTGAGAA	0	0	NHEJ
hg19	64	14	97741197	-	2	14	95253674	+	TTCCTTAGGAGATAAGACAGC/GCATGGCAGCGTCAGCGCTGCATGTTTGAGA ACACTGCATTAGACTCAATGTCATCTCTTTAACTCC	0	0	NHEJ
hg19	65	14	95498795	+	5	14	96957263	-	ATTGAAGACTATTTTTCCCTCAATGAAATTCCTGGACCAATGTTGAAAATCA/CA TTGATGGACTTTAAAGAAAATTTCTCTGATAGCATGTCAC	2	0	NHEJMMBIR/FoTeS
hg19	65	14	96957261	-	1047 60	14	95498793	+	AAATATCTGGCACAATGTTGAAAATCACATGATGACTCTTAAAGAAAAATTTCTC TGAGCATGCACGTGTTGATGGCAITTTACCACATTA	2	0	MMBIR/FoTeS
hg19	66	14	96852500	+	1047 60	14	100741566	-	TGGACAACTTTGTAGTAAATTAAGTCTACCAGCCTTTTGGAAAACAAATTAACA GTTCC/ACTTTGAGATATTTGTTGGGGCTATTTCAAAC	0	0	NA
hg19	66	14	100741566	-	3	14	96852500	+	TTGAAAACAAATTAACAAGTTTCC/ACTTTGAGATAATTTGGGGGCTATTTTCA AACTAATTTGGATGAGACTAACAGAAAATCAAGTATCA	0	0	NHEJ
hg19	67	14	98892903	+		14	105290364	+	TTATCAATTTCTAAGAAACCCCACTCAATAGTATTTGCTCTCCAAGCC/TCATG AGATCTGATGGTTTTAAGGAGAAATTTCCCTGCATAAG	1	0	NHEJ
hg19	67	14	105290365	+	6721	14	98892902	+	GGAAGGAACCTTAGCAGGAAATTCCTCTTATAAAACCACTAGATCTCATGAG/GCT TGGAGAGCAATACTATTAGGTATTGATGGGTTTTCTTTAGG /ATTTTCAACAGCTTATAAAGATAAAG	1	0	NA
hg19	68	6	48320565	-						0	0	NHEJ
hg19	68	14	100457981	-		6	48320565		GTGATTCCTGTTTTTATCTTTATATAAAGCTGGTGAANAAT/ATGCGAAAATAATA ATAAAAAGCAAGGTTTCGTGGCAAAGTTATCTCGGGTAA	0	0	NHEJ
hg19	69	17	56418313	-		5	174016897		CTCCAAAGCTCTAGTCAGAAAAGCCACAGGGATGTTGCTTTTGAAGT/ACTGCGG CTCTGCTGGAAAACAGAACAGACACTGCACCAGCCTACA	3	0	NHEJMMBIR/FoTeS
hg19	69	5	174016894	-		17	56418310		AGTACTTCAAAAGAAATGCACCATCCCTGTGGC	3	0	NHEJMMBIR/FoTeS
hg19	70	14	85039194	-	2149 05	14	89846269	+	TCTGACCTCAAGTGTCCACCTGCGCTCAGCCTCCCAAAGT/GGCAACAATGACTGAG AGGTGGTAGGCTATTACAAAATCAAAAGGGCAGGTAATTTT	1	0	NA

Ref.	Junction ID	Chr. A	Chr. A Position	chr. A strand	CNA	Chr. B	Chr. B Position	chr. B strand	Split-read sequence	Homo-logy	Insertion	Inferred mechanism
hg19	70	14	89846269	+		14	85039194		NA	NA	NA	NA
hg19	71	3	171614023	-		15	70133358	+	CTAGCACAGTGAACCCCTCCAGCTCTAACATGTTCTTGGTTCCGGTTCTCCCACTGA/T GTTTACCAGCTGTGAAAATGTGGCTCTGGGCACCCACA	2	0	NHEJMMBIR/FoSTeS
hg19	71	15	70133360	+		3	171614025		ATCCTGGTAGCACAGTGAACCCCTCCAGCTCTAACATGTTCCATGTTCCGGTCTCCCC ACTGA/Tg/TTTACCAGCTGCTGGAATGTCGGCTCTGGGC	2	0	NHEJMMBIR/FoSTeS
hg19	72	1	246568942	+		1	181300781		AAAGATTAATGATTAAGAAATGGAAGCTCTGAGCCAGGTGCA/Tg/CTTTACAAGTTCT CAAGGTATCTCTAATCCGTATCTAGCTGCCTGGTTGTTCAATA	2	0	NHEJMMBIR/FoSTeS
hg19	72	1	181300781			1	246568942		NA	2	0	NHEJMMBIR/FoSTeS
hg19	73	3	1095163	+		8	35911084		TCATGTTTTCTGGGAAAGCCACCCAGTCAAAAAGTCTTTTGTAT/GGATCACGGA GGTCAAGAGATCGAGACAAACAAATAGTAGCTATGTTCTAGG	0	26	MMBIR/FoSTeS
hg19	73	8	35911084						NA	NA	NA	NA
hg19	74	14	83088061	+		15	70133314	-	TGTAATCAGCACTTTGGAGGCCAGGCGGATCACGAGGTGAGGATCGA GACCATCTTGTAGCACAGTGAACCC/TCCAGCTCTATCAT	2	0	NHEJMMBIR/FoSTeS
hg19	74	15	70133316	-		14	83088059		GAGATCGAGCCATCTGGCTAGCACAGTGAAC/CTTCCAGCTCTAACATGTTCCAT GTTCCGGTCTCCACTGATTTTACCAGCTGTGGAAATG	2	0	NHEJMMBIR/FoSTeS
hg19	75	6	121792834	+		8	36053591		GCATGAAATCACTTGAACCTGGAGGAGAGATTTGCAGTGAAGTCAAGATCG/GCCTA GAACATAGTACTTATTTGTTGCTCGATCTTTGACCTCGTG	0	0	NHEJ
hg19	75	8	36053591			6	121792834		NA	NA	NA	NA
hg19	76	14	84163263	+	- 4316 47				GTATATCTATATTTTTGTATATACTAAATATCGTATATTTTCAAAATC/GGGC CGATTTGGCTCACTGCAAGGCTCCGCTCCGGGTTCCAG	0	0	NA
hg19	76	14	91656323	-					TGAAACCCCTCAATCAACAAATGGGGTTTCGCTCGTTTTTTTGCCTCCCTCTTTTA TTTT/TTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGACGGAG	0	0	NHEJ
hg19	77	4	49107462			14	85985071		NA	NA	NA	NA
hg19	77	14	85985071	+		4	49107462		CTGTAAGACATATGAAATTTCTTTCTTTTTTTTTTTTCTTTGAGACAGC/CCCAAGTGCA ATGGAATGGAATGGAATGCAATGGAATCCACCAGTGGAAatggaat	0	0	NHEJ
hg19	78	4	49107432			4	49107302		CCCAAGTGCATGGAATGGAATGGAATGCAA/TGGAATCCACCCGAGTGGAAatggaat CTGTATGCCACAAACGGCC/ATCCAGCATAGAGGAGATGGAAATACAAAATCTCTGCTC TTGATGTTCCGGGACTTCCACCAACTGAACCAACTAAACTACT	NA	NA	NA
hg19	79	14	91400188	+	1	22	39007199		NA	NA	NA	NA
hg19	79	22	39007199			14	91400188		NA	NA	NA	NA
hg19	80	5	174016924	+		11	104839238		AGTACTTCAAAGAAATGCACCATCCCTGTGGC/CTTTCTGACTAGGACTTGGGAG	1	0	NHEJ
hg19	80	11	104839238			5	174016924		NA	1	0	NHEJ
hg19	81	6	144168785	-		4	49100835		ATCAACCCGAGTGAATGGAAT/AAAACCCGCGCCGGTGTATAAGACTA	0	3	NHEJ
hg19	81	4	49100835			6	144168785		NA	0	3	NHEJ

II – 4/ Conséquences fonctionnelles des chromoanagenesis

Ce dernier chapitre de résultats vise à étudier les conséquences des remaniements chromosomiques sur le transcriptome. En étudiant particulièrement l'effet de position il contribue à répondre à la question posée en introduction :

- **Une fois identifié, par quel(s) mécanisme(s) un remaniement de structure peut-il être responsable d'un phénotype ?**

Nota bene : Une partie des analyses statistiques pertinentes pour une publication, est encore en cours au moment de la rédaction.

II – 4/ A- Introduction

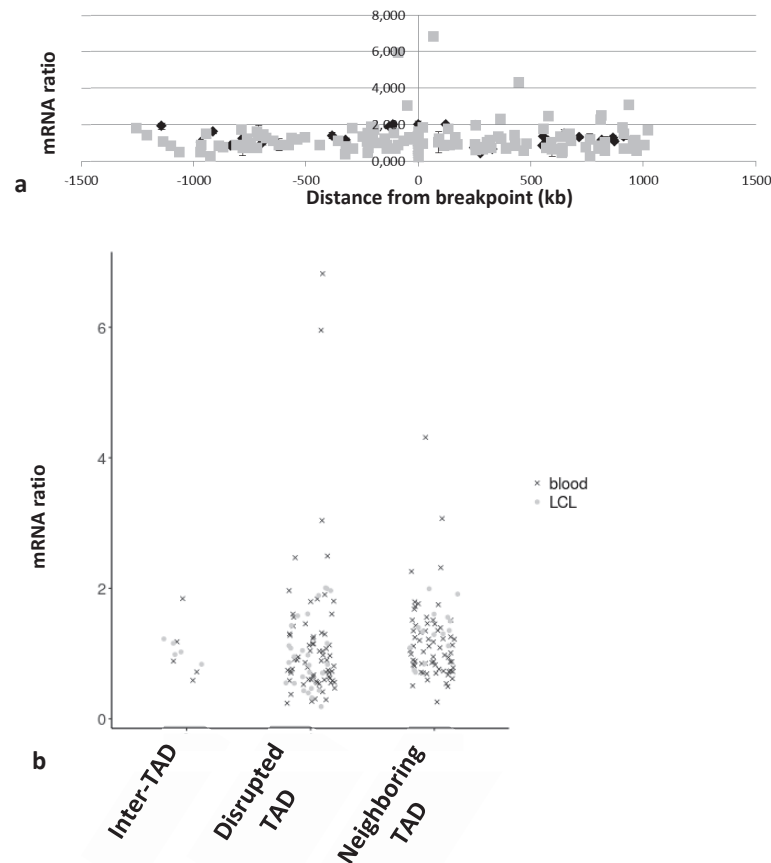
Les différentes hypothèses reliant un variant de structure chromosomique à un phénotype ont été décrits précédemment. Le cas d'un remaniement qui viendrait casser un gène sensible à l'haploinsuffisance connu en pathologie pour causer des phénotypes similaires à celui du patient étudié est très clair. L'interprétation est plus délicate lorsqu'aucun gène n'est interrompu ou qu'il n'y a pas de corrélation génotype/phénotype claire. En effet, démontrer un effet de position peut s'avérer particulièrement ardu avec plusieurs questions qui se posent.

- A quelle distance se situe(nt) le(s) gène(s) potentiellement concerné(s) par un effet de position ?
- Doit-on considérer la distance linéaire ou la distance tridimensionnelle entre le point de cassure et un gène ?

Nous avons réalisé un premier *screening* de ces questions lors de l'étude de validation du séquençage génome entier pour la caractérisation de remaniements chromosomiques chez des patients à phénotype anormal (Schluth-Bolard et al., 2019). Dans un intervalle de 1Mb de part et d'autre de chacun des points de cassure identifiés nous avons analysé par RT-qPCR le niveau d'expression des gènes. En explorant 189 gènes chez 47 patients *il n'avait pas été identifié de corrélation entre la distance au point de cassure et le niveau d'expression normalisé dans le sang* ($r=-0.187$, $p=0.097$) *ou dans les lignées lymphoblastoïdes* ($r=-0.0637$, $p=0.77$). *Il n'y*

avait pas non plus de corrélation entre ce niveau d'expression et la présence du gène dans le même TAD que le point de cassure, le TAD suivant ou l'espace inter-TAD ($p=0.7581$) (Figure 32) (Schluth-Bolard et al., 2019).

Figure 32 : Expression génique en fonction de la distance au point de cassure (A) ou de son appartenance au même TAD, TAD voisin ou espace interTAD (B). Figure supplémentaire de Schluth-Bolard et al. 2019)



Si une dérégulation locale ne semble donc pas évidente, nous avons souhaité savoir si une dérégulation plus globale du génome était observable et nous sommes pour cela appliqués à étudier le transcriptome complet des 7 cas de chromoanagenesis équilibrés caractérisés dans le chapitre précédent.

II – 4/ B- Matériel et Méthodes

Les caractéristiques des 7 cas de remaniements chromosomiques complexes équilibrés inclus dans cette étude sont rappelées dans le Tableau 3. Nous avons également inclus dans l'étude 8 cas de remaniements chromosomiques simples précédemment caractérisés (Schluth-Bolard et al., 2019) afin de quantifier une possible dérégulation génomique au-delà de la fenêtre de 2Mb

centrée sur le point de cassure précédemment étudiée. Pour chaque patient, l'ARN a été extrait à partir de 5 millions de cellules d'une lignée de cellules lymphoblastoïdes reprises dans une solution Trizol/chloroforme avec le kit RNeasy (Qiagen) selon les recommandations du fournisseur. Huit lignées lymphoblastoïdes contrôles (4 de chaque sexe) appariées sur l'âge et préparées dans les mêmes conditions ont été extraites selon le même protocole. Après contrôle de la qualité des acides nucléiques extraits (spectrophotométrie et analyse de fragment (RQN>9,5) 500 ng d'ARN étaient engagés dans la préparation de librairie. Celle-ci a été effectuée avec le kit TruSeq stranded mRNA (Illumina) (qui permet de ne pas capturer les ARN ribosomaux) avec des adaptateurs *Unique Dual* pour Illumina TruSeq (IDT). Les 23 librairies ont été poolés ensemble avant le séquençage réalisé en *single-end* 125 bases sur 3 *lanes* de *flowcell* pour séquenceur HiSeq4000 (Illumina) pour un total de 39 millions de lectures par échantillon en moyenne.

Après alignement des séquences par le logiciel BWA v0.7 (Li and Durbin, 2009) sur le génome GRCh38, le nombre de lectures par gène était reporté dans une matrice. A partir de la matrice de comptes par gènes, le reste de l'analyse s'est faite avec le logiciel R en rédigeant un script dédié. Contrairement à ce qui se fait habituellement nous n'avons pas supprimé de l'analyse les gènes pour lesquels il n'y avait pas de lectures pour un ou plusieurs échantillons. En effet, une des conséquences possibles des remaniements chromosomiques est l'expression aberrante d'un gène dans un tissu où il n'est habituellement pas exprimé. Dans ce cas, le transcrit n'aurait été lu que chez cet individu.

Après normalisation des données par la méthode *rld* (transformation logarithmique) pour mieux prendre en compte les valeurs extrêmes de profondeur de séquençage dans l'analyse statistique, une analyse de cluster a été réalisée pour observer une éventuelle dérégulation de l'expression génique à l'échelle du transcriptome entier. De plus, nous avons mené une analyse en composante principale pour identifier le(s) facteur(s) permettant de distinguer les échantillons les uns des autres.

