



**HAL**  
open science

# Exploration des troubles du neurodéveloppement : caractérisation des anneaux du chromosome 22 chez des patients atteints du syndrome de Phelan-McDermid et implication des gènes des contactines dans les troubles du neurodéveloppement

Myriam Rachid

## ► To cite this version:

Myriam Rachid. Exploration des troubles du neurodéveloppement : caractérisation des anneaux du chromosome 22 chez des patients atteints du syndrome de Phelan-McDermid et implication des gènes des contactines dans les troubles du neurodéveloppement. Génétique. Université Paris Cité, 2023. Français. NNT : 2023UNIP5286 . tel-04758059

**HAL Id: tel-04758059**

**<https://theses.hal.science/tel-04758059v1>**

Submitted on 29 Oct 2024

**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.

# Université Paris Cité

## Institut PASTEUR

École doctorale Bio Sorbonne Paris Cité (ED 562)

*Laboratoire de Génétique Humaine et Fonctions Cognitives*

### Exploration des troubles du neurodéveloppement :

*caractérisation des anneaux du chromosome 22 chez des patients atteints du syndrome de Phelan-McDermid et implication des gènes des contactines dans les troubles du neurodéveloppement*

Par Myriam RACHID

Thèse de doctorat de Génétique

Dirigée par Anne-Claude NOUCHI TABET  
Et co-encadrée par Claire LEBLOND

Soutenue publiquement le 11 Décembre 2023

Devant un jury composé de :

**Directrice de thèse** Anne-Claude NOUCHI TABET, Pr Ass-PH, CHU Robert Debré, Université Paris Cité

**Rapporteur** Franck PELLESTOR, PU-PH, CHU de Montpellier, Université de Montpellier

**Rapporteuse** Caroline DEMILY, PU-PH, CH le Vinatier, Université Lyon 1

**Examineur** Thomas BOURGERON, PhD, HDR, Institut Pasteur, Université Paris Cité

**Examineur** Patrick CALLIER, PU-PH, CHU Dijon-Bourgogne, Université de Dijon

**Examinatrice** Catherine YARDIN, PU-PH, CHU de Limoges, Université de Limoges

**Membre du jury invité, co-encadrante de thèse** Claire LEBLOND, MCU, Institut Pasteur, Université Paris Cité



**Titre :** Exploration des troubles du neurodéveloppement : caractérisation des anneaux du chromosome 22 chez des patients atteints du syndrome de Phelan-McDermid et implication des gènes des contactines dans les troubles du neurodéveloppement

**Résumé :**

Les troubles du neurodéveloppement (TND) touchent 9 à 18% de la population générale, et sont donc un enjeu majeur en santé publique. Les causes génétiques sont très variées, faisant parfois intervenir des mécanismes d'épistasie et rendant le diagnostic moléculaire difficile. Les gènes associés aux TND jouent un rôle majeur dans trois mécanismes biologiques principaux : le remodelage de la chromatine, la synthèse protéique et le fonctionnement de la synapse et la transmission de l'influx nerveux.

Mon premier objectif a été d'étudier l'apport du séquençage du génome entier dans l'identification des mécanismes chromosomiques responsables d'un syndrome de Phelan-McDermid (PMS) par formation d'un anneau du chromosome 22 (r(22)). La détection des r(22) est cruciale dans le parcours de soin du fait du risque de développer une neurofibromatose de type 2 suite à la perte somatique de l'anneau et donc du gène *NF2*. L'analyse des génomes de 7 patients porteurs d'un r(22) a permis de préciser les points de cassure terminaux au niveau de la formation des anneaux, mais également d'identifier des variations additionnelles, ponctuelles ou de structures, qui pourraient influencer sur le phénotype des patients. Par ailleurs, une analyse systématique des points de cassure en 22qter utilisant la récente version du génome *Telomere-to-telomere* (T2T), nous a permis de mettre en évidence un anneau additionnel du chromosome 22 chez un patient avec PMS, jusque-là non identifiable via les pipelines d'analyses utilisant la référence GRCh37 ou 38. La mise au point a d'abord été manuelle en visualisant des splits-reads et des reads discordants s'alignant à la fois dans la région 22qter et dans la région du bras court d'un acrocentrique chez les patients porteurs d'un anneau connu. Cette mise au point nous a permis de développer un outil en python capable d'identifier de façon automatique ces alignements 22qter-bras court d'acrocentrique et pourra à terme être utilisé pour identifier les anneaux des 5 paires de chromosomes acrocentriques. Ces nouvelles avancées pourraient permettre de placer le séquençage de génome entier en première intention avec la possibilité d'identifier des anomalies jusque-là mises en évidence par les seules techniques de cytogénétique conventionnelle (Caryotype et FISH).

Le deuxième objectif de ma thèse se pose dans la continuité de travaux ayant mis en évidence une association entre des variations des gènes *CNTN5* et *CNTN6* et le risque de troubles neurosensoriels chez des patients autistes. Nous avons analysé le lien entre la présence de variations touchant au moins l'un des gènes de la famille des contactines et des *contactin associated proteins* et l'apparition de TND.

Cette étude a porté sur l'analyse des variations du nombre de copies (CNV) dans une cohorte de 7 495 individus ayant eu une puce à ADN de type SNP. L'analyse ne révèle pas d'association significative avec les TND chez 140/5 812 patients porteurs et 36/1 683 contrôles. Toutefois, ces résultats sont à considérer avec prudence puisque seul le trait phénotypique principal était disponible (trouble du développement intellectuel, autisme, malformation), sans le détail du phénotype global et des potentielles comorbidités. Les conséquences d'un variant touchant les contactines sont potentiellement associées à un phénotype plus fin, tel que l'hyperacousie. De plus, seuls les CNVs ont été étudiés, une analyse complémentaire portant sur l'ensemble des variations génomiques devrait être réalisée afin d'avoir une vue plus complète sur le rôle des contactines et des protéines associées dans le neurodéveloppement.

**Mots clefs** : Troubles du neurodéveloppement, séquençage du génome entier (WGS), anomalies de structure, anneaux du chromosome 22, contactines, syndrome de Phelan-McDermid, Telomere-to-telomere (T2T), chromosomes acrocentriques

## **Title : Exploration of neurodevelopmental disorders: characterization of chromosome 22 rings in patients with Phelan-McDermid syndrome and involvement of contactin genes in neurodevelopmental disorders**

### **Abstract :**

Neurodevelopmental disorders (NDD) affect 9-18% of the general population, making them a major public health issue. The genetic causes are highly diverse, sometimes involving epistasis mechanisms and making molecular diagnosis difficult. NDD-associated genes play a major role in three main biological mechanisms: chromatin remodeling, protein synthesis and synapse function, and nerve impulse transmission.

My first objective was to study the contribution of whole-genome sequencing in identifying the chromosomal mechanisms responsible for Phelan-McDermid syndrome (PMS) through the formation of a ring on chromosome 22 (r(22)). Detection of r(22) is crucial in healthcare pathways, given the risk of developing neurofibromatosis type 2 due to the somatic loss of the ring and therefore of the NF2 gene. Analysis of the genomes of 7 patients with r(22) has made it possible to pinpoint the terminal breakpoints in ring formation, but also to identify additional point or structural variations that could influence the patients' phenotype. In addition, a systematic analysis of 22qter breakpoints using the recent version of the Telomere-to-telomere (T2T) genome, enabled us to identify an additional ring on chromosome 22 in a patient with PMS, previously undetectable via analysis pipelines using GRCh37 or 38. First, manual detection has consisted of visualizing discordant splits-reads and reads aligning both in the 22qter region and in the short arm region of an acrocentric in patients with a known ring. Then, we developed a Python tool able to automatically identify these 22qter-short-arm acrocentric alignments, which could eventually be used to identify the rings from any acrocentric chromosomes. These new advances could place whole-genome sequencing as the first-tier test to identify abnormalities previously detected only by conventional cytogenetic techniques (karyotype and FISH).

The second objective of my thesis follows on a work that has shown an association between variations in the *CNTN5* and *CNTN6* genes and the risk of neurosensory disorders in autistic patients. Thus, we analyzed the link between the presence of variations altering any genes from contactin or contactin associated proteins families and the onset of NDD.

This study involved the analysis of copy number variations (CNVs) in a cohort of 7,495 individuals detected by SNP array. The analysis revealed no significant association with NDD in 140/5,812 carrier patients and 36/1,683 controls. However, these results should be considered with caution as only the main phenotypic trait was available (intellectual disability, autism, malformation), without details of the overall phenotype

and potential comorbidities. The consequences of a contactin variant are potentially associated with a more refined phenotype, such as hyperacusis. Moreover, only CNVs have been studied, and a complementary analysis of all genomic variations should be carried out to gain a more comprehensive view of the role of the contactins and the contactin associated proteins in neurodevelopment.

**Keywords** : Neurodevelopmental disorder, whole genome sequencing (WGS), structural variations, ring chromosome 22, contactins, Phelan-McDermid syndrome, Telomere-to-telomere (T2T), acrocentric chromosomes

# Remerciements

Je tiens à adresser mes premiers remerciements aux membres de mon jury. Au Pr Caroline Demily et au Pr Franck Pellestor, qui ont accepté d'être rapporteurs de ce projet. Au Pr Catherine Yardin et au Pr Patrick Callier qui ont accepté d'être membre de ce jury. Merci pour le temps que vous aurez consacré à mon projet.

Je tiens à remercier chaleureusement ma directrice de thèse, Anne-Claude Tabet, pour m'avoir proposé ce projet et pour sa supervision tout au long de ces quatre années. Merci aussi de m'avoir accueillie en tant qu'AHU dans cette belle équipe à l'hôpital Robert Debré.

Je tiens également à remercier sincèrement Thomas Bourgeron de m'avoir accueillie dans son unité de recherche, et de m'avoir offert l'opportunité de mener ce beau projet au sein de l'équipe de GHFC.

Je souhaite exprimer ma profonde gratitude à Claire Leblond pour m'avoir encadré, pour ses conseils éclairés et son engagement dans la réussite de ce travail de recherche. Ses contributions ont été essentielles à la réalisation de cette thèse.

Je remercie Benjamin pour avoir permis que cette thèse aboutisse! Pour toutes ces petites choses qu'on ne voit pas mais qui font que tout fonctionne. Pour toutes les petites pauses et moments de détente qui sont importants pour que le moral tienne! Merci aussi de m'avoir fait prendre conscience des choses importantes dans la vie. Et je n'oublie pas que je suis tombée dans le monde de Teyvat grâce ou à cause de toi, alors je ne manquerai pas de te donner les dernières nouvelles croustillantes!

Un grand merci à Alexandre et Aline V. qui se sont tous les deux investis dans mon projet et m'ont aidée à y voir le bout!

Aline, tu as été (malgré tout!) ma compagne de bureau ces dernières années, et je te remercie pour ta gentillesse, pour m'avoir encouragée et pour tous les petits moments hors du travail. Merci pour tout le mal que tu t'es donné pour moi surtout cette dernière année. Je te souhaite le meilleur pour la suite!

Alex, tu as cru en ce projet et on a réussi à le mener à bout! Merci à toi pour tout, pour l'aide, l'écoute et la gentillesse que tu m'as offertes. J'espère que le soleil du sud vous apportera ce qu'il y a de mieux!

Nathalie, tu as été un grand soutien ces quatre dernières années. Merci pour le temps passé, ta patience qui a réussi à me supporter tout ce temps, et ta gentillesse qui m'ont



aidé à tenir bon. Merci de m'avoir laissé occuper un petit coin de ton labo pour me concentrer dans la dernière ligne droite, et d'avoir pensé à me nourrir sainement!

Je tiens également à remercier Sabrina, Fredo, Cécile, Lisa, Benoît, pour votre gentillesse et tous les bons moments que vous avez contribué à apporter au labo.

A Robin, merci pour nos échanges, qu'il s'agisse de râler ou de refaire la France gauloise avec Alex!

Merci à toute l'équipe de GHFC, à Freddy pour ton aide et pour toutes les choses que tu as pris le temps de m'apprendre, mais aussi à Aline L, Thomas R et Eli. Merci aussi à Jean-Pierre Bourgeois pour nos échanges et son aide à propos des synapses.

J'ai une pensée pour Amaury, Julien et Simon avec qui j'ai partagé de très moments, que ce soit à la ludothèque ou ailleurs, tout ce temps passé à l'Institut Pasteur.

Un grand merci aussi à une équipe qui m'a donné du courage à l'hôpital. Céline, merci pour tout, pour ton aide, pour ton écoute, pour ton amitié. Elodie et Peggy merci de m'avoir écoutée et fait sourire, vos touches d'humour vont me manquer. Merci Delphine pour les passions que l'on partage et tous les memes échangés, Edwina pour les fous rires, Sabatini pour ton énergie et la motivation que tu m'as donnée. Benitha merci pour toutes nos confidences et nos fous rires aussi. Elka, merci pour ton investissement et m'avoir motivée à apprendre une nouvelle langue (Благодаря много). Kate, merci pour ta patience et ton écoute. Cécile merci pour ta rigueur et ta gentillesse. Aurélie E, merci pour toutes nos petites discussions passionnées. Sandra merci pour ton écoute.

Merci à toutes pour votre gentillesse et pour m'avoir sans cesse remotivée!

J'ai une pensée aussi pour Aurélie A avec qui j'ai apprécié travailler et Evelyne qui m'a appris les bases de la cytogénétique. A Brigitte et Eva qui m'ont fait découvrir cette fabuleuse discipline qu'est la cytogénétique.

Un grand merci à Guillaume, Sophie, Bichr, Issam, Fabien, Suonavy, Marie-Hélène, Giulia, Adeline, Lyse et Louise avec qui j'ai pris plaisir à travailler.

A Angélique qui est partie trouver sa Bretagne, je te remercie pour tous tes encouragements et l'amitié qui s'est tissée au fil de nos gardes. A Léonie qui est toujours là depuis que l'on s'est rencontrées au début de mon internat, merci pour tous les bons moments que l'on a passés ensemble.

Aux internes, dont le passage est éphémère mais qui laissent un souvenir fort, et ont adoucies ces journées de travail. Bérénice, pour tous nos échanges littéraires et culturels. Cyrielle pour ta bonne humeur tout le long de ces six mois. Cécile pour ta douceur et ta gentillesse. Brian, pour la dose de bonne humeur également et l'énergie

que tu as apportée. Sixtine pour ce dernier semestre plein d'émotions, et toutes les solutions que tu sais trouver à chaque problème. Merci à vous!

Un grand merci à ma grande famille! A ma grand-mère, à mon parrain, mes tantes, mes oncles. Rachel, Agnès, Réjane, Philippe, Jacqueline, Fatiha, Boualem, Larbi, Salem, Idris, merci à vous d'être toujours là. A Chloé ma très chère consœur et Camille ma jolie *fleur*, à Natan, Clara, Yohan, Sophie, Marion, Christine, Loïc, Corinne, Eric G, Carole et Eric B.

A mes amis qui sont toujours à mes côtés depuis toutes ces années, merci à vous. A Laura qui m'a donné du courage et Marion qui m'a donné l'inspiration d'aller plus loin. A Alexandra qui est là depuis le début et me soutient toujours, et Karima qui m'a offert la plus belle preuve de confiance (et j'ai une pensée pour mon filleul Jules et son frère Louis). A Anar qui m'a soutenue et offert une précieuse écoute, et à Sara pour sa gentillesse et son énergie.

Aïcha, merci pour tes mots toujours sages et tes encouragements.

Cathy, merci pour ta gentillesse et ta générosité, et pour ces moments partagés à Teyvat et dans la vie. Merci à Tony, Ori et TJ qui ont été là pour m'aider à m'aérer l'esprit et oublier le travail quelques instants.

Aloïs, tu as été un grand soutien cette dernière année, merci pour ces moments cocooning que tu m'as offert, et tous nos échanges!

Erwan, merci d'avoir été là, de m'avoir motivée à apprendre tant de nouvelles choses, et pour m'avoir aidé à garder le moral. Et n'oublie pas d'être une *fleur* pour moi.

Ma plus grande pensée va vers mes parents, qui ont toujours été là pour moi, et qui sont mon plus grand soutien depuis toujours. Merci à vous de croire en moi et de me pousser vers le meilleur.

# Liste des principales abréviations

## A

<b>ABM</b>	Agence de Biomédecine
<b>ACMG</b>	American College of Medical Genetics
<b>ACPA</b>	Analyse chromosomique par puce à ADN
<b>ADI-R</b>	Autism Diagnostic Interview Revised
<b>ADOS</b>	Autism Diagnostic Observation Schedule
<b>AP</b>	Associated Proteins

## B

<b>BAF</b>	B Allele Frequency
<b>BCCA</b>	Bras courts des chromosomes acrocentriques
<b>BLAT</b>	BLAST-Like Alignment Tool

## C

<b>CGH</b>	Comparative genomic hybridization
<b>CIA</b>	CNV inheritance analysis
<b>CNV</b>	Copy Number Variation
<b>CRAM</b>	Compressed Reference-oriented Alignment Map
<b>Ctrl</b>	Contrôles

## D

<b>DI</b>	Déficiência intellectuelle
<b>DI Iso</b>	Déficiência intellectuelle isolée
<b>DI Sd</b>	Déficiência intellectuelle syndromique
<b>DSM-V</b>	Diagnostic And Statistical Manual Of Mental Disorders 5ème édition

## F

<b>FISH</b>	Fluorescence in situ hybridization
<b>FoSTeS</b>	Fork stalling and template switching

## G

<b>GRCh</b>	Genome Reference Consortium Human
<b>GPI</b>	Glycosylphosphatidylinositol
<b>GTG</b>	Giemsa

## H

**HGP** Human Genome Project  
**HC-NDD** High confidence neurodevelopmental

## I

**IC** Intervalle de confiance  
**IgCAM** immunoglobulins cell adhesions molecule  
**IGV** Integrative Genomics Viewer  
**Indel** insertions/délétion  
**Inv dup del** Inverted duplication/deletion

## L

**LCR** low copy repeats  
**LoF** Loss of Function  
**LRR** Log R Ratio

## M

**MCA** Multiple congenital anomalies  
**MLPA** Multiplex Ligation-dependent Probe Amplification  
**MMBIR** Microhomology-Mediated Break-Induced Replication  
**mTOR** mechanistic/mammalian target of rapamycin

## N

**NAHR** Non Allelic Homologous Recombination  
**NF** Neurofibromatose  
**NHEJ** Non Homologous End Joining  
**NOR** Nucleolus Organizer Regions  
**NRXN** Neurexine

## O

**OR** Odd Ratio

## P

**PHR** Pseudo Homologous Regions  
**PIEV** Pénétrance incomplète et expressivité variable  
**PMS** Syndrome de Phelan-McDermid

## Q

**QI** Quotient intellectuel

## R

**rDNA** ADN ribosomal

## S

<b>SFARI</b>	Simons Foundation Autism Research Initiative
<b>SHANK</b>	SH3 and multiple ankyrin repeat domains
<b>SHD</b>	Séquençage à haut débit
<b>SNP</b>	Single Nucleotid Polymorphism
<b>SNV</b>	Single Nucleotid Variant
<b>SPARK</b>	Simons Foundation Powering Autism Research for Knowledge
<b>SRS</b>	Social Responsiveness Scale
<b>SSC</b>	Simons Simplex Collection

## T

<b>T2T</b>	Telomere-to-Telomere
<b>TDAH</b>	Troubles déficitaires de l'attention avec ou sans hyperactivité
<b>TDI</b>	Trouble du Développement intellectuel
<b>TND</b>	Trouble du neurodéveloppement
<b>TSA</b>	Trouble du spectre autistique

## U

<b>UTR</b>	Untranslated Transcribed Region
------------	---------------------------------

## V

<b>VSI</b>	Variation de signification inconnue
------------	-------------------------------------

## W

<b>WES</b>	Whole exome sequencing
<b>WGS</b>	Whole genome sequencing
<b>WF</b>	Wave Factor

# Sommaire

<b>REMERCIEMENTS.....</b>	<b>7</b>
---------------------------	----------

<b>SOMMAIRE.....</b>	<b>13</b>
----------------------	-----------

<b>CHAPITRE 1 : INTRODUCTION.....</b>	<b>16</b>
---------------------------------------	-----------

<b>1. TROUBLES DU NEURODEVELOPPEMENT.....</b>	<b>16</b>
1.1. DEFINITION.....	16
1.2. INCIDENCE ET PREVALENCE.....	17
1.3. DIAGNOSTIC.....	18
<b>2. ARCHITECTURE GENETIQUE DES TND.....</b>	<b>20</b>
2.1. LES ANOMALIES CHROMOSOMIQUES ASSOCIEES.....	20
2.2. TAUX DE DIAGNOSTIC.....	21
2.3. GENES IMPLIQUES DANS LES TND ET VARIATIONS GENOMIQUES.....	22
2.4. VOIES COURAMMENT IMPLIQUEES DANS LES TND.....	24
2.4.1. LA SYNAPSE.....	24
2.4.2. REMODELAGE DE LA CHROMATINE.....	26
2.4.3. VOIE <i>mTOR</i> ET SYNTHESE PROTEIQUE.....	28
2.5. ARCHITECTURE GENETIQUE DES TND.....	29
2.6. EXEMPLE DE DEUX FAMILLES DE GENES.....	30
2.6.1. <b><i>PATHOLOGIE A PENETRANCE FORTE MAIS EXPRESSIVITE VARIABLE : SHANKOPATHIES</i></b>	
30	
2.6.1.1. DESCRIPTION DE LA FAMILLE SHANK.....	30
2.6.1.2. SHANK3 ET SYNDROME DE PHELAN-MCDERMID.....	33
2.6.1.3. GENETIQUE DU PMS.....	34
2.6.1.4. PARTICULARITE DES ANNEAUX DU CHROMOSOME 22.....	36
2.6.2. <b><i>PATHOLOGIE A PENETRANCE VARIABLE : FAMILLE DES CONTACTINES</i></b>	38
2.6.2.1. LES GENES DES CONTACTINES.....	38
2.6.2.2. LES CONTACTINES ASSOCIATED PROTEINS.....	40
2.6.2.3. LES ANOMALIES DES CONTACTINES DANS LA LITTERATURE.....	42
<b>3. VARIATIONS GENOMIQUES ET TND.....</b>	<b>45</b>
3.1. LES ANOMALIES CHROMOSOMIQUES.....	45
3.1.1. TYPES D'ANOMALIES.....	45
3.1.1.1. LES ANOMALIES DE NOMBRE OU ANEUPLOÏDIES.....	45
3.1.1.2. LES ANOMALIES DE STRUCTURE.....	45
3.1.2. PRINCIPAUX MECANISMES DE FORMATION.....	47
3.1.2.1. REMANIEMENTS RECURRENTS.....	47
3.1.2.2. REMANIEMENTS NON-RECURRENTS.....	48
3.1.3. FORMATION DES <i>INV DUP DEL</i> .....	50
3.1.4. FORMATION DES ANNEAUX CHROMOSOMIQUES.....	53
3.1.4.1. DESCRIPTION.....	53
3.1.4.2. INSTABILITE MITOTIQUES ET LEURS CONSEQUENCES.....	54
3.2. CHROMOSOMES ACROCENTRIQUES.....	54
3.3. VARIATIONS NUCLEOTIDIQUES.....	55
3.4. CRITERES DE PATHOGENICITE.....	56
<b>4. LES OUTILS DIAGNOSTICS.....</b>	<b>57</b>
4.1. CYTOGENETIQUE CONVENTIONNELLE ET MOLECULAIRE.....	57
4.1.1. CARYOTYPE ET <i>FISH</i> .....	57
4.1.2. ANALYSE CHROMOSOMIQUE PAR PUCE A ADN.....	58
4.2. SEQUENÇAGE HAUT DEBIT.....	59

4.2.1. PRINCIPE DU SEQUENÇAGE HAUT DEBIT .....	59
4.2.2. SEQUENÇAGE CIBLE ET GENOME ENTIER .....	62
4.2.3. OUTILS D'ANALYSES EN WGS .....	62
4.2.4. DEUX TYPES DE SHD : SHORT VS LONG READS.....	65
4.2.5. GENOME DE REFERENCES.....	65
4.2.5.1. GRCH37 ET 38.....	65
4.2.5.2. TELOMERE TO TELOMERE .....	66

## **CHAPITRE 2 : ANNEAUX DES CHROMOSOMES 22 ET DONNEES DE WGS..... 68**

1. OBJECTIFS.....	68
2. METHODOLOGIE .....	69
2.1. COHORTE DE PATIENTS AVEC PMS .....	69
2.2. SEQUENÇAGE DE GENOME ENTIER .....	70
2.3. SEQUENÇAGE SANGER.....	71
2.4. RECHERCHE DE REGIONS D'HOMOLOGIE .....	72
3. ARTICLE – <i>GOLLUM, A NEW TOOL USING SHORT READ SEQUENCING TO DETECT RING CHROMOSOMES IN PMS PATIENTS</i> .....	73
4. PMS ASSOCIE AUX REMANIEMENTS COMPLEXES DU CHROMOSOME 22 .....	94
4.1. UNE INV DUP DEL 22 CLASSIQUE.....	94
4.2. INV DUP DEL COMPLEXE .....	96
4.3. DERIVE 13 D'UNE TRANSLOCATION T(13;22)(P13;Q13.3).....	101
4.4. DERIVE D'INVERSION DU CHROMOSOME 22 .....	102
5. ÉLÉMENTS REPETES A L'ORIGINE DES REARRANGEMENTS .....	104
6. VARIABILITE CLINIQUE ET GENETIQUE DU PMS : VARIATIONS ADDITIONNELLES .....	107
6.1. CHEZ DES PATIENTS PORTEURS D'ANNEAU DU CHROMOSOME 22 .....	107
6.1.1. VARIABILITE CLINIQUE.....	107
6.1.2. VARIATIONS GENOMIQUES ADDITIONNELLES MODULANT LE PHENOTYPE .....	110
6.2. VARIABILITE CLINIQUE ET GENETIQUE DE PATIENTS PMS NON PORTEURS D'ANNEAU DU CHROMOSOME 22 .....	113

## **CHAPITRE 3 : ÉTUDE DES FAMILLES DE GENES DES CONTACTINES ..... 119**

1. OBJECTIFS.....	119
2. METHODES D'ANALYSE DE LA COHORTE DE PUCE A ADN.....	120
2.1. COHORTE CLINIQUE FRANÇAISE NDD.....	120
2.2. COHORTES AMERICAINES D'INDIVIDUS AUTISTES ET LEURS APPARENTES .....	121
2.3. ACPA ET ANALYSE DES CNVs.....	122
2.4. ANALYSES STATISTIQUES.....	123
3. LES CNVs IMPACTANT LES GENES DES CONTACTINES .....	125
3.1. ANALYSE GLOBALE DES CNVs.....	125
3.2. ANALYSE DES DONNEES EN FONCTION DU PHENOTYPE.....	130
3.3. META-ANALYSE.....	134

## **CHAPITRE 4 : DISCUSSION..... 137**

## **BIBLIOGRAPHIE ..... 141**

## **TABLE DES ILLUSTRATIONS..... 163**

## **TABLES..... 165**





# Chapitre 1 : Introduction

## 1. Troubles du neurodéveloppement

### 1.1. Définition

Les troubles du neurodéveloppement (TND) s'inscrivent dans une définition médicale récente apparue avec le *Diagnostic And Statistical Manual Of Mental Disorders* 5ème édition (DSM-V) en 2013, ouvrage de référence pour la classification des pathologies psychiatriques. Ce terme désigne un large spectre de pathologies liées à une ou des atteintes ayant un impact sur le développement cérébral, avec pour conséquence des symptômes neurologiques et/ou psychiatriques. Elles sont donc à l'origine de retard de développement, de retard des acquisitions, d'atteintes cognitives, sensorimotrices et/ou comportementales. Ces troubles sont à différencier des simples difficultés scolaires de par leur caractère durable, c'est à dire qu'ils vont persister partiellement malgré une rééducation classique (Des Portes, 2020).

Parmi le vaste spectre de pathologies incluses sous le terme de TND, nous trouvons les déficiences intellectuelles (DI) ou trouble du développement intellectuel (TDI), les troubles du spectre autistiques (TSA), les troubles envahissants du développement, l'épilepsie, les atteintes DYS (dyslexie, dyspraxie...) ou encore les troubles déficitaires de l'attention avec ou sans hyperactivité (TDAH) (Morris-Rosendahl & Crocq, 2020; Parenti et al., 2020; Thapar et al., 2017). Les signes de TND apparaissent précocement, habituellement dans l'enfance, et peuvent évoluer jusqu'à l'âge adulte. De plus ces atteintes sont fréquemment associées entre elles et se présentent sous des degrés de sévérité très variables d'un patient à l'autre.

Dans ce projet nous nous concentrerons principalement sur la DI et/ou les TSA.

**La déficience intellectuelle** est caractérisée par des limitations significatives dans le fonctionnement intellectuel global et dans les compétences adaptatives de la vie quotidienne. Les individus atteints de DI présentent un fonctionnement intellectuel inférieur à la moyenne, mesuré par un quotient intellectuel (QI) généralement en dessous de 70, c'est à dire deux écarts-type en dessous par rapport à la moyenne (-2DS), ainsi que des difficultés à acquérir des compétences essentielles pour l'autonomie, telles que la communication, la gestion de l'argent, ou la capacité à se

soigner (Battaglia et al., 1999). La DI peut être présente dès l'enfance et persiste tout au long de la vie, bien que l'intensité des limitations puisse varier considérablement d'un individu à l'autre.

Les **troubles du spectre autistique** représentent un large continuum de troubles neurodéveloppementaux caractérisés par deux éléments majeurs en accord avec le DSM-5 : des troubles de la communication et des interactions sociales, et des comportements stéréotypés et/ou des intérêts restreints.

Les personnes autistes peuvent présenter un large éventail de symptômes et de niveaux de fonctionnement, allant de troubles graves de l'autisme à des formes plus légères. Différentes capacités peuvent être touchées, parfois toutes, parfois juste le langage, parfois juste la perception, les émotions, etc... donnant ainsi à chaque individu atteint un spectre unique.

Les TSA se manifestent généralement dès l'enfance et persistent tout au long de la vie.

## **1.2. Incidence et prévalence**

Les TND concernent 9 à 18 % de la population générale (Shen et al., 2018; Zablotsky et al., 2018) et sont ainsi un enjeu de santé publique majeur. Une récente étude américaine a recensé le nombre de diagnostic posés de TND issu de deux bases de données médicales américaines, l'une publique collectant les données médicales des patients éligibles à *Medicaid*, un programme de l'état pour une couverture médicale gratuite, et l'autre privée : *Medicaid Analytic eXtract* entre 2000 et 2014 et l'*IBM MarketScan Research Database* (MarketScan) entre 2003 et 2015, comptant au final plus de 3 millions d'enfants. L'incidence d'un TND est estimée entre 11 et 23,9% dans ces deux sets de données (Straub et al., 2022).

Parmi les TND, la prévalence de la DI touche environ 1,5 à 2% des populations occidentales (Mefford et al., 2012; Tărlungeanu & Novarino, 2018). Les TSA ont une prévalence estimée elle aussi entre 1 et 2% de la population générale (Baio et al., 2018; Lai et al., 2014; Zeidan et al., 2022). Le sex-ratio est très souvent déséquilibré avec une majorité de garçons atteints : 2,1:1 en moyenne pour tous TND confondus, environ 4:1 pour les TSA et entre 1,2 et 1,9 pour la DI.

La co-occurrence de plusieurs pathologies est très fréquente, par exemple 70% des patients avec autisme présentent au moins un autre TND (Gillberg, 2010; Saito et al., 2020). Ces diagnostics sont parfois rendus difficiles du fait de la pathologie elle-même, un TSA sera difficilement identifiable chez un enfant atteint de DI sévère de par son incapacité à répondre aux outils de diagnostic de TSA.

### 1.3. Diagnostic

Le diagnostic d'un TND est avant tout clinique.

#### Clinique

Le diagnostic médical **de TDI** est un processus complexe qui implique généralement une évaluation multidisciplinaire approfondie menée par des professionnels de la santé et de l'éducation. Il se fait le plus souvent lors d'une évaluation neuropsychologique qui va associer une évaluation médicale préliminaire, une évaluation psychométrique qui va évaluer le QI, une évaluation des compétences adaptatives et une évaluation du contexte social. Les professionnels prennent en compte le contexte social de la personne, y compris son environnement familial, son accès aux soins et à l'éducation, ainsi que d'autres facteurs sociaux qui pourraient influencer le fonctionnement de la personne (Zablotsky et al., 2019).

En accord avec le DSM-5, **le diagnostic de TSA** repose sur une expertise clinique assurée par des praticiens spécialistes (généticiens cliniciens, neuropédiatre ou pédopsychiatres) complétée par des tests standardisés et considérés comme le « *gold standard* » qui sont : l'*Autism Diagnostic Interview Revised* (ADI-R), qui consiste en une interview anamnestique des parents de l'enfant atteint, et l'*Autism Diagnostic Observation Schedule* (ADOS) qui est une observation de l'enfant en conditions standardisées (Le Couteur et al., 2008; Lord et al., 1994). Plusieurs tests sont réalisés, visant à objectiver la présence d'un trouble de la communication et de la réciprocité sociale (*Childhood Autistic Rating Scale* CARS, l'échelle d'évaluation des comportements autistiques révisée ECA-R ou la *Social Responsiveness Scale*, SRS), ou des comportements répétitifs (*Repetitive Behaviors Scale-Revised*, RBS-R), qui participent au diagnostic.

Devant tout TND, il sera indispensable également d'évaluer le niveau de langage, le fonctionnement intellectuel et le développement de processus sensoriels à l'aide d'outils adaptés maîtrisés par des cliniciens experts.

## Moléculaire

L'origine génétique des TND a été mise en évidence depuis les années 70 grâce à de nombreuses études familiales où ségrégent des TND (Y. Huang et al., 2019; Parenti et al., 2020; Radley et al., 2019; Willfors et al., 2017). Ce lien avec la génétique a d'abord été suggéré par trois critères : l'association des TSA à des syndromes génétiques bien connus, les études sur des jumeaux et le risque relatif plus élevé de développer un TSA dans une famille.

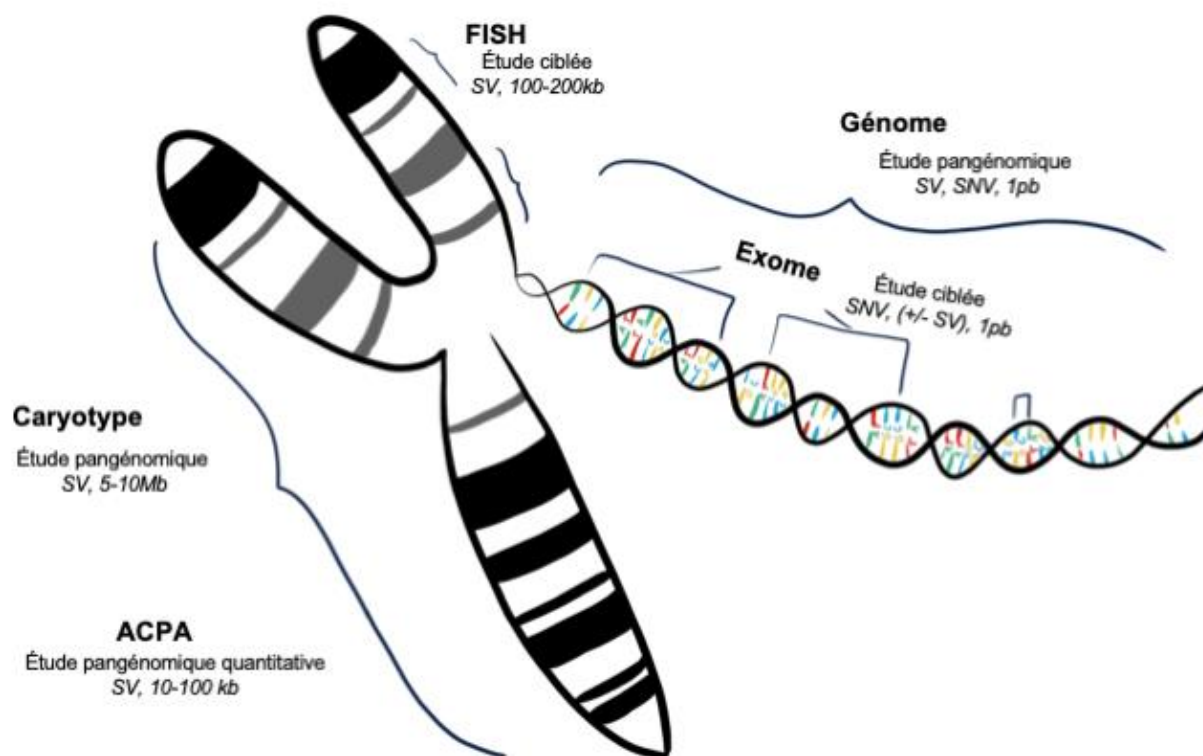
Les études de jumeaux ont en effet montré que l'autisme avait une très grande composante génétique (A. Bailey et al., 1995; Tick et al., 2016), avec une concordance phénotypique allant jusqu'à 91% chez les jumeaux monozygotes contre 10% chez les jumeaux dizygotes (Hallmayer et al., 2011; Rosenberg et al., 2009; Sandin et al., 2017). Les études familiales ont pu mettre en évidence que le risque de développer un TSA est d'autant plus élevé que l'on est génétiquement proche d'un individu avec autisme, avec une héritabilité estimée entre 50 et 90% (Alibutud et al., 2023; Sandin et al., 2017). Le risque relatif de TSA varie de l'ordre de 0,5 pour les apparentés les plus éloignés, à 150 pour les jumeaux monozygotes (Bourgeron, 2015). Il a également été montré dans la DI que l'héritabilité du QI variait de 71 à 87% (Wright et al., 2001).