Dans un second temps, nous avons comparé le nombre de lectures obtenues pour chaque transcrit pour un individu à la moyenne des valeurs observées chez les contrôles : *fold-change*. Nous avons également calculé un deuxième *fold-change* en utilisant cette fois que les 4 contrôles du même sexe que l'individu testé. Considérant qu'aucun patient n'est porteur du même remaniement il n'est pas possible de réaliser de tests statistiques plus en avant sur la significativité de ce ratio.

De manière ciblée, nous avons cherché à quantifier un éventuel « effet de position » dans la même fenêtre de 2Mb que celle étudiée précédemment. Pour cela, nous avons considéré l'ensemble des transcrits chevauchant au moins partiellement cet intervalle. Ayant inclus des remaniements particulièrement complexes nous avons cependant dû filtrer certains loci pour lesquels les points de cassure étaient distants de moins d'1Mb puisqu'il n'était alors pas possible de déterminer lequel des points de cassure était responsable de l'éventuelle dérégulation génique observée.

Enfin, à l'échelle individuelle, nous avons établi la liste des gènes manifestement dérégulés pour chaque patient avec pour objectif d'identifier un possible gène candidat responsable du phénotype du patient. Nous avons considéré comme dérégulés les gènes pour lesquels le ratio patient/contrôle calculé était inférieur à -0,5 ou supérieur à 0,38, correspondant à une déviation de +/- 30% du niveau d'expression.

Tableau 3 (page suivante): Liste des patients analysés par RNA-Seq sur lignées lymphoblastoïdes. CCR : complex chromosomal rearrangement ; ID : intellectual disability ; ASD : atrial septal defect ; VSD : ventricular septal defect. Données et identifiants tirés de (Schluth-Bolard et al., 2019). Les 6 patients porteurs d'un chromoanagenesis sans déséquilibre génomique ont été inclus dans cette étude. Pour ces patients, l'identifiant utilisé dans l'article 3 (chapitre précédent) est rapporté dans ce tableau (6 premières lignes)

Patient	Sex	Rearrangement	Karyotype	Inheritance	Phenotype
IL/1901 Patient 1	F	CCR	46,XX,inv(3)(p13p22),inv(3)(p12q26.3)	familial	Mild ID, speech delay, cleft palate, single palmar crease, strabismus. Translocation transmitted to her daughter, with the same phenotype.
EB/0401 Patient 5	M	Insertion	46,XY,ins(5)(q15q23.3q34)	<i>de novo</i>	IUGR, post-natal growth retardation, microcephaly, ID, sleeping and feedings difficulties, epilepsy (hyperthermic seizures), choreic and dystonic abnormal movements, strabismus
MD/2203 Patient 14	F	CCR	46,XX,t(1;3;5)(p21;p24;q734)	<i>de novo</i>	Severe ID, absent speech, autistic spectrum disorder, epilepsy, facial dysmorphism, constipation
NM/0201 Patient 15	M	Translocation	46,XY,t(1;2)(p13.2;q31.2)	<i>de novo</i>	ID, speech delay, hypotonia, anxiety
KT/1403 Patient 16	M	CCR	46,XY,t(8;22;13)(q12;q12;q31.1),t(14;15)(q21;q24)	<i>de novo</i>	Macrocephaly, ID, psychomotor delay, speech delay, left club-foot, café-au-lait spot
TS/1101 Patient 18	M	CCR	46,XY,t(3;4;10;14;18)(p23;q24;p13;q32;q23)	<i>de novo</i>	Mild ID, behavioral troubles, facial dysmorphism
JB/0102	M	CCR	46,XY,t(6;8;9;13)(q26;p23;p21;q21)	<i>de novo</i>	Mild ID, thin motor skills troubles, autistic spectrum disorder (lonely, irritable), repeated otitis media
CS/1002	F	Insertion	46,XX,ins(15;12)(q15 or q21;q24.1q24.2)	<i>de novo</i>	Relative macrocephaly (OFC +2SD), moderate ID, facial dysmorphism, café-au-lait spots
CL/2001	F	Translocation	46,XX,t(11;15)(p14;q14)	<i>de novo</i>	Facial dysmorphism, microcephaly, ID, absent speech, abnormal corpus callosum, ASD, VSD, epilepsy growth delay
MD/0104	F	Translocation	46,X,t(X;7)(p11.1;q36.1)	<i>de novo</i>	Moderate ID, hyperactivity attention deficit disorder, pharmaco-resistant epilepsy, ataxia, auto-immune hepatitis
CG/0106	F	CCR	46,X,t(X;13;15)(p21;q22;q22),t(6;11)(q21;q24)	<i>de novo</i>	Severe ID, absent speech, attention deficit, Duchenne muscular dystrophy
AD/1601	M	Translocation	46,XY,t(9;17)(q33;q22)	maternal	Spondylo-epiphyseal dysplasia, acro-osteolysis
LD/0108	F	Translocation	46,X,t(X;1)(p12;p36.1)	<i>de novo</i>	ID, facial dysmorphism, axial hypotonia, dyschromia Blaschko line, strabismus
OL/2202	F	Translocation	46,XX,t(1;14)(q32;q22)	<i>de novo</i>	Severe ID, absent speech, sterotypy, epileptic encephalopathy, constipation, myopia
CF/2304	M	CCR	46,XY,t(2;8;10;15)(q32.3;q26;q21.3;q23)	<i>de novo</i>	Mild ID, speech delay, major atopic dermatitis

II – 4/ C- Résultats

L'analyse de *clustering* à l'échelle transcriptomique n'a pas mis en évidence de regroupement significatif d'échantillons entre eux. Les remaniements complexes, simples ou les contrôles ne sont pas réunis (Figure 33). Il n'apparaît donc pas de dérégulation massive du génome en cas de remaniement chromosomique aussi complexe soit-il.

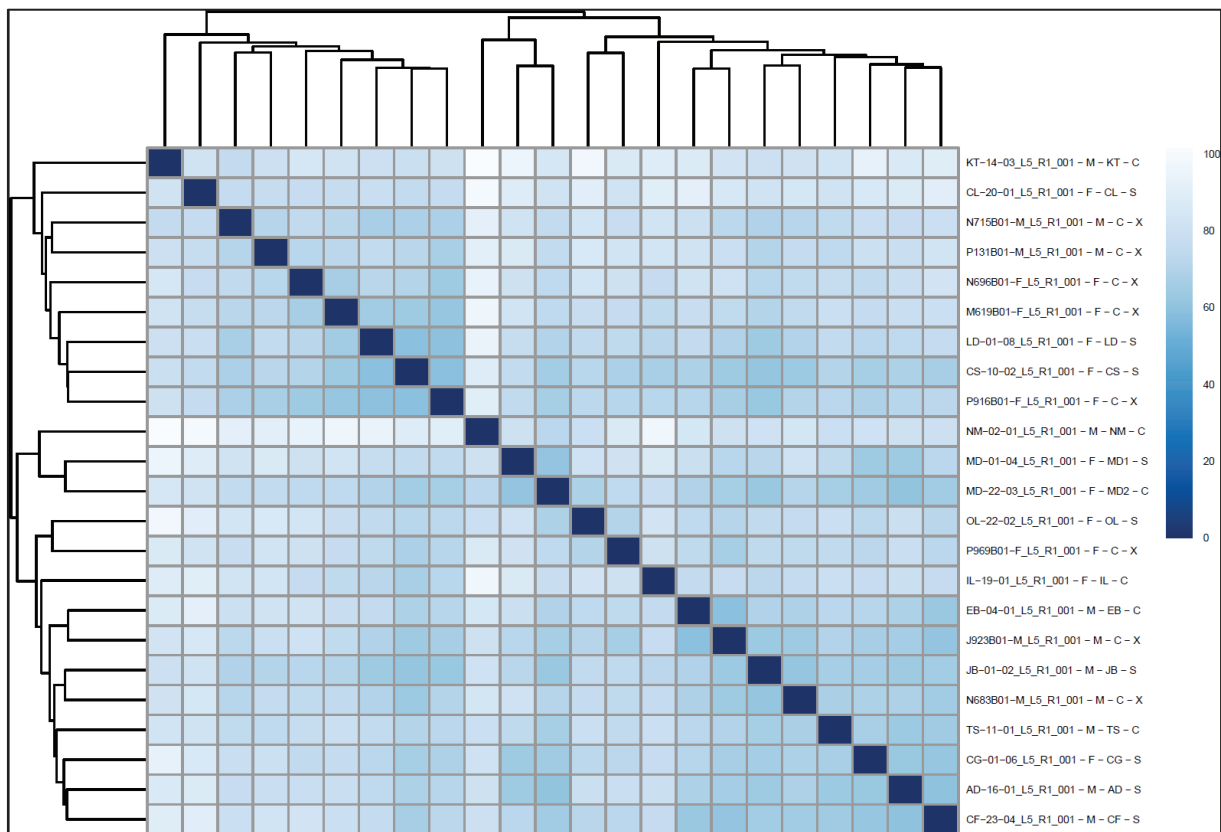


Figure 33 : Heatmap des distances inter-échantillons. Le dernier caractère du code échantillon désigne les échantillons contrôles (X), les remaniements simples (S) et les remaniements complexes (C).

Lors de l'analyse en composante principale le seul élément qui est ressorti comme discriminant les échantillons est le sexe (Figure 34). Aussi, dans les analyses réalisées en aval nous avons systématiquement réalisés les analyses en parallèle utilisant comme groupe de référence tantôt tous les contrôles ensemble (n=8), tantôt uniquement les contrôles appariés sur le sexe de l'échantillon analysé (n=4).

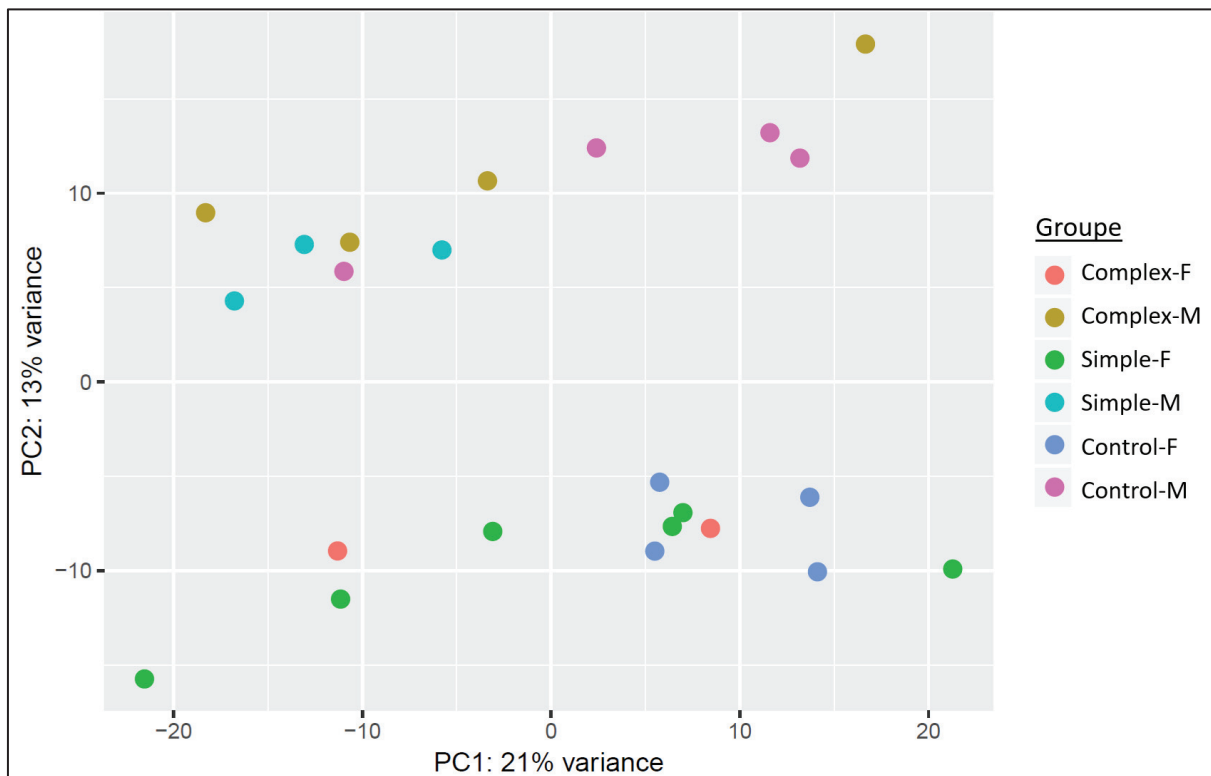


Figure 34 : Analyse en composante principale des distances inter-échantillons. Les groupes de sexe sont séparés selon la deuxième composante qui explique 13% de la variance observée.

Nous n'avons pas observé de corrélation entre la distance au point de cassure et la probabilité de dérégulation (Figure 35).

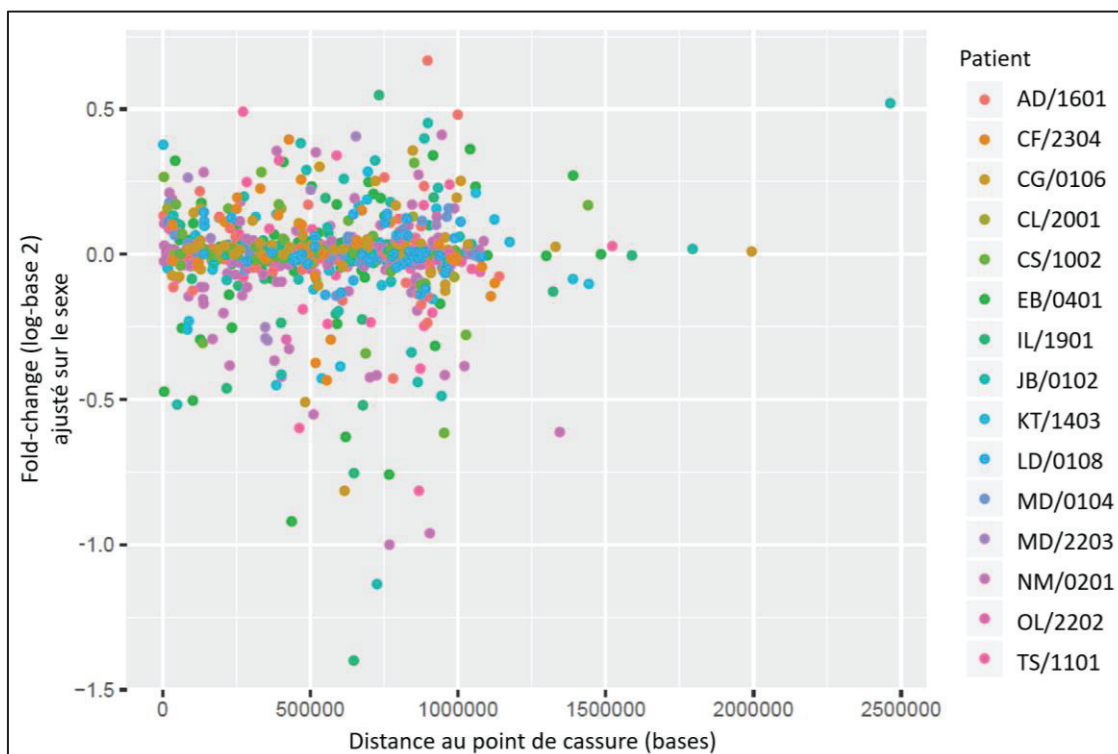
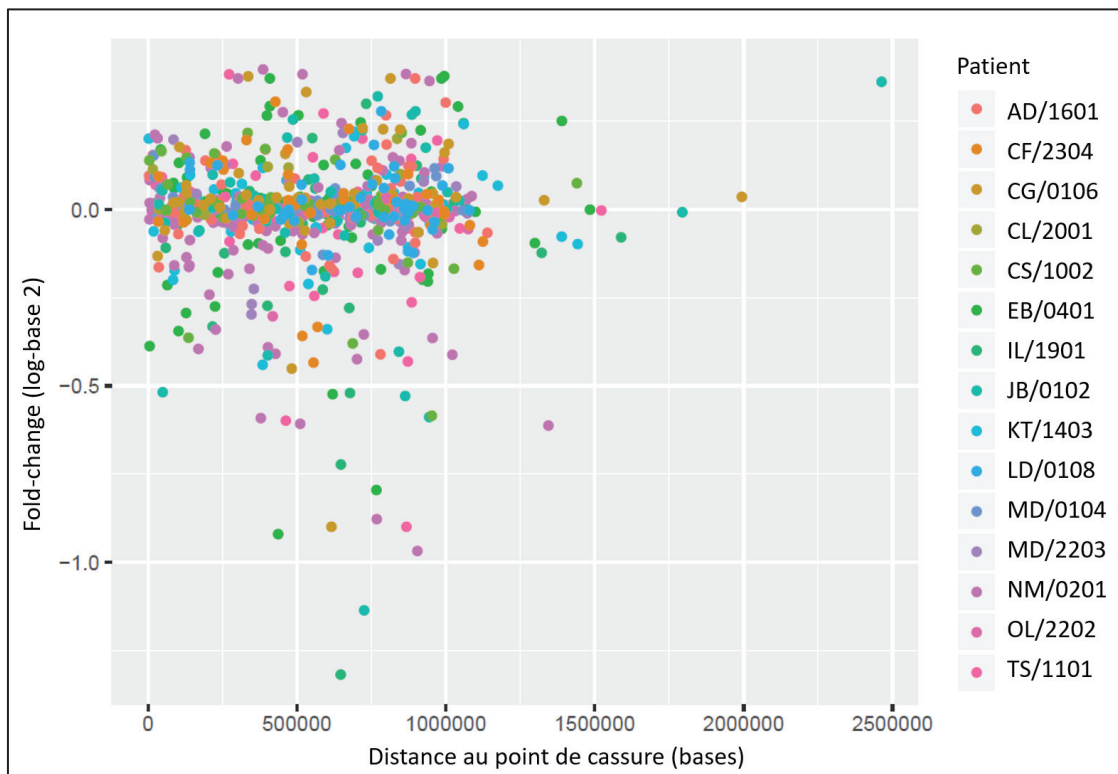