Aussi devant tout diagnostic de TND, il est recommandé de demander des explorations génétiques. Les examens demandés en première intention sont classiquement une étude cytogénétique pangénomique, l'analyse chromosomique par puce à ADN (ACPA) associée à une recherche du syndrome X-fragile en génétique moléculaire, comme recommandé par l'*American College of Medical Genetics (ACMG)* en 2010 (Manning et al., 2010). Selon le phénotype du patient, un séquençage par panel de gènes pouvait être demandé, c'est à dire l'exploration moléculaire de groupe de gènes impliqués dans un même phénotype (c'est-à-dire ciblé pour l'épilepsie, DI,...). De nouvelles recommandations sont préconisées avec l'avènement du séquençage haut débit (SHD), dont le séquençage d'exome (*whole exome sequencing*, WES) (Srivastava et al., 2019) et plus récemment le séquençage de génome entier (*whole genome sequencing*, WGS) (Souche et al., 2022).

## 2. Architecture génétique des TND

### 2.1. Les anomalies chromosomiques associées

Les premiers diagnostics génétiques associés aux TND impliquaient de gros remaniements chromosomiques visibles par les outils diagnostic dont la résolution était de 5 à 10 Mb (caryotype, [chapitre 1/4.1.1](#)). Il s'agissait d'aneuploïdies (la trisomie 21 étant la première cause de DI syndromique (Mefford et al., 2012)), ou de marqueurs chromosomiques surnuméraires comme le chromosome isodicentrique du 15. Les pathologies moléculaires liées à l'X occupent elles aussi une grande partie des causes génétiques de TND, principalement le syndrome de l'X fragile, lié à une perte de fonction du gène *FMR1* en Xq28, qui était la première cause monogénique de TSA (Mefford et al., 2012). L'introduction de la technique de FISH (*Fluorescent in situ hybridization*) a permis l'identification de syndromes microdélétionnels récurrents associés à une symptomatologie clinique reconnaissable par les experts cliniques. Dans les années 2010, l'utilisation diagnostique en première intention de l'ACPA a permis de détecter des pertes et gains de copies d'ADN, appelés variations du nombre de copies ou CNVs (*copy number variation*), parfois associés à des syndromes microdélétionnels et microduplicationnels récurrents.



**Figure1.** Différentes techniques de diagnostic en génétique

Le taux diagnostique a ainsi augmenté de façon significative chez des patients présentant un trouble du neurodéveloppement, et des anomalies congénitales (Miller et al., 2010). Les syndromes microdélétionnels récurrents identifiés par ACPA ou FISH les plus fréquents étant la délétion 22q11.2 associée au syndrome de Di George (entre 1/2000 et 1/4000 naissances vivantes), la délétion 7q11.23 associée au syndrome de Williams Beuren (prévalence d'environ 1/7 500), les microdélétions de la région 16p11.2 (1 à 5/10 000) ou les syndromes de Prader-Willi et Angelman associés au locus 15q11q13 (1/15 000-30 000 et 1/20 000 à 1/10 000 respectivement) (Smajlagić et al., 2020). Plus récemment, des CNVs à Pénétrance incomplète et expressivité variable ont été décrits, communément dénommé PIEV par le réseau français AChro-Puce (<https://acpa-achropuce.com/>). Ils sont présents dans la population générale, mais plus fréquents chez les patients atteints de TND. Leur présence chez un individu est donc considérée comme un facteur de susceptibilité à développer un TND (Coe et al., 2012, 2014; Girirajan et al., 2013; Rosenfeld et al., 2013; Stefansson et al., 2014). Ce sont par exemple les microdélétions 16p11.2 (Redaelli et al., 2019), 15q11.2 BP1-PB2 (Rafi & Butler, 2020) ou 16p12.2 (Huguet et al., 2018).

## 2.2. Taux de diagnostic

Le taux de diagnostic génétique dans les TND a augmenté avec l'évolution des technologies. Initialement, le **caryotype** associé ou non à la FISH, permettait de poser un diagnostic chez 4 à 34,1% des patients atteints de DI (de Ravel et al., 2007; Xu & Chen, 2003) et chez 2 à 5% des patients autistes (Mendelsohn & Schaefer, 2008; Reddy, 2005). La FISH, permettant le diagnostic rapide de syndromes microdélétionnels a permis de poser un diagnostic génétique dans 2,5 à 5% des DI non expliquées par le caryotype (Mefford et al., 2012).

Le taux diagnostique en **ACPA** atteint 11 à 28% chez des patients présentant un TND (Cheng et al., 2019; Hochstenbach et al., 2009; Jang et al., 2019; Miller et al., 2010; Srivastava et al., 2019), 13,8% pour les cas de DI (Hochstenbach et al., 2009, 2011; Manning et al., 2010), et 11% pour les patients atteints de TSA (Marshall et al., 2008; Sebat et al., 2007; Velinov, 2019).

Le SHD apporte un nouvel essor dans le taux de diagnostic posé pour des patients souffrant de TND. Le **WES** permet aujourd'hui de poser un diagnostic dans 23 à 57% des cas (Clark et al., 2018; Martin & Ledbetter, 2017; Meng et al., 2017;

Sánchez-Luquez et al., 2022; Srivastava et al., 2019; Stark et al., 2016; Stefanski et al., 2021; Vasudevan & Suri, 2017; Wiszniewski et al., 2018), ce taux variant en fonction de la taille et des critères de recrutement des cohortes d'exome. Environ 22 à 35% des patients avec une DI ont un diagnostic génétique identifié par WES, 24% pour les patients avec épilepsie et 17% pour les patients avec TSA (Hochstenbach et al., 2011; Stefanski et al., 2021).

En ce qui concerne le **WGS**, le taux de diagnostic atteint 24% à 33% pour des cohortes mélangeant TND et malformations congénitales (Bhatia et al., 2021; Runheim et al., 2023), entre 8 et 27% pour les TND (Abe-Hatano et al., 2021; Álvarez-Mora et al., 2022; Gross et al., 2019; Lindstrand et al., 2019; Scocchia et al., 2019).

### 2.3. Gènes impliqués dans les TND et variations génomiques

A ce jour de nombreuses régions chromosomiques et gènes sont connus pour être impliqués dans les TND, référencés dans plusieurs bases de données.

La base de données **SFARI** (*Simons Foundation Autism Research Initiative*, <https://gene.sfari.org/database/human-gene/>) recense 1140 gènes impliqués dans l'autisme d'après la littérature scientifique et propose un *score de confiance* basé sur le niveau de preuve : 243 gènes sont impliqués dans un cadre syndromique (annotés « S »), 766 de scores 1 à 2, c'est-à-dire qu'ils ont une association très forte et prouvée dans les TSA, et 87 de score 3 (c'est à dire que trop peu de cas ont été décrits ou qu'il ne s'agit que d'un effet faible).

La liste de la **SPARK** (la *Simons Foundation Powering Autism Research for Knowledge*, <https://sparkforautism.org/portal/page/spark-gene-list/>, mise à jour en septembre 2022) recense 162 gènes et 43 CNVs ayant un impact fort et le plus souvent impliqués dans des formes monogéniques très pénétrantes pour les TSA. Cette liste correspond à des gènes de la SFARI de score 1 majoritairement. La SPARK s'appuie sur les données de 50 000 familles avec au moins un membre portant un diagnostic de TSA, et offre un rendu diagnostique aux patients (SPARK Consortium, 2018).

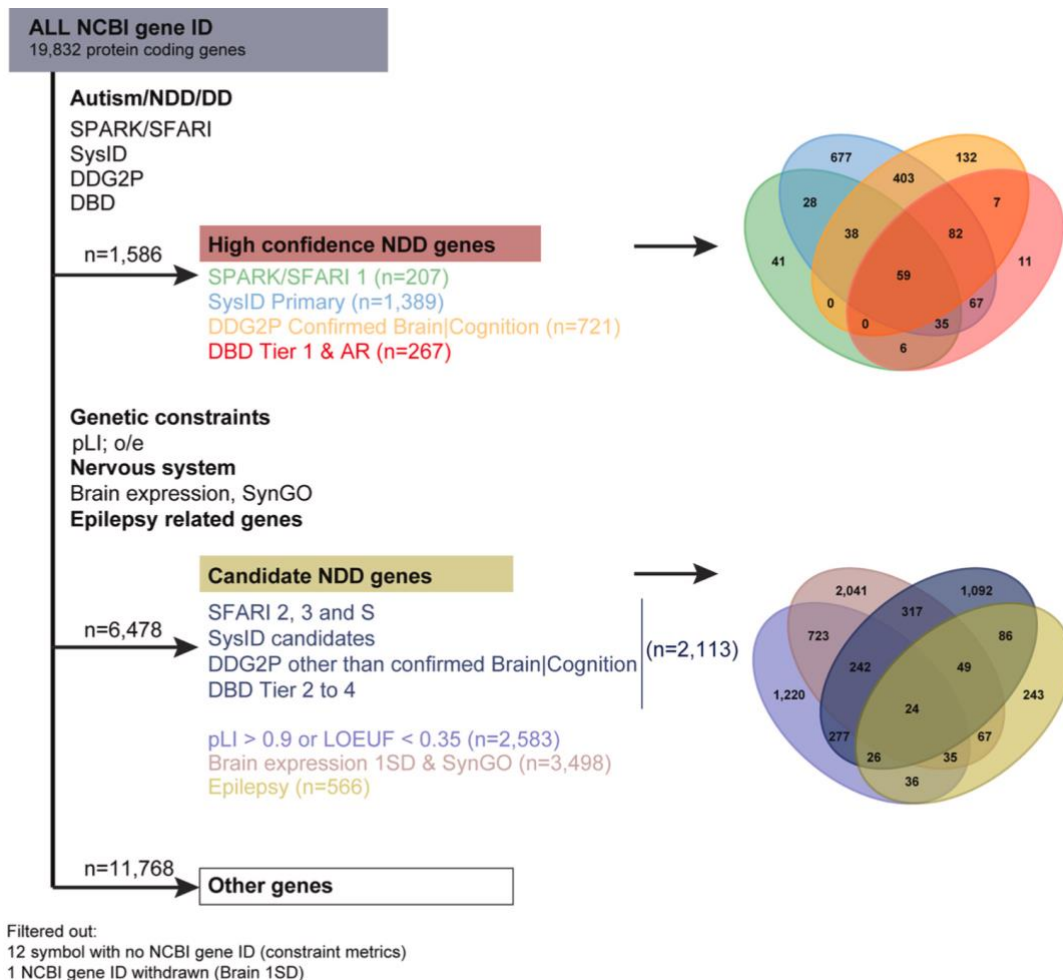
Une autre base de données du Royaume-Uni, la *Deciphering Developmental Disorders* (**DDD**, dont les données sont disponibles sur le site Decipher, <https://www.deciphergenomics.org/>), a collecté les données génétiques (puces à ADN et séquençage d'exome) et cliniques de plus de 14 000 patients atteints de TND et leurs parents. De leurs données ils ont également développé une liste de gènes

impliqués dans les TND, **DDG2P** (*Developmental Disorder gene panel of Gene2Phenotype*, <https://www.ebi.ac.uk/gene2phenotype>).

La base de données SysID (*Systems Biology Approaches to Intellectual Disability*) aujourd'hui devenue **SysNDD** (<https://sysndd.dbmr.unibe.ch/>) recense un catalogue de gènes impliqués dans la DI isolée ou syndromique (Kochinke et al., 2016).

Enfin, la base de données **DBD** (*Developmental Brain Disorder*, <https://dbd.geisingeradmi.org/>) propose des données de génotype et phénotype de différents types de TND vérifiés dans la littérature.

Par ailleurs, une récente étude a établi une liste de gènes impliqués dans les TND en s'appuyant sur ces nombreuses bases de données génétiques : les données d'expressions, les score de tolérance à la perte de fonction (pLI ou LOEUF), la SPARK et la SFARI, la SysID, DDG2P ou encore la DBD (citées plus haut) (Leblond et al., 2021).



**Figure 2.** Sélection des gènes pour les *High Confidence* et *Candidate NDD genes*

Source : C. Leblond, 2021, *Mol Cell Neurosci*



Leurs travaux a permis d'aboutir à une liste de gènes fortement impliqués dans les TND, les *High confidence neurodevelopmental* (HC-NDD) genes, correspondant globalement aux gènes primaires des précédentes bases de données, et une liste de gènes candidats pour les TND, les *Candidate NDD genes* (figure 2). Cette liste est régulièrement mise à jour et disponible sur le site <https://genetrek.pasteur.fr/>. Très souvent, ces gènes impliqués dans les TNDs se révèlent être fortement exprimés dans le cerveau en développement.

## **2.4. Voies couramment impliquées dans les TND**

On observe aujourd'hui un fort recouvrement dans l'étiologie génétique des différents TND, et plus particulièrement entre les TSA et la DI. Il ressort que la très grande majorité des gènes associés convergent dans deux voire trois voies biologiques, souvent intriquées entre elles, qui sont la plasticité et le fonctionnement de la **synapse**, le **remodelage de la chromatine**, et la **voie de signalisation mTOR** (Bourgeron, 2015; Parenti et al., 2020).

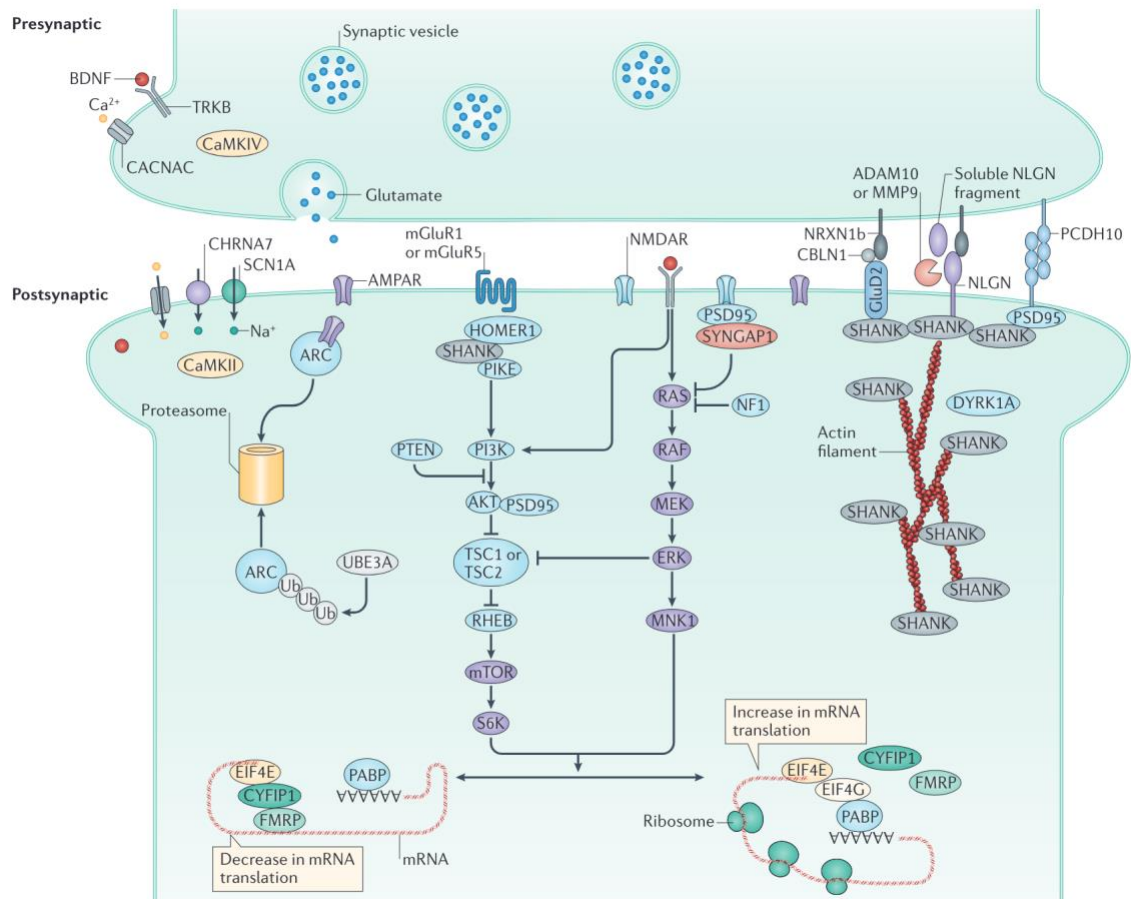
### **2.4.1. La Synapse**

La synapse chimique est une jonction cellulaire qui permet la transmission d'un signal entre les neurones via la transmission d'un neuromédiateur. Le nombre de synapses dans le système nerveux central humain est estimé à  $10^{14}$  et un neurone contient entre 1000 et 10000 synapses.

Elle est composée de trois parties : la région pré-synaptique, la région post-synaptique et entre les deux, la fente synaptique (figure 3). De nombreuses protéines contribuent au bon fonctionnement de ces synapses et leur atteinte peut entraîner des pathologies plus ou moins sévères.

#### **Élément présynaptique**

Dans cette région sont localisées des molécules d'adhésion cellulaire comme les **neurexines (NRXM)** (Südhof, 2008), qui vont assurer l'assemblage synaptique en se liant à leurs ligands situés dans la région postsynaptique (figure 3).



**Figure 3.** Schématisation d'une synapse  
 Source : T. Bourgeron, 2015, Nat. Rev. Neurosci.

### Élément postsynaptique

Le plus souvent, la région postsynaptique se situe au niveau d'une épine dendritique, qui est un renflement localisé sur les dendrites. On va trouver dans cette région des protéines membranaires comme les **neurologines**, qui sont le ligand des **neurexines** avec lesquelles elles vont former un complexe trans-synaptique (Südhof, 2008), ainsi que des protéines d'échafaudage comme **SHANK3** dont nous discuterons plus tard (voir [chapitre 1/2.5.1](#)) (figure 3). Des atteintes des **neurologines** sont associées aux TSA et aux TDAH (Bay et al., 2023; Jamain et al., 2003; Vieira et al., 2021; Yumoto et al., 2020).

Des atteintes des récepteurs aux neuromédiateurs ont aussi été décrites dans des TND, comme les **récepteurs NMDA** organisés en hétérotétramères sur la membrane de la région post-synaptique. Des variations dans des gènes codants leurs sous-unités, comme les gènes **GRIN2A** (OMIM \*138253) ou **GRIN2B** (OMIM \*138252) sont fortement impliqués dans des épilepsies focales avec ou sans DI (OMIM #245570) ou

dans des formes d'encéphalopathies épileptiques (OMIM #613970, 616139), respectivement (Vieira et al., 2021).

### **2.4.2. Remodelage de la chromatine**

La chromatine correspond à un complexe nucléo-protéique dans lequel l'ADN s'enroule autour de nombreuses protéines et se compacte à l'intérieur du noyau cellulaire. Elle est composée :

- de la molécule d'ADN
- de protéines majeures : les histones
- de protéines secondaires, impliquées dans le remodelage

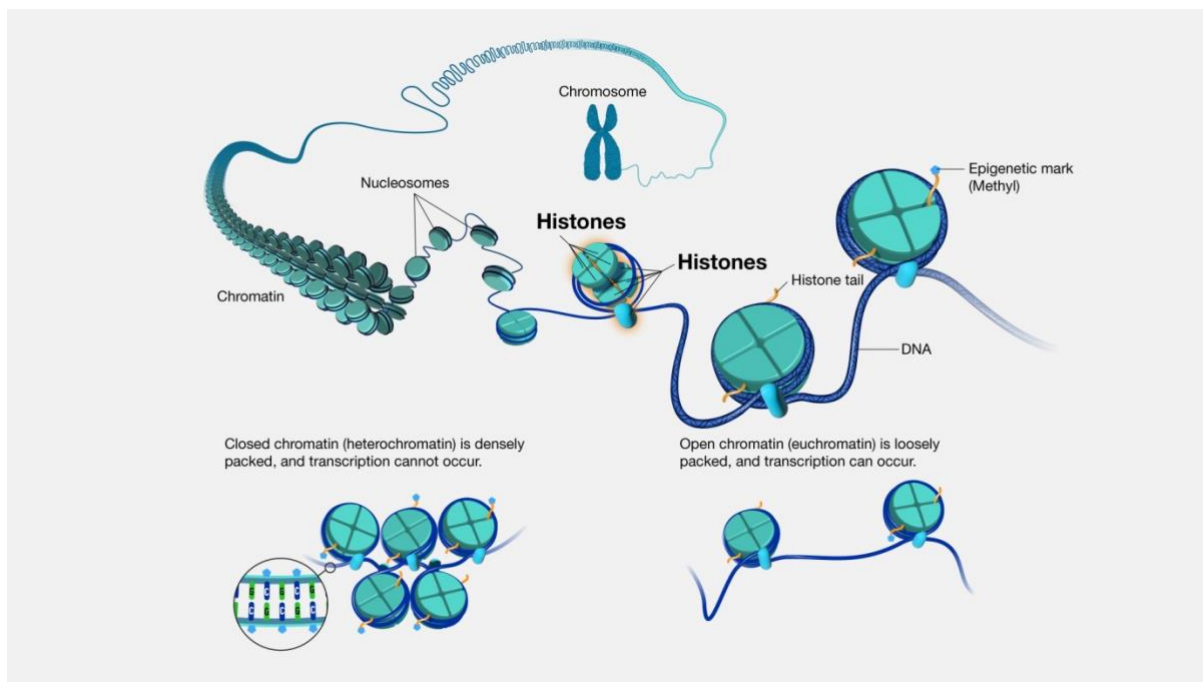
L'unité fonctionnelle de la chromatine est le **nucléosome**, un octamère constitué de deux fois 4 **histones** « cœur » (H3, H4, H2A, H2B, chacune par paire) autour duquel s'enroule la molécule d'ADN à raison de 147 paires de bases par octamère (Kouzarides, 2007) (figure 4). L'ensemble de l'ADN chez l'homme se compose d'environ 30 millions de nucléosomes.

D'autres protéines sont par la suite recrutées, notamment des protéines chromatiniennes comme les chromoboxes *CBX5*, *CBX1* et *CBX3* (anciennement nommées HP1-alpha, beta et gamma), qui vont interagir avec différents facteurs.

La modification de la chromatine se fait notamment par la modification des histones la constituant, ce qui va jouer sur son degré de compaction. Les modifications les plus courantes sont l'**acétylation** et la **méthylation**. Ces modifications vont elles aussi permettre l'ancrage d'autres protéines modificatrices, principalement les familles SWI/SNF, ISWI, CHD et INO80. La régulation peut également se faire via des méthylations directement sur l'ADN au niveau d'îlots CpG.

La structure de la chromatine, c'est à dire son niveau de compaction, régule l'accessibilité de l'ADN aux processus biologiques tels que la transcription et la réplication. Deux états sont visibles en microscopie et au caryotype :

- l'**euchromatine** : structure peu condensée, liée à l'acétylation des histones. L'ADN est accessible, permettant une meilleure activité de transcription et donc une expression génique active
- l'**hétérochromatine** : structure très condensée, liée à la méthylation des histones. L'ADN et donc les gènes sont peu accessibles pour la transcription, l'expression génique est réduite.



**Figure 4.** Organisation de la chromatine et des nucléosomes

Source: National Human Genome Research Institute, D.A. Gilchrist

<https://www.genome.gov/genetics-glossary/histone>

Le gène **MECP2** par exemple (OMIM \*300005), est un modulateur de la transcription essentiel pour la maturation des neurones. Il joue son rôle par la méthylation des îlots CpG. Les pertes de fonctions de ce gène, localisé sur le chromosome X en Xq28, sont bien connues et associées au syndrome de Rett chez les filles (OMIM #312750), un TND progressif et sévère, et à une encéphalopathie néonatale sévère chez les garçons (Kaur & Christodoulou, 1993).

Le gène **CHD8** pour *Chromodomain Helicase DNA-Binding Protein* (OMIM \*610528) est connu pour ses variations responsables de TND associant DI, TSA et excès de croissance (macrocéphalie). Il contribuerait pour 0,2% des causes d'autisme avec DI (Bourgeron, 2015; Mitchel et al., 1993; Smol et al., 2020).

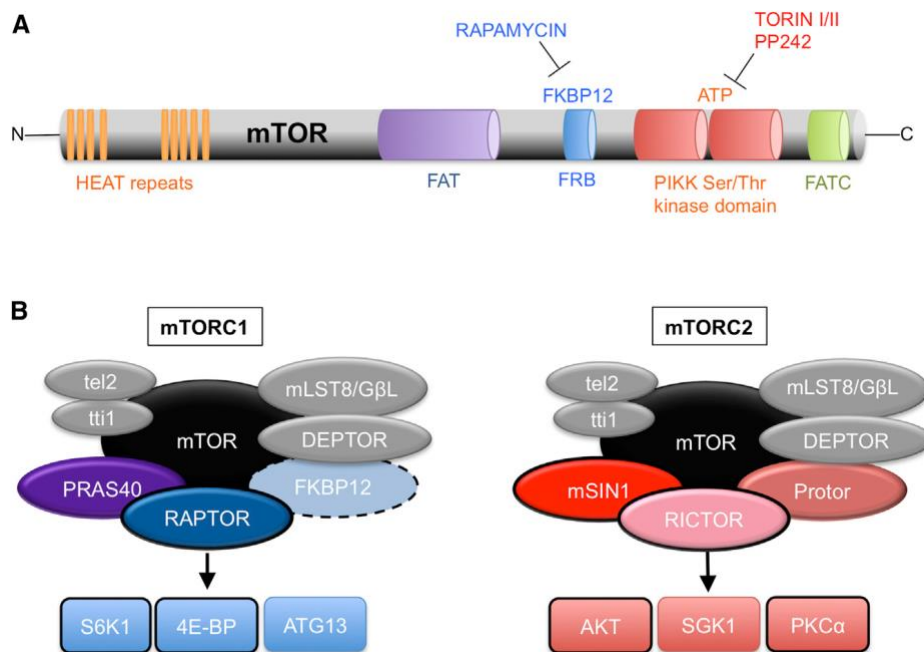
Un autre exemple est le gène **KMT2D** (\*602113) qui code une histone-lysine méthyltransférase qui s'imbrique dans un complexe protéique et a pour rôle majeur la méthylation de la lysine 4 de l'histone 3 (Froimchuk et al., 2017). Les variations perte de fonction de ce gène sont responsables du syndrome de Kabuki qui associe une DI syndromique, une dysmorphie caractéristique et d'autres malformations congénitales (OMIM #147920).

### 2.4.3. Voie mTOR et synthèse protéique

La voie de signalisation du *mechanistic/mammalian target of rapamycin* ou **mTOR** régule de très nombreux processus cellulaires tels que la synthèse protéique, la transcription, l'autophagie, la prolifération et la survie cellulaire (Lipton & Sahin, 2014). mTOR est une sérine/thréonine kinase très conservée et existe sous forme de deux complexes actifs (Brotelle & Bay, 2016) (figure 5) :

- **mTORC1**, qui régule la croissance cellulaire en favorisant la synthèse de protéines

- **mTORC2**, qui a des rôles plus diversifiés, notamment la régulation du cytosquelette et la modulation de la signalisation de l'insuline. Il est également impliqué dans la régulation de la croissance cellulaire et de la survie.



**Figure 5.** Schéma de la protéine mTOR et des complexes mTORC1 et mTORC2  
A. schéma des domaines de la protéine mTOR B. protéines impliquées dans les complexes mTORC1 et mTORC2

Source : Lipton & Sahin 2014, Cells

Les voies de la synthèse protéiques sont étroitement liées à la fonction de la synapse qui dépend de nombreuses protéines synaptiques (Jung et al., 2023). Au niveau cérébral, mTOR est impliqué dans la différenciation et l'architecture neuronale (Lipton & Sahin, 2014).

Les atteintes de la voie mTOR ont initialement été décrites dans des atteintes cancéreuses. Mais on sait aujourd'hui que des variations touchant directement **mTOR**

sont également impliquées dans des DI syndromiques comme le syndrome de Smith-Kingsmore (#616638).

L'atteinte d'autres protéines participant à la voie mTOR et la synthèse protéique est également impliquée dans des TND. C'est le cas de du gène ***PTEN*** (OMIM \*601728), un gène suppresseur de tumeur qui a un rôle de phosphatase et régule la voie de signalisation contrôlée par la phosphatidylinositol-3 kinase. Des variations du gène sont classiquement associées à des Syndrome tumoraux hamartomateux (OMIM #158350). Toutefois, une association entre des variations de *PTEN* plus fréquente chez les patients avec TSA et/ou DI a été rapportée (OMIM #605309) (Cummings et al., 2022; Frazier, 2019).

Le gène ***NF1*** (OMIM \*613113) est un répresseur de mTOR et code pour la neurofibromine 1, essentiellement exprimée dans les neurones, les cellules de Schwann, les oligodendrocytes et les leucocytes. Les pertes de fonction de ce gène sont associées à la Neurofibromatose de type I (OMIM #162200), caractérisée par des tâches café-au-lait, des neurofibromes cutanés et un risque accru de développer des tumeurs bénignes, associée à une prédisposition aux difficultés d'apprentissage (Wolkenstein, 2001).

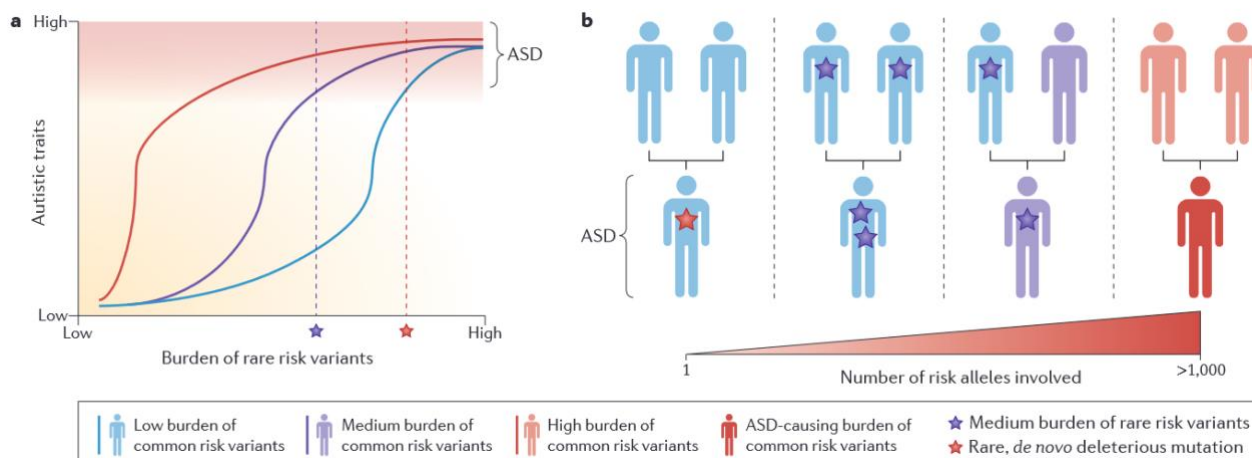
## 2.5. Architecture génétique des TND

Les TND sont hétérogènes dans leur expression clinique et ont une architecture génétique complexe allant de formes à transmission mendéliennes à des formes oligogéniques voire multifactorielles (Figure 6).

Dans les cas les plus simples, une seule variation rare est responsable du phénotype. Il s'agit des causes dites **monogéniques**, où une variation entraîne une modification de l'expression du gène à pénétrance forte, et qui touchent le plus souvent des gènes intolérants aux mutations perte de fonction (Iossifov et al., 2015; Parenti et al., 2020). Cette seule variation entraîne l'apparition de la pathologie. Il peut s'agir de mutations ponctuelles, d'insertions/délétion de quelques nucléotides ou Indel, ou de variations de structures impliquant un ou plusieurs gènes. Dans ces cas-là, la variation survient le plus souvent *de novo* et le poids du fond génétique est faible. Cette situation est la moins fréquente dans les TND (Parenti et al., 2020).

La majorité des cas de TND s'expliquent par des causes polygéniques ou multifactorielles (Parenti et al., 2020). Différentes situations peuvent se produire : la présence de plusieurs variations communes, à pénétrance faible et qui sont le plus

souvent héritées d'un parent, peuvent s'associer et déclencher la pathologie. Il a également été démontré l'existence d'un fond génétique à risque de TND, qui peut s'associer à un variant commun à faible risque pour déclencher une maladie, ou encore deux fonds génétiques à risque peuvent, sans variation génétique identifiée, s'associer et être à l'origine du phénotype anormal (Bourgeron, 2015; Parenti et al., 2020).



**Figure 6.** Relation entre fond génétique et variation génétique dans l'expression des TSA

Source : T. Bourgeron, 2015, Nature

On parle d'**épistasie** pour décrire des interactions entre gènes qui peuvent masquer ou modifier l'expression d'autres gènes, il s'agit d'une notion majeure aujourd'hui dans les phénotypes à transmission non mendélienne. L'épistasie participe très certainement au développement de TND.

## 2.6. Exemple de deux familles de gènes

### 2.6.1. Pathologie à pénétrance forte mais expressivité variable : SHANKopathies

#### 2.6.1.1. Description de la famille SHANK

La famille SHANK, pour *SH3 and multiple ankyrin repeat domains*, contient trois gènes, *SHANK1*, *SHANK2*, et *SHANK3*, qui codent pour des protéines d'échafaudage de la région dense post-synaptique (Leblond et al., 2012; Sheng & Kim, 2000), c'est à dire qu'elles jouent un rôle essentiel dans la transduction de signaux en favorisant un contact stable entre deux ou plusieurs molécules de signalisation.



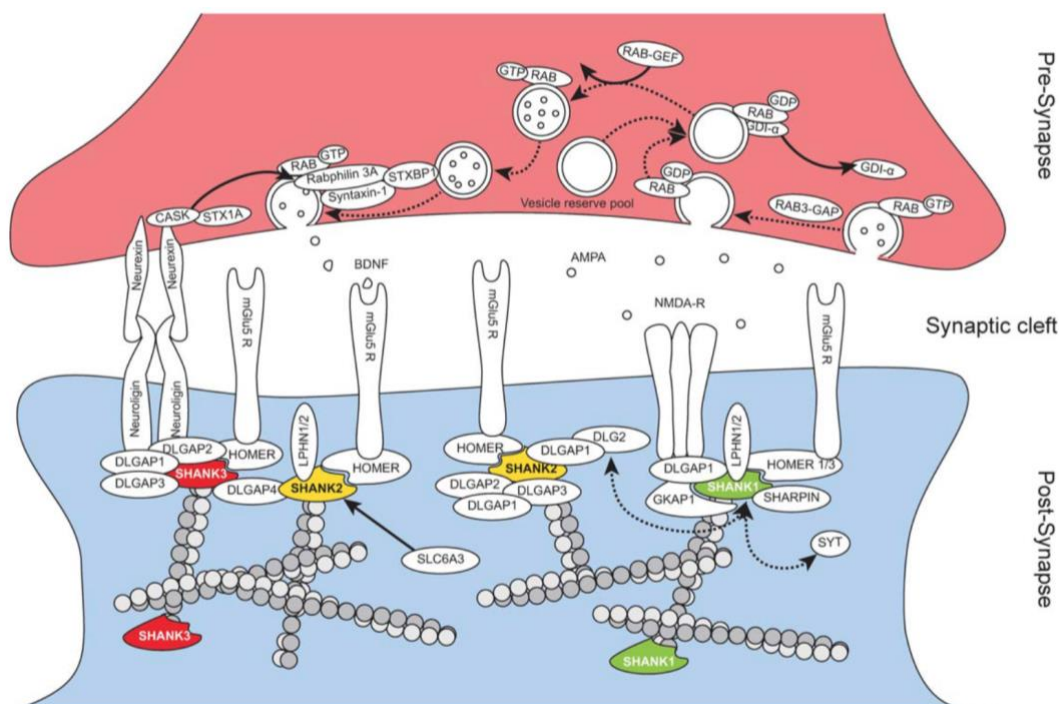
Cette famille partage des domaines qui permettent ces interactions protéine-protéine, notamment un domaine de répétition d'ankyrines (**ANK**), le domaine **SH3** (*SRC homology 3*), le domaine **PDZ** (PSD-95/DLG/ZO1), un long domaine riche en **proline** (qui occupe la moitié de la protéine) et le domaine **SAM** (*sterile alpha motif*) (figure 7). L'ensemble forme de longues protéines de plus de 200 kDa (Sheng & Kim, 2000).