Figure 35 : Degré de dérégulation génique, exprimé en fold-change en fonction de la distance linéaire à un point de cassure chromosomique. Le fold-change est calculé sur tous les contrôles (partie supérieure) ou uniquement sur les contrôles de même sexe (partie inférieure). La distance est mesurée depuis le site d'initiation de la transcription. Seuls les gènes chevauchant l'intervalle +/-1Mb autour du point de cassure sont considérés.

Nous ne retrouvons pas de différence significative entre le degré de dérégulation des gènes directement interrompus par rapport à ceux de l'intervalle de 1Mb ou du reste du génome.

Neuf des 15 patients inclus dans cette étude disposaient déjà d'un diagnostic avant l'étude des données de transcriptome suite à la caractérisation de leur anomalie chromosomique par séquençage de génome. Dans aucun des cas, le gène considéré comme causal n'a pu être directement identifié comme dérégulé lors de l'analyse RNA-Seq (Tableau 4). Neuf des 15 patients inclus dans cette étude disposaient déjà d'un diagnostic avant l'étude des données de transcriptome suite à la caractérisation de leur anomalie chromosomique par séquençage de génome. Dans aucun des cas, le gène considéré comme causal n'a pu être directement identifié comme dérégulé lors de l'analyse RNA-Seq (Tableau 5).

Tableau 4 : Valeur de fold-change (log-base2) pour les différents gènes retenus comme responsables de tout ou partie du phénotype des patients suite à la caractérisation de leur remaniement chromosomique par séquençage de génome.

Patient	Gène retenu pathogène avant RNA-Seq	Fold-change (tous contrôles)	Fold-change (contrôles de même sexe)	Ratio RT-qPCR (déviaton standard)	Matériel utilisé pour RT-qPCR
EB/0401	<i>MEF2C</i>	-0,002	-0,010	1,075 (0,059)	Sang
MD/2203	<i>MEF2C</i>	-0,005	0,003	0,821 (0,131)	LCL
OL/2202	<i>MEF2C</i>	-0,014	-0,006	0,375 (0,0346)	Sang
CL/2001	<i>MEIS2</i>	0,114	0,158	0,598 (0,116)	Sang
CS/1002	<i>SPRED1</i>	-0.003	-0.04	1,420 (0,138)	LCL
CG/0106	<i>DMD</i>	0.036	0.01	1,579 (0,116)	LCL
CF/2304	<i>SATB2</i>	-0,434	-0,434	NA	NA
LD/0108	<i>CAMTA1</i>	-0,016	-0,026	NA	NA
IL/1901	<i>FOXP1</i>	-0.08	-0.004	NA	NA

La liste des gènes présents dans l'intervalle de 2Mb centré sur le point de cassure et identifiés comme dérégulés, *fold change* inférieur à -0,5 ou supérieur à 0,38, est présentée dans le **Erreur ! Source du renvoi introuvable.** Pour les patients MD/0104 et KT/1403 qui ne disposaient pas de diagnostic aucun gène n'apparaît dérégulé. Pour les patients TS/1101, NM/0201, AD/1601 et JB/0102 il n'apparaît pas de gène dérégulé clairement pathogène. Le seul gène dérégulé chez le patient AD/1601 est associé à un phénotype féminin chez ce sujet masculin. Les deux gènes dérégulés chez le patient TS/1101 ne sont aujourd'hui pas associés à un phénotype pathologique et aucun d'entre eux n'est mesuré comme sensible à

l'haploinsuffisance par le score pLI. Pour le patient JB/0102 le seul gène sensible à l'haploinsuffisance à être sous-exprimé est le gène *ELAVL2*. Le niveau d'expression dans les lignées lymphoblastoïdes est cependant très faible et ce résultat est soumis à caution. Ce gène, exprimé quasi exclusivement dans le cerveau code une protéine de fonction méconnue capable de se lier à l'ARN (King et al., 1994). Il apparaît comme un candidat pour le diagnostic de ce patient présentant une déficience intellectuelle légère.

Enfin, le patient NM/0201 présente une sous-expression du gène *AFF3*. Ce gène est cependant lui aussi très peu exprimé dans le sang et ce résultat semble devoir être soumis à réserve. Des variants gain de fonction ont récemment été rapportés responsables d'une encéphalopathie épileptique syndromique (KINSSHIP syndrome : KI for horseshoe Kidney, NS for Nievergelt/Savarirayan type of mesomelic dysplasia, S for Seizures, H for Hypertrichosis, I for Intellectual disability and P for Pulmonary involvement) (Voisin et al., 2019).

II – 4/ D- Discussion

Il s'agit de la première étude du transcriptome de *chromoanagenesis*. Celle-ci n'a pas mis en évidence de dérégulation massive de l'expression génique. Ces 7 remaniements constitutionnels équilibrés ont été observés chez des patients vivants et il est possible qu'une certaine pression de sélection participe au résultat observé. Une dérégulation massive du génome serait trop défavorable pour être viable.

S'agissant de l'effet de position, nos résultats mettent également en avant l'absence de corrélation entre dérégulation génique et distance linéaire au point de cassure. Considérant les différents niveaux de repliement de la chromatine il semble cependant pertinent de reproduire cette analyse en considérant cette fois une distance tridimensionnelle dans le noyau interphasique. Pour cela nous allons considérer l'appartenance des gènes avoisinants un point de cassure au même TAD ou à l'espace inter-TAD. La redondance entre les éléments *enhancers* explique probablement en partie cette relative résistance aux remaniements de structure (Will et al., 2017).

Au niveau individuel, l'analyse ciblée du transcriptome avoisinant le point de cassure de remaniements non récurrents ne suffit pas à identifier un gène participant au phénotype. Un seul gène candidat, *AFF3*, peut être discuté. Néanmoins, dans la mesure où seules des variants

gain de fonction ont été décrits à ce jour, considérant l'haploinsuffisance supposé ici, il ne peut s'agir à ce stade que d'un candidat pour le diagnostic de ce patient qui ne présente pas d'éléments de ce syndrome. Nous prévoyons d'étendre notre analyse en dehors des loci de points de cassure pour identifier des gènes dérégulés à distance et potentiellement responsables du phénotype. Cette dérégulation à distance pourrait être l'aboutissement d'une cascade impliquant initialement un transcrit au point de cassure dont une dérégulation très faible non détectée suffit à aboutir à une dérégulation plus importante en aval. Une association fortuite reste également possible.

Chez ces patients présentant principalement des troubles du neuro-développement, l'utilisation de lignées lymphoblastoïdes pour étudier l'expression de gènes responsables du phénotype est une limite de notre étude.

Aussi la stratégie évoquée en introduction générale combinant une digitalisation des données cliniques et l'utilisation de données d'organisation chromatinienne de divers types cellulaires (Middelkamp et al., 2019) combine les avantages de coût et de rendement diagnostique par rapport au RNA-Seq dans cette situation.

Tableau 5 : Liste des gènes identifiés comme dérégulés par RNA-Seq.

Patient	Coordonnées du point de cassure	Gène	Distance au point de cassure (kb)	Fold-change (tous contrôles)	Fold-change (contrôles de même sexe)	gnomAD pLi	Phénotype OMIM (Transmission)	Diagnostic posé avant RNA-Seq	Commentaire
IL/ 1901	chr3:68220850	<i>FAM1944</i>	732	0,30	0,55	0,45	NA	<i>FOXP1</i>	Fold-change pour <i>FOXP1</i> : -0.08 (tous contrôles) et -0.004 (contrôles de même sexe)
	chr3:71107266	<i>EIF4E3</i>	648	-0,72	-0,75	0,02	NA		
	chr3:71107266	<i>PROK2</i>	678	-0,52	-0,52	0,01	Hypogonadotropie hypogonadisme 4 (AD)		
	chr3:71107266	<i>GPR27</i>	647	-1,32	-1,40	0,35	NA		
CS/ 1002	chr12:115361274	<i>TBX5</i>	953	-0,58	-0,62	1	Holt-Oram syndrome (AD)	<i>SPRED1</i>	Pas de signe de syndrome de Holt-Oram. Expression faible de <i>SPRED1</i> dans les lignées lymphoblastoïdes. Fold-change pour <i>SPRED1</i> : -0.003 (tous contrôles) et -0.04 (même sexe)
CG/ 0106	chr11:113443829	<i>ZBTB16</i>	616	-0,90	-0,81	1	NA	<i>DMD</i>	Expression très faible de <i>DMD</i> dans les lignées lymphoblastoïdes
		<i>NCAMI</i>	483	-0,45	-0,51	1	NA		
EB/ 0401	chr5:91148960	<i>ADGRV1</i>	620	-0,52	-0,63	0	?Febrile seizures, familial, 4 ; Usher syndrome (AR)	<i>MEF2C</i>	<i>MEF2C</i> bien exprimé. Fold-change = -0.002 (tous contrôles)

	chr5:166518112	TENM2	767	-0,79	-0,76	1	NA		et FC=-0.01 (contrôles de même sexe)
	chr5:177001595	HK3	102	-0,34	-0,50	0	NA		
TS/	chr3:31519270	ZNF860	463	-0,60	-0,60	NA	NA	-	
1101	chr14:86397652	FLRT2	868	-0,90	-0,81	0,82	NA		
NM/	chr2:101047114	AFF3	904	-0,97	-0,96	1	KINSSHIP	-	Mutations gain de fonction de <i>AFF3</i> responsables de encéphalopathie épileptique syndromique. Pas de variant perte de fonction rapporté.
0201	chr2:155216319	GALNT13	1344	-0,61	-0,61	0	NA		
	chr2:215993961	MARCH4	379	-0,59	-0,37	0,13	NA		
	chr2:238369580	RAB17	768	-0,88	-1,00	0,07	NA		
	chr2:238369580	RAMP1	511	-0,61	-0,55	0,13	NA		
AD/	chr17:47730066	WNT9B	897	0,37	0,67	0,99	Mayer-Rokitansky-Küster-Hauser syndrome	-	Patient masculin
1601	chr8:2531693	CSMD1	2463	0,36	0,52	1	NA		
JB/	chr9: 23875496	ELAVL2	49	-0,52	-0,52	1	NA		Pas de signe d'amélogénèse imparfaite.
0102	chr11:103569278	MMP20	944	-0,59	-0,49	0	Amelogenesis imperfecta, type IIA2 (AR)		
		MMP27	863	-0,53	-0,44	0	NA		
		MMP3	726	-1,14	-1,14	0	Association with coronary heart disease		

III - Discussion

III – 1/ Complexité des remaniements chromosomiques complexes

Notre travail sur la plus grande cohorte de chromoanagenesis constitutionnels a permis de donner une image plus globale de ces remaniements précédemment rapportées dans de petites séries ou individuellement (Kloosterman et al., 2011b; Masset et al., 2016; Redin et al., 2017). Il a permis de montrer la difficulté à se conformer à des définitions de la littérature s'agissant de sous-familles de chromoanagenesis. En effet, si le profil de gains ou de pertes peut contribuer à distinguer *chromothripsis* et *chromoanasythesis*, une association de mécanismes est très fréquemment observée avec finalement peu de remaniements étant purement « NHEJ » et d'autres purement « réplcatifs ». Ajouté à ça le caractère relativement arbitraire des définitions utilisées pour ces mécanismes et la difficulté à inférer certaines jonctions, ce critère semble finalement délicat d'utilisation.

Alors que la littérature définit deux cadres bien distincts pour le *chromothripsis* (pulvérisation chromosomique suivi d'un recollement par NHEJ) et le *chromoanasythesis* (emballage d'un processus réplcatif de type MMBIR/FoSTeS) nous décrivons ici une zone grise avec des situations où des mécanismes réplcatifs viennent « prendre en charge » un chromosome paraissant avoir subi une pulvérisation.

Pour révéler les mécanismes à l'origine des *chromoanagenesis* à partir de ces données de patients nous avons étudié la répartition de leurs points de cassure. Pour être le plus robuste possible nous avons élargi notre analyse à 20 autres cas de *chromoanagenesis* constitutionnels, ainsi qu'à la plus grande base de données de *chromoanagenesis* tumoraux (Yang et al., 2015). Deux simulations de points de cassure suivant des règles différentes nous ont permis d'interpréter ces résultats par rapport à une hypothèse nulle de répartition aléatoire. Enfin, nous avons étudié les points de cassure de 234 remaniements simples pour élucider l'existence d'un continuum liant remaniements simples et complexes.

Cette analyse a révélé un enrichissement des points de cassure des remaniements chromosomiques simples et complexes dans les régions de la chromatine à réplcation tardive. Il s'agit là du premier élément obtenu en dehors d'un modèle cellulaire appuyant l'hypothèse

d'une condensation prématurée d'un chromosome (PCC : *Premature Chromosome Condensation*) nous permettant de formuler un scénario des origines de ces remaniements.

Pour la majorité des cas, les points de cassure affectent très majoritairement ou exclusivement un chromosome. Celui-ci aurait été séquestré dans un micronoyau suite à un défaut de mise en place du fuseau de division. Ce micronoyau est le siège d'une réplication déficiente, retardée. En fin de phase G2, alors que la réplication est terminée dans le noyau principal, ce n'est pas le cas dans le micronoyau notamment au niveau des régions de réplication physiologiquement tardives. L'asynchronie entre les deux compartiments entraîne une condensation de la chromatine comprise dans le micronoyau. Cette condensation exerce un stress sur la fibre chromatinienne qui bloque les fourches de réplication encore actives. Si ces fourches sont complètement déconnectées, une cassure double brin va apparaître. C'est le mauvais rattachement des deux extrémités qui explique le remaniement visible *in fine*. Sinon, la fourche de réplication va pouvoir se déplacer vers un autre locus et démarrer un remaniement de structure. Ces déplacements de fourche de réplication peuvent même dans ce cas utiliser les chromosomes du noyau principal, la membrane du micronoyau ayant disparue. Certaines jonctions observées ne résultent alors pas *stricto sensu* de cassures double brin. Ces deux possibilités expliquent l'observation conjointe assez fréquente de mécanismes de réparation de type réplicatifs avec des mécanismes *non-homologous end-joining*.

Les *chromoanagenesis* constitutionnels et tumoraux présentent le même profil de distribution de points de cassure. Ceci nous permet de penser que les mécanismes en jeu sont les mêmes dans les deux cas. La littérature est aujourd'hui centrée sur les remaniements tumoraux avec des études réalisées sur des lignées cellulaires cancéreuses dont le génome est clairement hors de contrôle. Pour continuer à étudier ces phénomènes, aux vues de nos résultats, il semble pertinent de s'intéresser aux remaniements constitutionnels.

Si l'étude des divisions méiotiques semble difficile, s'intéresser aux premières divisions zygotiques paraît plus réaliste. Les premières divisions zygotiques sont connus pour être associées à de fréquentes anomalies chromosomiques (Vanneste et al., 2009). Si celles-ci sont en partie éliminées pour permettre à un embryon euploïde de se développer, certaines sont probablement incompatibles avec le développement embryonnaire. Dans nos données, cette pression de sélection est visible avec une déplétion de points de cassure dans les gènes en analyse univariée. Le suivi des premières divisions cellulaires d'un embryon obtenu par fécondation *in vitro* est aujourd'hui réalisé en quasi-continu avec des systèmes de photographies

timelapse et sert de critère pour choisir l'embryon à transférer lors de ces procédures. Une piste d'étude pourrait être d'utiliser ces images pour identifier les zygotes qui ont connu la formation d'un micronoyau au cours des toutes premières divisions pour ensuite les séquencer et confirmer l'association micronoyau/*chromoanagenesis* (Daughtry and Chavez, 2018). Cette approche a été utilisée pour étudier les premières divisions cellulaires d'embryons de primates. L'étude de la ségrégation des chromosomes et de la cinétique d'éventuelles micronoyau a également été réalisée sur des embryons de souris (Vázquez-Diez and FitzHarris, 2018).

Alternativement, des procédés exogènes pourraient être utilisés pour favoriser la formation de micronoyaux (irradiation) dans des lignées cellulaires (Terzoudi et al., 2015). Hors de l'hypothèse du micronoyau, il serait également intéressant de pouvoir provoquer un asynchronisme de réplication entre territoires chromosomiques d'un même noyau. Si notre hypothèse est juste, l'absence de barrière physique entre territoires chromosomiques devraient autoriser des remaniements impliquant un plus grand nombre de chromosomes. La fusion d'une cellule en interphase avec une cellule en mitose, phénomène très éloigné des remaniements constitutionnels, a déjà montré que les chromosomes alors en interphase étaient pulvérisés, cette démonstration étant à l'origine de l'hypothèse de condensation prématurée (Sperling and Rao, 1974). L'outil permettant une condensation prématurée localisée d'un nombre limité de chromosomes hors d'un micronoyau reste à inventer.

Un autre élément freinant l'observation de remaniements impliquant plusieurs chromosomes n'est étonnamment pas discuté dans la littérature. En admettant que plusieurs chromosomes soient réarrangés entre eux dans un évènement « catastrophique », outre le risque grandissant d'anomalies déséquilibrées, s'ensuivra la formation de structure très complexes lors de la division cellulaire suivante. Là où 4 chromosomes viennent former un tétravalent de pachytène lors d'une translocation réciproque de très nombreux chromosomes pourraient se retrouver imbriquer lors d'un remaniement chromosomique aussi complexe. Les multiples ségrégations alors possibles sont dès lors nombreuses à être à risque de déséquilibre non compatible avec le développement cellulaire. Si un réarrangement complexe affectant plusieurs territoires chromosomiques venait néanmoins à passer la phase M, lors de l'interphase suivante la remise en place de la conformation tridimensionnelle du génome serait également compliquée par le remaniement. Pour maintenir l'expression attendue des gènes il faut pouvoir les maintenir dans leur état de condensation d'origine (hétérochromatine, euchromatine) alors même que deux fragments possiblement opposés se retrouvent maintenant joints. Au total, la pression de

sélection rendant déjà rares les *chromoanagenesis* affectant un seul chromosome rendrait exceptionnels les *chromoanagenesis* en affectant plusieurs. Ici, l'étude très précoce d'embryons pourrait également être contributive, et ceci indépendamment de la visualisation d'un micronoyau.

La détection de variants de structure sur cellule embryonnaire unique est cependant difficile. Le matériel de départ, en très faible quantité, oblige un passage par une étape d'amplification génome entier qui peut potentiellement introduire des biais. Si la détection de variations du nombre de copie est aujourd'hui maîtrisée dans le cadre des procédures de dépistages pré-implantatoires par étude de la profondeur de séquençage (Parikh et al., 2018), la détection de variants de structure équilibrée est également possible (Voet et al., 2013).

III – 2/ Remaniements simples : vraiment ?

On savait déjà que nombre de remaniements chromosomiques apparemment simples au caryotype étaient en fait beaucoup plus complexes (Chiang et al., 2012).

Pour confirmer ou infirmer l'hypothèse d'un continuum mécanistique entre remaniements simples et complexes, nous avons également étudié la répartition des points de cassure de plus de 200 remaniements simples vrais (après caractérisation par séquençage). L'utilisation d'un modèle de régression logistique nous a permis de nous affranchir des conséquences de la pression de sélection a priori différentes entre un remaniement simple et un remaniement complexe et biaisant forcément la distribution des points de cassure par rapport à une réparation aléatoire. En identifiant une différence dans la répartition des points de cassure avec une diminution de moitié des points de cassure de remaniements simples dans les domaines associées à la lamine nous infirmons l'hypothèse d'un continuum mécanistique.

L'indépendance de la distribution des points de cassure de *chromoanagenesis* vis-à-vis des LADs vient corroborer l'hypothèse du micronoyau, ces derniers en étant dépourvus. En revanche, cette déplétion de points de cassure de remaniements simples dans les LADs est à ce stade inexplicée et inattendue. En effet, les cassures chromosomiques observées chez des patients vivants devraient être, intuitivement, plutôt repoussées vers l'hétérochromatine la moins « nécessaire » au bon fonctionnement du génome et donc vers les LADs du fait d'une certaine pression de sélection.

L'observation faite ici est donc inattendue et reste à expliquer. L'organisation nucléaire est connue pour être un paramètre clé dans la formation des remaniements de structure chromosomiques (Roukos and Misteli, 2014). Les translocations réciproques et les inversions nécessitent la survenue de deux cassures double brins suivis de la mobilisation des extrémités libres jusqu'à l'appariement avec un nouveau partenaire. Si on ignore si cette mobilisation se fait de manière active ou passive, elle est dans tous les cas assez restreinte de telle sorte que ces remaniements impliquent plus fréquemment des territoires chromosomiques proches (Soutoglou et al., 2007). Il a pu être montré que les cassures survenant dans les LADs étaient moins rapidement détectés et moins souvent réparés par recombinaison homologue (Lemaître et al., 2014). De plus ces extrémités libres ne sont pas mobilisées vers des environnements de réparation plus favorables et sont donc plutôt réparés par NHEJ. Etant incapable de détecter *a posteriori* les cassures réparées par recombinaison homologue allélique il était donc attendu d'observer plutôt un enrichissement des points de cassure de remaniements simples dans les LADs, là encore à l'inverse de nos résultats.

Une cassure survenant dans un LAD est mécaniquement en contact avec un nombre réduits de « partenaires » chromosomiques potentiels. De plus, il est possible que sa mobilisation soit plus restreinte qu'ailleurs dans le noyau par l'interaction existante avec le réseau de lamines. Ainsi les cassures doubles brins survenant dans les LADs si elles sont plus souvent réparées par NHEJ, le seraient entre les partenaires natifs de la cassure double brin et ne formeraient donc pas de remaniement chromosomique visible mais laisserait uniquement une cicatrice à l'échelle nucléotidique. Cette hypothèse doit évidemment être investiguée plus en avant.

L'étude de la mobilité des cassures localisées dans les LADs semble une première étape. Une piste pourrait être de modifier la technologie DamID utilisé pour identifier les domaines associées à la lamine (Guelen et al., 2008). Il s'agirait de remplacer la méthyltransférase utilisée pour marquer les LADs par une nucléase capable de générer des cassures dans la chromatine en contact et insérer un marqueur fluorescent pour suivre la mobilité de la cassure dans le noyau dans le temps (Qian et al., 1991). Alternativement, l'emploi de la technologie CRISPR-Cas9 ciblant une séquence située dans un LAD constitutif permettrait de sonder quelques loci bien définis.

III – 3/ Nouveaux outils : nouvelles questions !

Passée la question de la détection des variants, vient la problématique de l'interprétation. La technologie n'apparaît plus comme une limitation mais ce sont les connaissances sur la biologie de notre génome qui font maintenant défaut. Si l'interruption par un point de cassure d'un gène sensible à l'haploinsuffisance déjà impliqué en pathologie dans des phénotypes similaires à celui du patient analysé ne pose que peu de problème, la situation est plus complexe si c'est une région intergénique qui est concernée. La corrélation génotype/phénotype peut permettre de rapidement s'orienter vers un gène voisin mais un phénotype peu spécifique telle que la déficience intellectuelle isolée pose plus de questions. Notre travail sur le séquençage RNA-Seq de patients porteurs de remaniements chromosomiques n'a pas mis en évidence de dérégulation massive du transcriptome en cas de remaniement chromosomique complexe. De plus, il n'est pas apparu de corrélation entre la probabilité de dérégulation d'un gène et la distance entre son site d'initiation de la transcription et le point de cassure.

Ce résultat négatif, même s'il peut en partie être le fruit du design de l'étude et du matériel biologique utilisé, n'en reste pas moins pertinent. En effet, il confirme l'importance de la corrélation génotype/phénotype pour identifier un (ou plusieurs) gène(s) responsable(s) du phénotype. Très récemment la déconnexion entre modifications de la structure chromatinienne et dérégulation génique a pu être montrée chez la drosophile. En utilisant des drosophiles porteurs de remaniements chromosomiques complexes hétérozygotes, quelques gènes ont effectivement présenté une dérégulation mais une large majorité était résistante à la perturbation des TADs (Ghavi-Helm et al., 2019). Ce(s) mécanisme(s) de résistance, ou plutôt le(s) autre(s) facteur(s) organisant l'interaction entre un *enhancer* et son gène, restent à découvrir mais il semble possible que les analyses encore en cours sur nos données relatives à l'appartenance ou non à un TAD montrent la même absence de corrélation.

En l'état actuel des connaissances et hormis quelques loci bien caractérisés (*SOX9*, *IHH*, *EPHA4*) plusieurs éléments doivent donc probablement être pris en compte pour parvenir à pondérer l'importance des TADs dans l'interprétation des variants de structure.

Le premier élément réside dans la définition même d'un TAD, intervalle chromosomique au sein duquel la chromatine est statistiquement plus en contact que ne le voudrait une proximité linéaire. Les TADs sont des unités d'organisation physiques de notre génome et pas d'interactions fonctionnelles. L'image prise par certains auteurs est celle de deux

individus qui font tous les jours le même trajet en train. S'ils sont statistiquement présents aux mêmes endroits aux mêmes moments, rien ne permet d'affirmer qu'ils interagissent. Récemment, Delaneau et al. ont décrit les *cis-regulatory domains* (CRDs) à partir de données transcriptomiques et de l'étude des modifications d'histones (Delaneau et al., 2019). Ces domaines sont relativement similaires aux TADs confirmant l'importance de la proximité physique pour l'interaction fonctionnelle. Cependant, le chevauchement est loin d'être complet, certains TADs ne recouvrant aucun CRD (Figure 36). Les CRDs sont aussi plus petits (médiane = 32kb vs 185kb pour les TADs). Après l'analyse des TADs, nous prévoyons d'étudier nos données RNA-Seq en utilisant ces CRDs, avec l'objectif de « quantifier » l'effet de position en moyenne dans le génome et ainsi aider à l'interprétation de variants de structure.

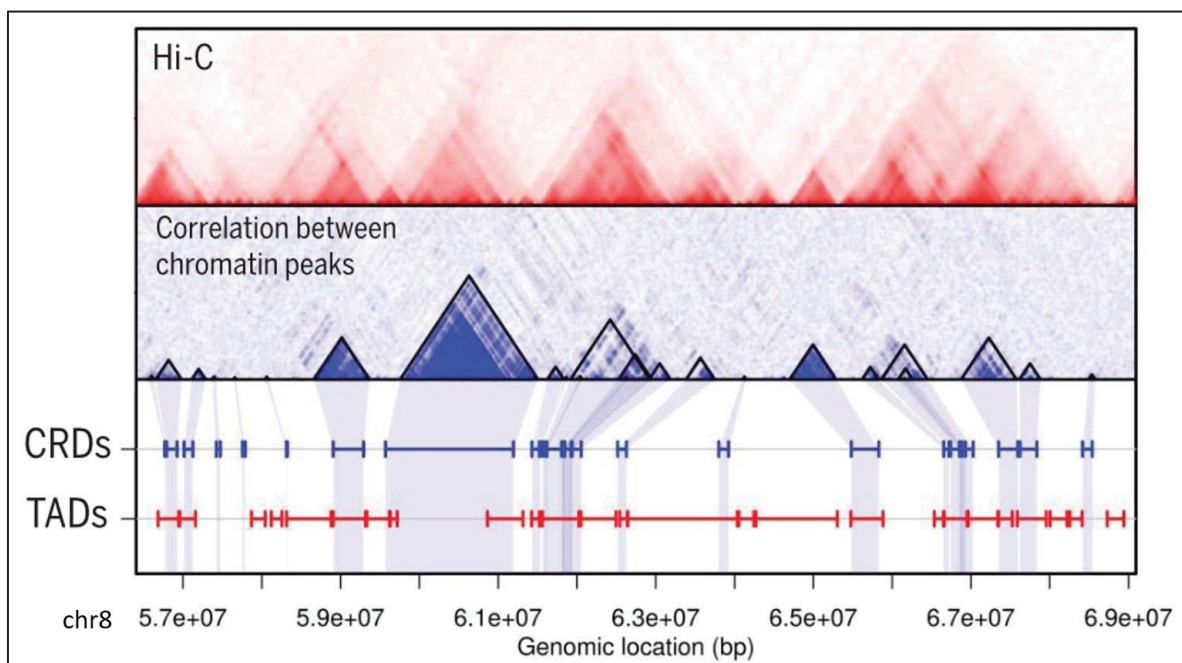


Figure 36 : Comparaison des TADs aux CRDs sur un intervalle du bras court du chromosome 8. Figure extraite de Delaneau et al., 2019.