**Figure 7.** Schéma des domaines protéiques de *SHANK3*

Source : C. Leblond et al., 2014, *PLoS genetics*

Les protéines SHANK interagissent au niveau des synapses excitatrices glutamatergiques avec plus d'une trentaine de protéines. Elles forment un complexe avec les protéines HOMER et DLGAP (anciennement GKAP) qui vont interagir avec les récepteurs mGluR5 et NMDA ou avec les Neurologines. Elles interagissent également avec le cytosquelette d'actine des épines dendritiques via les cortactines, ainsi qu'avec des molécules de signalisation intracellulaire (figure 8) (Guilmatre et al., 2014; Monteiro & Feng, 2017). Cette interaction permet de structurer et de stabiliser la synapse. Ainsi, elles vont jouer un rôle dans l'organisation structurale et fonctionnelle de l'épine dendritique et de la jonction synaptique.



**Figure 8.** Protéines SHANK et leurs interactions dans la synapse glutamatergique

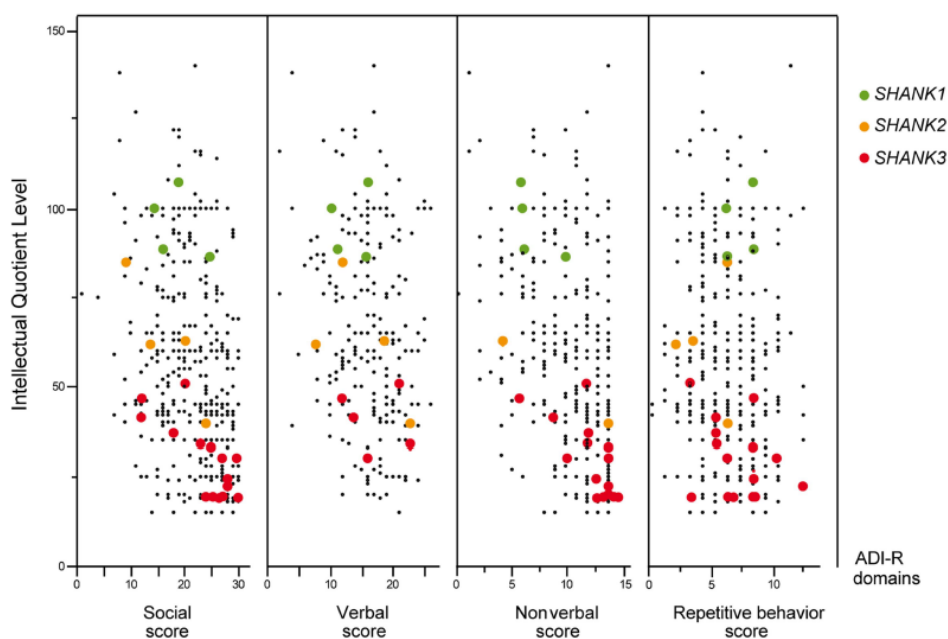
Source : A. Guilmatre, 2013, *Dev Neurobiol*



Les trois protéines se différencient par la temporalité de leur expression : *SHANK2* est le premier des trois gènes à être exprimé dans le cerveau en développement, suivi par *SHANK3* (Guilmatre et al., 2014; Monteiro & Feng, 2017). Il y a également une différence dans les régions où elles sont le plus exprimées : *SHANK1* est plus présent dans le cortex cérébrale et l'hippocampe, *SHANK2* dans la couche des cellules de Purkinje du cervelet, et *SHANK3* dans le thalamus et le striatum (Guilmatre et al., 2014).

Une méta-analyse de 2014 montre que les gènes *SHANK* sont associés à l'autisme et qu'il existe un gradient de sévérité selon le gène muté (Leblond et al., 2014). Les individus porteurs de variations délétères du gène *SHANK1* ont 3 fois plus de risque de présenter un TSA avec un QI dans la moyenne. Des variations délétères touchant *SHANK2* sont 1,76 fois plus présents chez des patients avec TSA que les témoins (4,58% vs 2,66%), avec un phénotype associant TSA, DI légère et troubles du langage et troubles moteurs. Les variations délétères de *SHANK3* sont associées à un TSA avec une DI modérée à sévère et ne sont *a priori* pas retrouvées chez des contrôles.

La figure 9 issue de l'étude de Leblond et al. représente le QI en fonction des scores de l'ADI-R qui est un élément du diagnostic de TSA, chez des patients mutés dans *SHANK1*, 2 ou 3, et montre une différence entre les 3 cohortes de patients.



**Figure 9.** QI rapporté à l'ADI-R chez des patients autistes selon le gène muté

Source : C. Leblond et al., 2014, *PLoS Genet*

Ainsi parmi les trois gènes, c'est l'atteinte du gène *SHANK3* qui présente une pénétrance pour l'autisme la plus élevée et le phénotype le plus sévère. Ce phénotype est associé au syndrome de Phelan-McDermid.

Depuis, de nouvelles études montrent que des atteintes du gène *SHANK2* (mutation perte de fonction, SNV (ou variation d'un seul nucléotide) ou Indels) sont aussi associées à des formes de TSA avec DI légère à modérée, voire sévère, avec une pénétrance de 100% pour la DI (11/11) et les troubles du langage (9/9) sur les 11 patients rapportés (Caumes et al., 2020).

Plus récemment une étude sur des cohortes de population générale (la UK-Biobank, <https://www.ukbiobank.ac.uk/>) montre que des individus sont porteurs de variations perte de fonction rares touchant *SHANK2* et *SHANK3*. Les individus, plus de 112 000, sont « non diagnostiqués pour TSA » à l'âge adulte. Il peut s'agir d'individus qui n'ont pas reçu les explorations aujourd'hui nécessaires pour diagnostiquer un TSA, ou des individus porteurs de variations localisées dans des régions du gène plus tolérantes face au phénotype de TSA, ou encore de vrai résilients (Rolland et al., 2023).

### **2.6.1.2. *SHANK3* et Syndrome de Phelan-McDermid**

#### **2.6.1.2.1. Clinique du PMS**

Le **syndrome de Phelan-McDermid** (OMIM #606232) est une pathologie du neurodéveloppement qui associe le plus souvent une hypotonie néonatale, un retard des acquisitions motrices avec un TDI de sévérité variable et un trouble de l'acquisition du langage allant d'un retard à l'absence de langage, ainsi que des éléments de dysmorphie faciale mineure (Kolevzon et al., 2014; K. Phelan & McDermid, 2011; M. C. Phelan, 2008). La clinique est très hétérogène d'un individu à l'autre et de nombreuses comorbidités peuvent être associées, notamment des TSA qui sont associés au syndrome dans 50 à 90% des cas selon les études, le diagnostic étant rendu difficile lorsque la DI est sévère. Des troubles digestifs sont très fréquents, avec notamment un reflux gastro-oesophagien, des troubles de l'alimentation ou une constipation (Vitrac et al., 2023).

Les atteintes sensorielles sont également fréquentes dans le PMS, touchant jusqu'à 80% des patients, et peuvent concerner l'hyposensibilité à la douleur, des troubles de la régulation thermique ou une hypersensibilité aux stimuli auditifs, tactiles ou visuels. L'épilepsie touche 17 à 70% des patients et des malformations cérébrales

sont observées dans 32 à 70% des cas sans éléments spécifiques au syndrome. On retrouvera des défauts de myélinisation, des anomalies du corps calleux, une ventriculomégalie ou une atrophie. On peut observer une microcéphalie chez 16% des patients et une macrocéphalie chez 17% d'entre eux (Vitrac et al., 2023). D'autres comorbidités sont également rapportées : troubles ophtalmologiques, pathologies cardiaques ou rénales. L'évolution longitudinale du syndrome est mal connue mais une régression des acquis est rapportée, notamment autour de la puberté ou au passage à l'âge adulte (Kolevzon et al., 2014).

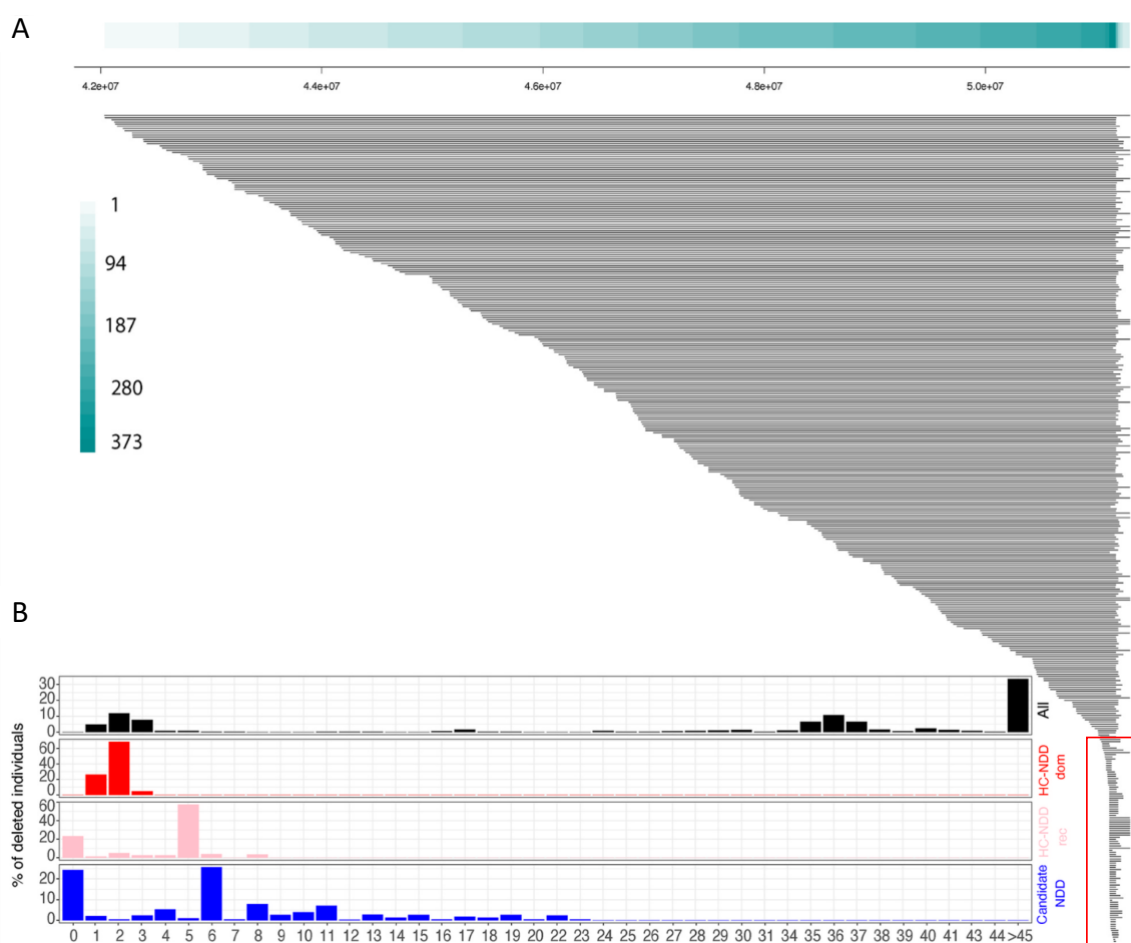
L'incidence de la pathologie est mal connue, mais serait estimée à 1/30 000. On estime qu'entre 0,25 et 3,33% des patients avec une DI non spécifique ont en fait un PMS, et que le syndrome serait présent chez 0,19 à 4,48% des patients atteints de TSA (De Rubeis et al., 2018; Schön et al., 2023).

### **2.6.1.3. Génétique du PMS**

La variabilité du syndrome se voit autant sur le plan phénotypique que génétique. Le PMS est initialement associé à des délétions de la région terminale du chromosome 22, en 22q13.3 (Phelan & McDermid). Parmi les 3 gènes impliqués dans la région critique du syndrome, *SHANK3* a été ciblé comme étant gène majeur du phénotype neurodéveloppemental dès 2007 (Durand et al., 2007). En effet, des variations perte de fonction de *SHANK3* ont été identifiées chez des patients avec DI et TSA (Durand et al., 2007). Des travaux ont montré par la suite que la perte d'expression d'une copie du gène *SHANK3* va modifier la région dense postsynaptique en modifiant le développement et la forme des épines dendritiques et ainsi jouer sur la formation des synapses, et la plasticité cérébrale (Durand et al., 2012; Roussignol et al., 2005).

La majorité des cas de PMS identifiés à ce jour sont porteurs de délétions emportant *SHANK3* (78-80%) (Koza et al., 2023; Levy et al., 2022), terminales ou plus rarement interstitielles. Le plus souvent isolées, les délétions terminales peuvent être en rapport avec des dérivés de translocation et s'associer alors à des duplications segmentaires issues d'un autre chromosome dans environ 11% des cas. Elles peuvent aussi être en lien avec des anneaux du chromosome 22 dans 10-20% des cas (Koza et al., 2023; Landlust et al., 2023; Levy et al., 2022; Tabet et al., 2017).

Dans le cas des délétions, les points de cassure s'étalent de façon homogène dans la région 22qter, hormis une région plus fréquemment touchée entre 22:51121293 et 22 :51161458 (GRCh37, figure 10) (Durand et al., 2007; Vitrac et al., 2023).



**Figure 10.** Répartition des délétions 22qter chez des patients PMS

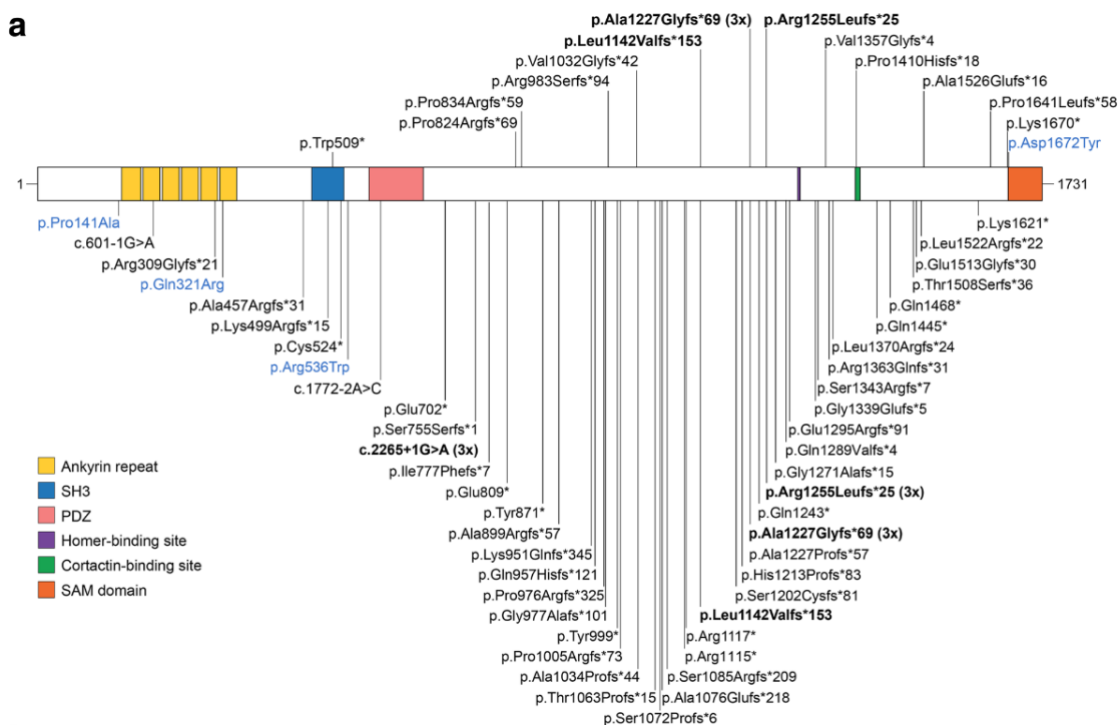
**A.** Représentation de la région 22qter. Chaque ligne correspond à la région délétée d'un patient. Encadré rouge = point de cassure distal fréquent. **B.** Nombre d'individus en fonction du nombre de gènes délétés (noir : tous gènes confondus ; rouge HC-NDD dominants ; rose : HC-NDD récessifs ; bleu : Candidate NDD)

Source : A. Vitrac et al., 2022, *European Journal of Medical Genetics*

Des études ont permis d'associer certaines régions avec un trait phénotypique : la région située entre 43 et 46Mb du chromosome 22 serait associée aux troubles du langage, les anomalies ophtalmologiques significativement plus fréquentes dans les délétions emportant la région 42,2-44,6 Mb et les reflux gastro-œsophagiens dans la région 48,9-49,9 Mb (Tabet et al., 2017; Vitrac et al., 2023).

Dans 12 à 20% des cas, les patients portent une variation ponctuelle ou indel délétère de *SHANK3* (Vitrac et al., 2023), mais ce taux pourrait évoluer avec la mise en place du séquençage en première intention. Les mutations ponctuelles et indel

délétères décrites dans le PMS se répartissent sur l'ensemble de la protéine (figure 11), avec quelques mutations récurrentes et un enrichissement sur l'exon 22 (NM\_001372044.2) (anciennement exon 21 avant correction de la séquence du gène) (De Rubeis et al., 2018; Leblond et al., 2014).



**Figure 11.** Répartition des mutations ponctuelles et indel sur la protéine *SHANK3*  
 Source : S. De Rubeis et al., 2018, *Molecular Autism*

Ces événements surviennent le plus souvent *de novo*, mais il existe de rares cas hérités, comme le cas d'une famille où une maman transmet une délétion interstitielle intragénique de 67kb de *SHANK3* à 5 de ses 6 filles (Tabet et al., 2017). Les cinq filles porteuses présentent un retard de langage associé à une DI pour trois d'entre elles. La mère ne présente pas de TSA mais aurait eu de légères difficultés d'apprentissage et une efficacité limitée avec un QI entre 70 et 80.

#### 2.6.1.4. Particularité des anneaux du chromosome 22

Certains patients atteints du PMS sont porteurs d'un anneau du chromosome 22, l'anneau se formant après une délétion terminale du chromosome qui aura entraîné la perte du gène *SHANK3* (voir [chapitre 1/3.1.4](#) pour les mécanismes de formation). La particularité des patients avec un PMS porteurs d'un anneau du chromosome 22, ou r(22), est leur risque accru de développer une **neurofibromatose**

**de type 2** (NF2, OMIM #101000). En effet, la région proximale 22q12.2 porte le gène **NF2** (OMIM \*607379) dont la perte d'expression est responsable de la pathologie NF2, qui se caractérise par la formation de tumeurs bénignes du système nerveux, principalement des **schwannomes vestibulaires**, le long du nerf auditif, ou des méningiomes (Lyons-Warren et al., 2017; Ziats et al., 2020). Les premiers symptômes sont les conséquences de ces tumeurs et apparaissent le plus souvent entre 18 et 24 ans (perte auditive, troubles moteurs ou sensoriels).

L'incidence de la NF2 est de 1/25 000 à 1/40 000 individus dans la population générale (Ardern-Holmes et al., 2017; Dg. R. Evans, 2009). Environ 1 à 1,6% des patients atteints de NF2 sont porteurs d'un r(22) (D. G. Evans et al., 2020; Koza et al., 2023), et on estime entre 2 et 4% le nombre de patients avec un PMS qui vont développer une NF2 (Koza et al., 2023).

La perte du gène *NF2* à l'origine de la pathologie est liée à la mosaïque dynamique de l'anneau, c'est à dire la perte de l'anneau liée à son instabilité en division de mitose (voir [chapitre 1/3.1.4.2](#)). L'instabilité du r(22) va générer une population de cellules en mosaïque avec une monosomie 22. Une hypothèse de double hit suggère un second événement somatique, qui entraîne la perte de fonction de l'allèle *NF2* controlatéral (par mutation ou délétion) dans les cellules de Schwann et conduire à la formation des tumeurs (Tsilchorozidou et al., 2004; Ziats et al., 2020; Zirn et al., 2012). Le phénotype de NF2 chez ces patients est souvent plus modéré, du fait de cette mosaïque cellulaire.

La mortalité associée à la NF2 est augmentée par rapport à la population générale. L'âge moyen de décès est estimé entre 32 et 62 ans, avec une médiane de survie de 15 après le diagnostic (Aboukais et al., 2015; D. G. Evans et al., 1992). Le risque relatif de décès est corrélé à l'âge du diagnostic (plus les premiers signes sont précoces, plus la maladie est sévère), mais aussi, de façon moins puissante à la présence de méningiomes (Baser et al., 2002). Les décès sont liés aux conséquences des tumeurs (hypertension intracrânienne, méningiome de haut grade), aux complications post-opératoire et dans quelques cas à la survenue de tumeurs malignes. L'origine génétique de la maladie est également un facteur de risque, avec les variations constitutionnelles pertes de fonctions associées à une plus grande mortalité, tandis que les autres types de variations et les mosaïques sont associées à des formes moins sévères. Il a également été montré qu'une prise en charge dans une structure de soin spécialisée améliore l'espérance de vie (Baser et al., 2002).

Par ailleurs, un autre gène suppresseur de tumeur localisé en 22q11.23, **SMARCB1** (OMIM \*601607), peut être responsable, par le même mécanisme de double hit suite à la perte de l'anneau, du développement de méningiome ou de tumeur rhabdoïde tératoïde atypique, rare mais de très mauvais pronostic (Byers et al., 2017; Cho et al., 2014; Koza et al., 2023).

Il est donc essentiel, pour tout patient avec un PMS par délétion terminale du chromosome 22, de réaliser un caryotype afin de préciser si cette délétion pourrait être en rapport avec la présence d'un r(22).

## **2.6.2. Pathologie à pénétrance variable : Famille des contactines**

### **2.6.2.1. Les gènes des contactines**

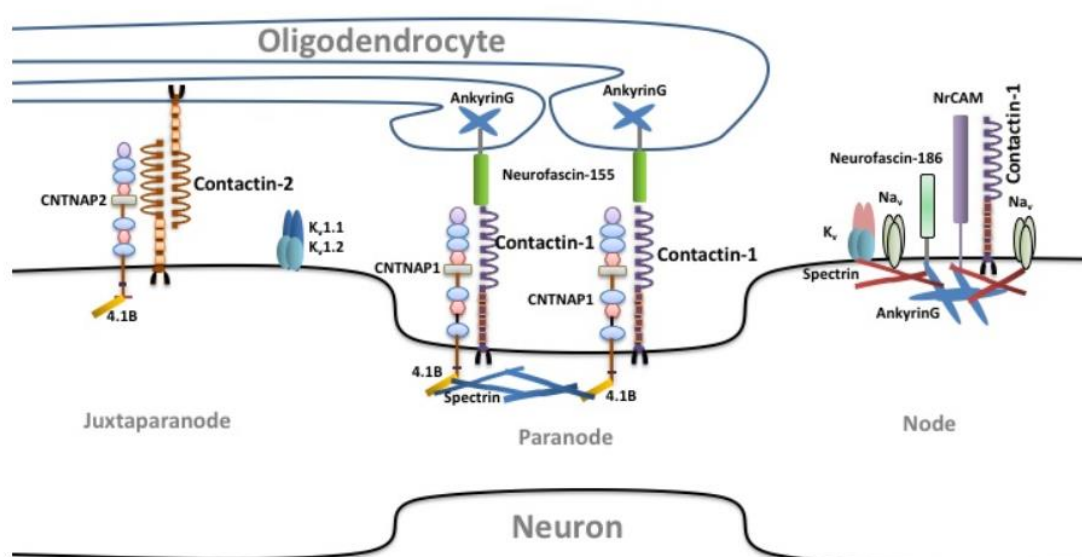
La famille des gènes des contactines comprend 6 gènes, *CNTN1* à *CNTN6*. Les contactines sont une famille de protéines impliquées dans l'adhésion cellulaire et la signalisation neuronale. Plus précisément, elles appartiennent à une sous famille de protéines de la superfamille des immunoglobulines, des IgCAM (*cell adhesions molecule*). Elles jouent un rôle important dans le développement du système nerveux par leur rôle de protéines d'adhésion neurale, mais aussi dans la formation des synapses, la myélinisation et le guidage neuronal.

Elles ne possèdent pas de domaine transmembranaire et sont reliées à la membrane cellulaire par une ancre glycosylphosphatidylinositol (GPI), ce qui leur confère une meilleure mobilité qu'un ancrage classique. Elles sont localisées à la fois dans les régions pré et postsynaptiques et jouent globalement un rôle dans l'organisation de l'axone, la myélinisation, le développement neuronal et la synaptogenèse (Chatterjee et al., 2019; Murai et al., 2002; Poliak & Peles, 2003; Rudenko, 2017; Shimoda & Watanabe, 2009; Stoeckli et al., 1991). Elles existent sous une forme liée à la membrane, et une forme soluble obtenue après le clivage du GPI (Furley et al., 1990; Yoshihara et al., 1995).

Les **contactine 1** et **contactine 2** sont situées sur la membrane cellulaire de l'axone au niveau du nœud de Ranvier notamment, où elles vont permettre l'ancrage de l'oligodendrocyte et donc de la myéline à l'axone. Elles vont organiser le nœud de Ranvier en différents domaines isolés (Chatterjee et al., 2019; Lubetzki et al., 2020; Zuko et al., 2013) (figure 12) :

- le **nœud**, région amyélinique de l'axone, qui contient les canaux sodiques,
- la **région paranodale**, entre les complexes formés par CNTN1 et CNTNAP1. C'est la zone d'ancrage de la myéline à l'axone.
- la **région juxtapanodale**, entre les complexes avec CNTN1 et avec CNTN2, où sont localisés les canaux potassiques

La **contactine 1** est particulièrement impliquée dans l'organisation de la membrane axonale, notamment dans le positionnement des canaux potassiques, mais aussi dans la stimulation de la myélinisation en interagissant positivement avec la voie Notch. Elle permet des interactions axo-gliales par son interaction avec la neurofascin-155. En formant un complexe avec *CNTNAP1*, elle va créer une structure isolante au niveau des nœuds de Ranvier. Son interaction avec la protéine PTPRZ va également jouer sur la maturation et la différenciation des oligodendrocytes.



**Figure 12.** *CNTN1* et *CNTN2* au niveau des régions paranodales et juxtapanodales  
Source : M. [Chatterjee](#) et al., 2019, *Neural Regen Res*.

La **contactine 2** joue également un rôle dans la régénération neurale, ainsi que dans la croissance des neurites via sa forme soluble. Les **contactines 1 et 2** sont toutes les deux exprimées dans le système nerveux central mais aussi dans le système nerveux périphérique, contrairement aux 4 autres protéines.

La **contactine 3** est exprimée dans les cellules de Purkinje, cellules du cervelet, contrairement aux contactines 1 et 2 (Chatterjee et al., 2019), mais elle participe à la formation et la croissance des neurites tout comme les contactines 1 et 2. Elle joue un rôle dans la formation des jonctions entre les cellules gliales, les cellules qui



soutiennent les neurones. Elle contribue à la myélinisation, un processus essentiel pour l'isolation électrique des axones neuronaux.

La **contactine 4** est impliquée dans le développement du cerveau, en particulier dans la formation de structures telles que le cervelet (Kaneko-Goto et al., 2008). Elle est particulièrement exprimée dans les neurones olfactifs (Chatterjee et al., 2019).

La **contactine 5** est essentielle à la formation et à la maturation des synapses. Elle joue également un rôle dans le guidage des axones, contribuant à la navigation des neurones pendant le développement. Elle est également impliquée dans le maintien des circuits neuronaux dans le système auditif comme elle a la particularité d'être exprimée dans les neurones des voies auditives (H. Li et al., 2003; Ogawa et al., 2001).

La **contactine 6** a un rôle important dans la régulation de la croissance axonale et de la guidance des neurones pendant le développement (Oguro-Ando et al., 2017).

#### **2.6.2.2. Les Contactines associated proteins**

Les *contactines associated proteins*, *CNTNAP1* à *CNTNAP5*, sont des protéines transmembranaires, appartenant à la famille des neurexines, qui jouent un rôle dans la formation et la stabilité des axones myélinisés. Elles sont co-localisées avec les canaux potassiques dans la région juxtapanodale des nœuds de Ranvier, ce qui leur donne un rôle crucial dans le "clustering" des canaux potassiques et la conduction des potentiels d'action axoniques (Lubetzki et al., 2003; Zou et al., 2017).

Elles vont jouer leur rôle notamment à travers la formation de complexes avec les contactines : *CNTN1* avec *CNTNAP1*, *CNTN2* avec *CNTNAP2*, qui vont permettre l'organisation des nœuds de Ranvier comme décrit au chapitre précédent, ou encore *CNTN5* et *CNTNAP4* ou *CNTNAP5*, dont le rôle est encore mal défini (Chatterjee et al., 2019; Poliak & Peles, 2003). Elles vont donc jouer un rôle dans la formation et le maintien des jonctions paranodales dans le système nerveux central pour ***CNTNAP1***, ***2*** et ***3***, ou périphérique pour ***CNTNAP4***. ***CNTNAP3*** aurait un rôle dans l'activité neuronale associée au contrôle moteur (Zou et al., 2017). ***CNTNAP4*** favorise la différenciation des cellules en neurones et cellules gliales. Le rôle précis de ***CNTNAP5*** est lui à ce jour moins compris (Traut et al., 2006; Zou et al., 2017). Récemment avec la mise en place de la référence du génome GRCh38, le gène *CNTNAP3* s'est décliné en *CNTNAP3A*, avec deux autres gènes proches, *CNTNAP3B* et *CNTNAP3C* qui ont été identifiés. Les caractéristiques, positions génomiques des *CNTN* et *CNTNAP* sont décrites dans le tableau suivant (table 1).

**Table 1.** Caractéristiques des gènes des contactines et contactines *associated-proteins*

Gène	Alias	Localisation	taille (kb - hg19)	Bornes hg19	Bornes hg38	Canonique Ensembl	Aa	Exons	pLI	LOEUF	NDD Group	SFARI
CNTN1	<i>F3, GP135, MYPCN</i>	12q12	379.977	chr12:41,086,244-41,466,220	chr12:40,692,439-41,072,415	ENST00000551295.7	1018	24	0,15	0,373	Candidate	-
CNTN2	<i>TAX1, AXT, TAG1</i>	1q32.1	35.303	chr1:205,012,325-205,047,627	chr1:205,042,937-205,078,289	ENST00000331830.7	1040	23	0	0,437	Candidate	-
CNTN3	<i>BIG-1, PANG</i>	3p12.3	258.625	chr3:74,311,719-74,570,343	chr3:74,262,568-74,614,445	ENST00000263665.6	1028	23	0	0,514	Candidate	2
CNTN4	<i>BIG-2</i>	3p26.3p26.2	959.149	chr3:2,140,497-3,099,645	chr3:2,098,813-3,057,961	ENST00000397461.5	1026	24	0	0,519	Candidate	2
CNTN5	<i>NB-2, hNB-2</i>	11q22.1	1337.934	chr11:98,891,683-100,229,616	chr11:99,020,949-100,358,885	ENST00000524871.6	1100	25	0	0,485	Candidate	2
CNTN6	<i>NB-3</i>	3p26.3	311.642	chr3:1,134,260-1,445,901	chr3:1,008,135-1,404,217	ENST00000446702.7	1028	23	0	1,153	Candidate	2
CNTNAP1	<i>CNTNAP, Caspr, NRXN4, p190</i>	17q21.2	17.381	chr17:40,834,631-40,852,011	chr17:42,682,531-42,699,993	ENST00000264638.9	1384	24	0	0,422	HC-NDD	2S
CNTNAP2	<i>Caspr2, KIAA0868, NRXN4</i>	7q35q36	2304.638	chr7:145,813,453-148,118,090	chr7:146,116,002-148,420,998	ENST00000361727.8	1331	24	0	0,547	HC-NDD	2
CNTNAP3 (CNTNAP3A)	<i>CASPR3, CNTNAP3A, FLJ14195, KIAA1714</i>	9p12	215.549	chr9:39,072,764-39,288,312	chr9:39,070,195-39,288,315	ENST00000297668.11	1288	24	0	0,767	Candidate	2
CNTNAP3B	<i>CASPR3B</i>	9p11.2	239.165	chr9:43,684,885-43,924,049	chr9:41,890,314-42,129,510	ENST00000377564	1288	24	0	1,666	-	-
CNTNAP3C		9q12	130.751	-	chr9:61,330,592-61,461,342	ENST00000636699.1	492	9			-	-
CNTNAP4	<i>CASPR4, KIAA1763</i>	16q23.1	281.960	chr16:76,311,176-76,593,135	chr16:76,277,278-76,560,884	ENST00000478060.5	1235	23	0	0,516	Candidate	2
CNTNAP5	<i>FLJ31966, caspr5</i>	2q14.3	890.001	chr2:124,782,864-125,672,864	chr2:124,025,287-124,921,201	ENST00000431078.1	1306	24	0,99	0,29	Candidate	2

### 2.6.2.3. Les anomalies des contactines dans la littérature

#### **Auto-immunité et hypomyélination – pathologie récessive**

Des variations délétères touchant les gènes des contactines ou des CNTNAP sont décrites dans des atteintes neuromusculaires, le plus souvent dans des formes autosomiques récessives, ou encore dans des maladies auto-immunes.

On peut citer par exemple une auto-immunité dirigée contre les protéines des nœuds de Ranvier, via des anticorps anti-contactine-1, anti-CNTNAP1 et anti-CNTNAP2, qui sont responsables de neuropathies périphériques auto-immunes aussi dénommées « nodopathies » ou « nodoparanodopathie » (Gupta et al., 2023; Querol et al., 2023). Des mutations homozygotes de *CNTN1* entraînent une Myopathie létale congénitale de type Compton-North (OMIM #612540), ou un Syndrome de neuropathie hypomyélinisante-arthrogrypose (OMIM # 618186). Les mutations homozygotes de *CNTN2* sont associées à des épilepsies myocloniques bénignes familiales (#615400), mais pourraient également être impliquées dans des pathologies démyélinisantes (Singh et al., 2015). Par ailleurs, le gène *CNTNAP1* serait impliqué dans des pathologies neurodégénératives comme la maladie d'Alzheimer (Zou et al., 2017).

#### **Association à des TND**

Leur association dans des TND reste plus discutée, bien que des CNVs ou SNV impliquant certains gènes des contactines aient été décrits.

Le gène *CNTN4* a été associé aux TND à travers son potentiel rôle dans le syndrome de délétion 3p- (OMIM #613792), un syndrome des gènes contiguës qui associe retard de développement, retard de croissance et une dysmorphie aspécifique. La région critique de ce syndrome serait de 4,5 Mb en 3p25p26 (Cargile et al., 2002). En 2004, Fernandez et al. suggèrent le rôle de *CNTN4* dans le phénotype neurodéveloppemental et la dysmorphie du syndrome à travers le cas d'un patient porteur d'une translocation équilibrée t(3;10)(p26;q26), interrompant le gène *CNTN4* (Fernandez et al., 2004). De récentes études pointent le gène *CHL1*, plus distal, comme ayant un rôle important dans le phénotype cognitif et le langage (Cuoco et al., 2011; Tsuboyama & Iqbal, 2021), ou considèrent les délétions 3p26 comme un facteur de risque de TSA (Cottrell et al., 2011; Cuoco et al., 2011; Gandawijaya et al., 2020).