Le second élément important pour l'utilisation de ces données chromatinienne (TADs) ou plus fonctionnelles (CRDs) est de garder à l'esprit la variabilité entre les types cellulaires. Idéalement, il faudrait parvenir à interpréter des données pertinentes compte tenu du phénotype et du tissu malade chez le patient. Clairement, l'utilisation de lignées lymphoblastoïdes dans notre travail sur le séquençage RNA-Seq, si elle était justifiée par sa faisabilité, a probablement limité notre capacité à retrouver les gènes déjà impliqués dans le phénotype des patients diagnostiqués et à identifier un gène causal chez les patients sans diagnostic.

Il faut donc, comme toujours, revenir à l'examen clinique et plus globalement au phénotype pour travailler à la corrélation génotype/phénotype. C'est ce travail de corrélation qui permettra de construire des nouvelles hypothèses de travail autour des remaniements chromosomiques et des variants génétiques plus globalement. Le lien supposé entre phénotype et génotype doit ensuite faire l'objet de démonstrations fonctionnelles. Si le génotypage est de plus en plus simple et rapide, nos connaissances sur la biologie du génome restent incomplètes et la capture du phénotype de chaque patient un défi. Les nouveaux usages de numérisation des phénotypes cliniques (codage HPO) (Bone et al., 2016), biologiques (LOINC2HPO) (Zhang et al., 2019), intelligence artificielle à partir de photos de patients, de résultats de radiologie,...) (Hsieh et al., 2019) ont montré leur intérêt dans l'amélioration du rendement des tests génétiques diagnostiques et l'identification de nouveaux gènes. Ils doivent être implémentés pour aider à l'interprétation de variants génomiques et réduire encore le nombre de patients sans diagnostic.

IV - Conclusion

En réponse aux questions initiales, ce travail a montré que les remaniements les plus complexes (*chromoanagenesis*) et les plus méconnus (rétrocopies) étaient aujourd'hui accessibles. S'agissant des rétrocopies il s'agit d'évènements à connaître de par la variété des mécanismes par lesquels elles peuvent être impliquées en pathologie et par le piège diagnostique qu'elles constituent. Notre travail permet de reconnaître ces évènements.

Si le séquençage *short-read* est au cœur de notre travail, nous présentons également **l'apport de technologies émergentes** (*long-read, linked-read*) dans l'arsenal cytogénétique.

Par l'étude de **la plus large cohorte de *chromoanagenesis* constitutionnels**, notre travail vient valider une hypothèse mécanistique claire. Elle débute par la séquestration d'un chromosome dans un micronoyau, où celui-ci tarde à se répliquer entraînant une condensation prématurée de ce chromosome. Cette condensation est un stress appliqué sur les fourches de réplication encore actives, concentrées dans les régions chromatiniennes qui se répliquent tardivement. C'est ce stress répliatif qui va engendrer des sauts de fourche de réplication voire des cassures double brin. Ces cassures ne seront réparées que lors de la réintégration dans le noyau principal. Ce scénario semble être **partagé entre *chromoanagenesis* constitutionnels et *chromoanagenesis* tumoraux** dans la mesure où la distribution des points de cassure suit les mêmes règles dans les deux cas. Il s'agit de la **première démonstration de ce scénario** réalisée à partir de données de patients et non par un modèle cellulaire.

En travaillant, sur les remaniements les plus complexes, nous avons également révélé de **nouveaux éléments** régissant les remaniements les plus simples. La déplétion en points de cassure observées dans les domaines liés à la lamine est une première ligne de **démarcation entre remaniements simples et complexes**. Les raisons de cette déplétion sont aujourd'hui inexplicables et doivent faire l'objet d'une poursuite de ce travail mais il apparaît clairement que remaniements simples et remaniements complexes ne résultent pas des mêmes mécanismes biologiques sous-jacents et ne font donc pas partie d'un même continuum.

Les mécanismes discutés ici (transcrit de fusion (Article 1), insertion d'élément mobile (Article 2), réorganisation chromatinienne (Article 3)) constituent, locus par locus, d'importantes pistes de recherche pour être en mesure d'expliquer de nouvelles corrélations génotype/phénotype et *in fine* améliorer le rendement diagnostique des analyses génétiques.

V - Bibliographie

- Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 21:974–984.
- Ader F, Heide S, Marzin P, Afenjar A, Diguët F, Chantot Bastaraud S, Rollat-Farnier P-A, Sanlaville D, Portnoï M-F, Siffroi J-P, Schluth-Bolard C. 2019. A 14q distal chromoanagenesis elucidated by whole genome sequencing. *Eur J Med Genet* 103776.
- Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, Rocco JW, Teknos TN, Kumar B, Wangsa D, He D, Ried T, et al. 2014. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res* 24:185–99.
- Anderson ND, Borja R de, Young MD, Fuligni F, Rosic A, Roberts ND, Hajjar S, Layeghifard M, Novokmet A, Kowalski PE, Anaka M, Davidson S, et al. 2018. Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science* 361:.
- Anderson SE, Kamath A, Pilz DT, Morgan SM. 2016. A rare example of germ-line chromothripsis resulting in large genomic imbalance. *Clin Dysmorphol* 25:58–62.
- Andrey G, Schöpflin R, Jerković I, Heinrich V, Ibrahim DM, Paliou C, Hochradel M, Timmermann B, Haas S, Vingron M, Mundlos S. 2017. Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res* 27:223–233.
- Aneichyk T, Hendriks WT, Yadav R, Shin D, Gao D, Vaine CA, Collins RL, Domingo A, Currall B, Stortchevoi A, Multhaupt-Buell T, Penney EB, et al. 2018. Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. *Cell* 172:897-909.e21.
- Antonarakis SE, Rossiter JP, Young M, Horst J, Moerlose P de, Sommer SS, Ketterling RP, Kazazian HH, Négrier C, Vinciguerra C, Gitschier J, Goossens M, et al. 1995. Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood* 86:2206–12.
- Athanasiadis A, Rich A, Maas S. 2004. Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome. *PLoS Biol* 2:e391.
- Auger J, Bonnet C, Valduga M, Philippe C, Bertolo-Houriez E, Beri-Dexheimer M, Schweitzer C, Leheup B, Jonveaux P. 2013. De novo complex X chromosome rearrangement unmasking maternally inherited *CSF2RA* deletion in a girl with pulmonary alveolar proteinosis. *Am J Med Genet Part A* 161A:n/a-n/a.
- Baala L, Briault S, Etchevers HC, Laumonier F, Natiq A, Amiel J, Boddaert N, Picard C, Sbiti A, Asermouh A, Attié-Bitach T, Encha-Razavi F, et al. 2007. Homozygous silencing of T-box transcription factor EOMES leads to microcephaly with polymicrogyria and corpus callosum agenesis. *Nat Genet* 39:454–456.
- Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, Allen E Van, Kryukov G V, et al. 2013. Punctuated evolution of prostate cancer genomes. *Cell* 153:666–77.
- Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* 11:1005–17.
- Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 129:823–837.

- Beckmann JS, Estivill X, Antonarakis SE. 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 8:639–646.
- Behjati S, Tarpey PS, Haase K, Ye H, Young MD, Alexandrov LB, Farndon SJ, Collord G, Wedge DC, Martincorena I, Cooke SL, Davies H, et al. 2017. Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat Commun* 8:15936.
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
- Bertelsen B, Nazaryan-Petersen L, Sun W, Mehrjouy MM, Xie G, Chen W, Hjermand LE, Taschner PEM, Tümer Z. 2015. A germline chromothripsis event stably segregating in 11 individuals through three generations. *Genet Med*.
- Binsbergen E van, Hochstenbach R, Giltay J, Swinkels M. 2012. Unstable transmission of a familial complex chromosome rearrangement. *Am J Med Genet Part A* 158A:2888–2893.
- Black BE, Bassett EA. 2008. The histone variant CENP-A and centromere specification. *Curr Opin Cell Biol* 20:91–100.
- Boer M de, Leeuwen K van, Geissler J, Weemaes CM, Berg TK van den, Kuijpers TW, Warris A, Roos D. 2014. Primary immunodeficiency caused by an exonized retroposed gene copy inserted in the CYBB gene. *Hum Mutat* 35:486–96.
- Bone WP, Washington NL, Buske OJ, Adams DR, Davis J, Draper D, Flynn ED, Girdea M, Godfrey R, Golas G, Groden C, Jacobsen J, et al. 2016. Computational evaluation of exome sequence data using human and model organism phenotypes improves diagnostic efficiency. *Genet Med* 18:608–617.
- Bonev B, Cavalli G. 2016. Organization and function of the 3D genome. *Nat Rev Genet* 17:661–678.
- Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot J-P, Tanay A, Cavalli G. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171:557-572.e24.
- Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* 10:211–219.
- Bragg DC, Mangkalaphiban K, Vaine CA, Kulkarni NJ, Shin D, Yadav R, Dhakal J, Ton M-L, Cheng A, Russo CT, Ang M, Acuña P, et al. 2017. Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in *TAF1*. *Proc Natl Acad Sci* 114:E11020–E11028.
- Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran J V., Kazazian HH. 2003. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci* 100:5280–5285.
- Burnside RD, Harris A, Speyer D, Burgin WS, Rose DZ, Sanchez-Valle A. 2016. Constitutional Chromoanagenesis of Distal 13q in a Young Adult with Recurrent Strokes. *Cytogenet Genome Res* 150:46–51.
- Carapito R, Ivanova EL, Morlon A, Meng L, Molitor A, Erdmann E, Kieffer B, Pichot A, Naegely L, Kolmer A, Paul N, Hanauer A, et al. 2019. ZMIZ1 Variants Cause a Syndromic Neurodevelopmental Disorder. *Am J Hum Genet* 104:319–330.
- Caspersson T, Zech L, Johansson C. 1970. Differential binding of alkylating fluorochromes in human chromosomes. *Exp Cell Res* 60:315–9.
- Chakraborty A, Ay F. 2018. Identification of copy number variations and translocations in cancer cells

from Hi-C data. *Bioinformatics* 34:338–345.

Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, et al. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 6:677–81.

Chen L, Zhou W, Zhang L, Zhang F. 2014a. Genome Architecture and Its Roles in Human Copy Number Variation. *Genomics Inform* 12:136.

Chen W, Ullmann R, Langnick C, Menzel C, Wotschovsky Z, Hu H, Döring A, Hu Y, Kang H, Tzschach A, Hoeltzenbein M, Neitzel H, et al. 2010. Breakpoint analysis of balanced chromosome rearrangements by next-generation paired-end sequencing. *Eur J Hum Genet* 18:539–43.

Chen X, Yammine S, Shi C, Tark-Dame M, Göndör A, Ohlsson R. 2014b. The visualization of large organized chromatin domains enriched in the H3K9me2 mark within a single chromosome in a single cell. *Epigenetics* 9:1439–1445.

Chen Y-C, Liu T, Yu C-H, Chiang T-Y, Hwang C-C. 2013. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One* 8:e62856.

Chiang C, Jacobsen JC, Ernst C, Hanscom C, Heilbut A, Blumenthal I, Mills RE, Kirby A, Lindgren AM, Rudiger SR, McLaughlan CJ, Bawden CS, et al. 2012. Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat Genet* 44:390–7, S1.

Ciavarra G, Zacksenhaus E. 2011. Multiple pathways counteract cell death induced by RB1 loss: Implications for cancer. *Cell Cycle* 10:1533–1539.

Ciccio A, Elledge SJ. 2010. The DNA Damage Response: Making It Safe to Play with Knives. *Mol Cell* 40:179–204.

Claverie-Martín F, Flores C, Antón-Gamero M, González-Acosta H, García-Nieto V. 2005. The Alu insertion in the CLCN5 gene of a patient with Dent's disease leads to exon 11 skipping. *J Hum Genet* 50:370–374.

Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Khera A V., Francioli LC, Gauthier LD, Wang H, Watts NA, Solomonson M, O'Donnell-Luria A, et al. 2019. An open resource of structural variation for medical and population genetics. *bioRxiv* 578674.

Collins RL, Brand H, Redin CE, Hanscom C, Antolik C, Stone MR, Glessner JT, Mason T, Pregno G, Dorrani N, Mandrile G, Giachino D, et al. 2017. Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. *Genome Biol* 18:36.

Conroy JM, Grebe TA, Becker LA, Tsuchiya K, Nicholls RD, Buiting K, Horsthemke B, Cassidy SB, Schwartz S. 1997. Balanced Translocation 46, XY, t(2;15)(q37.2;q11.2) Associated with Atypical Prader-Willi Syndrome. *Am J Hum Genet* 61:388–394.

Cooper GM, Coe BP, Girirajan S, Rosenfeld J a, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, Abdel-Hamid H, Bader P, et al. 2011. A copy number variation morbidity map of developmental delay. *Nat Genet* 43:838–46.

Corneo B, Wendland RL, Deriano L, Cui X, Klein IA, Wong S-Y, Arnal S, Holub AJ, Weller GR, Pancake BA, Shah S, Brandt VL, et al. 2007. Rag mutations reveal robust alternative end joining. *Nature* 449:483–486.

Cortés-Ciriano I, Lee J-K, Xi R, Jain D, Jung YL, Yang L, Gordenin D, Klimczak LJ, Zhang C-Z, Pellman DS, Park PJ, Group on behalf of the PSVW, et al. 2018. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *bioRxiv* 333617.

Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova E V, Pan Y, Nezi L, Protopopov A,

- Chowdhury D, Pellman D. 2012. DNA breaks and chromosome pulverization from errors in mitosis. *Nature* 482:53–8.
- Cremer T, Cremer C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2:292–301.
- Cremer T, Cremer C, Schneider T, Baumann H, Hens L, Kirsch-Volders M. 1982. Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum Genet* 62:201–209.
- Cremer T, Kreth G, Koester H, Fink RH, Heintzmann R, Cremer M, Solovei I, Zink D, Cremer C. 2000. Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. *Crit Rev Eukaryot Gene Expr* 10:179–212.
- Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. 1999. Differences in the Localization and Morphology of Chromosomes in the Human Nucleus. *J Cell Biol* 145:1119–1131.
- Daughtry BL, Chavez SL. 2018. Time-Lapse Imaging for the Detection of Chromosomal Abnormalities in Primate Preimplantation Embryos. p 293–317.
- de Pagter MS, van Roosmalen MJ, Baas AF, Renkens I, Duran KJ, van Binsbergen E, Tavakoli-Yaraki M, Hochstenbach R, van der Veken LT, Cuppen E, Kloosterman WP. 2015. Chromothripsis in Healthy Individuals Affects Multiple Protein-Coding Genes and Can Result in Severe Congenital Abnormalities in Offspring. *Am J Hum Genet* 96:651–656.
- De S, Michor F. 2011. DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. *Nat Biotechnol* 29:1103–1108.
- Delaneau O, Zazhytska M, Borel C, Giannuzzi G, Rey G, Howald C, Kumar S, Ongen H, Popadin K, Marbach D, Ambrosini G, Bielser D, et al. 2019. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* (80-) 364:eaat8266.
- Deriano L, Roth DB. 2013. Modernizing the Nonhomologous End-Joining Repertoire: Alternative and Classical NHEJ Share the Stage. *Annu Rev Genet* 47:433–455.
- Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35:41–48.
- Dillinger S, Straub T, Németh A. 2017. Nucleolus association of chromosomal domains is largely maintained in cellular senescence despite massive nuclear reorganisation. *PLoS One* 12:e0178821.
- Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, Diao Y, Liang J, et al. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* 518:331–336.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–80.
- Dixon JR, Xu J, Dileep V, Zhan Y, Song F, Le VT, Yardımcı GG, Chakraborty A, Bann D V., Wang Y, Clark R, Zhang L, et al. 2018. Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet* 50:1388–1398.
- Dolgalev I, Sedlazeck F, Busby B. 2017. DangerTrack: A scoring system to detect difficult-to-assess regions. *F1000Research* 6:443.
- Drier Y, Lawrence MS, Carter SL, Stewart C, Gabriel SB, Lander ES, Meyerson M, Beroukhir R, Getz G. 2013. Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res* 23:228–35.
- Ebbert MTW, Jensen TD, Jansen-West K, Sens JP, Reddy JS, Ridge PG, Kauwe JSK, Belzil V,

- Pregent L, Carrasquillo MM, Keene D, Larson E, et al. 2019. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol* 20:97.
- Eisfeldt J, Pettersson M, Vezzi F, Wincent J, Källér M, Gruselius J, Nilsson D, Syk Lundberg E, Carvalho CMB, Lindstrand A. 2019. Comprehensive structural variation genome map of individuals carrying complex chromosomal rearrangements. *PLoS Genet* 15:e1007858.
- Elyanow R, Wu H-T, Raphael BJ. 2017. Identifying structural variants using linked-read sequencing data. *Bioinformatics* 34:353.
- English AC, Salerno WJ, Hampton OA, Gonzaga-Jauregui C, Ambreth S, Ritter DI, Beck CR, Davis CF, Dahdouli M, Ma S, Carroll A, Veeraraghavan N, et al. 2015. Assessing structural variation in a personal genome-towards a human reference diploid genome. *BMC Genomics* 16:286.
- Ernst A, Jones DTW, Maass KK, Rode A, Deeg KI, Jebaraj BMC, Korshunov A, Hovestadt V, Tainsky MA, Pajtler KW, Bender S, Brabetz S, et al. 2016. Telomere dysfunction and chromothripsis. *Int J Cancer* 138:2905–2914.
- Erson-Omay EZ, Henegariu O, Omay SB, Harmancı AS, Youngblood MW, Mishra-Gorur K, Li J, Özduman K, Carrión-Grant G, Clark VE, Çağlar C, Bakırcıoğlu M, et al. 2017. Longitudinal analysis of treatment-induced genomic alterations in gliomas. *Genome Med* 9:12.
- Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 24:363–367.
- Ewing AD, Ballinger TJ, Earl D, Harris CC, Ding L, Wilson RK, Haussler D, Haussler D. 2013. Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. *Genome Biol* 14:R22.
- Ewing AD, Kazazian HH. 2010. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res* 20:1262–70.
- Feng Q, Moran J V, Kazazian HH, Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87:905–16.
- Fontana P, Genesisio R, Casertano A, Cappuccio G, Mormile A, Nitsch L, Iolascon A, Andria G, Melis D. 2014. Loeys–Dietz syndrome type 4, caused by chromothripsis, involving the TGFB2 gene. *Gene* 538:69–73.
- Fortin J-P, Hansen KD. 2015. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol* 16:180.
- Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan W-L, Spielmann M, Timmermann B, et al. 2016. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 538:265–269.
- Fredga K, Hall B. 1970. A complex familial translocation involving chromosomes 5, 9 and 13. *Cytogenetics* 9:294–306.
- Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* 15:2038–2049.
- Fukami M, Shima H, Suzuki E, Ogata T, Matsubara K, Kamimaki T. 2017. Catastrophic cellular events leading to complex chromosomal rearrangements in the germline. *Clin Genet* 91:653–660.
- Fungtammasan A, Walsh E, Chiaromonte F, Eckert KA, Makova KD. 2012. A genome-wide analysis of common fragile sites: What features determine chromosomal instability in the human genome? *Genome Res* 22:993–1005.
- Furey TS, Haussler D. 2003. Integration of the cytogenetic map with the draft human genome

sequence. *Hum Mol Genet* 12:1037–1044.

Gal-Mark N, Schwartz S, Ast G. 2008. Alternative splicing of Alu exons--two arms are better than one. *Nucleic Acids Res* 36:2012–23.

Gallus GN, Cardaioli E, Rufa A, Pozzo P Da, Bianchi S, D'Eramo C, Collura M, Tumino M, Pavone L, Federico A. 2010. Alu-element insertion in an OPA1 intron sequence associated with autosomal dominant optic atrophy. *Mol Vis* 16:178–83.

Gamba BF, Richieri-Costa A, Costa S, Rosenberg C, Ribeiro-Bicudo LA. 2015. Chromothripsis with at least 12 breaks at 1p36.33-p35.3 in a boy with multiple congenital anomalies. *Mol Genet Genomics* 290:2213–6.

Ganem NJ, Godinho SA, Pellman D. 2009. A mechanism linking extra centrosomes to chromosomal instability. *Nature* 460:278–282.

Ganji M, Shaltiel IA, Bisht S, Kim E, Kalichava A, Haering CH, Dekker C. 2018. Real-time imaging of DNA loop extrusion by condensin. *Science* (80-) 360:102–105.

Gardner RJM, Amor DJ. 2018. *Gardner and Sutherland's Chromosome Abnormalities and Genetic Counseling*. Oxford University Press.

Genesio R, Fontana P, Mormile A, Casertano A, Falco M, Conti A, Franzese A, Mozzillo E, Nitsch L, Melis D. 2015. Constitutional chromothripsis involving the critical region of 9q21.13 microdeletion syndrome. *Mol Cytogenet* 8:96.

Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. 2019. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet* 51:1272–1282.

Gilissen C, Hehir-Kwa JY, Thung DT, Vorst M van de, Bon BWM van, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A, Leach R, Klein R, et al. 2014. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511:344–7.

Gisselsson D, Jonson T, Petersen A, Strombeck B, Dal Cin P, Hoglund M, Mitelman F, Mertens F, Mandahl N. 2001. Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. *Proc Natl Acad Sci* 98:12683–12688.

Glenn CC, Saitoh S, Jong MT, Filbrandt MM, Surti U, Driscoll DJ, Nicholls RD. 1996. Gene structure, DNA methylation, and imprinted expression of the human SNRPN gene. *Am J Hum Genet* 58:335–46.

Gregori M De, Ciccone R, Magini P, Pramparo T, Gimelli S, Messa J, Novara F, Vetro A, Rossi E, Maraschio P, Bonaglia MC, Anichini C, et al. 2007. Cryptic deletions are a common finding in "balanced" reciprocal and complex chromosome rearrangements: a study of 59 patients. *J Med Genet* 44:750–762.

Gregorio E Di, Bianchi FT, Schiavi A, Chiotto AMA, Rolando M, Verdun di Cantogno L, Grosso E, Cavalieri S, Calcia A, Lacerenza D, Zuffardi O, Retta SF, et al. 2013. 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* 50:543–551.

Gregory JJ, Wagner JE, Verlander PC, Levran O, Batish SD, Eide CR, Steffenhagen A, Hirsch B, Auerbach AD, Auerbach AD. 2001. Somatic mosaicism in Fanconi anemia: evidence of genotypic reversion in lymphohematopoietic stem cells. *Proc Natl Acad Sci U S A* 98:2532–7.

Gu H, Jiang J, Li J, Zhang Y, Dong X, Huang Y, Son X, Lu X, Chen Z. 2013. A Familial Cri-du-Chat/5p Deletion Syndrome Resulted from Rare Maternal Complex Chromosomal Rearrangements

(CCRs) and/or Possible Chromosome 5p Chromothripsis. *PLoS One* 8:e76985.

Gudjonsson T, Altmeyer M, Savic V, Toledo L, Dinant C, Grøfte M, Bartkova J, Poulsen M, Oka Y, Bekker-Jensen S, Mailand N, Neumann B, et al. 2012. TRIP12 and UBR5 Suppress Spreading of Chromatin Ubiquitylation at Damaged Chromosomes. *Cell* 150:697–709.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, Klein A de, Wessels L, Laat W de, Steensel B van. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453:948–951.

Haince J-F, McDonald D, Rodrigue A, Déry U, Masson J-Y, Hendzel MJ, Poirier GG. 2008. PARP1-dependent Kinetics of Recruitment of MRE11 and NBS1 Proteins to Multiple DNA Damage Sites. *J Biol Chem* 283:1197–1208.

Halgren C, Nielsen NM, Nazaryan-Petersen L, Silahtaroglu A, Collins RL, Lowther C, Kjaergaard S, Frisch M, Kirchoff M, Brøndum-Nielsen K, Lind-Thomsen A, Mang Y, et al. 2018. Risks and Recommendations in Prenatally Detected De Novo Balanced Chromosomal Rearrangements from Assessment of Long-Term Outcomes. *Am J Hum Genet* 102:1090–1103.

Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH. 2011. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet* 20:3386–3400.

Hatch EM, Fischer AH, Deerinck TJ, Hetzer MW. 2013a. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell* 154:47–60.

Hatch EM, Fischer AH, Deerinck TJ, Hetzer MW. 2013b. Catastrophic Nuclear Envelope Collapse in Cancer Cell Micronuclei. *Cell* 154:47–60.

Hehir-Kwa JY, Marschall T, Kloosterman WP, Francioli LC, Baaijens JA, Dijkstra LJ, Abdellaoui A, Koval V, Thung DT, Wardenaar R, Renkens I, Coe BP, et al. 2016. A high-quality human reference panel reveals the complexity and distribution of genomic structural variants. *Nat Commun* 7:12989.

Holland AJ, Cleveland DW. 2012. Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. *Nat Med* 18:1630–8.

Hsieh T-C, Mensah MA, Pantel JT, Aguilar D, Bar O, Bayat A, Becerra-Solano L, Bentzen HB, Biskup S, Borisov O, Braaten O, Ciaccio C, et al. 2019. PEDIA: prioritization of exome data by image analysis. *Genet Med* 1–8.

Hsu TC. 1952. Mammalian CHromosomes in vitro. *J Hered* 43:167–172.

Hsu TC. 1975. A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics* 79 Suppl:137–50.

Ira G, Haber JE. 2002. Characterization of RAD51-independent break-induced replication that acts preferentially with short homologous sequences. *Mol Cell Biol* 22:6384–92.

Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. 2013. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19:141–157.

Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, Balloux F, Dessimoz C, Bähler J, Sedlazeck FJ. 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* 8:14061.

Jost D, Vaillant C, Meister P. 2017. Coupling 1D modifications and 3D nuclear organization: data, models and function. *Curr Opin Cell Biol* 44:20–27.

Kabotyanski EB, Gomelsky L, Han JO, Stamato TD, Roth DB. 1998. Double-strand break repair in Ku86- and XRCC4-deficient cells. *Nucleic Acids Res* 26:5333.

- Kallioniemi A, Kallioniemi O, Sudar D, Rutovitz D, Gray J, Waldman F, Pinkel D. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* (80-) 258:818–821.
- Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332:164–166.
- Kent WJ. 2002. BLAT---The BLAST-Like Alignment Tool. *Genome Res* 12:656–664.
- Kind J, Pagie L, de Vries SS, Nahidiazar L, Dey SS, Bienko M, Zhan Y, Lajoie B, de Graaf CA, Amendola M, Fudenberg G, Imakaev M, et al. 2015. Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. *Cell* 163:134–147.
- King PH, Levine TD, Freneau RT, Keene JD. 1994. Mammalian homologs of *Drosophila* ELAV localized to a neuronal subset can bind in vitro to the 3' UTR of mRNA encoding the Id transcriptional repressor. *J Neurosci* 14:1943–52.
- Kleinjan D, Heyningen V van. 1998. Position effect in human genetic disease. *Hum Mol Genet* 7:1611–1618.
- Kloosterman WP, Guryev V, Roosmalen M van, Duran KJ, Bruijn E de, Bakker SCM, Letteboer T, Nesselrooij B van, Hochstenbach R, Poot M, Cuppen E. 2011a. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. *Hum Mol Genet* 20:1916–24.
- Kloosterman WP, Hoogstraat M, Paling O, Tavakoli-Yaraki M, Renkens I, Vermaat JS, Roosmalen MJ van, Lieshout S van, Nijman IJ, Roessingh W, 't Slot R van, Belt J van de, et al. 2011b. Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. *Genome Biol* 12:R103.
- Kloosterman WP, Koster J, Molenaar JJ. 2014. Prevalence and clinical implications of chromothripsis in cancer genomes. *Curr Opin Oncol* 26:64–72.
- Kloosterman WP, Tavakoli-Yaraki M, van Roosmalen MJ, van Binsbergen E, Renkens I, Duran K, Ballarati L, Vergult S, Giardino D, Hansson K, Ruivenkamp CAL, Jager M, et al. 2012. Constitutional Chromothripsis Rearrangements Involve Clustered Double-Stranded DNA Breaks and Nonhomologous Repair Mechanisms. *Cell Rep* 1:648–655.
- Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, Nomura Y, Segawa M, Yoshioka M, Saito K, Osawa M, Hamano K, Sakakihara Y, et al. 1998. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* 394:388–392.
- Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, Wong WSW, Sigurdsson G, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488:471–475.
- Koningsbruggen S van, Gierlinski M, Schofield P, Martin D, Barton GJ, Ariyurek Y, Dunnen JT den, Lamond AI. 2010. High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol Biol Cell* 21:3735–48.
- Korbel JO, Campbell PJ. 2013. Criteria for inference of chromothripsis in cancer genomes. *Cell* 152:1226–36.
- Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, et al. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318:420–6.
- Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, McCarroll SA. 2012. Differential Relationship of DNA Replication Timing to Different Forms of Human Mutation and Variation. *Am J*

Hum Genet 91:1033–1040.

Kragesteen BK, Spielmann M, Paliou C, Heinrich V, Schöpflin R, Esposito A, Annunziatella C, Bianco S, Chiariello AM, Jerković I, Harabula I, Guckelberger P, et al. 2018. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat Genet* 50:1463–1473.

Kronenberg ZN, Fiddes IT, Gordon D, Murali S, Cantsilieris S, Meyerson OS, Underwood JG, Nelson BJ, Chaisson MJP, Dougherty ML, Munson KM, Hastie AR, et al. 2018. High-resolution comparative analysis of great ape genomes. *Science* (80-) 360:eaar6343.

Kurtas N, Arrigoni F, Errichiello E, Zucca C, Maghini C, D'Angelo MG, Beri S, Giorda R, Bertuzzo S, Delledonne M, Xumerle L, Rossato M, et al. 2018. Chromothripsis and ring chromosome 22: a paradigm of genomic complexity in the Phelan-McDermid syndrome (22q13 deletion syndrome). *J Med Genet* 55:269–277.

Kurtas NE, Xumerle L, Giussani U, Pansa A, Cardarelli L, Bertini V, Valetto A, Liehr T, Clara Bonaglia M, Errichiello E, Delledonne M, Zuffardi O. 2019. Insertional translocation involving an additional nonchromothriptic chromosome in constitutional chromothripsis: Rule or exception? *Mol Genet Genomic Med* 7:e00496.

Kürten S, Obe G. 1975. Premature chromosome condensation in the bone marrow of Chinese hamsters after application of bleomycin in vivo. *Mutat Res Mol Mech Mutagen* 27:285–294.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Lappalainen T, Scott AJ, Brandt M, Hall IM. 2019. Genomic Analysis in the Age of Human Genome Sequencing. *Cell* 177:70–84.

LaSalle JM, Lalande M. 1996. Homologous Association of Oppositely Imprinted Chromosomal Domains. *Science* (80-) 272:725–728.

Lee JA, Carvalho CMB, Lupski JR. 2007. A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders. *Cell* 131:1235–1247.