En 2016, Mercati et al. montrent que les variations touchant les gènes *CNTN5* et *CNTN6* seraient des facteurs de risque des troubles neurosensoriels et notamment de l'hyperacousie chez les patients autistes (Mercati et al., 2017). En effet, ils montrent

un enrichissement des CNVs sur des cohortes de patients autistes (0,39% chez 1534 patients) par rapport à la population générale (0,01% sur 8936 individus contrôles). Dans leur cohorte, tous les CNVs étaient hérités, parfois de parents atteints. De plus, l'analyse des SNVs pour 429 individus montre également un enrichissement significatif de variations rares de *CNTN6* chez les patients TSA comparés aux contrôles ( $p=0,03$ ), mais pas pour *CNTN5* ( $p=0,36$ ). Cette association entre contactine et TSA est renforcée par d'autres études, comme Huang et al. en 2017 qui montrent que des duplications de *CNTN6* seraient un facteur de risque de syndrome de la Tourette (A. Y. Huang et al., 2017), ou plus récemment Schmilovich et al. qui montrent que des délétions dans le gène *CNTN5* seraient à risque intermédiaire de développer un TSA (Schmilovich et al., 2022). Guo et al. en 2017 étudient des duplications de *CNTN4* et les retrouve chez des patients avec TSA mais chez aucuns contrôles (Guo et al., 2017). De plus, d'autres études supportent le rôle de *CNTN4* dans le développement du langage (S. Q. Zhang et al., 2020). Récemment, des études de *genome-wide association studies* trouvent une association entre *CNTN5* et *CNTNAP5* pour les TSA (Narita et al., 2020) ou de *CNTNAP5* avec des troubles bipolaires (Djurovic et al., 2010). En 2021, Lengyel et al. rapportent une microdélétion de *CNTNAP5* impliquée dans le phénotype neurodéveloppemental d'un patient TSA avec retard sévère de langage (Lengyel et al., 2021).

### **Association aux TND controversée**

D'autres études semblent montrer l'absence d'association significatives entre CNVs des gènes des contactines et TND. En 2015, Murdoch et al. étudient 1030 patients avec un TSA contre 942 contrôles et ne trouvent aucune association significative entre la présence d'une variation délétère d'un gène soit des 6 contactines, soit de 4 des contactines *associated-proteins* (Murdoch et al., 2015). D'autres études montrent des résultats similaires, sur des cohortes de près de 20 000 individus pour la recherche de CNV touchant *CNTN6* (Repnikova et al., 2020), ou pour la recherche de CNV et de SNV touchant le gène *CNTNAP2* sur des cohortes regroupant des milliers d'individus (Toma et al., 2018).

Plus récemment, ces gènes ont été associés à des scores SFARI ou ont été classés dans les catégories *HC-NDD* ou *Candidate NDD gene* définies au [chapitre 1/2.3](#). Quatre gènes des contactines (*CNTN3* à *CNTN6*) et toutes les contactines *associated protein* sont de classes SFARI 2 (table 1). Par ailleurs, l'ensemble des gènes de ces

deux familles sont classés comme étant des *Candidate NDD*, sauf *CNTNAP1* et *CNTNAP2* qui sont considérés comme à haut risque de TND et dans le groupe des *HC-NDD genes*, du fait probablement de leur association à des encéphalopathies récessives.

L'implication des gènes des contactines et des contactines *associated proteins* dans le développement d'un TND n'est donc pas si tranchée, mais leur rôle au moins en tant que facteur de susceptibilité aux TND est à rediscuter.

## 3. Variations génomiques et TND

### 3.1. Les anomalies chromosomiques

#### 3.1.1. Types d'anomalies

Les anomalies chromosomiques impliquées dans les TND concernent l'ensemble des chromosomes et sont de taille très variable qui peuvent aller d'un chromosome entier, visible au caryotype, à quelques kilobases identifiables par les récentes techniques de cytogénétique moléculaire que nous expliciterons au [chapitre 1/4.1](#).

##### 3.1.1.1. Les anomalies de nombre ou aneuploïdies

Les anomalies de nombre de chromosome, ou aneuploïdies, ont été les premières identifiées grâce au caryotype, principalement la trisomie 21 associée à une DI syndromique. Les trisomies peuvent impliquer n'importe quel chromosome et sont le plus souvent associées à un phénotype léthal dans leur forme homogène. Les formes en mosaïques sont associées à un phénotype qui peut être plus modéré. On peut citer l'exemple de la trisomie 8 en mosaïque, dont l'incidence est estimée entre 1/25 000 et 1/50 000 et dont le phénotype associe principalement une DI modérée, un retard de langage, une dysmorphie faciale discrète et un syndrome polymalformatif touchant principalement les voies urinaires, cardiaques et ostéo-articulaires (ORPHA:96061). Les trisomies en mosaïques surviennent généralement suite à une malségrégation post-zygotique.

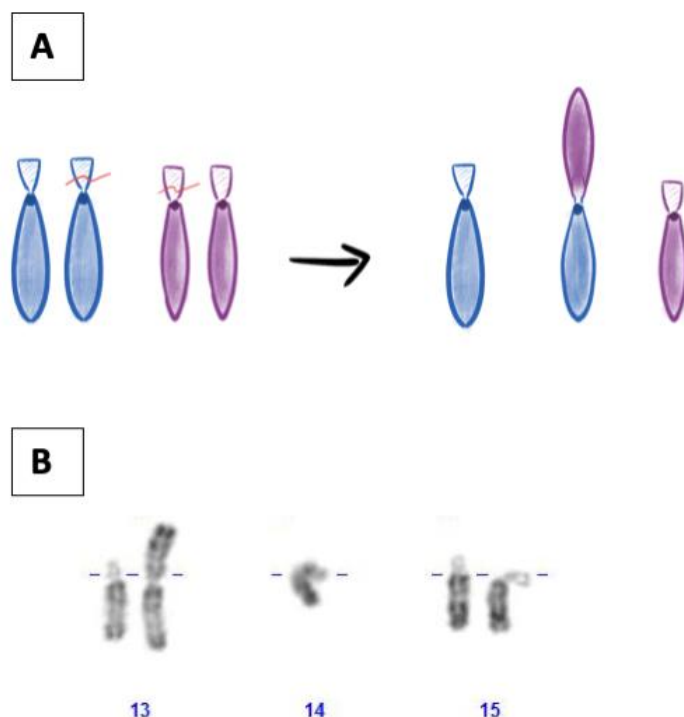
##### 3.1.1.2. Les anomalies de structure

#### Anomalies de structures équilibrées

Le caryotype a également permis d'identifier des remaniements équilibrés, comme des translocations réciproques apparemment équilibrées, des inversions ou des insertions, le plus souvent sans impact phénotypique. On sait toutefois qu'un point de cassure affectant l'expression d'un gène, modifiant l'architecture 3D du génome ou encore entraînant la formation d'un gène de fusion peut avoir un effet délétère (Schluth-Bolard et al., 2019).

Si des translocations peuvent toucher l'ensemble des chromosomes, les **chromosomes acrocentriques**, de par la nature de leur bras court qui sera développée au [chapitre 1/3.2](#), sont impliqués dans des anomalies de structure

particulières, appelées des **translocations robertsoniennes**. Ces translocations impliquent deux chromosomes acrocentriques qui ont recombinaison dans leur bras court (figure 13), entraînant la perte des deux bras courts.



**Figure 13.** Schéma et image d'une translocation robertsonienne

**A.** Schéma représentant la délétion des bras courts et l'inactivation d'un des deux centromères à la formation de la translocation **B.** image de translocation au caryotype impliquant les chromosomes 13 et 14.

La perte de deux bras courts de chromosomes acrocentriques a peu d'impact sur le développement cellulaire, et aboutit le plus souvent à la formation d'un chromosome dicentrique (un des deux centromères, très proches, sera inactivé, permettant une meilleure stabilité de ce chromosome recombinant).

Ce remaniement chromosomique est un des plus fréquents, avec une prévalence estimée à 0,8-1,2 pour 1 000 individus (Hamerton et al., 1975; Mack & Swisshelm, 2013; Poot & Hochstenbach, 2021; Zhao et al., 2015).

### **Anomalies de structure déséquilibrées**

Les anomalies de structure déséquilibrées de grande taille (>5-10 Mb) sont visibles au caryotype. Il peut s'agir de **dérivés chromosomiques** (dérivés de translocation, dérivés d'inversion par aneusomie de recombinaison) ou de **cassure**

**chromosomique interstitielle** (deux points de cassure à l'intérieur du chromosome) ou **terminale** (un point de cassure emportant le télomère).

D'autres anomalies déséquilibrées peuvent correspondre à la formation d'**anneaux chromosomiques**, (dont le mécanisme est détaillé au [chapitre 1/3.1.4](#)). L'incidence des anneaux chromosomiques est estimée entre 1/30 000 et 1/60 000 naissances vivantes (Caba et al., 2012; Kosztolányi, 2009). Tous les chromosomes peuvent être impliqués dans la formation d'anneaux avec des incidences variables : 10 à 12% des anneaux concernent les chromosomes 18 et 20, tandis que 5 à 9% des anneaux sont représentés par des chromosomes acrocentriques (P. Li et al., 2022). La technique standard de détection d'un anneau chromosomique est le caryotype plus ou moins complété par une technique de FISH (voir [chapitre 1/4.1.1](#)).

Les anomalies déséquilibrées infrachromosomiques, plus communément appelées **microremaniements** chromosomiques, ou **variations du nombre de copie (CNV)**, ont une taille > 1 000 nucléotides et sont mis en évidence par les techniques de puce à ADN ou ACPA.

Comme déjà précisé, les CNVs représentent une cause importante de pathologies génétiques des TND. Ils peuvent être récurrents impliquant des points de cassures récurrents, ou non récurrents. Ces CNVs peuvent être classés pathogènes ou probablement pathogènes, ou encore être considérés comme des facteurs de susceptibilité aux troubles du neurodéveloppement, ou PIEV comme cité plus haut ([chapitre 1/2.1](#)).

### **3.1.2. Principaux mécanismes de formation**

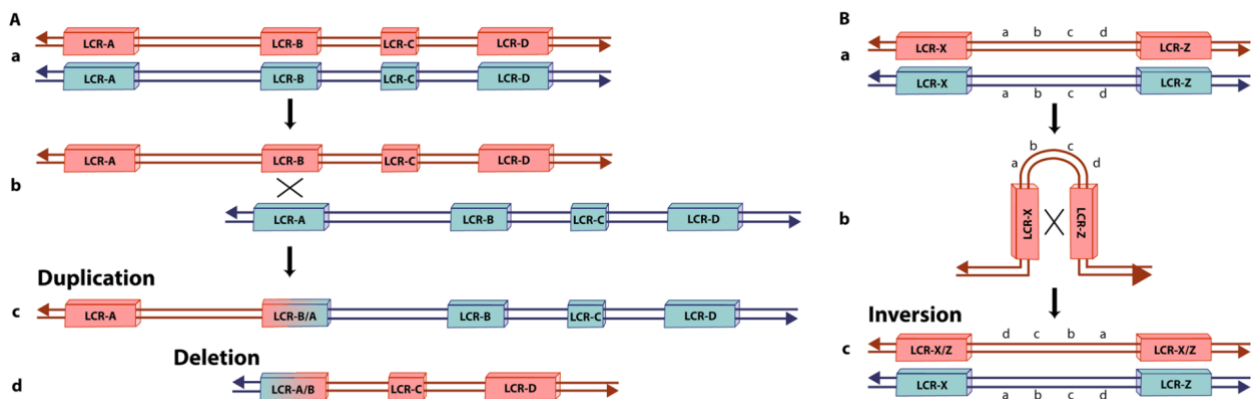
#### **3.1.2.1. Remaniements récurrents**

Un microremaniement récurrent correspond à un remaniement avec des points de cassure identiques entre des patients non apparentés. Un des exemples le plus fréquent est le syndrome de Di George (OMIM #188400), lié à une délétion récurrente en 22q11.2 entre les LCR22A et LCR22D (Burnside, 2015).

Ces remaniements sont essentiellement médiés par un mécanisme de recombinaison allélique non homologue, ou **Non Allelic Homologous Recombination (NAHR)**, favorisé par la présence de nombreuses duplications segmentaires ou **low copy repeats (LCR)** à travers le génome. Ces LCR correspondent à des régions d'ADN de plus de 10 à 1 000 kb avec plus de 90-97%



d'homologie de séquence (Burssted et al., 2022; Stankiewicz & Lupski, 2010; F. Zhang et al., 2009). Les LCR à même orientation et à une distance inférieure à 10 Mb peuvent entraîner des mésappariements entre chromatides sœurs ou chromosomes homologues, du fait de cette homologie de séquence, et provoquer des délétions ou des duplications réciproques (figure 14.A). Les points de cassure correspondent à la position de ces LCR et sont donc récurrents.



**Figure 14.** Création de microremaniements par NAHR

**A.** La présence de deux LCR de très forte homologie entraîne un mauvais alignement et le réarrangement des deux séquences (b). Création de deux fragments, l'un avec duplication (c), l'autre avec délétion (d). **B.** La présence de LCR proche et en sens inverse va favoriser la formation d'inversions (c)

Source : B.Burssted, et al., 2022, *Mol Cytogenet*

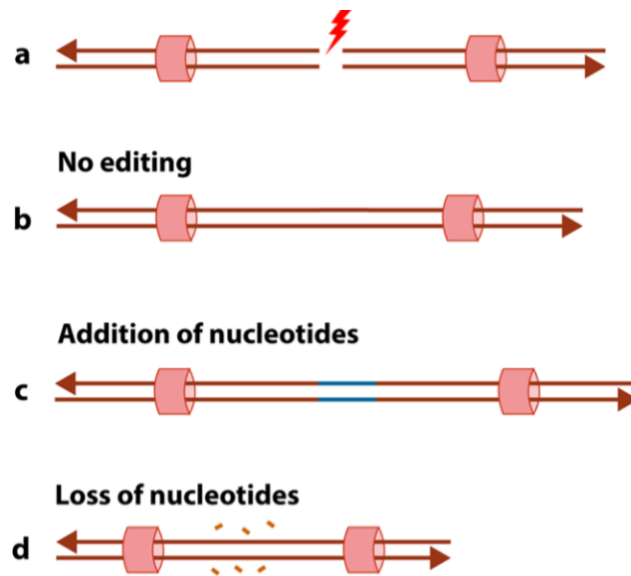
Des inversions peuvent être causées quand les LCR sont en orientations opposées (figure 13.B), voire des translocations quand les LCR qui s'apparient sont localisées sur des chromosomes différents.

### 3.1.2.2. Remaniements non-récurrents

Les remaniements non récurrents sont le plus souvent la conséquence de cassures de l'ADN suivies des différents mécanismes de réparation.

#### **Non Homologous End Joining (NHEJ)**

Le mécanisme de NHEJ permet la réparation des cassures d'ADN double brin. La réparation n'utilisant pas de séquence homologue comme moyen de repère, des délétions ou duplications aux points de cassure sont souvent induites par ce mécanisme (figure 15) (Burssted et al., 2022; Stankiewicz & Lupski, 2010; F. Zhang et al., 2009).



**Figure 15.** Mécanisme de réparation par NHEJ

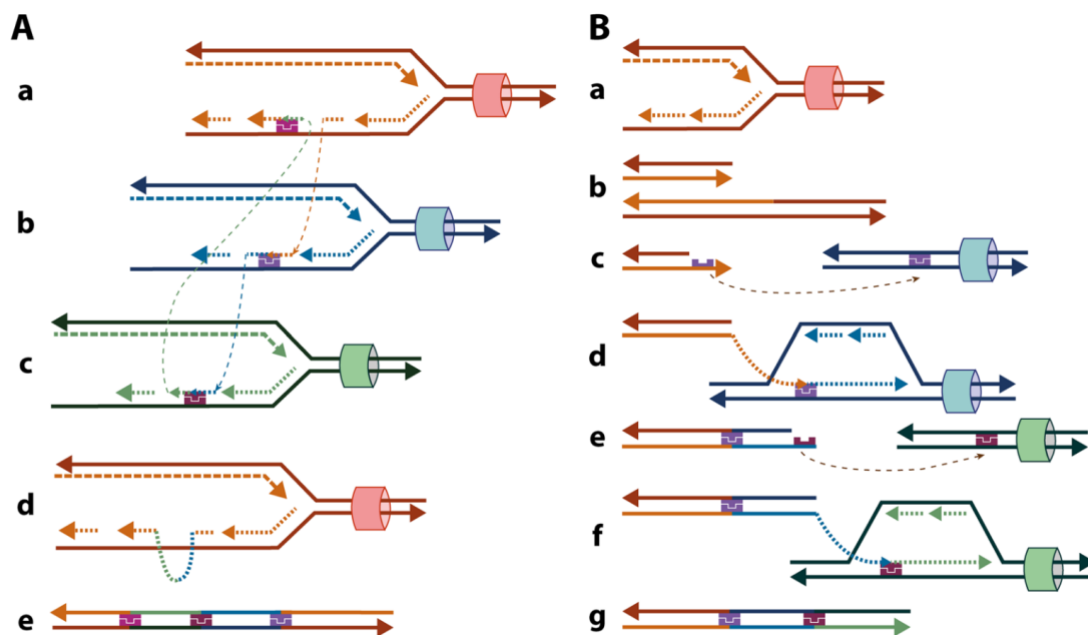
Une cassure d'ADN double brin (a) va provoquer une réparation de la séquence sans modèle, qui peut ne pas modifier la séquence (a) ou entraîner des microduplications au point de cassure (c) ou des microdélétions (d).

Source : B.Bursted, et al., 2022, *Mol Cytogenet*

### **FoSTeS et MMBIR**

Des remaniements plus complexes peuvent être induits par des anomalies au niveau de la fourche de réplication de l'ADN lorsque celle-ci rencontre une lésion de l'ADN, le **FoSTeS (*Fork stalling and template switching*)**. Dans ce mécanisme, le brin indirect (ou retardé) va se détacher du brin d'ADN qui lui sert de modèle et s'engager vers une autre fourche de réplication active et adjacente par un mécanisme d'hybridation de séquences microhomologues (Bursted et al., 2022; Weischenfeldt et al., 2013). L'élongation de l'ADN reprend alors, créant une molécule nouvellement synthétisée qui peut associer des séquences génomiques provenant de différentes régions chromosomiques, tout comme des délétions si le brin s'engage dans une fourche en aval ou des duplications si le brin s'engage dans une fourche de réplication située en amont de son origine (Figure 16.A).

Le mécanisme **MMBIR (*Microhomology-Mediated Break-Induced Replication*)** est également à l'origine de microremaniements complexes et non récurrents et suit le même mécanisme que FoSTeS en dehors d'une cassure double brin initiale (Bursted et al., 2022; Weischenfeldt et al., 2013) (figure 16.B)



**Figure 15.** Mécanismes de réparation par FoSTeS et MMBIR

**A. FoSTeS.** Le brin retard (pointillés) (a) a switché vers une autre fourche de réplication active (b) et peut répéter ce procédé plusieurs fois (c, d) et aboutir à une séquence remaniée (e).

**B. MMBIR.** Rupture de la fourche de réplication à la rencontre d'une lésion de l'ADN (b). Démasquage d'une région de microhomologie (violet) qui va servir de modèle à une nouvelle région du génome et redémarrer la synthèse d'ADN (d). Ce procédé peut se répéter (e, f), et aboutir à la formation d'une séquence complexe (g).

Source : B.Bursted, et al., 2022, *Mol Cytogenet*

### 3.1.3. Formation des *Inv dup del*

Les **duplications inversées avec délétion terminale**, ou *inverted duplication associated with terminal deletion* (*inv dup del*) sont des remaniements bien connus en cytogénétique. Ces remaniements ont initialement été décrits sur le bras court du chromosome 8, région riche en séquences avec une grande homologie du fait des clusters de gènes de la famille des récepteurs olfactifs (environ 400 gènes), mais également du fait d'une inversion paracentrique polymorphe retrouvée dans environ 30% de la population. Plus tard, tous les chromosomes se sont révélés susceptibles d'être impliqués dans des *inv dup del* (Bursted et al., 2022; Rowe et al., 2009; Zuffardi et al., 2009).

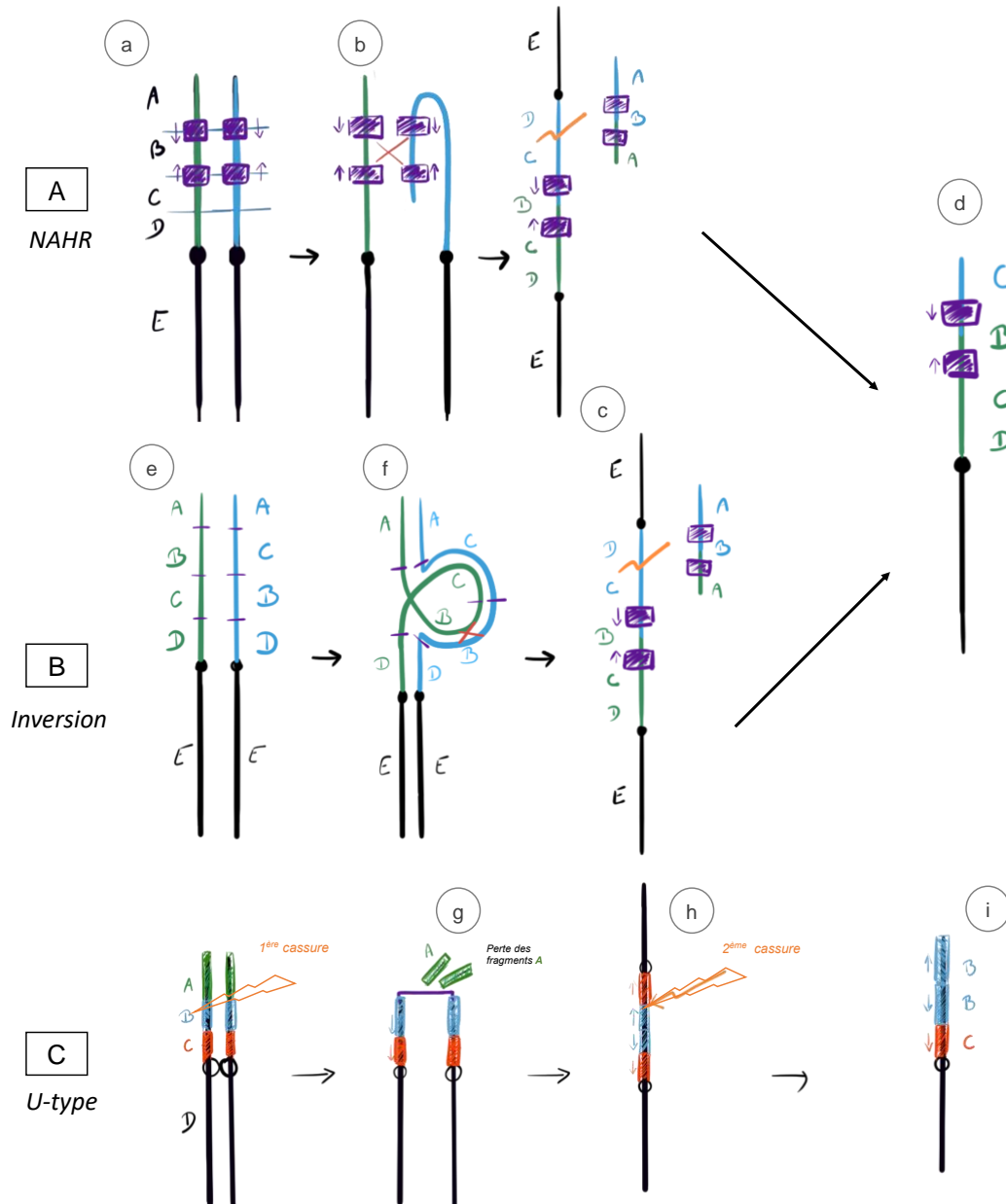
Les deux mécanismes principaux à l'origine des *inv dup del* sont des **NAHR** ou des mécanismes de stabilisation d'une délétion télomérique par un échange « en U » entre deux chromosomes homologues ou chromatides sœurs (« *U-type exchange* »). Le point commun entre les différents mécanismes aboutissant à une *inv dup del* est la

formation d'un chromosome dicentrique qui, instable au moment de la division cellulaire, va se casser.

**Les inv dup del par NAHR** surviennent principalement après un mauvais alignement et réarrangement de deux LCR orientés inversement. Si un échange chromatidien (ou crossing-over) survient, un chromosome dicentrique se forme (figure 17.A), avec perte d'un segment terminal. Au moment de l'anaphase, les pressions cellulaires exercées sur le chromosome dicentrique vont entraîner sa cassure et la formation du chromosome présentant une délétion terminale, une duplication adjacente inversée, les deux segments dupliqués étant séparés d'une région en séquence unique (Fig 17.A.d). En parallèle, un segment acentrique se forme et finira par se perdre dans les divisions successives.

De même, la présence d'une **inversion paracentrique** chez un parent (c'est-à-dire sur le même bras chromosomique) peut entraîner en méiose la formation d'une boucle d'inversion et conduire à une inv dup del en cas de crossing-over dans la boucle d'inversion par mécanisme de NAHR (Rowe et al., 2009) (figure 17.B). Là encore, le chromosome anormal présente une délétion terminale, une duplication adjacente inversée, les deux remaniements étant séparés par une région en séquence unique.

**Le mécanisme U-type** (ou **Breakage-fusion-bridge**) se produit lors d'une cassure double brin des deux chromatides sœurs (première étape induisant la délétion terminale) qui serait réparée par NHEJ (Rowe et al., 2009). La réparation au niveau des points de cassure induit la fusion des deux chromatides créant une structure symétrique dicentrique en forme de U (figure 17.C.g). On reconnaît les inv dup del par U-type par l'absence de copie unique entre les deux segments dupliqués. Ce mécanisme apparaît comme le plus fréquent dans la formation d'inv dup del (Rowe et al., 2009).



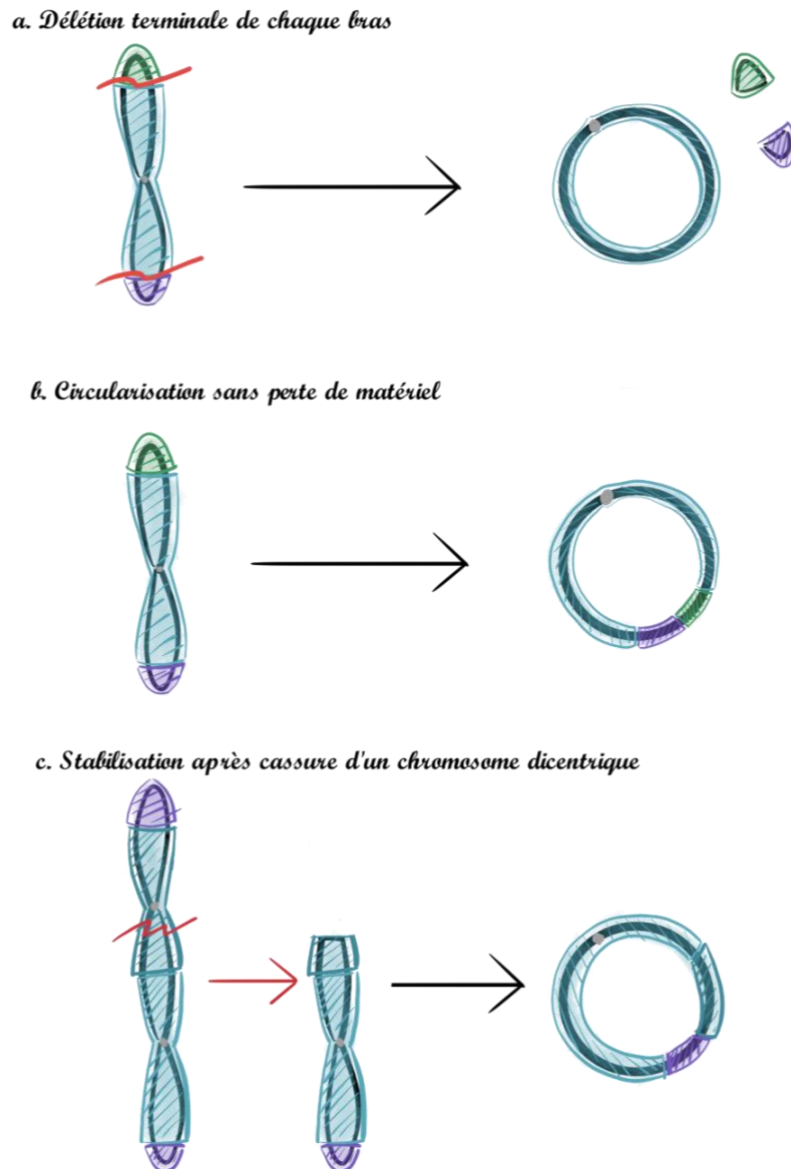
**Figure 17.** Formation d'un chromosome dicentrique aboutissant à une inv dup del

**A.** Mécanisme d'inv dup del par **NAHR**. La présence de deux LCR d'orientation inverse (a) va favoriser le mésappariement des deux chromosomes (b). En cas de crossing over dans la région (croix orange), formation d'un chromosome dicentrique avec perte des fragments terminaux A-B-A (c). Cassure du chromosome dicentrique qui aboutit à une inv dup del avec la région C dupliquée et inversée, séparée par une région en copie unique, B (d). **B.** Cas avec inversion paracentrique des segments B et C (e). Formation d'une boucle de recombinaison (f), qui va conduire à la formation d'un chromosome dicentrique (d) en cas de crossing over dans la boucle (croix orange) qui aboutit à une inv dup del (d). **C. U-type** : cassure terminale double brin puis stabilisation par fusion des deux chromosomes ou chromatides avec perte des segments terminaux A (g). Formation d'un dicentrique puis cassure (h) qui entraîne la formation d'une inv dup del avec deux segments dupliqués et inversés (segment B), sans séquence unique entre les deux segments (i).

### 3.1.4. Formation des anneaux chromosomiques

#### 3.1.4.1. Description

La formation d'anneau chromosomique est un mécanisme rare de stabilisation par circularisation d'un chromosome. Ce mécanisme survient généralement après une cassure terminale à la fois du bras court et du bras long d'un même chromosome, qui se suit par la fusion de chaque extrémité (figure 18.a).



**Figure 18.** Schéma de différentes situations aboutissant à la formation d'un anneau  
**a.** Circularisation d'un chromosome après double cassure et perte des fragments terminaux du bras court (vert) et long (violet). **b.** circularisation du chromosome sans perte de matériel génétique. **c.** un chromosome dicentrique qui a perdu son extrémité terminale (verte) se casse et se circularise pour se stabiliser.

Schéma inspiré de Pristyazhnyuk and Menzorov, 2018, *Protoplasma*

Parfois la circularisation peut faire suite à la cassure d'un chromosome dicentrique (quel que soit l'origine de la formation de ce dicentrique) instable au cours des divisions mitotiques (figure 18.c). Dans de rares cas, la circularisation peut survenir par « jonction télomérique », sans perte de matériel génétique et conduire à la formation « d'anneaux complets » (figure 18.b) (Pristyazhnyuk & Menzorov, 2018).

### **3.1.4.2. Instabilité mitotiques et leurs conséquences**

**Au niveau cellulaire**, la structure d'anneau en elle-même crée une instabilité et de multiples réarrangements possibles lors des divisions cellulaires successives. On peut alors voir apparaître des anneaux doubles ou des anneaux entrelacés s'il y a un échange chromatidien (duplication de tout le chromosome en anneau). De même, des microdélétions et microduplications sont fréquemment observées, probablement secondaires à des mécanismes de corrections de type FoSTeS (Carvalho & Lupski, 2016; Pristyazhnyuk & Menzorov, 2018). Au maximum, un anneau peut être le siège de plusieurs remaniements adjacents correspondant à un chromothripsis (Kurtas et al., 2018). Ces anneaux remaniés vont également devenir instables et finalement se perdre dans certaines cellules. Ces phénomènes peuvent être la cause de monosomie en mosaïque.

**Les conséquences cliniques** d'un anneau chromosomique sont très variables et dépendent d'une part du chromosome impliqué, et d'autre part de la taille et du contenu de la délétion associée. Le phénotype correspond habituellement au syndrome associé à la délétion. Plusieurs études parlent également d'un « syndrome de l'anneau », qui entraînerait un retard de croissance lié à la perte de l'anneau, instable au cours des divisions (Gardner et al., 2012; Kosztolányi, 1987; Rivera & Domínguez, 2010; Rossi et al., 2008). L'existence de ce « syndrome de l'anneau » est aujourd'hui controversée et la cause du retard de croissance pourrait être liée aux simples modifications d'expression d'un ou plusieurs gènes, soit par délétion soit par effet de position (Pristyazhnyuk & Menzorov, 2018; Rossi et al., 2008). La présence de régions d'hyperpigmentation cutanées est, elle, spécifique mais est plutôt le reflet d'un mosaïcisme chromosomique, plus que de la présence de l'anneau lui-même (Anderson, 2020).

## **3.2. Chromosomes acrocentriques**

La grande majorité des chromosomes ont un centromère situé en position médiane, séparant un bras court en amont, d'une taille inférieure (chromosomes

submétacentriques) ou de même taille (chromosomes métacentriques) qu'un bras long en aval. Cinq paires se caractérisent toutefois par un bras court très réduit et dont la séquence est globalement partagée, ce sont les paires 13, 14, 15, 21 et 22 dites « acrocentriques ». Le bras court de ces chromosomes acrocentriques (BCCA) est composé essentiellement d'ADN satellite (alpha, beta et gamma) et des organisateurs nucléolaires (NOR), contenant entre autres les ADN ribosomiaux (rDNA, à raison d'environ 300 copies dans la bande p12) (Antonarakis, 2022). Durant l'interphase, période où l'activité de transcription et de traduction est importante, les NOR vont favoriser le rapprochement de leurs séquences et former les nucléoles, où sont produites différentes sous-unités ribosomiales.

La présence de séquences répétées dans les rDNA et les séquences satellites, couplées à la proximité nucléolaire des différents chromosomes acrocentriques va favoriser les recombinaisons interchromosomiques. Il y a donc un haut risque de translocations dites robertsoniennes,

Grâce à la dernière version du génome proposée par *telomere-to-telomere* (T2T, voir [chapitre 1/4.2.5.2](#)), la structure des BCCA a pu être mieux décrite. Notamment par la présence de *pseudo homologous regions* (PHR) localisées dans les régions proximales des rDNA et de taille variable entre les chromosomes (de 719,25 kb pour le chromosome 15 à 6,48 Mb sur le chromosome 14). Ces PHR vont permettre des recombinaisons entre chromosomes non homologues (Guarracino et al., 2023).

### 3.3. Variations nucléotidiques

Les variations génomiques impliquent des modifications de la séquence de l'ADN, soit par la variation d'un seul nucléotide (SNV), qui peut changer l'acide aminé au niveau protéique (mutation faux sens) ou insérer un codon stop prématuré (mutation non-sens ou stop), soit par l'insertion ou la délétion d'un ou quelques nucléotides, des Indels, dont la définition est variable mais souvent <1 à 10kb (Mills et al., 2006; Mullaney et al., 2010; Sehn, 2015). Ces Indels peuvent conserver le cadre de lecture (In frame), ce qui causera la perte d'un ou plusieurs acides aminés (Indel multiple de 3), soit décaler le cadre de lecture (Frameshift, Indel différent d'un multiple de 3) (Ng et al., 2008).

Les variations sont de plusieurs type :



- Les **pertes de fonction** (*loss of function* LoF) : incluant les mutations non-sens (stop), les mutations avec décalage du cadre de lecture, les mutations sur les sites donneurs ou accepteurs, et certaines mutations faux-sens.

- Les mutations **faux-sens** sans effet perte de fonction.

- Les mutations aux **régions d'épissage**.

- Les mutations **synonymes**, où le changement de nucléotide n'entraîne pas de changement d'acide aminé. L'impact de ces variations est le plus souvent neutre, mais de récentes études montrent qu'elles peuvent avoir un effet sur la transcription ou l'épissage et être impliquées en maladie humaine (S. F. Bailey et al., 2021; Sarkar et al., 2022).

- Les mutation dans des **régions non codantes** introniques (sites d'épissage exclus).

Parfois, les variations génomiques consistent en l'**expansion de répétitions**, qui peut provoquer des modifications de la méthylation et donc de l'expression du gène. C'est le cas du syndrome de l'X fragile, où la séquence située dans la région 5'UTR (région transcrite non traduite) du gène *FMR1* sur le chromosome Xq28 possède une répétition d'un codon CGG allant normalement jusqu'à une trentaine de copies. Lorsque le nombre de répétitions dépasse 200, l'allèle est dit muté, le gène est méthylé et n'est plus exprimé, provoquant la pathologie neurodéveloppementale.

### 3.4. Critères de pathogénicité

Les variations sont interprétées puis classées selon leur degré de pathogénicité : **Bénin** et **probablement bénin** pour une variation sans implication sur le phénotype, **pathogène** et **probablement pathogène** quand elle est impliquée dans le phénotype d'un patient, et les **variations de signification inconnues** ou **VSI** quand les arguments ne permettent de les classer dans aucune des 4 autres catégories.

Des critères ont été établis afin d'homogénéiser les interprétations, initialement par Lee et al. en 2007 et ont été complétés par le collège américain de génétique médicale ou ACMG (C. Lee et al., 2007; Richards et al., 2015).

Les critères prennent en compte de nombreux éléments issus entre autres de données de population (fréquence dans la population générale ou dans les bases de données de patients), des données prédictives (impact sur la protéine ou sur l'expression), des données fonctionnelles, ou encore les études de ségrégation ou le nombre d'allèles atteints (figure 19).

	Benign			Pathogenic		
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
<b>Population data</b>	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
<b>Computational and predictive data</b>		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
<b>Functional data</b>	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
<b>Segregation data</b>	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
<b>De novo data</b>				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
<b>Allelic data</b>		Observed in <i>trans</i> with a dominant variant BP2 Observed in <i>cis</i> with a pathogenic variant BP2		For recessive disorders, detected in <i>trans</i> with a pathogenic variant PM3		
<b>Other database</b>		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
<b>Other data</b>		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

**Figure 19.** Critères d'évaluation de la pathogénicité d'une variation génomique selon l'ACMG

Source : S. Richards et al., ACMG, 2015, *Genet Med*.

Le réseau AChro-Puce (<https://acpa-achropuce.com/>) a aussi proposé des recommandations pour aider à homogénéiser les classifications et interprétation des CNVs (dernière mise à jour en septembre 2022 : <https://acpa-achropuce.com/diagnostic-postnatal/>).

## 4. Les outils diagnostics

### 4.1. Cytogénétique conventionnelle et moléculaire

#### 4.1.1. Caryotype et FISH

Les premières techniques de cytogénétique ont été mises au point dans les années 50, avec le premier caryotype réalisé en 1955. Le caryotype constitutionnel est

une des premières études pangénomique visant à analyser le nombre et la structure de l'ensemble des chromosomes d'un individu. Il est obtenu après la mise en culture de sang périphérique selon une technique peu robotisée. D'une résolution faible de 5 à 10 Mb, le caryotype a largement été supplanté par les techniques nouvelles d'analyse pangénomiques de plus forte résolution (ACPA et SHD). Toutefois, le caryotype conserve toute son importance en particulier pour les anomalies de structures équilibrées ou non, touchant les bras courts des chromosomes acrocentriques dont les chromosomes en anneaux.

La technique d'**hybridation in situ par fluorescence** ou **FISH**, apparue dans les années 80 est une technique de cytogénétique moléculaire ciblée. Elle consiste à hybrider des sondes faites de petites séquences d'ADN d'une taille allant de 100 à 300kb couplée à un fluorochrome, sur les cellules d'un patient. L'étude peut se faire en milieu interphasique (frottis sanguin, coupe de tissu), où seules des anomalies quantitatives seront identifiées (perte ou duplication d'un signal), ou après une culture cellulaire, où l'analyse des mitoses permet d'apprécier des anomalies à la fois quantitatives mais aussi qualitatives, par l'analyse de la structure chromosomique. La résolution de cette analyse dépend donc de la taille des sondes utilisées. Cette technique est l'une des techniques de choix dans la validation et la vérification d'anomalies de structure identifiées par d'autres techniques.

#### **4.1.2. Analyse chromosomique par puce à ADN**

L'**analyse chromosomique par puce à ADN** ou **ACPA** est utilisée en diagnostic depuis les années 2000/2010. Cette analyse est maintenant utilisée en première intention en diagnostic de cytogénétique pour la grande majorité des indications cliniques : DI isolée ou syndromique, malformations congénitales, TND, épilepsie.

L'ACPA est une technique de cytogénétique moléculaire quantitative pangénomique basée sur la comparaison d'un ADN de patient à l'ADN d'un témoin.

La **CGH array** (*Comparative genomic hybridization*) est un type d'ACPA dans lequel l'ADN d'un patient et d'un témoin sont marqués par deux fluorochromes différents, puis cohybridés en quantité égale sur une lame de verre sur laquelle sont fixés des milliers de fragments d'ADN couvrant l'ensemble du génome (la « puce à ADN »). Une mesure quantitative de la fluorescence permet de déterminer des variations du nombre de copies, délétions et duplications. La résolution de la puce

utilisée correspond à la distance moyenne entre les sondes, classiquement entre 1 Mb et quelques dizaines de kilobases.

La **puce de type SNP** (pour *Single Nucleotide Polymorphism*) est un autre type d'ACPA qui diffère de la précédente par l'absence de compétition entre l'ADN de l'individu à tester et un ADN témoin. Dans ce cas, l'ADN du patient est déposé sur la puce sur laquelle sont déjà déposées les sondes. Ces sondes sont des oligonucléotides qui se terminent par un SNP, des « polymorphismes de simple nucléotide », c'est à dire présents avec une fréquence allélique >1% dans la population générale. Les sondes SNP ont été choisies parmi des SNP identifiés par les projets HapMap ou *1000 genomes* (1000 Genomes Project Consortium et al., 2010; International HapMap Consortium et al., 2007; Kong et al., 2008). Ce type de puce permet une analyse quantitative (identification de CNVs) mais aussi qualitative, d'analyse de génotypage. Elle permet donc d'identifier des régions d'homozygoties (conséquence d'union consanguine quand réparties sur plusieurs chromosomes ou le reflet d'une isodisomie uniparentale), de préciser l'origine parentale d'un CNV quand la puce SNP des parents de l'individu est disponible, et permet également l'identification des triploïdies.

### **Limites de l'ACPA**

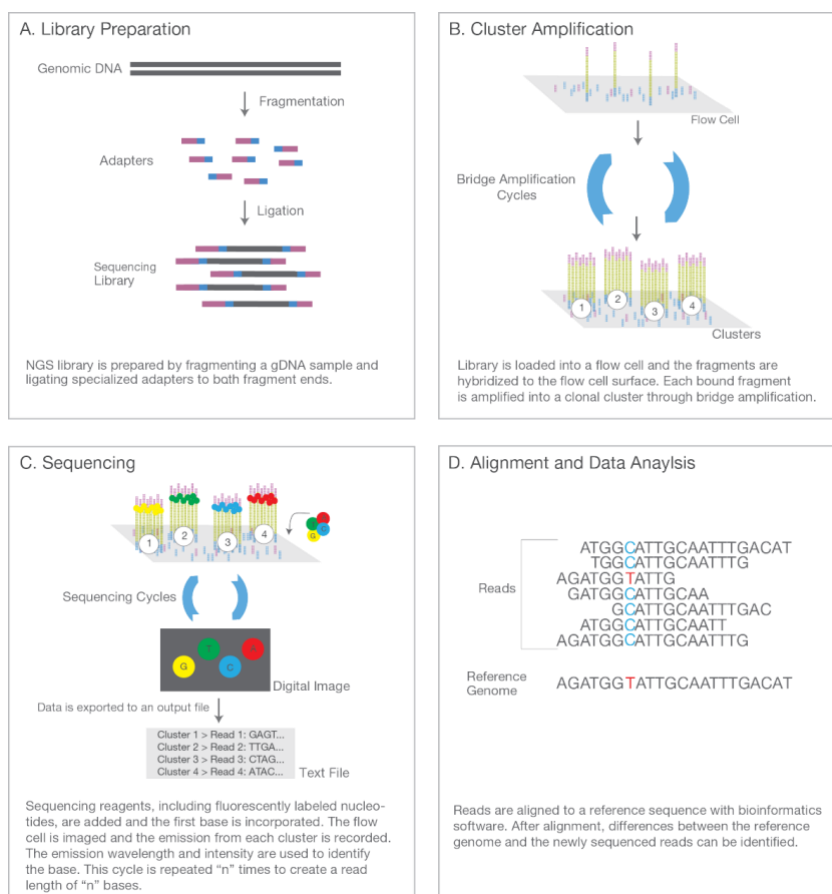
Cette technique ne permet pas d'apprécier la structure des chromosomes, qu'il s'agisse de la position d'un segment dupliqué, ou de remaniements de structure équilibrés comme les translocations, insertions ou inversions. La résolution des CNVs mis en évidence dépend du nombre et de la localisation des sondes présentes dans la puce. Certaines mosaïques chromosomiques faibles peuvent échapper aux deux types d'ACPA. De plus, ces techniques ne permettent pas l'identification de variations nucléotidiques.

## **4.2. Séquençage haut débit**

### ***4.2.1. Principe du séquençage haut débit***

Les techniques de séquençage permettent de déterminer la séquence nucléotidique de l'ADN. Initié par les Sanger, les techniques de SHD ont donné un nouvel essor à la génétique moléculaire, en accélérant le recueil des données et le nombre d'échantillons séquencés en parallèle. Aujourd'hui en diagnostic le séquençage peut cibler des régions prédéfinies du génome (panels ou encore exome),

ou l'ensemble du génome lui-même. Le SDH s'organise en plusieurs étapes (figure 20).



**Figure 20.** Les différentes étapes d'une technique de séquençage haut débit (technique Illumina)

**A.** Préparation de la librairie **B.** Amplification de la librairie sur la flow-cell **C.** Séquençage **D.** Alignement des séquences et analyses bio-informatiques

Source : documentation Illumina <https://www.illumina.com/>

### Préparation de la librairie (figure 20.A)

Une première étape avant le passage sur les plateformes de SHD et après la fragmentation de l'ADN consiste à sélectionner la partie du génome que l'on voudra séquençer pour les analyses ciblées, c'est-à-dire constituer ce que l'on appelle sa « **librairie** ». Une étape préalable pour toutes les techniques de SHD (en dehors du WGS), la « **capture** », utilise des sondes d'hybridation spécifiques pour cibler et capturer les régions d'ADN d'intérêt. Une fois que ces régions sont isolées, elles sont amplifiées et adaptées avec des séquences d'index uniques, permettant ainsi de générer une librairie d'ADN prête pour être ensuite séquençée à proprement dit (Audebert et al., 2014).

### Séquençage (figure 20.B et C)

Les techniques de séquençage diffèrent ensuite selon les sociétés de production, mais ont pour point commun de détecter un signal à chaque cycle d'incorporation d'un nouveau nucléotide pendant l'élongation. Elles peuvent s'appuyer sur différents mécanismes, les plus fréquents étant l'incorporation de nucléotides fluorescents terminateurs réversibles, ou *séquençage par synthèse* (Illumina), où les fragments d'ADN sont fixés sur une surface solide, puis se fait l'extension de l'ADN. Chaque cycle correspond à l'introduction d'un nucléotide qui émet un signal lumineux spécifique qui sera détecté par l'automate.

Une autre méthode consiste en la mesure du pH suite à la libération d'un ion H<sup>+</sup> après l'ajout d'un nucléotide (Ion Torrent).

L'étape d'extension va donc produire des fragments de plusieurs nucléotides dont la séquence est « lue » par l'automate. Ces fragments sont des **lectures** (ou **reads**), et ont une taille définie pour chaque technologie (100 à 300pb chez Illumina). Elles peuvent être générées par paire, on parle de séquençage en **paired-end** : les fragments d'ADN sont « lus » dans les deux sens (figure 21). La distance entre les deux reads d'une même paire est connue, et des variations de celle-ci pourront être analysées. Une autre forme de séquençage consiste à séquencer les reads d'une seule extrémité le : single-read.

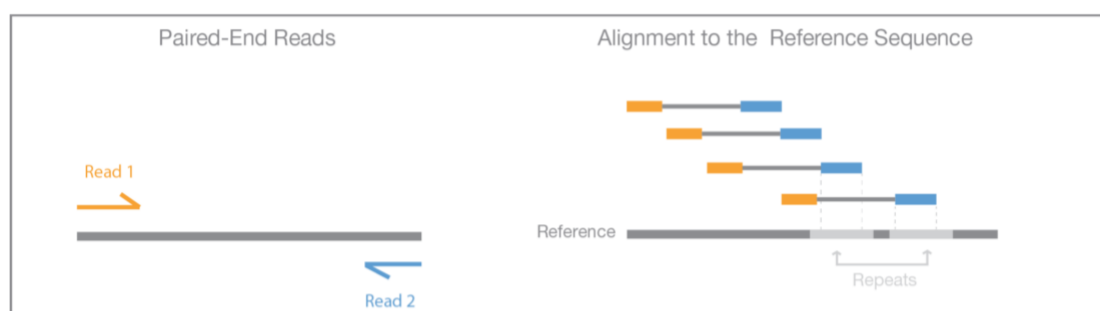


Figure 4: Paired-End Sequencing and Alignment—Paired-end sequencing enables both ends of the DNA fragment to be sequenced. Because the distance between each paired read is known, alignment algorithms can use this information to map the reads over repetitive regions more precisely. This results in better alignment of reads, especially across difficult-to-sequence, repetitive regions of the genome.

## Figure 21. Séquençage en paired-end

Source : documentation Illumina <https://www.illumina.com/>

### Alignement et analyse bio-informatique (figure 20.D)

Cette étape s'organise globalement en trois étapes.

La première est tout d'abord **d'aligner le génome** séquencé du patient sur le génome de référence en vigueur (les versions GRCh37 ou 38 actuellement ; voir [chapitre 1/4.2.5](#)).

La deuxième étape est le « **calling** » ou **l'appel des variant**, et consiste à comparer la séquence obtenue à la séquence de référence et d'en identifier les variations. Les outils utilisés seront détaillés au [chapitre 1/4.2.3](#).

La troisième est **l'annotation**, qui va déterminer la position des gènes, des éléments régulateurs et autres séquences sur le génome.

Deux caractéristiques vont par ailleurs définir la qualité de SHD : sa **couverture**, c'est le pourcentage de la librairie qui est couverte par des lectures ou reads, et sa **profondeur**, c'est-à-dire le nombre de fois qu'un nucléotide est séquencé. Ces paramètres sont inversement liés à la taille de la librairie.

Ensuite, de multiples outils bio-informatique vont aider à déterminer des variations dans la séquence ou dans la structure du génome.

#### **4.2.2. Séquençage ciblé et génome entier**

**Le séquençage ciblé** peut concerner un panel de gènes d'intérêt ou l'ensemble des régions codantes du génome, l'exome. **Le séquençage d'exome** (ou **WES**) est une étude ciblée de l'ensemble des séquences codantes du génome, c'est-à-dire tous les exons, plus ou moins les sites d'épissages et régions 5' et 3' UTR selon les kits utilisés. L'exome représente environ 30Mb, soit 1% du génome (Majewski et al., 2011). Comme précisé, les régions séquencées sont déterminées dès le départ par la constitution de la librairie puis par la Capture.

**Le séquençage de génome entier (WGS)** ne passe pas par une étape de capture préliminaire puisque l'ensemble du génome va être séquencé. Les reads sont donc répartis sur l'ensemble du génome.

#### **4.2.3. Outils d'analyses en WGS**

Les techniques de SHD ont pour objectif premier d'identifier des SNV, pathogènes, ou probablement pathogènes. Les panels et le WES ont pour avantage d'être plus rapides et moins coûteux en raison de leur caractère ciblé, bien que l'exome couvre la très grande partie des régions impliquées dans des variations probablement

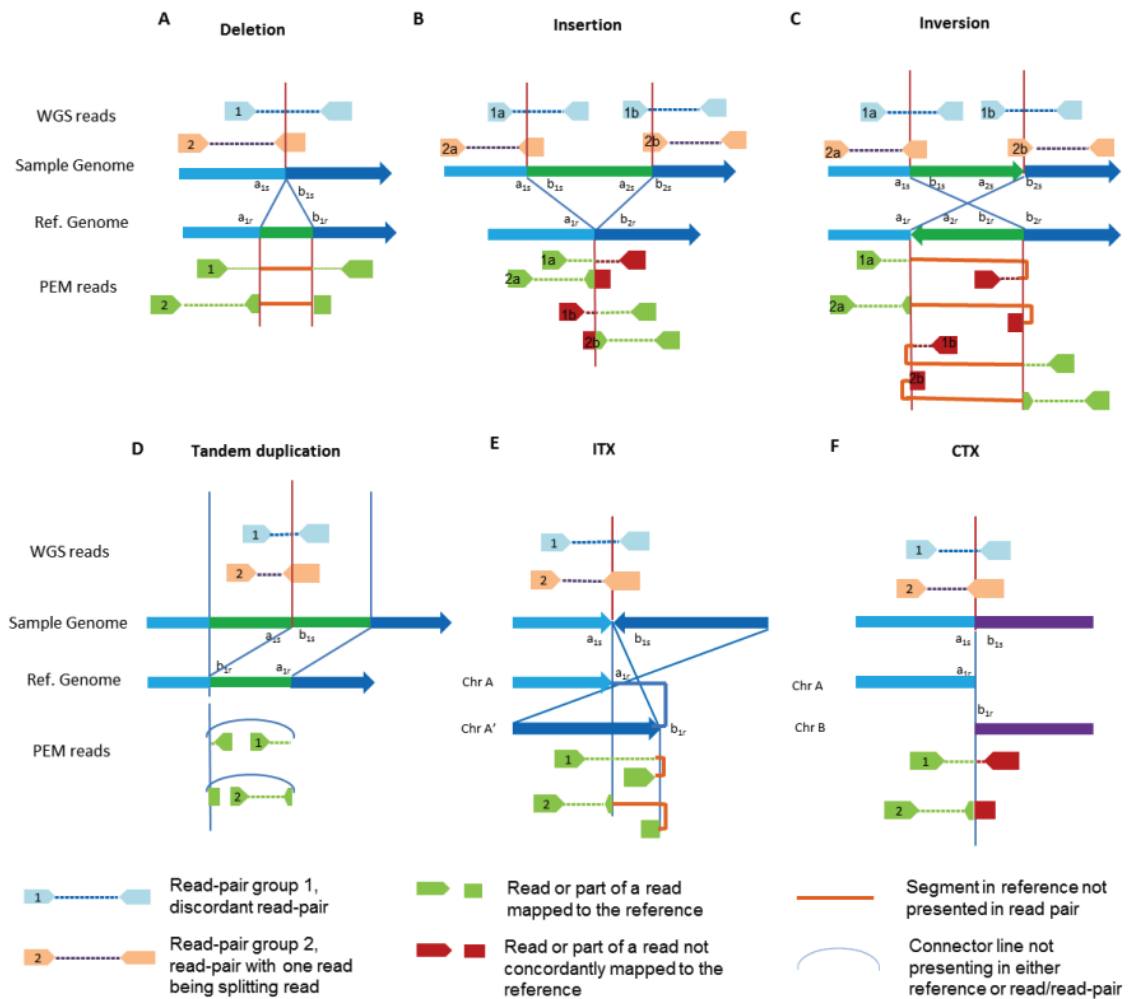
pathogènes, c'est-à-dire les régions codantes (Goh & Choi, 2012). Il est possible de déterminer des petits CNVs avec certains pipelines d'exome (Charng et al., 2016; Retterer et al., 2016; Srivastava et al., 2019; Vissers et al., 2017) mais l'analyse est limitée du fait de la couverture très ciblée des exons et ne peut identifier d'autres variations de structure (les introns, régions régulatrices et autres séquences non codantes ne sont pas couverts).

L'analyse du WGS permet en plus l'identification de variations de structure. En effet, des algorithmes qui associent l'analyse de la profondeur de lecture d'une région avec la détection d'anomalies d'appariements des paired-end reads va permettre d'objectiver des anomalies de structure avec des points de cassure précis. Ces anomalies des paired-end read sont les reads **discordants** et les **splits-reads** (Liu et al., 2015) (Figure 22).

Une paire de read est dites **discordante** quand certaines propriétés ne sont pas respectées : la distance entre les deux reads est trop longue (signe d'une délétion) ou trop courte (signe d'une duplication), l'orientation des deux reads n'est pas normale (classiquement un *forward* et un *reverse*), qui peut traduire une inversion, ou encore la position du deuxième read (le *mate*) est incohérente : trop loin sur le chromosome voir sur un autre chromosome, qui peut révéler une insertion ou une translocation.

Un read fractionné, ou **split read**, est un read qui se situe sur un point de cassure. La première partie s'aligne sur la séquence normale tandis que la deuxième partie s'aligne sur une autre région génomique.





**Figure 22.** Représentation des reads fractionnés et discordants avant et après alignement

**A.** Délétion **B.** Insertion **C.** Inversion **D.** Duplication en tandem **E.** Translocation intra et **F.** inter chromosomique. Les reads dans la partie supérieure (bleu pâle et orange) correspondent aux reads obtenus après le séquençage. Les reads dans la partie inférieure (verts et rouges) correspondent à ce qui est visualisé après l'alignement sur le génome de référence.

Source : B. Liu, 2015, *Oncotarget*

Il devient donc plus simple de détecter n'importe quelle variation de structure, qu'elle associe délétion, insertion, inversion ou translocation grâce à ces éléments et la combinaison de plusieurs algorithmes, tels que Lumpy (Layer et al., 2014) ou CNVnator (Abyzov et al., 2011). **Lumpy** combine de multiples signaux pour identifier ces remaniements : la taille entre les paires de reads, les split-read et la profondeur. **CNVnator** quant à lui va se concentrer plus particulièrement sur la profondeur de lecture.

D'autre part, le WGS permet d'identifier des expansions de répétition responsables de nombreuses pathologies neurologiques ou de TND, comme le syndrome de l'X fragile, la maladie de Huntington ou encore la Sclérose latérale amyotrophique, avec des régions pour le moment plus propices aux faux positifs,

comme dans le gène *FMR1* (Ibañez et al., 2022). Ces expansions sont détectées grâce à des outils tels que Expansion Hunter (Dolzhenko et al., 2020) mais restent limitées à de petites variations (les grands allèles sont encore moins précisément détectés).

#### **4.2.4. Deux types de SHD : Short VS Long reads**

L'analyse d'un séquençage (WES ou WGS) repose entre-autre sur ses reads, qui peuvent être de deux types :

**Les short reads** sont de courtes séquences d'ADN, généralement 100 à 300pb. Ils offrent une grande précision et une plus grande couverture du génome. Cependant, leur courte longueur peut rendre difficile l'assemblage de séquences génomiques complexes, notamment dans les régions répétées comme les LCR (Mastrososa et al., 2023). Ils peuvent manquer des variations de structure > 50pb.

**Les long reads** sont significativement plus longs, atteignant souvent une taille supérieur à 10kb, jusqu'à plusieurs Mb parfois (Warburton & Sebra, 2023). Ils sont meilleurs pour l'assemblage de génomes *de novo* et la caractérisation de régions génomiques complexes. Ils vont analyser l'ADN natif (sans amplification), mais peuvent avoir une précision légèrement inférieure (précision à 5 bases près) et être plus coûteux à produire. De plus leur taille permet de connaître la phase des allèles.

#### **4.2.5. Génome de références**

##### **4.2.5.1. GRCh37 et 38**

Le génome de référence sert de point de comparaison à tous les outils qui utilisent des données de génétiques. De nombreuses régions restent mal séquencées, notamment les régions riches en séquences répétées, les BCCA, les régions centromériques, qui sont difficiles à aligner sur le génome de référence où elles sont encore mal annotées, et en pratique non analysables sur les références habituelles.

La première version du génome de référence humain a été générée grâce à un projet de recherche appelé le Projet Génome Humain (*Human Genome Project*, HGP) et a marqué une étape majeure dans la compréhension de la génétique humaine. Le HGP a été un effort de collaboration international qui a duré plusieurs années et qui a abouti à la publication de la séquence complète du génome humain en 2001 (International Human Genome Sequencing Consortium, 2004; McPherson et al.,

2001). Cette première référence s'est basée sur des échantillons d'ADN collectés à partir de plusieurs donneurs anonymes. La séquence identifiée a ensuite été **annotée** par plusieurs projets, *Reference Sequence* ou RefSeq pour la base NCBI (Pruitt & Maglott, 2001; Wheeler et al., 2000) ou encore le projet GENCODE (Harrow et al., 2012). Ces données sont régulièrement corrigées et mises à jour dans les différentes bases de données.

Aujourd'hui, les génomes de référence GRCh37 (*Genome Reference Consortium Human Reference 37*) et GRCh38 (*Genome Reference Consortium Human Reference 38*) sont les deux versions de référence du génome humain les plus couramment utilisées.

**La version GRCh37** a été publiée en 2009, et contient des informations détaillées sur la séquence de l'ADN, les emplacements des gènes, les régions non codantes, les variations génétiques et d'autres éléments du génome humain. Elle est toutefois limitée dans l'annotation de certaines régions génomiques répétées et complexes.

**La version GRCh38**, propose une meilleure représentation de ces régions répétées et complexes que sont les régions génomiques télomériques et centromériques, ce qui en fait une référence plus complète et plus précise que GRCh37, mais encore incomplète avec 151 Mb de séquence non connue à travers le génome essentiellement dans les régions péri-centromériques et subtélomériques (Nurk et al., 2022).

#### **4.2.5.2. *Telomere to Telomere***

Le projet "**Telomere-to-Telomere**" (**T2T**) est une initiative de recherche visant à séquencer intégralement les chromosomes humains, en couvrant chaque extrémité télomérique. Ce projet est conçu pour produire une séquence complète, hautement précise et continue du génome humain, en améliorant la séquence du génome de référence existant (GRCh38) (Miga et al., 2020; Nurk et al., 2022). Les équipes se sont basées sur le génome d'une môle hydatiforme complète 46,XX, c'est à dire diploïdie mais avec la duplication du génome d'un seul parent, la rendant donc homozygote sur l'ensemble du génome (lignée cellulaire appelée *CHM13*).

Le projet T2T fait usage de multiples technologies de séquençage de nouvelle génération, telles que le séquençage Nanopore et le séquençage PacBio, des techniques de séquençage en long-read qui permettent de générer des lectures plus longues et de qualité supérieure que les méthodes précédentes.

Les points majeurs du projet sont tout d'abord l'inclusion des régions télomériques, qui sont les extrémités des chromosomes et sont souvent riches en séquences répétées et complexes, mais également d'autres régions encore peu détaillées, notamment les BCCA. Cette nouvelle version du génome a ainsi permis de corriger près de 238 Mb de séquences, absentes ou erronées sur la dernière version GRCh38, essentiellement comprises dans les séquences d'ADN satellite centromérique (76%), des duplications segmentaires et l'ADN ribosomal.

# Chapitre 2 : Anneaux des chromosomes 22 et données de WGS

## 1. Objectifs

Le PMS est un syndrome avec une grande hétérogénéité clinique, liée à la perte de fonction du gène *SHANK3* sur le chromosome 22. Ces pertes de fonctions peuvent être liées à des mutations perte de fonction ou des délétions qui peuvent être interstitielles ou terminales. Les délétions terminales peuvent être des délétions simples, secondaires à des dérivés de translocations impliquant un chromosome 22, ou encore dans 10% des cas environ des délétions stabilisées par la formation d'un anneau. Le risque de l'anneau du chromosome 22 (r(22)) réside dans son instabilité en division et le risque d'être perdu dans certaines populations cellulaires (création d'une monosomie 22 en mosaïque). Cet événement entraîne un risque de développer une neurofibromatose de type 2 (NF2) qui se caractérise par la formation de tumeurs du système nerveux central, notamment des schwannomes vestibulaires ou des méningiomes, avec un risque de complications et de décès non négligeable si non pris en charge.

Aujourd'hui, le diagnostic d'anneau ne peut être posé que par la réalisation d'un caryotype qui permet d'analyser la structure des chromosomes. Or cette technique n'est plus réalisée de façon systématique chez des patients avec un TND au profit de l'ACPA et du WES, et le nombre de r(22) est très probablement sous-estimé chez les patients avec un PMS. Pourtant il serait essentiel de pouvoir orienter dès les premiers examens génétiques les patients pour qui la réalisation d'un caryotype est nécessaire.

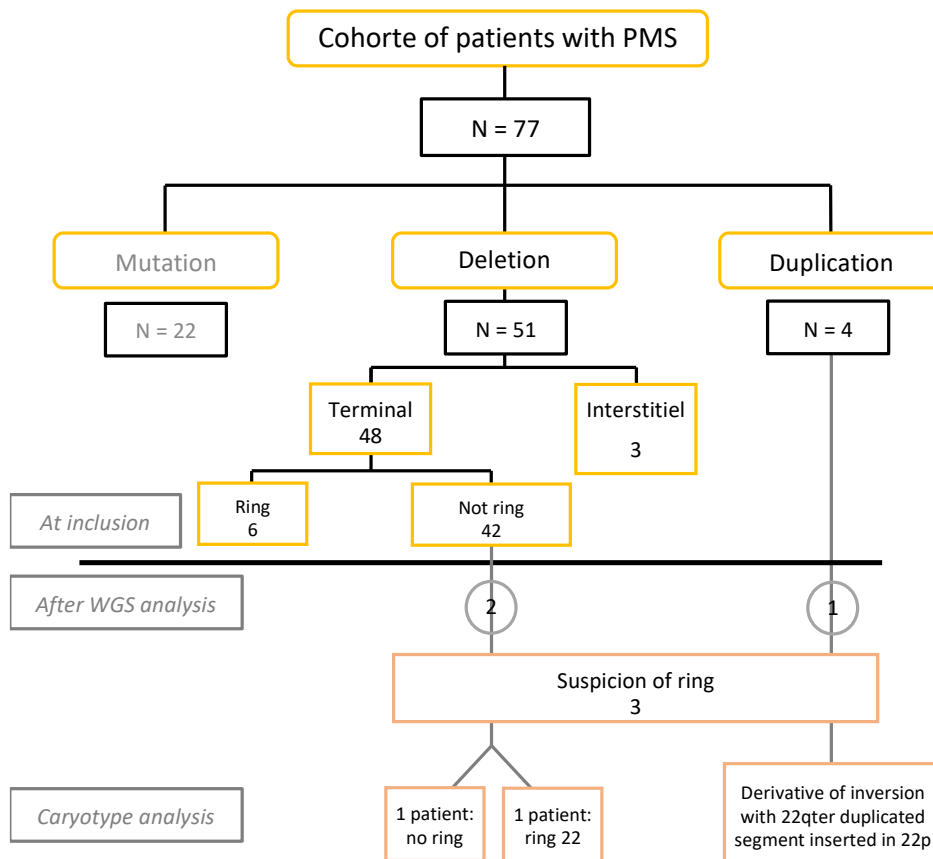
**Ainsi l'objectif principal de ce projet est d'étudier l'apport du séquençage du génome entier dans l'identification des mécanismes chromosomiques responsables d'un PMS par formation d'un r(22).** Secondairement, nous avons voulu évaluer l'apport du WGS dans le diagnostic génétique global des patients, c'est à dire l'identification de variations additionnelles significatives dans le phénotype des patients.

## 2. Méthodologie

### 2.1. Cohorte de patients avec PMS

Nous avons travaillé sur une cohorte de 77 patients atteints de PMS recrutés en France et issus de 39 centres hospitaliers. Le diagnostic génétique a été réalisé par ACPA, MLPA et/ou séquençage, plus rarement par FISH. Les données phénotypiques ont été recueillies via un tableur standardisé par les généticiens cliniciens initialement en charge des patients. Après présentation de l'étude et accord de la famille, chaque patient et ses parents, lorsque les prélèvements étaient disponibles, ont bénéficié d'un séquençage du génome entier.

Parmi eux, 22 portent une mutation dans le gène *SHANK3*, dont nous n'étudierons pas les données, 4 portent une duplication terminale 22qter et 51 une délétion (figure 23) impliquant *SHANK3* parmi lesquelles 3 sont interstitielles. Finalement, 48 patients présentent une délétion terminale 22qter parmi lesquels 6 patients présentent un anneau du chromosome 22.



**Figure 23.** Cohorte des patients avec PMS

## 2.2. Séquençage de génome entier

Le séquençage de génome a été réalisé par le Centre National de Recherche en Génomique Humaine (CNRGH). La plateforme, Illumina NovaSeq 6000, utilise un kit TruSeq DNA PCR-free kit d'Illumina, avec un séquençage en paired-end (2x150 bases). L'alignement se fait sur la référence GRCh37 via l'outil BWA, et les fichiers CRAM (*Compressed Reference-oriented Alignment Map*) sont produits grâce à l'outil samtools (Danecek et al., 2021). L'ensemble des variations identifiées, CNVs, SNVs et indels, sont décrites dans la version du génome **GRCh37**.

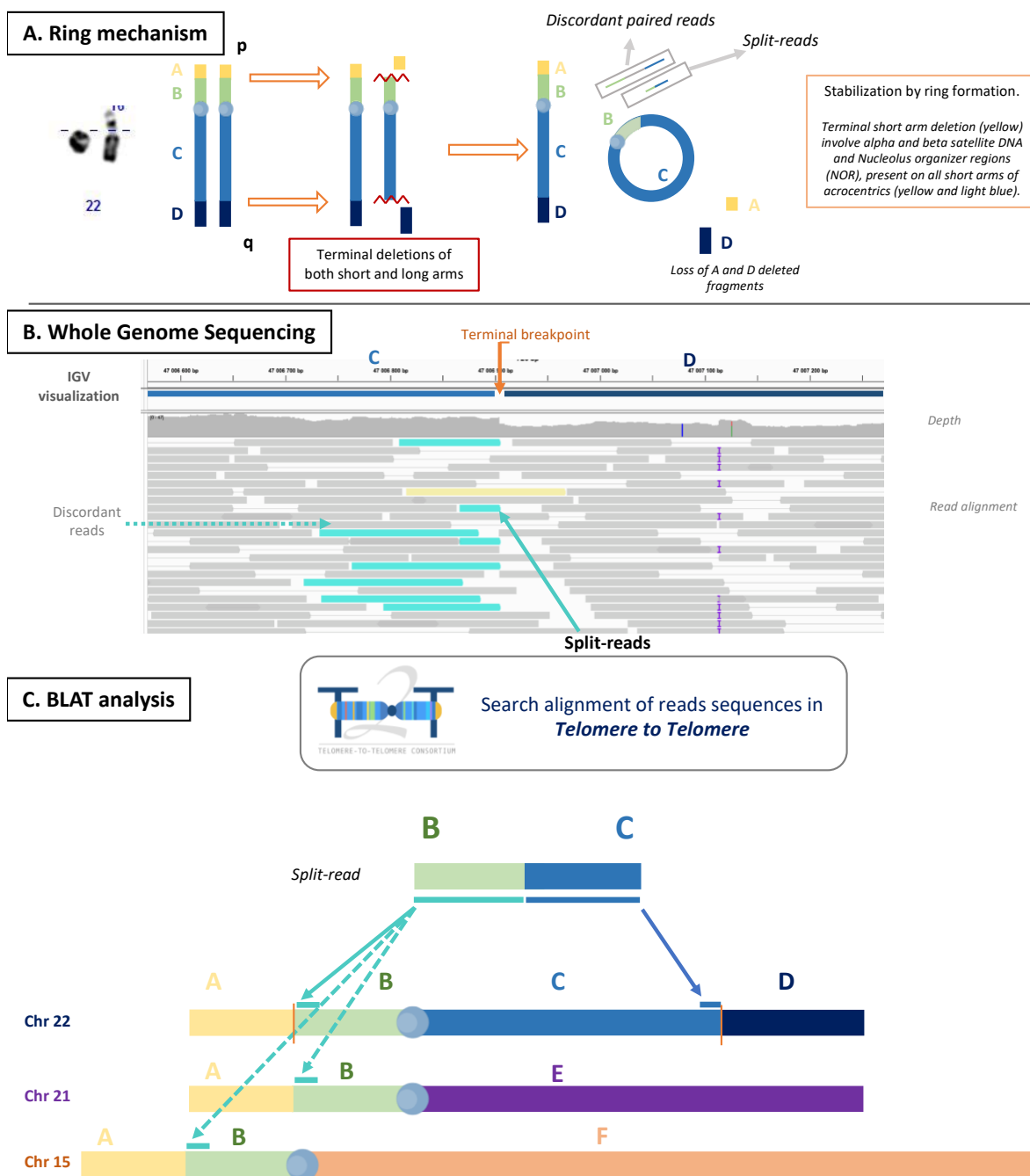
Les **CNVs** ont été identifiés avec l'outil ClinSV (v1.0) qui combine les deux algorithmes Lumpy et CNVnator. Les **SNVs** et **indels** l'ont été en utilisant l'outil GATK. L'annotation du fichier VCF obtenu a été faite par Ensembl Variant Effect Predictor (VEP v104) et la pathogénicité de ces variations a été déterminée selon les critères sus cités (ACMG et Réseau Achropuce) complétés de scores de prédictions. Ainsi l'interprétation prend en compte :

- la fréquence allélique dans la population d'après la base de données gnomAD (<https://gnomad.broadinstitute.org/>), qui recense des données de WES pour 125 748 individus et de WGS pour 15 708 individus
- Des scores d'intolérance à la perte de fonction du gène comme la pLI > 0,9 ou le score LOEUF < 0,35
- Des scores de prédiction sur l'effet de la protéine, tels que Polyphen (<http://genetics.bwh.harvard.edu/pph2/>) ou SIFT (<https://sift.bii.a-star.edu.sg/>)
- Des scores d'impact tels que le CADD (Rentzsch et al., 2019, 2021) ou le score MPC (Samocha et al., 2017) qui prennent en compte de nombreux éléments dont ceux cités ci-dessus

Les fichiers CRAM de tous les patients ont été analysés sur l'interface *Integrative Genomics Viewer* (IGV), versions v2.8.9 à v2.16 (Robinson et al., 2011).