Lejeune J, Gautier M, Turpin R. 1959. Etude des chromosomes somatiques de 9 enfants mongoliens. *C R Hebd Seances Acad Sci* 248:1721–2.

Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536:285–291.

Lemaître C, Grabarz A, Tsouroula K, Andronov L, Furst A, Pankotai T, Heyer V, Rogier M, Attwood KM, Kessler P, Dellaire G, Klaholz B, et al. 2014. Nuclear position dictates DNA repair pathway choice. *Genes Dev* 28:2450–2463.

Lessel D, Gehbauer C, Bramswig NC, Schluth-Bolard C, Venkataramanappa S, Gassen KLI van, Hempel M, Haack TB, Baresic A, Genetti CA, Funari MFA, Lessel I, et al. 2018. BCL11B mutations in patients affected by a neurodevelopmental disorder with reduced type 2 innate lymphoid cells. *Brain* 141:2299–2311.

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, et al. 2007. The Diploid Genome Sequence of an Individual Human. *PLoS Biol* 5:e254.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–60.

- Liang F, Jasin M. 1996. Ku80-deficient Cells Exhibit Excess Degradation of Extrachromosomal DNA. *J Biol Chem* 271:14405–14411.
- Lieber MR, Lu H, Gu J, Schwarz K. 2008. Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: relevance to cancer, aging, and the immune system. *Cell Res* 18:125–133.
- Lieberman-Aiden E, Berkum NL van, Williams L, Imakaev M, Ragozy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, et al. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* (80-) 326:289–293.
- Liu B, K.H Yip R, Zhou Z. 2012. Chromatin Remodeling, DNA Damage Repair and Aging. *Curr Genomics* 13:533–547.
- Liu L, De S, Michor F. 2013. DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. *Nat Commun* 4:1502.
- Liu P, Erez A, Nagamani SCS, Dhar SU, Kołodziejka KE, Dharmadhikari A V, Cooper ML, Wiszniewska J, Zhang F, Withers MA, Bacino CA, Campos-Acevedo LD, et al. 2011. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* 146:889–903.
- Lonfat N, Duboule D. 2015. Structure, function and evolution of topologically associating domains (TADs) at *HOX* loci. *FEBS Lett* 589:2869–2876.
- Luijten MNH, Lee JXT, Crasta KC. 2018. Mutational game changer: Chromothripsis and its emerging relevance to cancer. *Mutat Res Mutat Res* 777:29–51.
- Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, et al. 2015. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161:1012–1025.
- Ly P, Teitz LS, Kim DH, Shoshani O, Skaletsky H, Fachinetti D, Page DC, Cleveland DW. 2017. Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. *Nat Cell Biol* 19:68–75.
- Macera MJ, Sobrino A, Levy B, Jobanputra V, Aggarwal V, Mills A, Esteves C, Hanscom C, Pereira S, Pillalamarri V, Ordulu Z, Morton CC, et al. 2015. Prenatal diagnosis of chromothripsis, with nine breaks characterized by karyotyping, FISH, microarray and whole-genome sequencing. *Prenat Diagn* 35:299–301.
- Maciejowski J, Li Y, Bosco N, Campbell PJ, de Lange T. 2015. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* 163:1641–1654.
- Madan K. 2013. What Is a Complex Chromosome Rearrangement? *Am J Med Genet Part A* 161:1181–1184.
- Malan V, Lapierre J-M, Egloff M, Goidin D, Beaujard M-P, Maurin M-L, Attié-Bitach T, Bessières B, Bernard J-P, Roth P, Stirnemann J, Salomon L, et al. 2016. A French Approach to Test Fetuses with Ultrasound Abnormalities Using a Customized Microarray as First-Tier Genetic Test. *Cytogenet Genome Res* 147:103–110.
- Malhotra A, Lindberg M, Faust GG, Leibowitz ML, Clark RA, Layer RM, Quinlan AR, Hall IM. 2013. Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res* 23:762–776.
- Mao YS, Zhang B, Spector DL. 2011. Biogenesis and function of nuclear bodies. *Trends Genet* 27:295–306.

- Marchal C, Sima J, Gilbert DM. 2019. Control of DNA replication timing in the 3D genome. *Nat Rev Mol Cell Biol* 1–17.
- Mardin BR, Drainas AP, Waszak SM, Weischenfeldt J, Isokane M, Stütz AM, Raeder B, Efthymiopoulos T, Buccitelli C, Segura-Wang M, Northcott P, Pfister SM, et al. 2015. A cell-based model system links chromothripsis with hyperploidy. *Mol Syst Biol* 11:828.
- Marks P, Garcia S, Barrio AM, Belhocine K, Bernate J, Bharadwaj R, Bjornson K, Catalanotti C, Delaney J, Fehr A, Fiddes IT, Galvin B, et al. 2019. Resolving the full spectrum of human genome variation using Linked-Reads. *Genome Res* 29:635–645.
- Martin SL, Cruceanu M, Branciforte D, Wai-lun Li P, Kwok SC, Hodges RS, Williams MC. 2005. LINE-1 Retrotransposition Requires the Nucleic Acid Chaperone Activity of the ORF1 Protein. *J Mol Biol* 348:549–561.
- Masset H, Hestand MS, Esch H Van, Kleinfinger P, Plaisancié J, Afenjar A, Mollignier R, Schluth-Bolard C, Sanlaville D, Vermeesch JR. 2016. A Distinct Class of Chromoanagenesis Events Characterized by Focal Copy Number Gains. *Hum Mutat*.
- Massip F, Laurent M, Brossas C, Fernández-Justel JM, Gómez M, Prioleau M-N, Duret L, Picard F. 2019. Evolution of replication origins in vertebrate genomes: rapid turnover despite selective constraints. *Nucleic Acids Res* 47:5114–5125.
- Mathias S, Scott A, Kazazian H, Boeke J, Gabriel A. 1991. Reverse transcriptase encoded by a human transposable element. *Science* (80-) 254:1808–1810.
- McDermott DH, Gao J-L, Liu Q, Siwicki M, Martens C, Jacobs P, Velez D, Yim E, Bryke CR, Hsu N, Dai Z, Marquesen MM, et al. 2015. Chromothriptic Cure of WHIM Syndrome. *Cell* 160:686–699.
- McGowan-Jordan J. 2016. ISCN 2016: An International System for Human Cytogenomic Nomenclature (2016): Recommendations of the International Standing Committee on.
- Meuleman W, Peric-Hupkes D, Kind J, Beaudry J-B, Pagie L, Kellis M, Reinders M, Wessels L, Steensel B van. 2013. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res* 23:270–280.
- Middelkamp S, Heesch S van, Braat AK, Ligt J de, Iterson M van, Simonis M, Roosmalen MJ van, Kelder MJE, Kruisselbrink E, Hochstenbach R, Verbeek NE, Ippel EF, et al. 2017. Molecular dissection of germline chromothripsis in a developmental context using patient-derived iPSCs. *Genome Med* 9:9.
- Middelkamp S, Vlaar JM, Giltay J, Korzelius J, Besselink N, Boymans S, Janssen R, Fonteijne L de la, Binsbergen E van, Roosmalen MJ van, Hochstenbach R, Giachino D, et al. 2019. Prioritization of genes driving congenital phenotypes of patients with de novo genomic structural variants. *bioRxiv* 707430.
- Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, Church DM, Crolla J a, Eichler EE, Epstein CJ, Faucett WA, Feuk L, et al. 2010. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 86:749–64.
- Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye KK, Cheetham RK, Chinwalla A, Conrad DF, et al. 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* 470:59–65.
- Mitelman F, Johansson B, Mertens F, Schyman T, Mandahl N. 2019. Cancer chromosome breakpoints cluster in gene-rich genomic regions. *Genes, Chromosom Cancer* 58:149–154.
- Morisada N, Rendtorff ND, Nozu K, Morishita T, Miyakawa T, Matsumoto T, Hisano S, Iijima K,

- Tranebjærg L, Shirahata A, Matsuo M, Kusuhara K. 2010. Branchio-oto-renal syndrome caused by partial EYA1 deletion due to LINE-1 insertion. *Pediatr Nephrol* 25:1343–1348.
- Morishita M, Muramatsu T, Suto Y, Hirai M, Konishi T, Hayashi S, Shigemizu D, Tsunoda T, Moriyama K, Inazawa J. 2016. Chromothripsis-like chromosomal rearrangements induced by ionizing radiation using proton microbeam irradiation system. *Oncotarget* 7:10182–92.
- Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran J V. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* 31:159–165.
- Nakamura Y, Murata M, Takagi Y, Kozuka T, Nakata Y, Hasebe R, Takagi A, Kitazawa J, Shima M, Kojima T. 2015. SVA retrotransposition in exon 6 of the coagulation factor IX gene causing severe hemophilia B. *Int J Hematol* 102:134–139.
- Nazaryan L, Stefanou EG, Hansen C, Kosyakova N, Bak M, Sharkey FH, Mantziou T, Papanastasiou AD, Velissariou V, Liehr T, Syrrou M, Tommerup N. 2014. The strength of combined cytogenetic and mate-pair sequencing techniques illustrated by a germline chromothripsis rearrangement involving FOXP2. *Eur J Hum Genet* 22:338–343.
- Nones K, Waddell N, Wayte N, Patch A-M, Bailey P, Newell F, Holmes O, Fink JL, Quinn MCJ, Tang YH, Lampe G, Quek K, et al. 2014. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nat Commun* 5:5224.
- Nowak SJ, Corces VG. 2004. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet* 20:214–220.
- Nowell P, Hungerford D, Nowell P. 1960.
- Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AWC, Pippel M, Winkler S, Hastie AR, Young G, Roscito JG, Falcon F, Knapp D, et al. 2018. The axolotl genome and the evolution of key tissue formation regulators. *Nature* 554:50–55.
- Nussenzweig A, Nussenzweig MC. 2007. A Backup DNA Repair Pathway Moves to the Forefront. *Cell* 131:223–225.
- Obe G, Beek B. 1975. The human leucocyte test system. VII. Further investigations concerning micronucleus-derived premature chromosome condensation. *Humangenetik* 30:143–54.
- Ochalski ME, Engle N, Wakim A, Ravnán BJ, Hoffner L, Rajkovic A, Surti U. 2011. Complex X chromosome rearrangement delineated by array comparative genome hybridization in a woman with premature ovarian insufficiency. *Fertil Steril* 95:2433.e9-2433.e15.
- Pagter MS de, Roosmalen MJ van, Baas AF, Renkens I, Duran KJ, Binsbergen E van, Tavakoli-Yaraki M, Hochstenbach R, Veken LT van der, Cuppen E, Kloosterman WP. 2015. Chromothripsis in healthy individuals affects multiple protein-coding genes and can result in severe congenital abnormalities in offspring. *Am J Hum Genet* 96:651–6.
- Pai GS, Thomas GH, Mahoney W, Migeon BR. 1980. Complex chromosome rearrangements. Report of a new case and literature review. *Clin Genet* 18:436–44.
- Pampalona J, Roscioli E, Silkworth WT, Bowden B, Genescà A, Tusell L, Cimini D. 2016. Chromosome Bridges Maintain Kinetochore-Microtubule Attachment throughout Mitosis and Rarely Break during Anaphase. *PLoS One* 11:e0147420.
- Pantelias A, Karachristou I, Georgakilas AG, Terzoudi GI. 2019. Interphase Cytogenetic Analysis of Micronucleated and Multinucleated Cells Supports the Premature Chromosome Condensation Hypothesis as the Mechanistic Origin of Chromothripsis. *Cancers (Basel)* 11:1123.
- Parikh FR, Athalye AS, Naik NJ, Naik DJ, Sanap RR, Madon PF. 2018. Preimplantation Genetic

- Testing: Its Evolution, Where Are We Today? *J Hum Reprod Sci* 11:306–314.
- Payer LM, Steranka JP, Yang WR, Kryatova M, Medabalimi S, Ardeljan D, Liu C, Boeke JD, Avramopoulos D, Burns KH. 2017. Structural variants caused by *Alu* insertions are associated with risks for many human diseases. *Proc Natl Acad Sci* 114:E3984–E3992.
- Pellestor F. 2014. Chromothripsis: how does such a catastrophic event impact human reproduction? *Hum Reprod* 29:388–393.
- Pellestor F, Anahory T, Lefort G, Puechberty J, Liehr T, Hédon B, Sarda P. Complex chromosomal rearrangements: origin and meiotic behavior. *Hum Reprod Update* 17:476–94.
- Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. 2019. On the length, weight and GC content of the human genome. *BMC Res Notes* 12:106.
- Plaisancié J, Kleinfinger P, Cances C, Bazin A, Julia S, Trost D, Lohmann L, Vigouroux A. 2014. Constitutional chromoanasythesis: description of a rare chromosomal event in a patient. *Eur J Med Genet* 57:567–70.
- Pombo A, Dillon N. 2015. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol* 16:245–257.
- Pons L, Till M, Alix E, Abel C, Boggio D, Bordes A, Caloone J, Raskin FC, Chatron N, Cordier M-P, Fichez A, Labalme A, et al. 2017. Prenatal microarray comparative genomic hybridization: Experience from the two first years of activity at the Lyon university-hospital. *J Gynecol Obstet Hum Reprod* 46:.
- Qi H, Xuan Z-L, Du Y, Cai L-R, Zhang H, Wen X-H, Kong X-D, Yang K, Mi Y, Fu X-X, Cao S-B, Wang J, et al. 2018. High resolution global chromosomal aberrations from spontaneous miscarriages revealed by low coverage whole genome sequencing. *Eur J Obstet Gynecol Reprod Biol* 224:21–28.
- Qian H, Sheetz MP, Elson EL. 1991. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys J* 60:910–921.
- Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159:1665–1680.
- Rausch T, Jones DTW, Zapatka M, Stütz AM, Zichner T, Weischenfeldt J, Jäger N, Remke M, Shih D, Northcott PA, Pfaff E, Tica J, et al. 2012. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* 148:59–71.
- Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, Hanscom C, Pillalamarri V, Seabra CM, Abbott M-A, Abdul-Rahman OA, Aberg E, et al. 2017. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet* 49:36–45.
- Reijnders MRF, Miller KA, Alvi M, Goos JAC, Lees MM, Burca A de, Henderson A, Kraus A, Mikat B, Vries BBA de, Isidor B, Kerr B, et al. 2018. De Novo and Inherited Loss-of-Function Variants in *TLK2*: Clinical and Genotype-Phenotype Evaluation of a Distinct Neurodevelopmental Disorder. *Am J Hum Genet* 102:1195–1203.
- Rey J Del, Santos M, González-Meneses A, Milà M, Fuster C. 2016. Heterogeneity of a Constitutional Complex Chromosomal Rearrangement in 2q. *Cytogenet Genome Res* 148:156–64.
- Robin JD, Magdinier F. 2016. Physiological and Pathological Aging Affects Chromatin Dynamics, Structure and Function at the Nuclear Edge. *Front Genet* 7:153.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29:24–6.
- Robson MI, Ringel AR, Mundlos S. 2019. Regulatory Landscaping: How Enhancer-Promoter

Communication Is Sculpted in 3D. *Mol Cell* 74:1110–1122.

Rodríguez-Martín C, Cidre F, Fernández-Teijeiro A, Gómez-Mariano G, la Vega L de, Ramos P, Zaballo Á, Monzón S, Alonso J. 2016. Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene. *J Hum Genet* 61:463–466.

Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. 1998. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. *J Biol Chem* 273:5858–5868.

Roth DB, Wilson JH. 1986. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol Cell Biol* 6:4295–304.