Le but de cette première approche était d'analyser les points de cassure proximaux de chaque patient dans le but d'identifier des split reads et des reads discordants, qui s'aligneraient sur une région correspondant à un BCCA (figure 24.B) et seraient donc en faveur de la formation de l'anneau (figure 24.A). Un premier BLAT des séquences des **split-reads** sur les versions GRCh37 et 38 n'était pas probant. Nous avons ensuite fait un BLAT des séquences sur le récent génome de référence **T2T**, qui a montré des

alignements multiples sur différentes régions de BCCA de différents chromosomes acrocentriques (figure 24.C).



**Figure 24.** Principe d'analyse des points de cassure en 22qter

**A.** mécanisme de formation d'un anneau. **B.** recherche de split-reads sur les fichiers CRAM via IGV. **C.** Recherche de séquences qui s'alignent sur les bras courts de chromosomes acrocentriques en faisant un BLAT sur *Telomere to telomere*

### 2.3. Séquençage Sanger

Des examens complémentaires par **séquençage Sanger** ont été réalisés, notamment pour le patient P1. Les amorces utilisées ont été *designées* sur les plateformes *Primer Blast* de NCBI (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>)



(Ye et al., 2012) et *Primer3Plus* (<https://dev.primer3plus.com/index.html>) (Untergasser et al., 2012). Les différentes amorces utilisées sont listées dans la table 2.

Name	Séquence (5'→3')	Tm	Position (GRCh37)
ring22_P2_CtrlE_F	AGCTGTGTCAGGCAGAAGAC	59.97	chr22:43964764-43964783
ring22_P1_F1rev	GATTCGTCCAGTCCAGTGCA	60.04	chr22:43964097-43964116
ring22_P1_F4rev	GCCAGAACATAAGCAGCCAC	59.55	chr22:43964126-43964145
ring22_P1_F6_CtrlC	CTCCTGTCACTCTCGGTCAC	59.48	chr22:43095478-43095497
ring22_P1_R6	CATCTCCCTCTGTGCAGTCC	59.82	chr22:43095622-43095641
ring22_P1_F7	GGGTTGGAGGGGGACTCATA	60.33	chr22:43965000-43965019
ring22_P1_R7	CGTCTCCTGGGTAGATTGCC	59.89	chr22:43096100-43096119
ring22_P1_F8	AGGACCCGTGTAGTCACAGA	59.89	chr22:43965292-43965311
ring22_P1_R8e	GCCCCTTAGGACTGACACAA	59.31	chr22:43096209-43096228
ring22_P1_R8g	CACGCTAAGGCAGAAGGAGT	59.75	chr22:43965489-43965508
chr22_TYMP_F1_50964437	CAGCCTCTGACCCACGTC	59.7	chr22:50964437-50964454
chr22_TYMP_R1_50965018	CGCCAGACTTAAGGGACCTG	60.1	chr22:50965018-50965037
chr22_TYMP_F2_50964451	CGTCGACCAGCAGCTCTG	60,5	chr22:50964451-50964468
chr22_TYMP_R2_50964437	CATGGACAAGCCCCTGGG	60	chr22:50964437-50964454
chr22_TYMP_F3_50964576	GGGAGAGGGGCTGAGAGG	60,4	chr22:50964576- 50964593
chr22_TYMP_R3_50965173	CACGCTGAGCCTCCTCTC	59,5	chr22:50965173-50965190

**Table 2.** Amorces utilisées pour séquençage Sanger

Les profils du séquençage Sanger ont été analysés sur la plateforme *Alamut™ Visual Plus* de Sophia Genetics™, version 1.5.

## 2.4. Recherche de régions d'homologie

Les régions d'homologie ont été étudiées sur une fenêtre de +/- 100pb autour des points de cassure pour les 7 patients porteurs d'anneau et pour 45 individus porteurs de délétion terminale ou interstitielle. Nous avons utilisé le dataset Repeat Masker pour identifier les régions répétées (<https://www.repeatmasker.org/>) (Jurka, 2000).

### **3. Article – *gollum*, a new tool using short read sequencing to detect ring chromosomes in PMS patients**

**Titre : GOLLUM, A NEW TOOL USING SHORT READ SEQUENCING TO DETECT RING CHROMOSOMES IN PMS PATIENTS**

**Auteurs** : Myriam Rachid<sup>1,2</sup>, Freddy Cliquet<sup>2</sup>, Alexandre Mathieu<sup>2</sup>, Aline Vitrac<sup>2</sup>, Anne Boland<sup>3</sup>, Jean- Francois Deleuze<sup>3</sup>, Jonathan Lévy<sup>1</sup>, Damien Haye<sup>1</sup>, Sandrine Passemard<sup>1</sup>, Anna Maruani<sup>2,4</sup>, Frederique Amsellem<sup>2,4</sup>, Céline Dupont<sup>1</sup>, Catherine Yardin<sup>5</sup>, Damien Sanlaville<sup>6</sup>, Yline Capri<sup>1</sup>, Magalie Gorce<sup>7</sup>, Isabelle Marey<sup>8</sup>, Thomas Rolland<sup>2</sup>, Thomas Bourgeron<sup>2</sup>, Richard Delorme<sup>2, 4</sup>, Claire S Leblond<sup>2,\*</sup>, Anne-Claude Tabet<sup>1, 2,\*</sup>

\* *equal contribution*

**État** : soumis (en cours de review) dans la revue *NPJ Genomic Medicine*

**Affiliations :**

<sup>1</sup> Genetics Department, Robert Debré Hospital, APHP, Paris, France

<sup>2</sup> Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS, IUF, Université Paris Cité, Paris, France

<sup>3</sup> Université Paris-Saclay, Commissariat à l'Energie Atomique et aux Energies Alternatives, Centre National de Recherche en Génomique Humaine, 91057 Evry, France

<sup>4</sup> Department of Child and Adolescent Psychiatry, Robert Debré Hospital, APHP, Paris, France

<sup>5</sup> CNRS XLIM UMR 7252, Université de Limoges, Service de Cytogénétique, CHU Limoges

<sup>6</sup> Genetic Department, Hospices Civils de Lyon

<sup>7</sup> Clinical Genetic Unit, CH Le Mans

<sup>8</sup> Université Grenoble Alpes, INSERM, U1209, CHU Grenoble Alpes, Institute for Advanced Bioscience, Grenoble, France

# GOLLUM, A NEW TOOL USING SHORT READ SEQUENCING TO DETECT RING CHROMOSOMES IN PATIENTS WITH PHELAN-MCDERMID SYNDROME

## Authors:

Myriam Rachid<sup>1,2</sup>, Freddy Cliquet<sup>2</sup>, Alexandre Mathieu<sup>2</sup>, Aline Vitrac<sup>2</sup>, Anne Boland<sup>3</sup>, Jean-Francois Deleuze<sup>3</sup>, Jonathan Lévy<sup>1</sup>, Damien Haye<sup>1</sup>, Sandrine Passemard<sup>1</sup>, Anna Maruani<sup>2,4</sup>, Frederique Amsellem<sup>2,4</sup>, Céline Dupont<sup>1</sup>, Catherine Yardin<sup>5</sup>, Damien Sanlaville<sup>6</sup>, Yline Capri<sup>1</sup>, Magalie Gorce<sup>7</sup>, Isabelle Marey<sup>8</sup>, Thomas Rolland<sup>2</sup>, Thomas Bourgeron<sup>2</sup>, Richard Delorme<sup>2,4</sup>, Claire S Leblond<sup>2,\*</sup>, Anne-Claude Tabet<sup>1,2,\*</sup>

## Affiliations:

<sup>1</sup> Genetics Department, Robert Debré Hospital, APHP, Paris, France

<sup>2</sup> Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS, IUF, Université Paris Cité, Paris, France

<sup>3</sup> Université Paris-Saclay, Commissariat à l'Energie Atomique et aux Energies Alternatives, Centre National de Recherche en Génomique

Humaine, 91057 Evry, France

<sup>4</sup> Department of Child and Adolescent Psychiatry, Robert Debré Hospital, APHP, Paris, France

<sup>5</sup> CNRS XLIM UMR 7252, Université de Limoges, Service de Cytogénétique, CHU Limoges

<sup>6</sup> Genetic Department, Hospices Civils de Lyon

<sup>7</sup> Clinical Genetic Unit, CH Le Mans

<sup>8</sup> Université Grenoble Alpes, INSERM, U1209, CHU Grenoble Alpes, Institute for Advanced Bioscience, Grenoble, France

\* equal contribution

Corresponding author: Claire S leblond, [claire.leblond@pasteur.fr](mailto:claire.leblond@pasteur.fr)

## Running head

Ring chromosome 22 detection using gollum

## Abstract

Most of the individuals with Phelan-McDermid syndrome (PMS) present with hypotonia, developmental delay, mild to severe intellectual disability (ID), speech disorder and autism. PMS is caused by a deletion of 22q13.3 or a deleterious variation of the *SHANK3* gene. Ring chromosomes are rare cytogenetic anomalies due to the circularization of chromosomes mainly formed after breakage of both arms of a chromosome followed by the fusion of the broken ends. In PMS, rings of chromosome 22 (r(22)) are observed in 10 to 20% of the individuals carrying a chromosome 22 terminal deletion and are associated with increased risk of developing nervous tumors related to the loss of tumor-suppressor genes (e.g., *SMARCB1* and *NF2*) located on chromosome 22. Here, we investigated the whole genome sequence (WGS) of a cohort of individuals with PMS and developed “Gollum” a new bioinformatic tool to easily and robustly detect ring chromosome. Gollum analyzes the alignment of short paired-end read sequences on the recent Telomere to Telomere (T2T) human genome and detect rings of chromosome in any acrocentric chromosome. Gollum is easily added to the pipeline of WGS analyses and improve the management of individuals with PMS and ring chromosome to reduce the risk for cancer.

## Introduction

The vast majority of individuals with PMS (MIM #606232; ICD-11 #LD44.NY) have an haploinsufficiency of the *SHANK3* gene due to an interstitial/terminal deletion of 22q13.3 region or a loss-of-function mutations<sup>1</sup>. Individuals with PMS can have different clinical features, but in majority they present with hypotonia, developmental delay, mild to severe intellectual disability (ID) and speech disorder, from delayed to complete absence of speech. Approximately 80% of cases present with autism, and other medical conditions are variably associated<sup>2,3</sup> (gastro esophageal reflux disease, epilepsy, lymphedema, cardiopathy, ophthalmic disorder...). Such phenotypic variability is partially due to the size of the 22q13 deletions and/or to additional genomic variants across the genome<sup>3</sup>.

It is estimated that 10 to 20% of chromosome 22 terminal deletions are ring-related<sup>1</sup>. Ring chromosomes are rare cytogenetic anomalies due to the circularization of chromosomes. They are formed mainly after breakage of the distal part of both short and long arms of the same chromosome followed by the fusion of the broken ends. In rare cases, circularization occurs by telomere junction, without loss of DNA and leading to “complete rings”<sup>4</sup>. Incidence of ring chromosomes is estimated between 1/30,000 and 1/60,000 live birth<sup>5</sup>. All chromosomes can be involved in ring formation, with higher incidence for chromosomes 18 and 20 accounting for 10-12% of cases each, while rings of acrocentric chromosomes (chromosomes 13, 14, 15, 21, 22) account for 7 to 8% of cases each<sup>6</sup>. Patients with r(22) have an increased risk of developing nervous tumors related to tumor-suppressor genes located on chromosome 22, including *SMARCB1* at 22q11.2 and *NF2* at 22q12.2. Indeed, mitotic instability of the ring can lead to the loss of this chromosome. Such monosomic cells can have a second somatic hit on the *NF2* gene causing histologically benign tumors such as schwannoma or meningioma<sup>7,8</sup> or on *SMARCB1* leading to brain teratoid/rhabdoid tumors<sup>9</sup>. Ring chromosomes are mainly detected by karyotyping or Fluorescence in situ hybridization (FISH) but can also be suspected by Chromosomal Micro Array (CMA) when fragments of short and long arms of the same chromosome are deleted. CMA however cannot detect rings derived from acrocentric chromosomes since no probe targeting their short arm is available. Indeed, short arms of acrocentric chromosomes (SAACs) are composed of alpha and beta satellites DNA, and Nucleolar organizer regions with ribosomal DNA<sup>10</sup>. Therefore, the frequency of r(22) in PMS patients could be underestimated.

Whole genome sequencing (WGS) is becoming a standard practice in hospitals of high income countries given the cost effectiveness and the increase of diagnostic rate and also given the possibility to go back to the data for new analyses<sup>11,12</sup>. Currently, WGS analyses are mostly based on pipelines using human genome reference GRCh 37 or 38, for which some specific sequences are still missing, such as centromeric regions or SAACs, and thus failed to identify rings in acrocentric chromosomes. Even optical genomic mapping, which is a promising molecular technology to detect structural variations (SV), failed to identify SV lying within large repetitive, unmappable regions, constitutive heterochromatin and p arm of acrocentric including rings derived from acrocentric chromosomes<sup>13</sup>. So far, rings can be only identified using karyotype and/or FISH, which cannot be straightforwardly used in clinic. Recently,

thanks to the Telomere-to-Telomere (T2T) consortium<sup>14</sup>, a large part of the remaining human genome has been sequenced including the SAACs. This should bring changes in the possibilities of identifying these SV.

Investigating a cohort of individuals with PMS and WGS data, we developed a new tool that analyzes the alignment of short paired-end read sequences on the recent T2T human genome that can robustly detect rings of chromosome 22 and more generally of all acrocentric chromosomes. This method has direct clinical implication on the management of the individuals with PMS with ring chromosome.

## **METHOD**

### **Recruitment**

Individuals were first explored by conventional cytogenetic and molecular tools including standard karyotype (GTG and RHG banding), FISH and CMA depending on the lab and year at diagnosis. When a copy-number variation (CNV) covering the 22q13 region was detected, individuals and their parents were invited to participate in a research protocol including WGS. Fifty-five individuals with a 22q13 CNV were included in our study. In addition, 27 individuals carrying loss-of-function variants (SNV or indel) altering *SHANK3* were also included in our cohort, as well as their relatives.

### **Whole genome sequencing**

WGS was performed in collaboration with the Centre National de Recherche en Génomique Humaine (CNRGH) using Illumina TruSeq DNA PCR-free kits. The sequencing was performed in paired-end (2x150 bases) on the Illumina NovaSeq 6000 platform of the CNRGH. The average depth of sequencing is 30x for each sample. Read mapping was initially performed at CNRGH onto the GRCh37 human reference genome version using BWA, with mark duplicates using sambamba and file compression into CRAM using samtools.

### **Manual detection of ring 22**

Short read sequencing data from six patients with previously known r(22) were used to develop a pipeline to detect acrocentric rings. Visual inspection of each 22qter breakpoint was performed using Integrative Genomics Viewer (IGV)<sup>15</sup> and corresponding CRAM files in order to isolate split reads (reads with a portion of the sequence mapping on 22qter and a portion of the sequence that mapped outside the 22q arm). Realignment of the split read sequences was performed using BLAT<sup>16</sup> against the T2T human genome reference (Jan. 2022 T2T CHM13v2.0/hs1 assembly)<sup>14</sup> in order to confirm that the targeted sequences mapped to a SAAC.

## **Automatic detection of ring 22 using Gollum**

Gollum is a python tool that we developed to detect rings on acrocentric chromosomes (Supplementary Methods and Supplementary Figure 1). Input data used by Gollum are short read sequencing previously aligned to T2T-CHM13 reference. Gollum first extracts discordant read pairs (R1 & R2) with R1 on the long arm of the targeted acrocentric chromosome and then filters R2 based on mapping quality and on location on any SAAC. A density clustering is used to remove isolated discordant read pairs and to keep read pairs clustering in the same region (surrounding the putative ring breakpoint region). The remaining R2 are extracted using samtools<sup>17</sup> and the R2 sequences are aligned on T2T-CHM13 using BLAT. BLAT results are filtered to keep only alignments with e-value < 0.01 and a bit score > 150. The read pairs are further filtered so that all the R2 are mapping on the short arm of the targeted acrocentric chromosome. A last density clustering analysis allowed the identification of the ring breakpoints.

## **RESULTS**

Our cohort included 55 French individuals with PMS carrying a 22q13.3 CNV previously investigated by CMA. Fifty-one had a 22q13 deletion and four carried a 22q13 duplication. Only 6 individuals (P1-P6) carrying 22qter deletions out of 51 were previously detected as r(22) by FISH and/or karyotype at their inclusion in our study. All 55 patients were further analyzed using WGS.

### **Refined breakpoints of r(22)**

WGS data alignment to GRCh37 confirmed that P2 to P6 showed a r(22) derived from simple 22q terminal (22qter) deletion (length between 1.2 and 7.2 Mb without recurrent breakpoint, supplementary Table 1 and Figure 1c). For P1, previous CMA study suggested a complex ring derived from an inversion deletion duplication mechanism (Inv dup del) associating a 7.2 Mb 22qter deletion with two adjacent gains of copies (788 kb et 1.5 Mb). Both terminal deletion and the large duplication were confirmed using FISH studies. Further analyses on the breakpoints of the duplicated segments using WGS data, revealed a higher complexity: four interstitial gains of copies for which the number of copies varied from 3 to 6 and their size ranged from 1.7 kb to 1.6 Mb (supplementary Figure 2). Moreover, two duplicated segments (regions E and G) appear to be inverted based on CRAM visualization of the split and discordant reads.

### **Manual detection of r(22) using BLAT**

With the aim of identifying r(22) within short paired-end reads sequencing data, we first manually analyzed the region of the 22qter breakpoints of the 6 patients with known r(22) (see material and methods). For each individual, we could observe from 5 to 15 split read pairs mapping close to the 22qter breakpoint regions. We then used an alignment against the human

T2T genome reference (Jan. 2022 T2T CHM13v2.0/hs1 assembly) and observed that all split read pair sequences from 5 patients out of 6 (P2-P6) were mapped on a SAAC. Mapping results were not restricted to 22 short arms since all acrocentric chromosomes share large pseudo homologous regions (PHRs)<sup>18</sup>. For P1, who had a more complex SV, we extended the analysis to every 22q breakpoints (including loss and gain) and 7 split read pairs mapping on a SAAC were found near the proximal breakpoint of the proximal duplication, consistent with the formation of a ring to stabilize the complex SV (inv-dup-del).

We then analyzed all the individuals from our cohort in the same way and a seventh patient (P7) showed five split paired-end reads mapping at the breakpoint region on chromosome 22q13 and on a SAAC. The r(22) of the seventh patient was then confirmed by FISH. Finally, two other suspicions of r(22) were excluded in two patients who presented split paired-end reads within the breakpoint region. Patient P8 carried a terminal duplication with five split paired-end reads supporting a derivative chromosome 22 with insertion of the duplicated 22qter segment in the short arm 22p. Patient P9 showed only one split paired-end read and had a normal karyotype, suggesting that it was most likely a false positive detection of a ring or a very low level of mosaicism with few cells carrying a ring chromosome. Other tissues were not available to study.

### **Automatic detection of r(22) using Gollum**

In order to efficiently detect r(22) and more generally all rings from acrocentric chromosomes, we developed a python tool called Gollum (see method and supplementary). We first aligned 235 WGS of our cohort (7 r(22), 44 deletions, 4 duplications, 27 SNVs/indels in *SHANK3*, and 153 relatives) on T2T-CHM13, and then ran Gollum on all of them. Gollum rapidly (less than 2 minutes per sample on average) allowed the identification of discordant read pairs (R1 and R2) with R1 mapping on the 22q and R2 on the 22p. From this systematic analysis of discordant read pairs, Gollum found supporting reads for all seven individuals with known r(22) and only three spurious hits in a different region of 22q (sensitivity 1.00, specificity 0.96, and F<sub>1</sub> score 0.82) (see figure 1).

## **DISCUSSION**

### **r(22) in PMS : new highlight from WGS**

Currently standard karyotyping or FISH technologies is necessary to detect rings that involve human acrocentric chromosomes. Such rings are not possibly or easily identified with current human genomic diagnosis technologies (including SNP array, short and long read high-throughput sequencing, optical genomic mapping) and pipelines based on GRCh37/38 reference mapping. As the diagnosis of r(22) in PMS patients is of medical importance, we developed Gollum that required short read sequencing alignment to human T2T. Gollum's strategy is based on the detection of discordant paired-end reads clustering within 22qter breakpoint and on the extraction of corresponding R2 mapping on the 22p. After a systematic examination of all individuals from this study, our tool confirmed the presence of r(22) in the

six patients known to carry a r(22) and allowed us to make an additional diagnosis in one patient initially known to have a terminal 22q13.3 deletion (P7). Moreover, in one patient of our cohort, Gollum allowed us to identify the mechanism of a 22qter duplication inserted on the short arm of the derivative chromosome 22 whose structure could only have been determined using FISH. With WGS more readily available, Gollum offers the opportunity to identify and explore rings and complex structural variants involving SAACs. In addition, WGS data allowed us to refine the complex structure of the ring of P1 that probably resulted from a two-steps mechanism, an inv-dup-del stabilized by the circularization of chromosome 22 followed by a mosaic dynamism favored by the Alu elements found at all breakpoints.

Consequences of carrying a ring chromosome can be highly variable, depending on the chromosome involved and the size and the content of the associated unbalanced segments. Phenotypes usually correspond to the deleted segment, but the possibility that a “ring syndrome” remains controversial<sup>6,7</sup>. The molecular consequences associated with some ring chromosomes can however have strong implications. Ring instability during mitosis may induce additional structural variations, from single microdeletion or microduplication to complex chromosomal structure such as chromothripsis<sup>19</sup>, or even lead to the loss of the ring leading to monosomic cells<sup>4,20</sup>. The mechanism of ring formation remains challenging to understand, particularly in case of complex molecular structure of the ring partly due to its mitotic instability. Indeed, sister chromatid exchanges arising during the S phase of the cell cycle can lead to dicentric, interlocked or complex structural rings that perturb the cell division. As a result, complex rings could be stable in daughter cells or finally lost and generate aneuploid cells<sup>4,21,22</sup>. To define further sequence complexity and mechanism formation of this unique classes of SV, long-read sequencing with appropriate pipeline based on T2T reference seems to be a promising solution<sup>23</sup>.

### ***Neurofibromatosis type 2***

PMS patients with r(22) are at risk of developing neurofibromatosis type 2 (NF2), a pathology characterized by benign tumors in the nervous system, mainly vestibular schwannomas along the auditory nerve which may cause hearing loss, or meningiomas. These tumors can develop following loss of the *NF2* tumor suppressor gene expression (MIM #607379), located in the proximal region 22q12.2. Incidence of NF2 is 1/25,000-1/40,000 individuals in the general population<sup>24,25</sup> and up to 1.6% of patients with NF2 carries a r(22)<sup>1,26</sup>. The hypothesis for the development of NF2 symptoms in r(22) patients is based on a two hits model: monosomy 22 in some cells due to instability of the ring, followed by a somatic loss-of-function mutation of the remaining *NF2* allele in Schwann cells, leading to schwannoma or meningiomas<sup>8,27,28</sup>.

The first symptoms of NF2 in patients with r(22) usually appear at 20 years<sup>25</sup>, and consist in uni or bilateral hearing loss, focal motor deficits, tinnitus, seizures or sensory deficits or visual disturbances as a consequence of nervous tumors. Moreover, mortality is higher in patients with NF2. Mean survival after diagnosis was 15 years in a 1992 study<sup>29</sup>, mean age between 39 and 62. Death results from complications of tumors (intracranial hypertension, high grade



meningioma, low-grade central nervous system malignancies, post-surgery complications)<sup>25,29–31</sup> and is positively correlated with younger age at diagnosis, the presence of meningiomas and loss-of-function mutation<sup>30</sup>. Care in specialized structures improves survival in these patients. In our study, ages of patients at PMS diagnosis varies from < 1 to 23 years-old. Among the seven patients with r(22), P4 and P2 present bilateral hearing loss diagnosed around 21 years and 6 years, respectively. Cerebral MRI was available for five patients: two were normal (P3, P4), one had microcephaly and two had a short corpus callosum. No schwannoma was identified. No patient was a carrier of a constitutional loss-of-function nor missense variation of *NF2*. *NF2* diagnosis is of clinical importance in PMS patients carrying r(22) and medical follow-up should be performed in adulthood including MRI of the spine, head, and internal auditory meatus<sup>1</sup>.

### ***Conclusion***

A ring chromosome 22 can be observed in individuals with PMS and such ring can increase the occurrence of cancer. We developed “Gollum” a new efficient tool based on WGS data and T2T annotation to identify rings involving acrocentric chromosomes, and more generally to identify all structural variants involving SAACs. Based on the WGS of seven individuals with PMS and r(22), we showed that such rearrangements remain largely undetected using microarrays and can be more complex than expected. Robust detection of ring chromosome has a direct clinical impact including increase surveillance (in childhood but also in the transition to adulthood) of these individuals who are at higher risk of tumor formation.

## REFERENCES

1. Koza, S. A. *et al.* Consensus recommendations on counselling in Phelan-McDermid syndrome, with special attention to recurrence risk and to ring chromosome 22. *Eur. J. Med. Genet.* **66**, 104773 (2023).
2. Phelan, K. & McDermid, H. E. The 22q13.3 Deletion Syndrome (Phelan-McDermid Syndrome). *Mol. Syndromol.* (2011) doi:10.1159/000334260.
3. Tabet, A.-C. *et al.* A framework to identify contributing genes in patients with Phelan-McDermid syndrome. *Npj Genomic Med.* **2**, 32 (2017).
4. Pristyazhnyuk, I. E. & Menzorov, A. G. Ring chromosomes: from formation to clinical potential. *Protoplasma* **255**, 439–449 (2018).
5. Kolevzon, A. *et al.* Phelan-McDermid syndrome: a review of the literature and practice parameters for medical assessment and monitoring. *J. Neurodev. Disord.* **6**, 39 (2014).
6. Li, P. *et al.* The past, present, and future for constitutional ring chromosomes: A report of the international consortium for human ring chromosomes. *HGG Adv.* **3**, 100139 (2022).
7. Lyons-Warren, A. M., Cheung, S. W. & Holder, J. L. Clinical Reasoning: A common cause for Phelan-McDermid syndrome and neurofibromatosis type 2: One ring to bind them. *Neurology* **89**, e205–e209 (2017).
8. Ziats, C. A. *et al.* Neurofibromatosis type 2 in Phelan-McDermid syndrome: Institutional experience and review of the literature. *Eur. J. Med. Genet.* **63**, 104042 (2020).
9. Byers, H. M. *et al.* Description of a new oncogenic mechanism for atypical teratoid rhabdoid tumors in patients with ring chromosome 22. *Am. J. Med. Genet. A.* **173**, 245–249 (2017).
10. Antonarakis, S. E. Short arms of human acrocentric chromosomes and the completion of the human genome sequence. *Genome Res.* **32**, 599–607 (2022).
11. Stranneheim, H. *et al.* Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. *Genome Med.* **13**, 40 (2021).
12. Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96–102 (2020).
13. Mantere, T. *et al.* Next generation cytogenetics: genome-imaging enables comprehensive structural variant detection for 100 constitutional chromosomal aberrations in 85 samples. <http://biorxiv.org/lookup/doi/10.1101/2020.07.15.205245> (2020) doi:10.1101/2020.07.15.205245.
14. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
15. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
16. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
17. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
18. Guarracino, A. *et al.* Recombination between heterologous human acrocentric

chromosomes. *Nature* **617**, 335–343 (2023).

19. Kurtas, N. *et al.* Chromothripsis and ring chromosome 22: a paradigm of genomic complexity in the Phelan-McDermid syndrome (22q13 deletion syndrome). *J. Med. Genet.* **55**, 269–277 (2018).

20. Carvalho, C. M. B. & Lupski, J. R. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* **17**, 224–238 (2016).

21. Nikitina, T. V. *et al.* Complex biology of constitutional ring chromosomes structure and (in)stability revealed by somatic cell reprogramming. *Sci. Rep.* **11**, 4325 (2021).

22. Guilherme, R. *et al.* Human Ring Chromosomes - New Insights for their Clinical Significance. *Balk. J. Med. Genet. BJMG* **16**, 13–20 (2013).

23. Mostovoy, Y. *et al.* Resolution of ring chromosomes, Robertsonian translocations, and complex structural variants from long-read sequencing and telomere-to-telomere assembly. 2023.09.07.555775 Preprint at <https://doi.org/10.1101/2023.09.07.555775> (2023).

24. Ardern-Holmes, S., Fisher, G. & North, K. Neurofibromatosis Type 2: Presentation, Major Complications, and Management, With a Focus on the Pediatric Age Group. *J. Child Neurol.* **32**, 9–22 (2017).

25. Evans, Dg. R. Neurofibromatosis type 2 (NF2): A clinical and molecular review. *Orphanet J. Rare Dis.* **4**, 16 (2009).

26. Evans, D. G. *et al.* Incidence of mosaicism in 1055 de novo NF2 cases: much higher than previous estimates with high utility of next-generation sequencing. *Genet. Med.* **22**, 53–59 (2020).

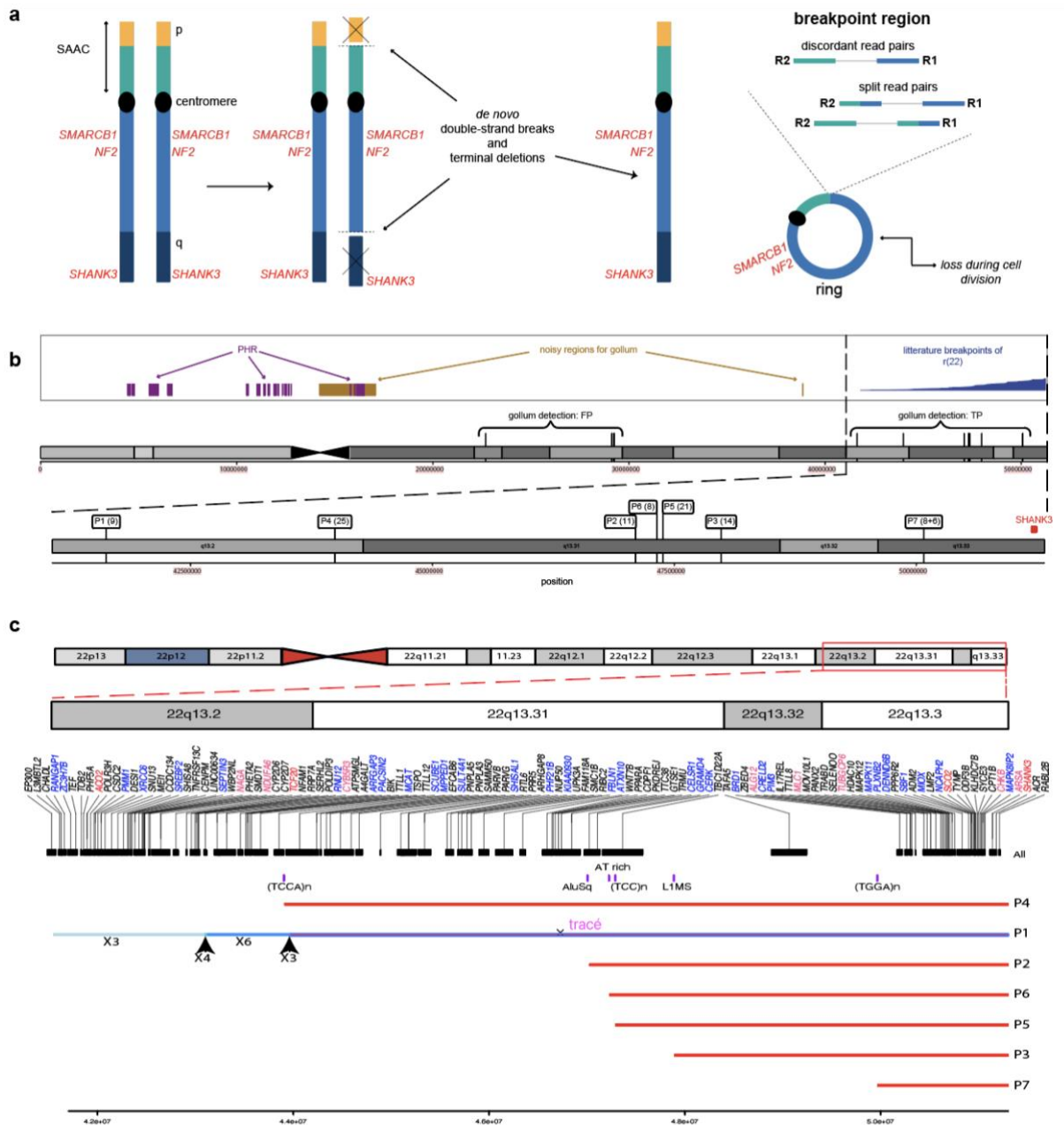
27. Tsilchorozidou, T. *et al.* Constitutional rearrangements of chromosome 22 as a cause of neurofibromatosis 2. *J. Med. Genet.* **41**, 529–534 (2004).

28. Zirn, B. *et al.* Ring chromosome 22 and neurofibromatosis type II: proof of two-hit model for the loss of the NF2 gene in the development of meningioma. *Clin. Genet.* **81**, 82–87 (2012).

29. Evans, D. G. *et al.* A genetic study of type 2 neurofibromatosis in the United Kingdom. I. Prevalence, mutation rate, fitness, and confirmation of maternal transmission effect on severity. *J. Med. Genet.* **29**, 841–846 (1992).

30. Baser, M. E. *et al.* Predictors of the Risk of Mortality in Neurofibromatosis 2. *Am. J. Hum. Genet.* **71**, 715–723 (2002).

31. Aboukais, R. *et al.* Causes of mortality in neurofibromatosis type 2. *Br. J. Neurosurg.* **29**, 37–40 (2015).



**Figure 1: Ring formation and discordant read pairs supporting ring breakpoint detected by Gollum along chromosome 22**

**(a) Ring formation.** After *de novo* double-strand breaks and terminal deletions, a r(22) is formed to stabilize the resulting 22 chromosome. In the region of the fusion of the broken ends, discordant and/or split read pairs are present and are the signature of putative ring breakpoint. **(b) Cluster of discordant read pairs detected by Gollum.** The figure shows the chromosome 22 with the pseudo-homologous regions (PHR) of the SAACs (purple), regions having a lot of false positive across multiple individuals from Gollum detection (dark orange), r(22) breakpoints from the literature (blue) and Gollum detection on our cohort (black solid line). r(22) breakpoints identified in our study are supported by discordant read pairs with R1 mapping on the long arm of the chromosome 22 and R2 on the 22p (Gollum detection: TP). Incorrect candidate breakpoints (Gollum detection: FP) were also detected by Gollum and are consequences of structural variants in low complexity region of the genome. At the bottom, a zoom on the r(22) breakpoints region (TP) shows the number of discordant read pairs detected by Gollum supporting the presence of r(22) in the tags next to the individuals labels. For P7, two very close breakpoints (2kb in between) were identified supporting by 8 and 6 discordant read pairs. **(c) 22qter regions of individuals with r(22) in this study.** The region containing the ring breakpoints along the chromosome 22 is indicated with a red

box. Genes and repeat elements are indicated in a zoom of the region. The genes are colored based on neurodevelopmental disorders (NDD) association: red for dominant high-confidence NDD, pink for recessive high-confidence NDD genes and blue for candidate NDD. SVs involved in r(22) formation are indicated for each patients (22qter deletion with red solid line). For the complex r(22) of patient P1 derived from an inv dup del, the duplications are indicated in blue with the number of copies. Inverted duplicated segments are pointed out by Arrows. SAAC, short arm of acrocentric chromosome; TP, true positive; FP, false positive; P, patients; x3 to x6, 3 to 6 copies.

## **Author contribution Statement**

T.B., C.S.L., M.R. and A.C.T. contributed to the conception and the design of the study. F.A., Y.C., R.D., M.G., D.H., Anna.M., I.M., S.P. and A.C.T were involved in the recruitment and the evaluation of the patients and the sample collection. C.D., J.L., D.S. and C.Y. were in charge of the initial cytogenetic testing. A.B. and J.F.D. were in charge of the library preparation and sequencing. F.C. and A.M. were in charge of the data processing of the sequencing data and developed a new python tool for the automatic detection of acrocentric ring. M.R. performed the analysis of processed data. T.B., F.C., C.S.L., M.R. and A.C.T. participated to the interpretation of the data. F.C., A.M., C.S.L., M.R., A.C.T. and A.V. contributed to write original draft and figures. T.B. and T.R. reviewed the manuscript. C.S.L and A.C.T. had jointly supervised the work.

## **Acknowledgements**

The authors are grateful to the “Banque ADN & cellules - Institut du Cerveau - ICM - Hôpital Pitié Salpêtrière” and to the participants of the study.

## **Funding**

This work was funded by Institut Pasteur, the Bettencourt-Schueller Foundation, Université de Paris Cité, the Conny-Maeva Charitable Foundation, the Cognacq Jay Foundation, the Eranet-Neuron (ALTRUISM), the GenMed Labex, AIMS-2-TRIALS which received support from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777394 and the INCEPTION program (Investissement d’Avenir grant ANR-16-CONV-0005). This project has received funding from the Horizon Europe programs CANDY and R2D2-MH under grant agreement nos. 847818 and 101057385. The views expressed here are the responsibility of the author(s) only. The EU Commission takes no responsibility for any use made of the information set out.

## **Code availability**

The underlying code for this study is available in the gollum github repository and can be accessed via this link <https://github.com/bourgeron-lab/gollum>.

## SUPPLEMENTARY APPENDIX

### GOLLUM, A NEW TOOL USING SHORT READ SEQUENCING TO DETECT RING CHROMOSOMES IN PATIENTS WITH PHELAN-MCDERMID SYNDROME

#### Authors:

Myriam Rachid<sup>1,2</sup>, Freddy Cliquet<sup>2</sup>, Alexandre Mathieu<sup>2</sup>, Aline Vitrac<sup>2</sup>, Anne Boland<sup>3</sup>, Jean-Francois Deleuze<sup>3</sup>, Jonathan Lévy<sup>1</sup>, Damien Haye<sup>1</sup>, Sandrine Passemard<sup>1</sup>, Anna Maruani<sup>2,4</sup>, Frederique Amsellem<sup>2,4</sup>, Céline Dupont<sup>1</sup>, Catherine Yardin<sup>5</sup>, Damien Sanlaville<sup>6</sup>, Yline Capri<sup>1</sup>, Magalie Gorce<sup>7</sup>, Isabelle Marey<sup>8</sup>, Thomas Rolland<sup>2</sup>, Thomas Bourgeron<sup>2</sup>, Richard Delorme<sup>2,4</sup>, Claire S Leblond<sup>2,\*</sup>, Anne-Claude Tabet<sup>1,2,\*</sup>

#### Affiliations:

<sup>1</sup> Genetics Department, Robert Debré Hospital, APHP, Paris, France

<sup>2</sup> Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS, IUF, Université Paris Cité, Paris, France

<sup>3</sup> Université Paris-Saclay, Commissariat à l'Energie Atomique et aux Energies Alternatives, Centre National de Recherche en Génomique

Humaine, 91057 Evry, France

<sup>4</sup> Department of Child and Adolescent Psychiatry, Robert Debré Hospital, APHP, Paris, France

<sup>5</sup> CNRS XLIM UMR 7252, Université de Limoges, Service de Cytogénétique, CHU Limoges

<sup>6</sup> Genetic Department, Hospices Civils de Lyon

<sup>7</sup> Clinical Genetic Unit, CH Le Mans

<sup>8</sup> Université Grenoble Alpes, INSERM, U1209, CHU Grenoble Alpes, Institute for Advanced Bioscience, Grenoble, France

\* equal contribution

Corresponding author: Claire S leblond, [claire.leblond@pasteur.fr](mailto:claire.leblond@pasteur.fr)

<b>Supplementary Table 1.....</b>	<b>p2</b>
<b>Supplementary Figure 1.....</b>	<b>p3</b>
<b>Supplementary Figure 2.....</b>	<b>p4</b>
<b>Supplementary information.....</b>	<b>p5-p7</b>
<b>References.....</b>	<b>p8</b>

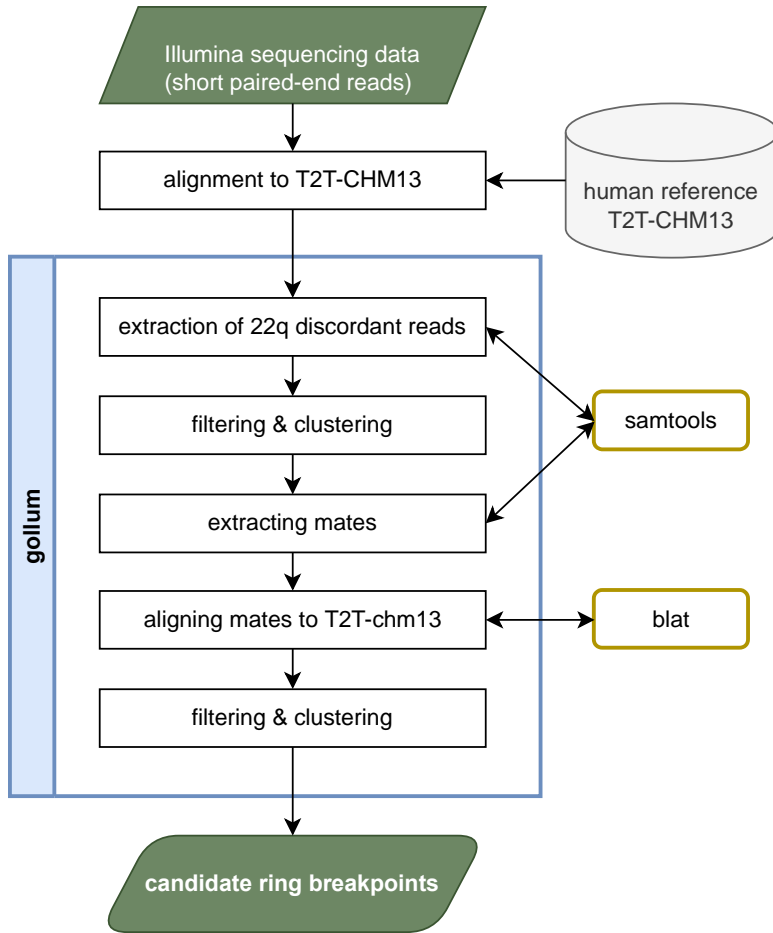
## Supplementary Table 1

**Supplementary Table 1**

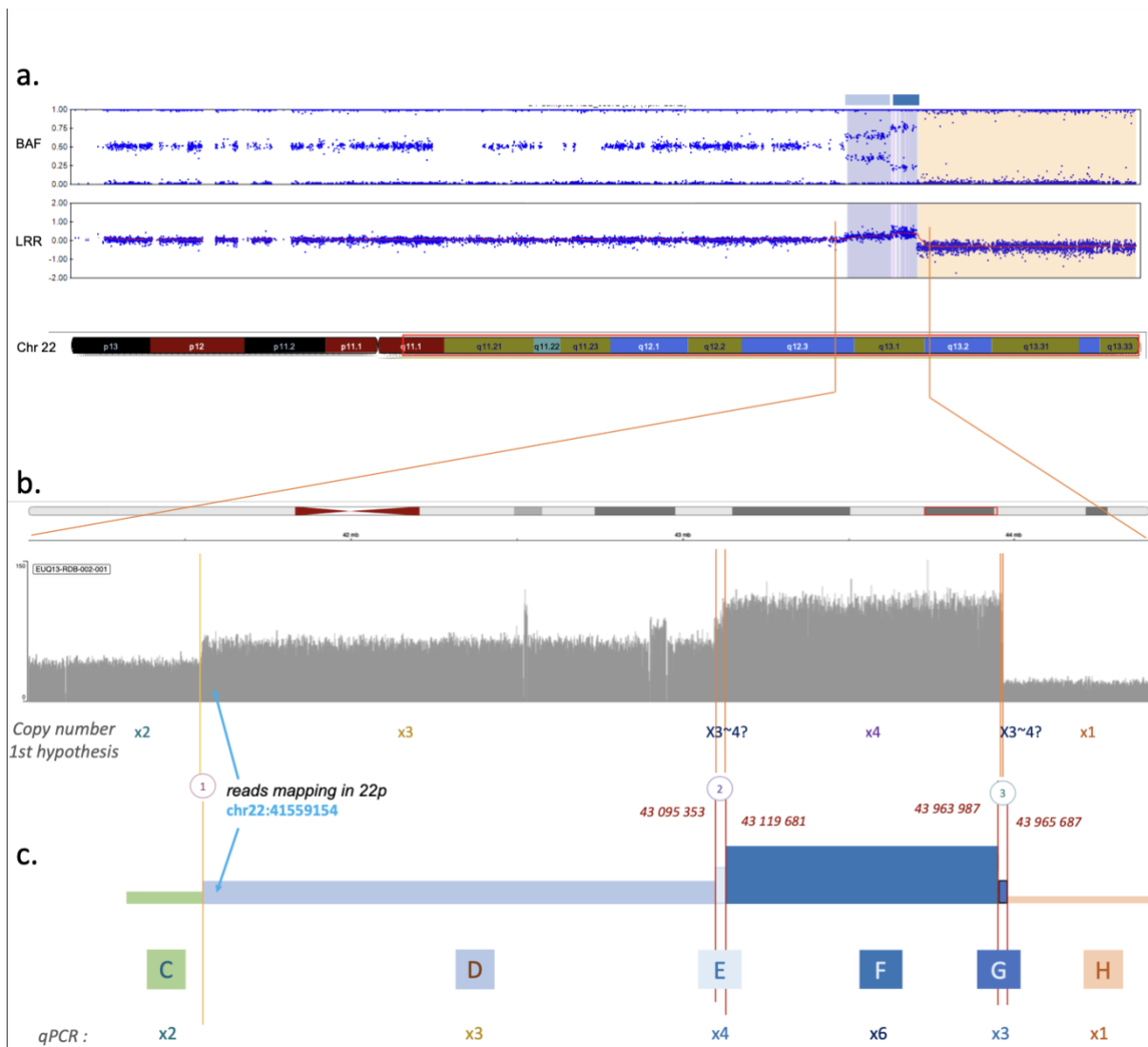
Patient	22qter deletion size	22qter deletion BP	split read pairs	22q additional SVs
<b>P1</b>	7.2Mb	chr22:43,965,687	7	22q13.2(41,551,360_43,095,353)x3 22q13.2(43,095,353_43,119,681)x4 22q13.2(43,119,681_43,963,987)x6 22q13.2(43,963,987_43,965,687)x3
<b>P2</b>	4.2 Mb	chr22:47,006,904	6	
<b>P3</b>	3.3 Mb	chr22:47,888,459	7	
<b>P4</b>	7.2 Mb	chr22:43,905,471	15	
<b>P5</b>	3.9 Mb	chr22:47,289,652	5	
<b>P6</b>	4 Mb	chr22:47,227,651	11	
<b>P7</b>	1.2 Mb	chr22:49,966,834	5	

Coordinates are based on human GRCh37 reference. BP, breakpoint; WGS, whole-genome sequencing; CNV, Copy Number Variant.





**Supplementary Figure 1: Workflow of Gollum**



**Supplementary Figure 2: Further characterization of complex r(22) of P1 using WGS**  
**(A) CMA profile from Genome Studio.** Signals of the B allele frequency (upper plot) and the Log R ratio (lower plot) have allowed the identification of two adjacent gain of copies (in blue, D and F segments) and one 22qter deletion (orange, H segment). **(B) Coverage of the region of the complex r(22).** Depth of coverage of the region has been visualized using CRAM file and IGV. Two additional inverted gain of copies (E and G segment) flanking the small duplication (F segment) were detected. **(C) Schematic representation of the 22q13.1q13.3 region of P1.** Copy number analysis was performed using quantitative PCR to precisely assess the number of gain of copies. E and G segments are inverted. x1 to x6, 1 to 6 copies. The position of the split read pairs of the r(22) breakpoint are indicated with a blue arrow.

## **Clinical presentation of the patient carrying r(22)**

### **Patient 1**

P1 is a 2 years old boy at the time of diagnosis. He is the fifth child of healthy parents. The pregnancy was uneventful except for the detection of bilateral hydronephrosis by ultrasonography. A neonatal renal ultrasound discovered bilateral caliper dilation and cysts of the right kidney. He also had a malformative uropathy with vesicoureteral reflux which needed multiple surgeries. Ophthalmologically he had severe congenital glaucoma, myopia, and astigmatism. He presented with a neurodevelopmental delay (in particular motor and speech delay), microcephaly, and facial dysmorphism. At the age of 2 years old, he presented a neurodevelopmental stagnation. At the neurological evaluation, he had hypotonia, pyramidal, and axonal lesions, generalized amyotrophy, and abolition of reflexes. Cerebral MRI showed cortical and subcortical atrophy, bilateral ventriculomegaly, and corpus callosum hypoplasia.

### **Patient 2**

P2 is a 12 years old boy with global developmental delay and absent speech. The twin pregnancy was uneventful, resulting from IVF because of paternal azoospermia. P2 has a healthy twin, and he was hospitalized for acute respiratory distress at birth, resulting from a right diaphragmatic hernia. He had hypermetropia and strabismus. The audiological evaluation shows bilateral hearing loss (mild to profound). He had gastroesophageal reflux and an operated testicular ectopia. He had neonatal hypotonia, but has a normal neurologic examination at 12. He presented with hetero- and self-aggression, intolerance to frustration, and nocturnal awakenings. Cerebral MRI was normal at age 1 month and EEG at same age showed discreetly impoverished sleep and wake patterns without significant abnormalities and without signs of focus.

### **Patient 3**

P3 is a 6 years old boy. He is the first child of unrelated parents and is born after an uneventful pregnancy. He presented neonatal hypotonia, swallowing difficulties, and a discrete supraaortic stenosis. At the age of 6, he had no speech and had difficulties falling asleep (treated by melatonin). He also presented behavior troubles with major anxiety and intolerance to frustration and wears glasses for hypermetropia and astigmatism. He was diagnosed with autism.

### **Patient 4**

P4 is a 19 years old girl, the first child of unrelated parents. She had a global developmental delay, sat at 1 year old, and walked at 2.5 years old. She presented with myopia and astigmatism, a diaphragmatic hernia, and scoliosis. She had an ataxic gait, and one hyperthermic convulsion. Cerebral MRI performed at 9 months and EEG was normal. She presented intellectual disability, and some stereotypies but did not meet the criteria for autism. She did not acquire verbal language. An auditive evaluation was normal in her childhood. She has a few hypopigmented spots in the arm and the abdomen.

### **Patient 5**

P5 is a boy (age at evaluation) with intellectual disability and delayed speech. He has normal neuropsychomotor development and didn't have seizures. EEG and MRI were not available.

### **Patient 6**

P6 is a 3 years-old boy at evaluation. He presented IUGR and was born by Cesarean section for fetal heart rhythm abnormalities. Then he presented feeding difficulties, a global developmental delay (walk at 21 months of age) and severe delayed speech. He also had dysmorphic facial features with plagiocephaly. Neurologic examination shows cerebellar syndrome, with mega- cisterna magna of the posterior cerebral fossa visualized at MRI. EEG was performed and was normal. P6 also presented esophagitis, white line hernia, atopic dermatitis, and astigmatism.

### **Patient 7**

P7 is a 8 -month-old boy at the time of diagnosis. Pregnancy was uneventful. He was referred for global developmental delay. Evaluations did not meet criteria for intellectual disability nor autism. He had microcephaly (-2SD), dysmorphic facial features with receding forehead, receding chin, slightly upturned nostrils, wide eyebrows. He had eating troubles in the first months of life because of a cow's milk protein allergy. He never had a seizure nor EEG performed. Cerebral MRI showed a thin corpus callosum. In addition, he presented a bilateral glaucoma, asymmetric deafness, a syndactyly 2-3 on the right foot, an angioma of the right hand, mild hydrocele and recurrent ear, nose and throat infection.

### **Automatic detection of r(22)**

A python tool called *Gollum* was designed to detect rings on acrocentric chromosomes. The detection is based on the analysis of short paired-end reads from Illumina sequencing data. This tool requires having both samtools<sup>1</sup> and BLAT<sup>2</sup> available on the command line. We first align the FASTQ files on the T2T-CHM13 reference<sup>3</sup>. The alignment on T2T-CHM13 enable the process all the unmapped reads using other human genome reference. It also includes the read previously mapped to unknown chromosomes and contaminants which would increase the noise in the approach and the detection time. For this, we used bwa-mem2 (v2.2.1)<sup>4</sup> for the alignment followed by sambalster (v0.1.26)<sup>5</sup> to mark duplicates, then sambamba (v0.8.0)<sup>6</sup> to convert to bam files and then sort by position before finally compressing to cram format using samtools (v1.16).

Gollum work in several steps:

1. Extraction of the discordant read pairs (R1 & R2) with R1 mapping on the long arm of the chromosome 22 and R2 on the short arm of any acrocentric chromosome (SAAC)
2. Filtering discordant read pairs to keep only R1 with a mapping quality at 60 (MQ >= 60) and with the corresponding R2 located on the (SAAC)
3. A density clustering (DBSCAN) is performed to find clusters of R1 on 22qter region supporting putative ring breakpoints.

4. Extraction of all the R2 belonging to the clusters from step 3 using samtools.
5. R2 sequences are realigned on the T2T-CHM13 using blat.
6. Filtering the BLAT results on its e-value (user parameter, by default  $< 0.01$ ) and bit score (default value  $> 150$ ).
7. Extracting the R1 having a R2 (from step 6) located within 22p.
8. Performing another density clustering (DBSCAN) on the R1 from step 7.
9. Each cluster found denotes a potential breakpoint for a ring on the studied chromosome. If no cluster are found, then no conclusive evidence of a ring was detected on the WGS.

Genomic regions having candidate ring breakpoints detection across multiple individuals were excluded (noise).  $r(22)$  breakpoints reported in the literature, point out a region where candidate ring breakpoints are more susceptible to be observed. This information is not used by Gollum, but is there to help to interpret results.

Additionally, we chose not to provide a breakpoint on the short arm because reads tend to align to multiple positions.

It is to be noted that we designed and tested Gollum on chromosome 22, but it is easily adaptable to work on the other acrocentric chromosomes.

## References:

1. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
2. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656–664 (2002).
3. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
4. Vasimuddin, Md., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (2019). doi:10.1109/IPDPS.2019.00041.
5. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
6. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

## 4. PMS associé aux remaniements complexes du chromosome 22

Parmi la cohorte de patients atteints d'un PMS, nous nous sommes également intéressés d'une part à deux patients porteurs d'une inv dup del 22 : le premier est une inv dup del classique, le deuxième est une inv dup del plus complexe, stabilisée par la formation d'un anneau et intégré dans l'étude précédente sous la référence P1. D'autre part, après avoir identifié des points de cassure s'alignant sur un BCCA pour les 6 patients porteurs d'un r(22) initialement étudiés de notre cohorte (voir article [chapitre 2/3](#)), nous avons entrepris de reprendre l'analyse sur BAM de 45 patients dans la cohorte avec une délétion ou une duplication terminale 22qter afin d'identifier des anneaux non connus ou non précisés au moment de l'inclusion. Nous avons ainsi pu identifier un anneau additionnel qui a été confirmé par l'examen du caryotype (patient P7 de l'article) mais aussi deux autres remaniements impliquant des BCCA : un dérivé 13 d'une translocation t(13;22)(p13;q13.3) connue, et un dérivé d'inversion entraînant une duplication 22qter insérée dans le bras court du chromosome 22.

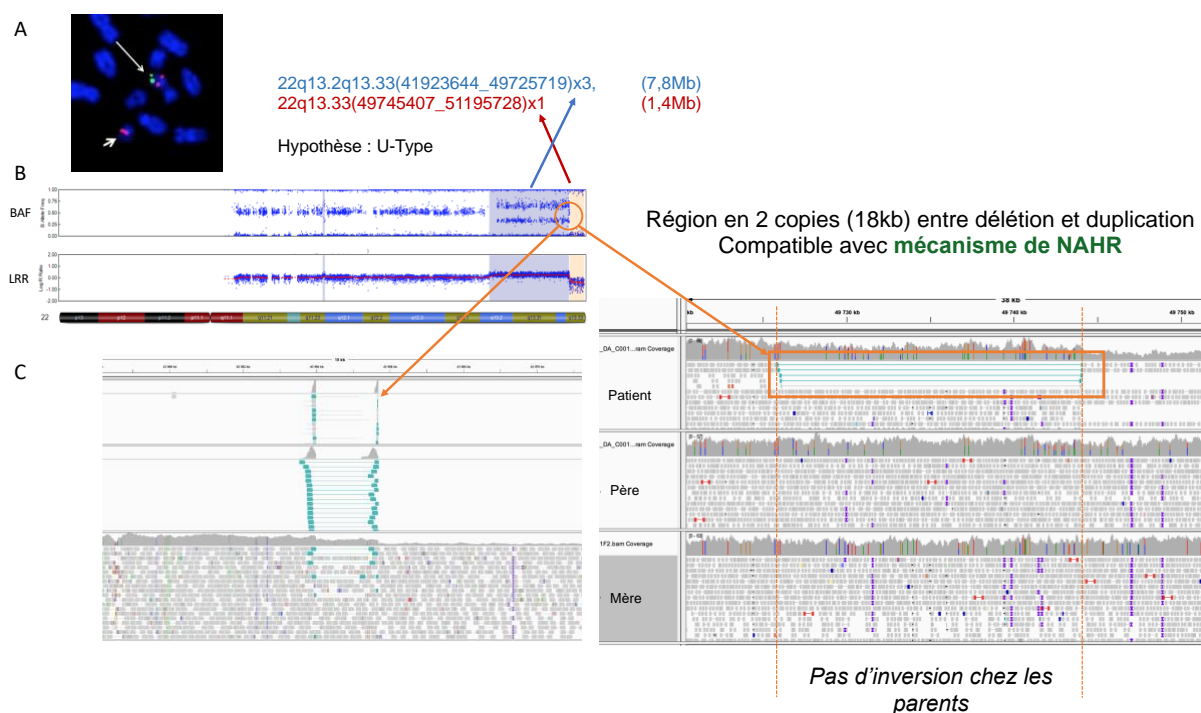
### 4.1. Une inv dup del 22 classique

Le premier patient avec une inv dup del classique est un garçon de 10 ans au moment du diagnostic. Il est le troisième enfant issu de parents sains. La grossesse s'est déroulée sans particularité. A la naissance, l'examen montre une hypotonie et des troubles de la succion. A l'âge de 1 an, il présentait des stéréotypies. A 10 ans, on retrouve un enfant avec un retard du développement sévère, une absence totale de langage, une DI sévère, et un diagnostic de TSA. Les données cliniques révélaient par ailleurs des troubles du sommeil et des troubles de l'alimentation. Les comorbidités rapportées incluaient un reflux gastro-œsophagien, un retard de croissance à -5DS, un retard de puberté. L'examen neurologique retrouvait une épilepsie et l'EEG montrait la persistance de points ondes continues.

Les explorations génétiques ont consisté en la réalisation d'une puce de type SNP qui retrouve une délétion terminale de 1,4 Mb associée à une duplication adjacente de 7,8 Mb. La formule chromosomique selon l'ISCN est : [GRCh37] arr 22q13.2q13.33(41923644\_49725719)x3, 22q13.33(49745407\_51195728)x1, sans région en copie unique entre les deux suggérant ainsi un mécanisme de U-Type (Figure 25.B). La FISH confirme la délétion de *SHANK3* (figure 25.A). L'analyse parentale était normale et fait conclure à une inv dup del (22) *de novo*. L'analyse des

SNP du patient et ses parents nous ont permis de comparer les génotypes et montrer que les deux anomalies chromosomiques étaient en phase et survenues sur l'allèle maternel.

Dans le cadre de notre étude, le patient et ses parents ont pu être analysés en WGS. Comme précédemment décrit, nous avons dans un premier temps analysé la région 22q distale. L'analyse précise des points de cassure a montré la présence d'une petite région de 18kb entre la délétion et la duplication, en 2 copies et inversée (figure 25.C). Cette nouvelle donnée nous conduit à revoir le mécanisme évoqué sur les données d'ACPA. En effet la présence d'une région en séquence unique séparant la délétion terminale d'une duplication adjacente est plutôt en faveur d'un mécanisme de recombinaison homologue non allélique (figure 14). L'analyse de ces régions chez les parents nous a permis d'éliminer une inversion paracentrique parentale ayant pu favoriser la survenue de cette anomalie complexe, définitivement survenue *de novo*.



**Figure 25.** Représentation des analyses génétiques réalisées

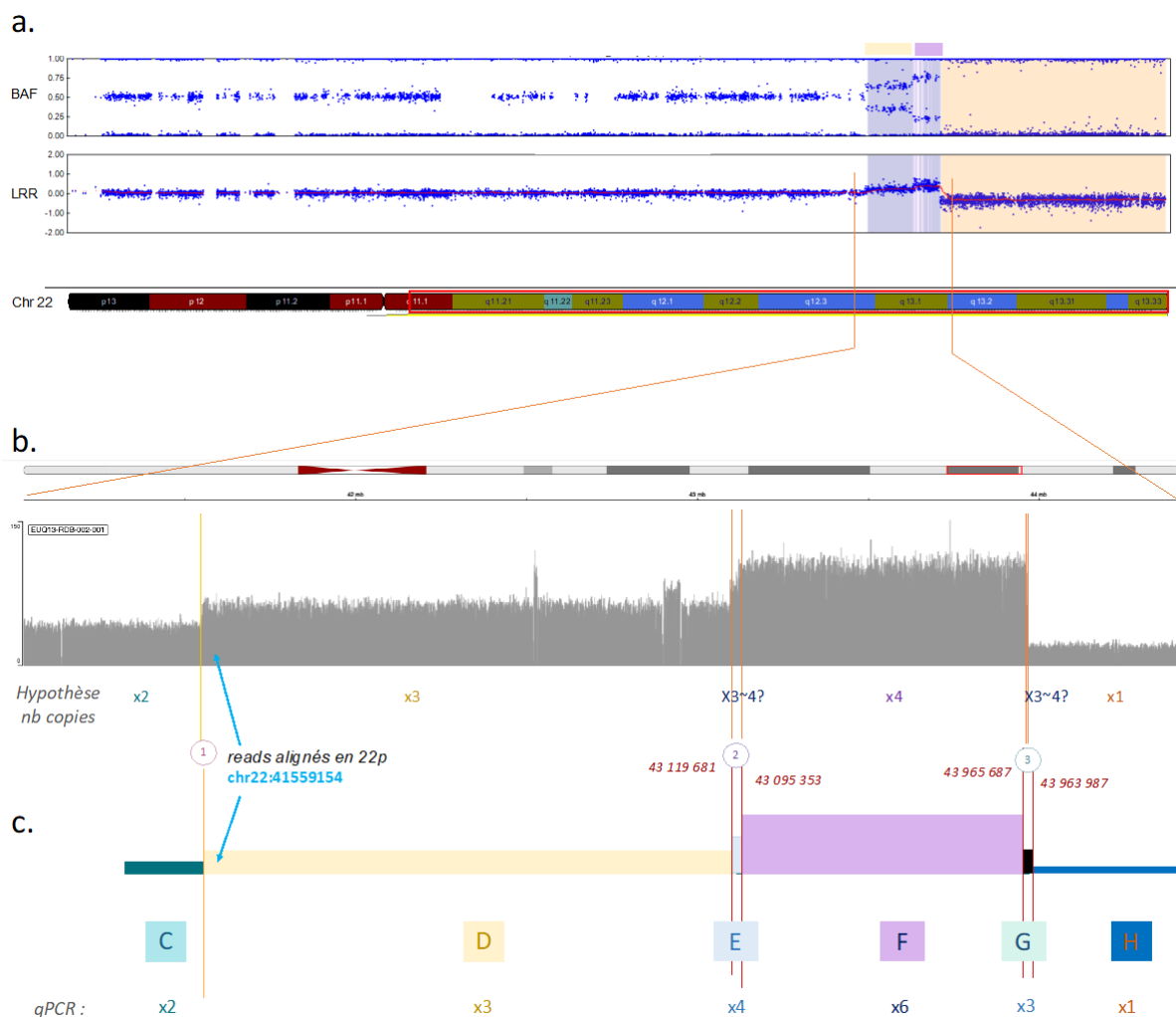
**A.** Image de FISH, sonde TBX1 en rouge, SHANK3 en vert. **B.** Profil de puce SNP sur GenomeStudio. **C.** Capture des profils CRAM du WGS sur patient (gauche) avec ses parents (droite). Les reads discordants et splits reads sont marqués en bleu.

Ce cas montre l'intérêt du WGS dans la précision du mécanisme de formation de l'inv dup del qui aurait pu changer le conseil génétique à cette famille.



## 4.2. Inv Dup Del complexe

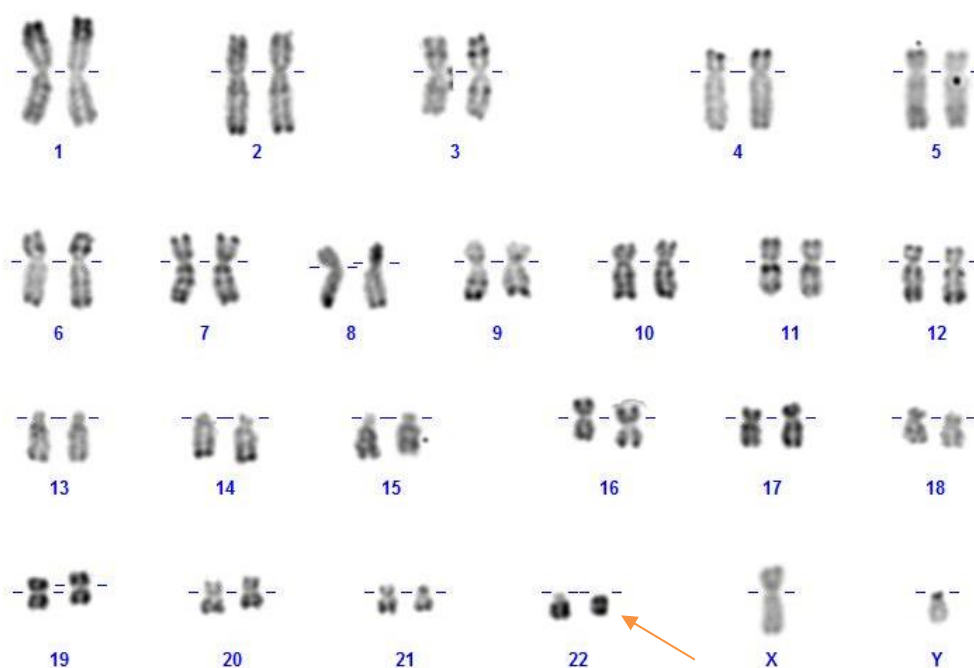
Parmi les 7 patients PMS porteurs de r(22) rapportés dans l'article dédié, le patient P1 porte un remaniement complexe (voir article [chapitre 2/3](#)). L'ACPA réalisée en première intention avait mis en évidence une délétion terminale de 7,2 Mb associée à deux gains de copie adjacents, le plus distal de 788kb en mosaïque entre 3 et 4 copies et le plus proximal a priori en 3 copies de 1,5 Mb (figure 26.a) (formule ISCN : [GRCh37]arr 22q13.2(41591770\_43158867)x3, 22q13.2(43159090\_43947103)x3~4, 22q13.2q13.33(43973229\_51195728)x1). La FISH a confirmé la délétion (sonde SHANK3, Leica) et la duplication distale (sonde RP11-2E2). Toutefois, cette technique ne nous a pas permis de confirmer le nombre de copies de chaque segment dupliqué.



**Figure 26.** Différentes visualisations et schématisation du chromosome 22 de P1  
**a.** Profil de puce SNP sur Genome Studio, montrant la délétion terminale (orange) et deux gains de copie adjacents (bleu). **b.** Visualisation de la profondeur de lecture par WGS sur l'interface IGV web, zoomé sur les segments dupliqués. **c.** Représentation schématique de la région 22q13.1q13.3 et position des splits reads alignées sur un BCCA. qPCR : PCR quantitative

L'analyse du caryotype complet a révélé la présence d'un chromosome 22 en anneau (figure 27), homogène. L'analyse parentale par FISH a montré que la délétion et la duplication étaient survenues *de novo*, avec *a priori* absence d'inversion parentale.

L'analyse des données de WGS ciblées sur les régions d'intérêt, en particulier sur les trois points de cassure, a mis en évidence deux gains de copie supplémentaires : le premier entre les deux duplications (région E), et le deuxième entre la duplication distale et la délétion terminale (région G), de 24,3 kb et 1,4 kb respectivement (Figure 26.b). L'analyse IGV montre que ces deux régions sont inversées, et en nombre de copies intermédiaire entre la région D et la région F d'après la profondeur (figure 25b). Ainsi, si l'ACPA identifiait 3 CNVs dans la région distale du chromosome 22, les données de WGS ont montré un remaniement beaucoup plus complexe, fait de 5 CNVs dont le nombre de copies semblait varier de 1 à 4. (Figure 26.b). De plus, l'analyse du point de cassure sur la duplication le plus distal en chr22:41551360 montre la présence de split reads et discordants en aval, en position chr22 :41559154 dans le segment D s'alignant sur un BCCA sur T2T (figure 26.c), compatible avec l'anneau.