Rothkamm K, Krüger I, Thompson LH, Löbrich M. 2003. Pathways of DNA double-strand break repair during the mammalian cell cycle. *Mol Cell Biol* 23:5706–15.

Roucher-Boulez F, Mallet D, Chatron N, Dijoud F, Gorduza DB, Bretones P, Morel Y. 2019. Reversion SAMD9 Mutations Modifying Phenotypic Expression of MIRAGE Syndrome and Allowing Inheritance in a Usually de novo Disorder. *Front Endocrinol (Lausanne)* 10:625.

Roukos V, Misteli T. 2014. The biogenesis of chromosome translocations. *Nat Cell Biol* 16:293–300.

ROWLEY JD. 1973. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature* 243:290–293.

Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM. 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res* 20:761–770.

Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science (80-)* 270:467–470.

Schluth-Bolard C, Delobel B, Sanlaville D, Boute O, Cuisset J-M, Sukno S, Labalme A, Duban-Bedu B, Plessis G, Jaillard S, Dubourg C, Henry C, et al. 2009. Cryptic genomic imbalances in de novo and inherited apparently balanced chromosomal rearrangements: array CGH study of 47 unrelated cases. *Eur J Med Genet* 52:291–6.

Schluth-Bolard C, Diguët F, Chatron N, Rollat-Farnier P-A, Bardel C, Afenjar A, Amblard F, Amiel J, Blesson S, Callier P, Capri Y, Collignon P, et al. 2019. Whole genome paired-end sequencing elucidates functional and phenotypic consequences of balanced chromosomal rearrangement in patients with developmental disorders. *J Med Genet* jmedgenet-2018-105778.

Schluth-Bolard C, Labalme A, Cordier M-P, Till M, Nadeau G, Tevissen H, Lesca G, Boutry-Kryza N, Rossignol S, Rocas D, Dubruc E, Edery P, et al. 2013. 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* 50:144–50.

Schluth-Bolard C, Till M, Edery P, Sanlaville D. 2008. Syndromes chromosomiques émergents. *Pathol Biol* 56:380–387.

Schueler MG, Sullivan BA. 2006. Structural and Functional Dynamics of Human Centromeric Chromatin. *Annu Rev Genomics Hum Genet* 7:301–313.

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Haeseler A von, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 15:461–468.

Sellou H, Lebeau-pin T, Chapuis C, Smith R, Hegele A, Singh HR, Kozlowski M, Bultmann S, Ladurner AG, Timinszky G, Huet S. 2016. The poly(ADP-ribose)-dependent chromatin remodeler Alc1 induces local chromatin relaxation upon DNA damage. *Mol Biol Cell* 27:3791–3799.

- Simsek D, Jasin M. 2010. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4–ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol* 17:410–416.
- Singer RH, Green MR. 1997. Compartmentalization of eukaryotic gene expression: causes and effects. *Cell* 91:291–4.
- Slade I, Stephens P, Douglas J, Barker K, Stebbings L, Abbaszadeh F, Pritchard-Jones K, Cole R, Pizer B, Stiller C, Vujanic G, Scott RH, et al. 2010. Constitutional translocation breakpoint mapping by genome-wide paired-end sequencing identifies HACE1 as a putative Wilms tumour susceptibility gene. *J Med Genet* 47:342–7.
- Slamova Z, Nazaryan-Petersen L, Mehrjouy MM, Drabova J, Hancarova M, Marikova T, Novotna D, Vlckova M, Vlckova Z, Bak M, Zemanova Z, Tommerup N, et al. 2018. Very short DNA segments can be detected and handled by the repair machinery during germline chromothriptic chromosome reassembly. *Hum Mutat* 39:709–716.
- Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, LePichon J-B, Miller NA, Thiffault I, Dinwiddie DL, Twist G, Noll A, et al. 2014. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci Transl Med* 6:265ra168.
- Solovei I, Kreysing M, Lanctôt C, Kösem S, Peichl L, Cremer T, Guck J, Joffe B. 2009. Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. *Cell* 137:356–368.
- Soutoglou E, Dorn JF, Sengupta K, Jasin M, Nussenzweig A, Ried T, Danuser G, Misteli T. 2007. Positional stability of single double-strand breaks in mammalian cells. *Nat Cell Biol* 9:675–682.
- Sperling K, Rao PN. 1974. The phenomenon of premature chromosome condensation: its relevance to basic and applied research. *Hum Genet* 23:235–258.
- Spielmann M, Lupiáñez DG, Mundlos S. 2018. Structural variation in the 3D genome. *Nat Rev Genet* 19:453–467.
- Stankiewicz P, Lupski JR. 2002. Genome architecture, rearrangements and genomic disorders. *Trends Genet* 18:74–82.
- Startek M, Szafranski P, Gambin T, Campbell IM, Hixson P, Shaw CA, Stankiewicz P, Gambin A. 2015. Genome-wide analyses of LINE–LINE-mediated nonallelic homologous recombination. *Nucleic Acids Res* 43:2188–2198.
- Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin M-L, et al. 2011. Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* 144:27–40.
- Storchová Z, Kloosterman WP. 2016. The genomic characteristics and cellular origin of chromothripsis. *Curr Opin Cell Biol* 40:106–113.
- Stracker TH, Petrini JHJ. 2011. The MRE11 complex: starting from the ends. *Nat Rev Mol Cell Biol* 12:90–103.
- Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Hsi-Yang Fritz M, Konkel MK, Malhotra A, et al. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature* 526:75–81.
- Sun H Bin, Shen J, Yokota H. 2000. Size-Dependent Positioning of Human Chromosomes in Interphase Nuclei. *Biophys J* 79:184–190.
- Sung P, Klein H. 2006. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol* 7:739–50.

- Suzuki E, Shima H, Toki M, Hanew K, Matsubara K, Kurahashi H, Narumi S, Ogata T, Kamimaki T, Fukami M. 2016. Complex X-Chromosomal Rearrangements in Two Women with Ovarian Dysfunction: Implications of Chromothripsis/Chromoanasythesis-Dependent and -Independent Origins of Complex Genomic Alterations. *Cytogenet Genome Res* 150:86–92.
- Swystun LL, James PD. 2017. Genetic diagnosis in hemophilia and von Willebrand disease. *Blood Rev* 31:47–56.
- Szabo Q, Bantignies F, Cavalli G. 2019. Principles of genome folding into topologically associating domains. *Sci Adv* 5:eaaw1668.
- Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, Kirby A, Liu S, Muddukrishna B, Ohsumi TK, Shen Y, Borowsky M, et al. 2011. Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am J Hum Genet* 88:469–81.
- Tartof KD, Hobbs C, Jones M. 1984. A structural basis for variegating position effects. *Cell* 37:869–78.
- Taşkesen M, Collin GB, Evsikov A V., Güzel A, Özgül RK, Marshall JD, Naggert JK. 2012. Novel Alu retrotransposon insertion leading to Alström syndrome. *Hum Genet* 131:407–413.
- Terradas M, Martín M, Genescà A. 2016. Impaired nuclear functions in micronuclei results in genome instability and chromothripsis. *Arch Toxicol* 90:2657–2667.
- Terzoudi GI, Karakosta M, Pantelias A, Hatzi VI, Karachristou I, Pantelias G. 2015. Stress induced by premature chromatin condensation triggers chromosome shattering and chromothripsis at DNA sites still replicating in micronuclei or multinucleate cells when primary nuclei enter mitosis. *Mutat Res Genet Toxicol Environ Mutagen* 793:185–98.
- The 1000 Genomes Project Consortium . 2015. A global reference for human genetic variation. *Nature* 526:68–74.
- Tijo JH, Levan A. 1956. The chromosome number of Man. *Hereditas* 42:1–6.
- Tipton AR, Wang K, Oladimeji P, Sufi S, Gu Z, Liu S-T. 2012. Identification of novel mitosis regulators through data mining with human centromere/kinetochore proteins as group queries. *BMC Cell Biol* 13:15.
- Titen SWA, Golic KG. 2008. Telomere loss provokes multiple pathways to apoptosis and produces genomic instability in *Drosophila melanogaster*. *Genetics* 180:1821–32.
- Tubio JMC, Estivill X. 2011. When catastrophe strikes a cell. *Nature* 470:476–477.
- Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, Beck S, Hurles ME. 2008. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet* 40:90–5.
- Van Meter M, Simon M, Tomblin G, May A, Morello TD, Hubbard BP, Bredbenner K, Park R, Sinclair DA, Bohr VA, Gorbunova V, Seluanov A. 2016. JNK Phosphorylates SIRT6 to Stimulate DNA Double-Strand Break Repair in Response to Oxidative Stress by Recruiting PARP1 to DNA Breaks. *Cell Rep* 16:2641–2650.
- Vanneste E, Voet T, Caignec C Le, Ampe M, Konings P, Melotte C, Debrock S, Amyere M, Vikkula M, Schuit F, Fryns J-P, Verbeke G, et al. 2009. Chromosome instability is common in human cleavage-stage embryos. *Nat Med* 15:577–583.
- Vázquez-Diez C, FitzHarris G. 2018. Correlative Live Imaging and Immunofluorescence for Analysis of Chromosome Segregation in Mouse Preimplantation Embryos. *Methods in Molecular Biology* (Clifton, N.J.), p 319–335.

- Velagaleti GVN, Bien-Willner GA, Northup JK, Lockhart LH, Hawkins JC, Jalal SM, Withers M, Lupski JR, Stankiewicz P. 2005. Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. *Am J Hum Genet* 76:652–62.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, et al. 2001. The Sequence of the Human Genome. *Science* (80-) 291:1304–1351.
- Verschure PJ, Kraan I van Der, Manders EM, Driel R van. 1999. Spatial relationship between transcription sites and chromosome territories. *J Cell Biol* 147:13–24.
- Vignoli A, Bisulli F, Darra F, Mastrangelo M, Barba C, Giordano L, Turner K, Zambrelli E, Chiesa V, Bova S, Fiocchi I, Peron A, et al. 2016. Epilepsy in ring chromosome 20 syndrome. *Epilepsy Res* 128:83–93.
- Visser AE, Jaunin F, Fakan S, Aten JA. 2000. High resolution analysis of interphase chromosome domains. *J Cell Sci* 113 (Pt 14):2585–93.
- Voet T, Kumar P, Loo P Van, Cooke SL, Marshall J, Lin M-L, Zamani Esteki M, Aa N Van der, Mateiu L, McBride DJ, Bignell GR, McLaren S, et al. 2013. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res* 41:6119–38.
- Vogel MJ, Peric-Hupkes D, Steensel B van. 2007. Detection of in vivo protein–DNA interactions using DamID in mammalian cells. *Nat Protoc* 2:1467–1478.
- Vogt J, Bengesser K, Claes KB, Wimmer K, Mautner V-F, Minkelen R van, Legius E, Brems H, Upadhyaya M, Högel J, Lazaro C, Rosenbaum T, et al. 2014. SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. *Genome Biol* 15:R80.
- Voisin N, Schnur RE, Douzgou S, Hiatt SM, Rustad CF, Brown NJ, Earl DL, Keren B, Levchenko O, Geuer S, Amor D, Brusco A, et al. 2019. Variants in the degron of AFF3 cause a multi-system disorder with mesomelic dysplasia, horseshoe kidney and developmental and epileptic encephalopathy. *bioRxiv* 693937.
- Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS. 1991. A de novo Alu insertion results in neurofibromatosis type 1. *Nature* 353:864–866.
- Wang J-C, Fisker T, Sahoo T. 2015. Constitutional chromothripsis involving chromosome 19 in a child with subtle dysmorphic features. *Am J Med Genet A* 167A:910–3.
- Wang L, Norris ET, Jordan IK. 2017. Human Retrotransposon Insertion Polymorphisms Are Associated with Health and Disease via Gene Regulatory Phenotypes. *Front Microbiol* 8:1418.
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh T-Y, Peng W, Zhang MQ, Zhao K. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40:897–903.
- Warburton D. 1991. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet* 49:995–1013.
- Waters LS, Walker GC. 2006. The critical mutagenic translesion DNA polymerase Rev1 is highly expressed during G(2)/M phase rather than S phase. *Proc Natl Acad Sci U S A* 103:8971–6.
- Weckselblatt B, Hermetz KE, Rudd MK. 2015. Unbalanced translocations arise from diverse mutational mechanisms including chromothripsis. *Genome Res* 25:937.
- Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017. Direct determination of diploid

genome sequences. *Genome Res* 27:757–767.

Will AJ, Cova G, Osterwalder M, Chan W-L, Wittler L, Brieske N, Heinrich V, Villartay J-P de, Vingron M, Klopocki E, Visel A, Lupiáñez DG, et al. 2017. Composition and dosage of a multipartite enhancer cluster control developmental expression of *Ihh* (Indian hedgehog). *Nat Genet* 49:1539–1545.

Wilson JH, Berget PB, Pipas JM. 1982. Somatic cells efficiently join unrelated DNA segments end-to-end. *Mol Cell Biol* 2:1258–69.

Wold MS. 1997. REPLICATION PROTEIN A: A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. *Annu Rev Biochem* 66:61–92.

Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, Zhou Q, Kirkness EF, Levy S, Batzer MA, Jorde LB. 2009. Mobile elements create structural variation: Analysis of a complete human genome. *Genome Res* 19:1516–1526.

Yan CT, Boboila C, Souza EK, Franco S, Hickernell TR, Murphy M, Gumaste S, Geyer M, Zarrin AA, Manis JP, Rajewsky K, Alt FW. 2007. IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* 449:478–482.

Yang J, Deng G, Cai H. 2015. ChromothripsisDB: a curated database of chromothripsis. *Bioinformatics*.

Zakharyevich K, Ma Y, Tang S, Hwang PY-H, Boiteux S, Hunter N. 2010. Temporally and Biochemically Distinct Activities of Exo1 during Meiosis: Double-Strand Break Resection and Resolution of Double Holliday Junctions. *Mol Cell* 40:1001–1015.

Zanardo ÉA, Piazzon FB, Dutra RL, Dias AT, Montenegro MM, Novo-Filho GM, Costa TVMM, Nascimento AM, Kim CA, Kulikowski LD. 2014. Complex structural rearrangement features suggesting chromoanagenesis mechanism in a case of 1p36 deletion syndrome. *Mol Genet Genomics* 289:1037–1043.

Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. 2009. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet* 41:849–53.

Zhang XA, Yates A, Vasilevsky N, Gouridine JP, Callahan TJ, Carmody LC, Danis D, Joachimiak MP, Ravanmehr V, Pfaff ER, Champion J, Robasky K, et al. 2019. Semantic integration of clinical laboratory tests from electronic health records for deep phenotyping and biomarker discovery. *npj Digit Med* 2:32.

Zhang Y, Gostissa M, Hildebrand DG, Becker MS, Boboila C, Chiarle R, Lewis S, Alt FW. 2010. The Role of Mechanistic Factors in Promoting Chromosomal Translocations Found in Lymphoid and Other Cancers. *Advances in Immunology*, p 93–133.

Zhu C, Mills KD, Ferguson DO, Lee C, Manis J, Fleming J, Gao Y, Morton CC, Alt FW. 2002. Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell* 109:811–21.

Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q, Heinzen EL, Cirulli ET, Pelak K, He M, Ruzzo EK, Gumbs C, et al. 2012. Using ERDS to infer copy-number variants in high-coverage genomes. *Am J Hum Genet* 91:408–21.

Zufferey M, Tavernari D, Oricchio E, Ciriello G. 2018. Comparison of computational methods for the identification of topologically associating domains. *Genome Biol* 19:217.