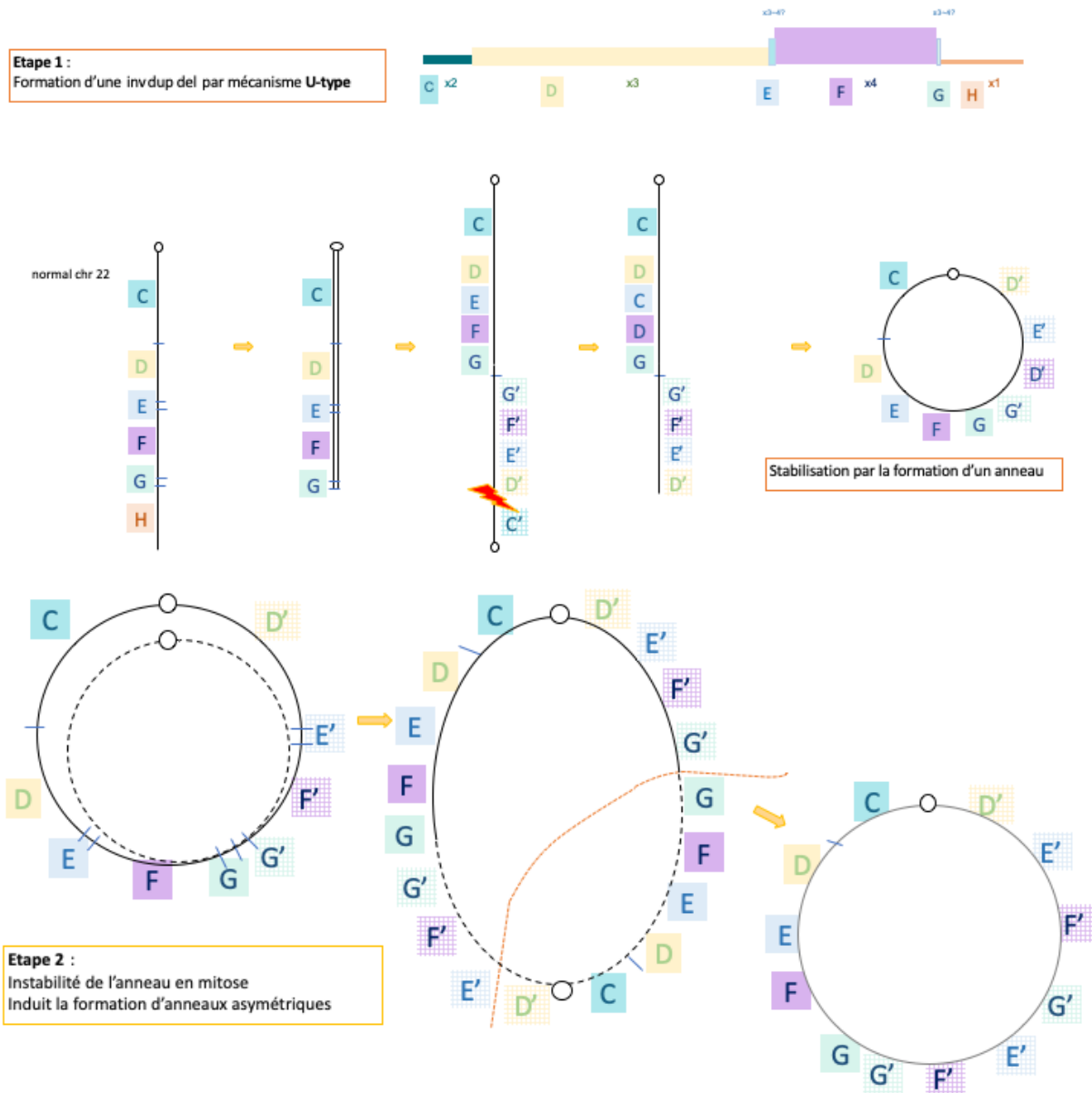


**Figure 27.** Caryotype du patient P1 avec anneau du chromosome 22  
Flèche : r(22)

Les mécanismes de formation des anneaux ont été évoqués en introduction ([chapitre 1/3.1.4](#)). Pourtant, aucun mécanisme déjà décrit ne concordait avec la

formation de cet anneau complexe. Nous avons donc tenté d'identifier le mécanisme en cause.

Notre première hypothèse était la suivante : le premier événement consistait en une inv dup del du 22 formée par un mécanisme de U-type, qui aboutit à la délétion du segment terminal H, la formation d'un dicentrique après fusion des chromatides sœurs puis cassure du dicentrique à l'origine de la duplication inversée des segments D, E, F et G. Ce chromosome remanié va se stabiliser par circularisation (figure 28. étape 1).

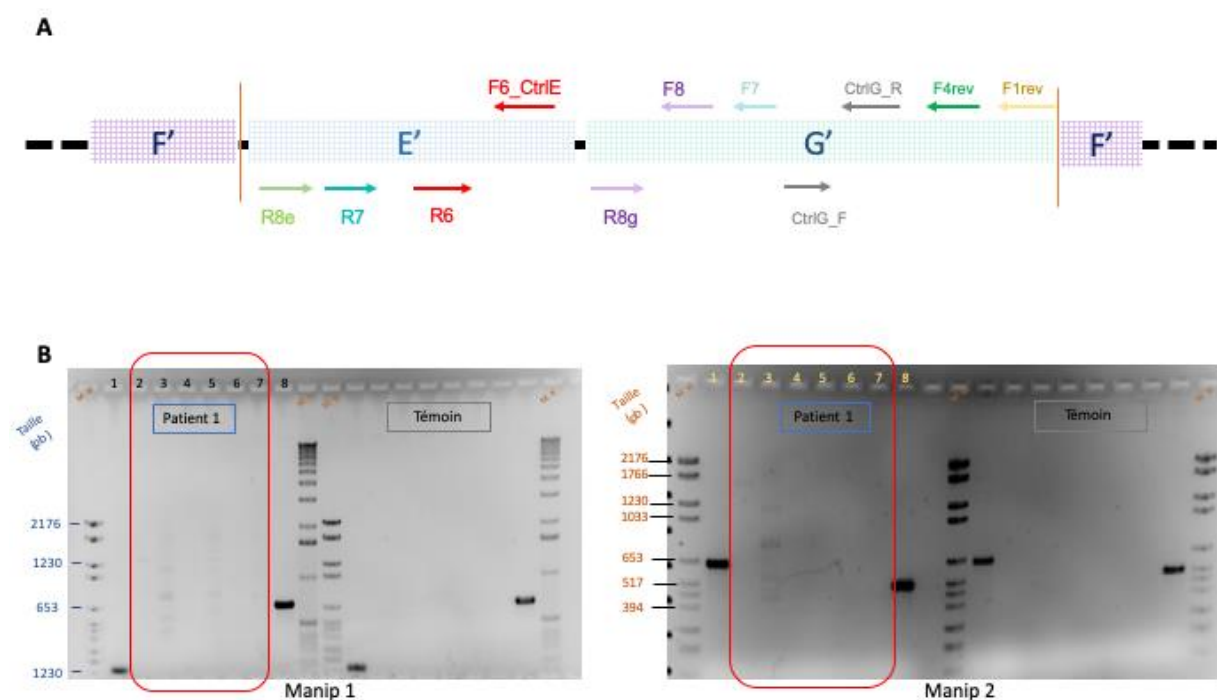


**Figure 28.** Première hypothèse de formation de l'anneau complexe de P1  
*Présenté en communication orale à l'ACLF 2021*

L'instabilité mitotique connue des anneaux pourrait avoir engendré les événements suivants. En effet, cet anneau peut se dupliquer puis se recombiner par homologie de

séquence aboutissant à la formation d'un nouvel anneau dicentrique asymétrique. Ce chromosome va à son tour subir un étirement au moment des divisions successives et se casser pour donner un chromosome circulaire fait des segments supplémentaires à la séquence normale G, F, E, G, F, E et D (figure 28. étape 2).

Pour aller plus loin dans le mécanisme, nous avons voulu déterminer si la séquence chimérique G-E, c'est à dire la fusion de ces deux séquences normalement séparées de 868,6 kb, était présente chez ce patient. Pour cela nous avons tenté d'amplifier cette séquence en *designant* une amorce *forward* sur le segment E et une amorce *reverse* sur le segment G, et vice versa, et ce pour plusieurs couples afin d'augmenter les chances d'amplification (figure 29.A, table 3). Un témoin sans remaniement du chromosome 22 a été utilisé pour s'assurer de la spécificité de l'amplification dans le cas positif.



**Figure 29.** Test d'amplification de la séquence chimérique

**A.** Représentation de la séquence chimérique et position des amorces utilisées (flèches). **B.** Gel après amplification par les couples d'amorces (liste des manips 1 et 2 en table 3). Les puits 1 et 8 correspondent à des couples contrôles. Les puits 2 à 7 correspondent à une amplification qui se produirait si la séquence chimérique existe.

Malgré plusieurs techniques, aucune des amplifications réalisées n'a permis de confirmer la présence de cette séquence chimérique (figure 29.B).

Puits	Couple de primers		P1	T	Taille (pb)	Puits	Couple de primers		P1	T	Taille (pb)
<b>Manip 1</b>						<b>Manip 2</b>					
1	F6_CtrlC	x R6	x	x	164	1	F6_CtrlC	x R7	x	x	642
2	F1rev	x R6	x	o	1880	2	CtrlE_R	x R6	x	o	1741
3	F4rev	x R6	x	o	1851	3	F4rev	x R6	x	o	1851
4	F1rev	x R7	x	o	2358	4	F7	x R7	x	o	1511
5	F4rev	x R7	x	o	2329	5	F8	x R7	x	o	1219
6	F1rev	x R8	x	o	2468	6	F7	x R8c	x	o	1620
7	F4rev	x R8	x	o	2439	7	F8	x R8c	x	o	1328
8	F1rev	x CtrlE_F	x	x	687	8	F7	x R8e	x	x	687

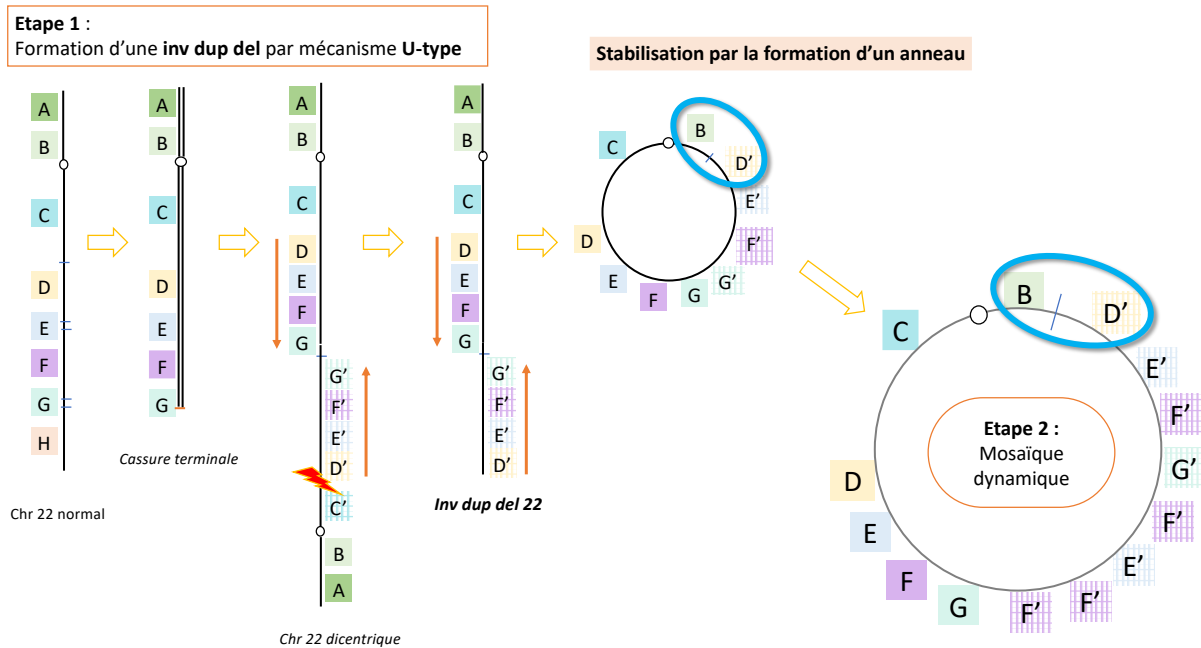
**Table 3.** Couples d'amorces utilisées pour la séquence chimérique

P1 : Patient 1 ; T : Témoin

Face à ces techniques non concluantes, nous avons réalisé une PCR quantitative de chacun des segments pour établir avec plus de précision leur nombre de copies respectif. Ainsi, le nombre de copies des segments dupliqués variait de 3 à 6, les segments D et G étaient en effet dupliqués (x3), le segment E tripliqué (x4) alors que le segment F apparaissait en 6 copies (figure 26.c). La précision du nombre de copies invalidait donc notre première hypothèse de formation de l'anneau.

Nous avons alors suggéré une nouvelle hypothèse basée également sur deux évènements successifs. Le premier évènement est identique à notre première hypothèse à savoir une inv dup del « classique » stabilisée par circularisation (figure 30). Le deuxième évènement, toujours lié à l'instabilité mitotique de l'anneau, pourrait faire intervenir un réarrangement favorisé par les points de cassure initiaux faisant intervenir un mécanisme de FoSTeS. Ce mécanisme favorisant la formation de microduplications pourrait expliquer le nombre de copies des segments E et G selon le schéma proposé en figure 30. Cette hypothèse est renforcée par la mise en évidence de régions d'homologie de type Alu au niveau de chacun des points de cassure (voir [chapitre 2/5](#)) et pourrait s'intégrer dans un mécanisme de chromoanasythésis.

De plus, l'analyse des CRAM montrant des splits reads proches du point de cassure distale du segment D suggère que la fusion entre le bras court et le bras long de l'anneau du chromosome 22 s'est fait entre les séquences B du bras court et D du bras long (figure 26.c), ce qui est concordant avec notre hypothèse de formation.



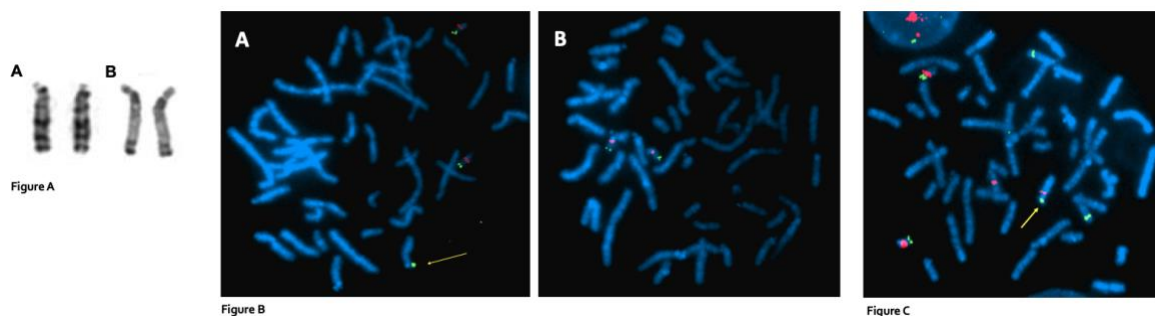
**Figure 30.** Proposition finale de mécanisme de formation de l'anneau complexe chez le patient P1

Formation d'une inv dup del (22) puis stabilisation par formation d'un anneau 22.

Flèches oranges : orientation du génome. Ronds bleus : séquences en continue sur les données de WGS.

### 4.3. Dérivé 13 d'une translocation t(13;22)(p13;q13.3)

Un des patients inclus dans notre étude présentait une duplication terminale de 135 kb liée à la présence d'un dérivé 13 d'une translocation t(13;22)(p13;q13.3) survenue *de novo* et en mosaïque (30%), responsable d'une duplication 22qter. Cette anomalie avait été identifiée en ACPA et confirmée par FISH au laboratoire du centre d'origine (figure 31).



**Figure 31.** Image de caryotype et FISH montrant le dérivé 13 et la duplication 22qter en mosaïque.

**A.** chromosomes 13 de l'enfant par coloration A- en bandes GTG et B- en bandes RHG. **B:** FISH avec les sondes TBX1 en rouge et SHANK3 en vert. **C.** FISH avec les sondes SHANK3 (22q13.33) et RP11-179O11 (14q31.2) en vert, et TBX1 (22q11.21) et RP11-56M2 (13q13.3) en rouge. Flèche jaune : der(13).

Source : Poster présenté à l'ACLF 2016 par V. Marquet, C. Thuillier, S. Bourthoumiou et C. Yardin

Après WGS du patient, l'analyse sur IGV du CRAM nous a permis de voir près du point de cassure un split read (en bleu sur la figure 32). Nous avons donc utilisé notre stratégie pour la mise en évidence d'anneau du chromosome 22 et avons montré que le BLAT de la séquence de ce split read sur T2T s'alignait en effet sur un BCCA (figure 32, encadré orange).



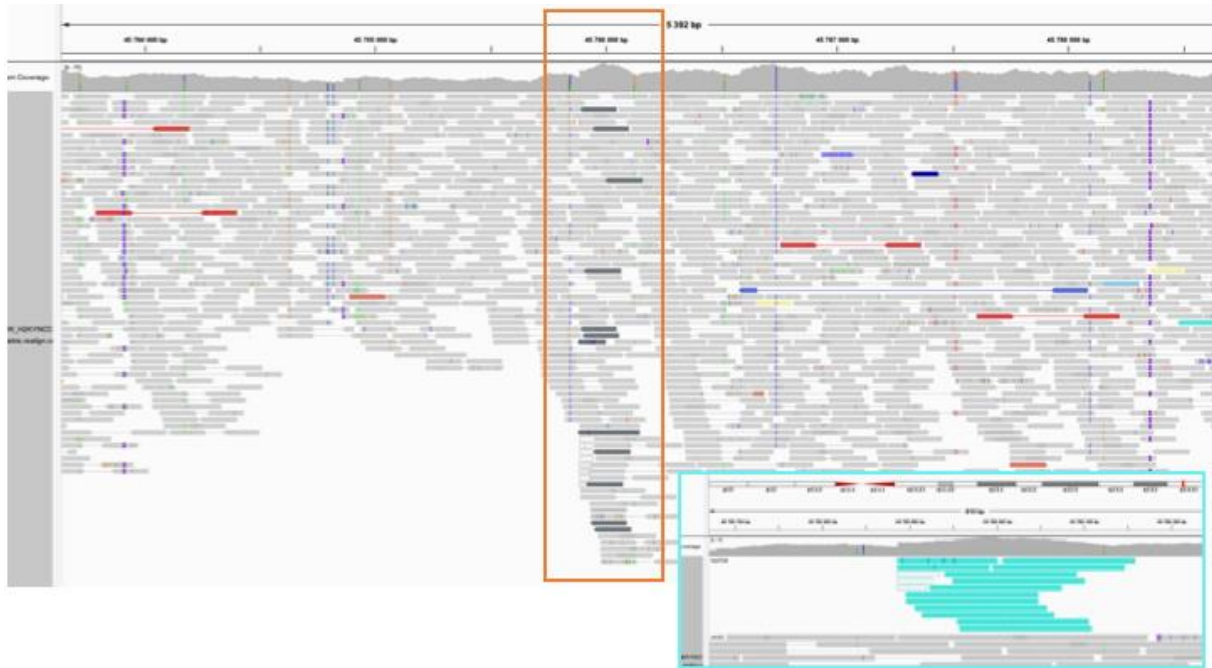
**Figure 32.** Visualisation du CRAM sur IGV au niveau du point de cassure 22qter  
Présence d'un split read (cercle bleu), qui s'aligne sur T2T sur une séquence de BCCA (encadré orange)

Bien que le remaniement soit survenu en mosaïque, l'analyse fine des *split reads* issus des données de séquençage nous a permis de confirmer que cette méthode d'analyse était efficace pour identifier des séquences insérées dans un BCCA.

#### 4.4. Dérivé d'inversion du chromosome 22

Le quatrième patient qui nous a interpellé est également porteur d'une duplication terminale 22qter. L'analyse sur IGV au point de cassure mettait en évidence 5 splits reads et 10 reads discordants (figure 33). Le BLAT des séquences des splits reads sur T2T montraient un alignement sur un BCCA. Initialement, ce patient avait été inclus dans l'étude en raison de la duplication terminale incluant *SHANK3* identifiée en ACPA. Nous n'avions pas récupéré les résultats de son caryotype.





**Figure 33.** Point de cassure 22qter avec dérivé d'inversion

Encadré orange : split-reads et reads discordants en gris foncé au niveau du point de cassure.  
Encadré bleu : zoom sur le point de cassure avec regroupement des reads d'intérêt.

Après analyse des données de WGS, le laboratoire du centre d'origine nous a confirmé que le caryotype retrouvait « la présence d'un dérivé 22 caractérisé d'une part par l'absence de séquences satellites sur les bras courts et d'autre part par la présence de deux signaux *SHANK3*, l'un en place sur les bras longs et l'autre localisé sur les bras courts. En conclusion, il s'agit d'un remaniement d'un chromosome 22 évocateur d'un dérivé d'inversion et entraînant une trisomie 22q partielle. »

Une nouvelle fois nous avons pu confirmer que notre méthode permet de mettre en évidence des remaniements chromosomiques impliquant des BCCA. Toutefois, malgré les progrès liés à l'utilisation de la version T2T du génome, la distinction entre les différents BCCA de chaque paire de chromosome acrocentrique reste très difficile. Il n'est en effet à ce jour pas encore possible de différencier les différentes séquences des 5 paires de BCCA, ni de différencier un anneau d'autres variations de structure conduisant à une fusion entre le bras court et le bras long d'un acrocentrique.

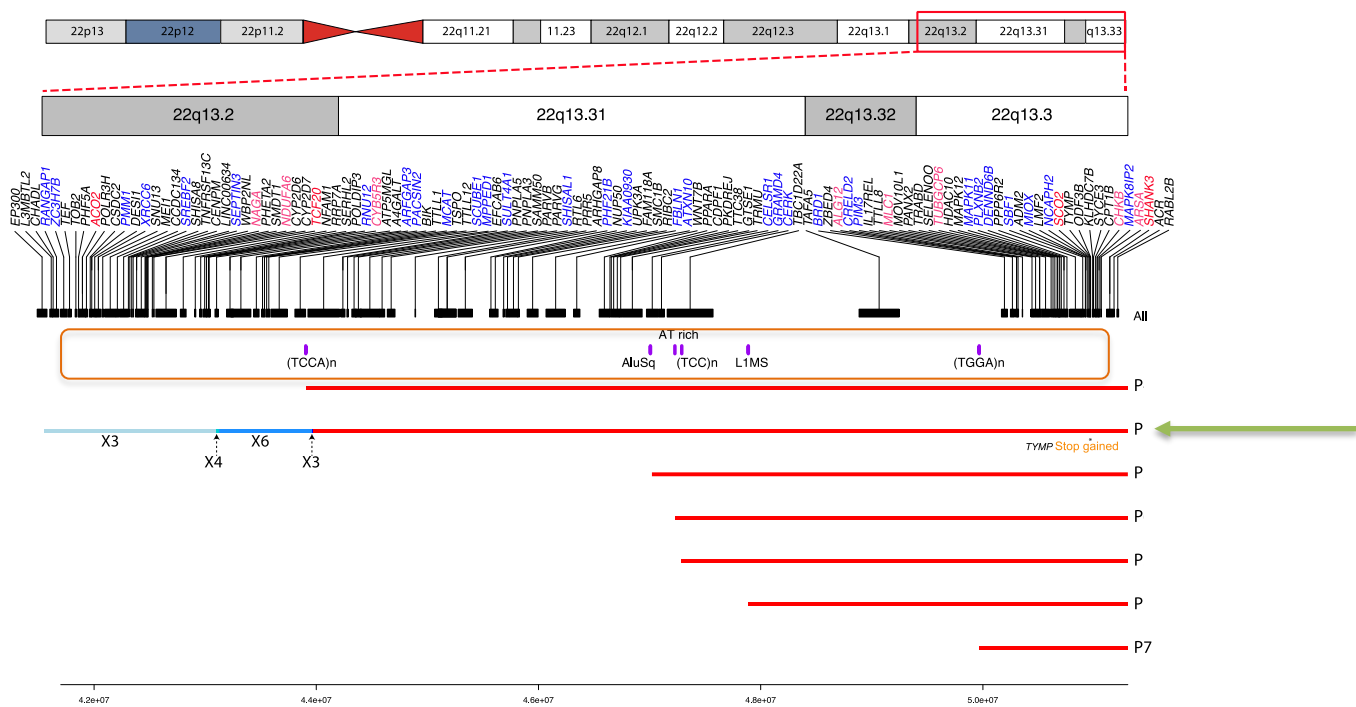
Ainsi, les techniques de cytogénétique conventionnelle, caryotype et FISH gardent une place importante dans la validation et la caractérisation d'un remaniement impliquant un BCCA.



## 5. Éléments répétés à l'origine des réarrangements

Les éléments répétés répartis dans le génome constituent environ 50% du génome humain, représentés majoritairement par les séquences Alu de la famille des SINEs. La forte homologie de séquence entre eux favorisent les échanges génomiques et de fait les microréarrangements (Bose et al., 2014; Carvalho & Lupski, 2016). Nous nous sommes donc intéressés aux séquences répétées présentes en 22q chez tous les individus de notre cohorte.

Ainsi, nous avons évalué la proportion de ces séquences répétées au niveau des points de cassure distaux en 22q et également dans une fenêtre de +/- 100 pb. Pour les 7 patients porteurs d'un anneau du 22, 4 points de cassure distaux tombent directement dans une séquence répétée, SINE (Alu) (P1 et P2), LINE (P3) ou des Tandem Repeat (P4, (TCCA)n) (figure 34, table 4).



**Figure 34.** Région 22qter avec représentation des séquences répétées aux points de cassure terminaux

Encadré orange : position des séquences répétées identifiées aux points de cassure terminaux.  
Flèche verte : patient P1

Les patients P5, P6 et P7 ont des séquences répétées localisées à 6 ((TCC)n), 26 (LC ATrich) et 12 nucléotides ((TGGA)n) respectivement du point de cassure, ce qui donne un taux de 100% de séquences répétées aux points de cassure dans une fenêtre de

100 pb. Nous nous sommes également intéressés aux points de cassures additionnels du chromosome 22 complexe de P1, et les 4 points de cassures bordant les segments dupliqués se trouvent effectivement à proximité de séquences répétées Alu.

	Fenêtre	TOTAL	SINE (Alu)	LINE	Simple Repeat	LTR	DNA repeat	Low complexity
<b>Ring</b> n = 7	<b>1pb</b>	<b>4</b>	2	1	1	0	0	0
	%	<b>57%</b>	<b>29%</b>	14%	14%	0%	0%	0%
	<b>100pb</b>	<b>7</b>	2	1	3	0	0	1
	%	<b>100%</b>	29%	14%	<b>43%</b>	0%	0%	14%
<b>Del</b> n = 45	<b>1pb</b>	<b>20</b>	7	4	3	4	1	1
	%	<b>44%</b>	<b>16%</b>	9%	7%	9%	2%	2%
	<b>100pb</b>	<b>33</b>	9	4	3	5	1	8
	%	<b>73%</b>	<b>20%</b>	9%	7%	11%	2%	18%

**Table 4.** Nombre et type d'éléments répétés au niveau des points de cassure 22q

Nous avons comparé ce taux chez les patients porteurs de délétion simple. Vingt patients sur 45 ont une séquence répétée à 1pb autour du point de cassure (44%), et 33/45 quand la fenêtre s'étend à 100 pb, soit 73% (table 4). Si ce taux est plus faible, notre échantillonnage dans les deux groupes de patient ne permet pas de réaliser de statistiques solides.

Si l'on s'intéresse au type de séquence impliquée, on observe pour les anneaux une majorité de cassure dans une séquence Alu (2 cas/7) mais dans une fenêtre plus large de 100 pb une majorité de *Simple repeat* ou séquences répétées en tandem. Pour les patients avec une délétion terminale, on trouve une majorité de séquences Alu tant à 1 et 100 pb du point de cassure. On observe également un plus grand nombre, à 100 pb, de régions de Low Complexity.

Dans une étude de 2011, Bonaglia et al. ont isolé et séquencé 22 points de cassure terminaux chez 20 patients porteurs d'une délétion terminale simple en 22qter (Bonaglia et al., 2011). Pour 19 des points de cassure identifiés chez 17 cas (86%), ils ont mis en évidence une répétition d'un hexamère GGTTAG, et 13 points de cassure (59%) tombaient dans des séquences répétées (SINE, LINE, simple repeat et DNA repeat).

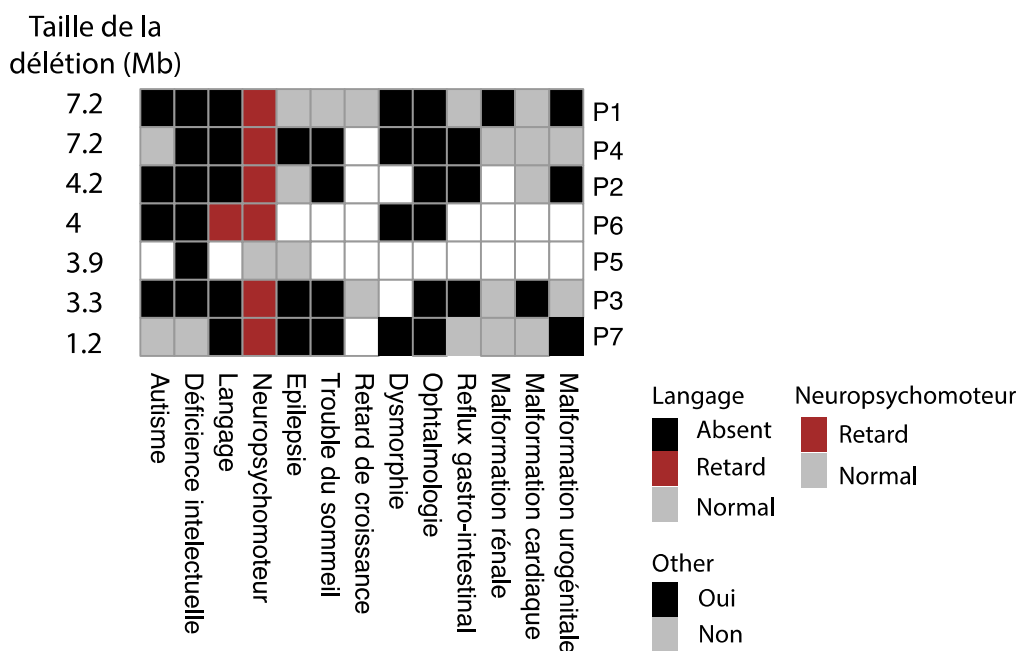
Nos résultats sont donc cohérents avec la littérature bien que non significatifs. Toutefois, bien que la présence de séquences répétées puisse favoriser les microremaniements chromosomiques et dans notre étude la survenue de délétions terminales, leur implication dans la formation des anneaux ne peut être affirmée. De plus, la présence de séquences répétées au niveau de chaque point de cassure des duplications 22q du patient P1 suggère qu'elles ont pu jouer un rôle dans le remaniement complexe de l'anneau et favoriser des microremaniements par FoSTeS par exemple.

## 6. Variabilité clinique et génétique du PMS : variations additionnelles

### 6.1. Chez des patients porteurs d'anneau du chromosome 22

#### 6.1.1. Variabilité clinique

Les données cliniques ont été recueillies pour chacun des patients et sont résumées dans la figure 35 pour les principaux traits associés au PMS. Nous retrouvons, comme décrits dans la littérature, une DI chez tous les patients et un TSA pour 4 d'entre eux. Le langage est touché pour tous les patients sauf un pour lequel nous n'avons pas eu l'information. Une atteinte ophtalmologique est observée pour 6 des 7 patients (pas d'information pour P5), et une épilepsie a été diagnostiquée pour 3 patients.



**Figure 35.** Représentation heatmap du phénotype des 7 patients

Heatmap montrant, pour chaque patient les données de 14 caractéristiques cliniques pertinentes. Les patients sont classés en fonction de la taille de la délétion. La présence et l'absence d'une caractéristique clinique sont représentées respectivement en noir ou en gris. Pour le langage, l'absence de parole, le retard de parole et la parole normale sont illustrés en noir, marron et gris respectivement. Les données manquantes sont représentées en blanc.

Figure réalisée par A. Vitrac, Institut Pasteur

#### **Patient 1**

P1 est un garçon de 2 ans au moment du diagnostic. Il est le cinquième enfant de parents en bonne santé. La grossesse s'est déroulée sans problème, à l'exception d'une hydronéphrose bilatérale diagnostiquée par échographie. Une échographie

rénale néonatale a révélé une dilatation pyélocalicielle bilatérale et des kystes du rein droit. Il présentait également une malformation urologique avec un reflux vésico-urétéral nécessitant plusieurs interventions chirurgicales. Sur le plan ophtalmologique, il souffre d'un glaucome congénital sévère, de myopie et d'astigmatisme. Sur le plan du développement, il est déficient intellectuel, avec un retard de développement sévère, une absence de langage, et ne marche pas. Il est microcéphale (-2,5 DS) et a une dysmorphie faciale. À l'âge de 2 ans, il a présenté un arrêt du développement. L'évaluation neurologique a révélé une hypotonie, des signes d'atteintes pyramidales et axonales, une amyotrophie généralisée et une abolition des réflexes. L'IRM cérébrale a révélé une atrophie corticale et sous-corticale, une ventriculomégalie bilatérale et une hypoplasie du corps calleux. Il a également une fossette sacrée.

### ***Patient 2***

P2 est un garçon de 12 ans présentant un DI sévère avec une absence de langage et un TSA (ADOS-2 = 16, SRS = 97, RBS-S = 9). La grossesse était gémellaire et s'est déroulée sans problème, résultant d'une fécondation in vitro en raison d'une azoospermie paternelle. P2 a un frère jumeau en bonne santé, et il a été hospitalisé pour un syndrome de détresse respiratoire aiguë à la naissance, et opéré d'une hernie diaphragmatique droite qui a nécessité une gastrostomie pendant 3 ans. Il a une hypermétropie et un strabisme. Les potentiels évoqués auditifs étaient normaux à 9 mois puis montrent une perte auditive bilatérale (de légère à profonde) à 2 ans et 3 mois. Il a souffert de reflux gastro-œsophagien et d'une ectopie testiculaire opérée. Il avait une hypotonie néonatale, un retard de développement avec une marche à 3 ans et demi et une propreté non acquise mais son examen neurologique à 12 ans est normal. Il présente des signes d'auto et d'hétéro-agressivité, une intolérance à la frustration et des réveils nocturnes. L'IRM cérébrale était normale à l'âge d'un mois, et l'EEG au même âge ne montre pas d'anomalies significatives et ni signes de focalisation.

### ***Patient 3***

P3 est un garçon de 6 ans. Il est le premier enfant de parents non apparentés et est né après une grossesse sans problème. Il a présenté une hypotonie néonatale, des difficultés à avaler et une sténose aortique supra-avalvulaire discrète. À l'âge de 6 ans, il ne parlait pas et avait des difficultés à s'endormir (traitées par la mélatonine). Il présentait également des troubles du comportement avec une anxiété majeure, une

intolérance à la frustration et porte des lunettes pour hypermétropie et astigmatisme. Il a été diagnostiqué avec un TSA.

#### **Patient 4**

P4 est une jeune fille de 19 ans à l'inclusion, première enfant de parents non apparentés. Elle a eu un retard global du développement, s'est assise à l'âge de 1 an et a marché à l'âge de 2,5 ans. Elle a une myopie et de l'astigmatisme, une hernie diaphragmatique et une scoliose. Elle a une démarche ataxique et a fait une convulsion hyperthermique dans l'enfance. Une IRM cérébrale réalisée à 9 mois ainsi qu'un EEG étaient normaux. Elle présente un TDI, quelques stéréotypies, mais ne répond pas aux critères de TSA. Elle n'a pas acquis le langage verbal. Un bilan auditif réalisé tôt dans son enfance était normal, mais a été renouvelé à 21 ans et retrouve des potentiels évoqués auditifs désynchronisés. Elle présente par ailleurs quelques tâches hypopigmentées sur le bras et l'abdomen, ainsi que des anomalies dentaires.

#### **Patient 5**

P5 est un garçon présentant un TDI avec retard de langage. Il avait un développement psychomoteur normal, et *a priori* pas d'épilepsie. L'EEG et l'IRM n'étaient pas disponibles.

#### **Patient 6**

P6 est un garçon de 3 ans au moment de l'évaluation. Il a présenté un retard de croissance intra-utérin et est né par césarienne en raison d'anomalies du rythme cardiaque fœtal. Par la suite, il a rencontré des difficultés alimentaires, un retard global du développement (a marché à l'âge de 21 mois) et un retard sévère du langage. Il présente également des traits dysmorphiques avec une plagiocéphalie. L'examen neurologique révèle un syndrome cérébelleux, avec une méga-citerne de la fosse cérébrale postérieure visualisée à l'IRM. L'EEG était normal. P6 a également présenté une œsophagite, une hernie de la ligne blanche, une dermatite atopique et de l'astigmatisme.

#### **Patient 7**

P7 est un garçon de 8 mois au moment du diagnostic. La grossesse s'est déroulée sans problème. Il a été orienté en génétique en raison d'un retard global du développement. Les évaluations n'ont pas révélé de DI ni de TSA. Il présente une

microcéphalie (-2DS) ainsi que des caractéristiques faciales dysmorphiques, notamment un front fuyant, un menton en retrait, des narines légèrement retroussées et des sourcils épais. Il a rencontré des difficultés alimentaires au cours des premiers mois de vie en raison d'une allergie aux protéines du lait de vache. Il n'a jamais présenté de crises et aucun EEG n'a été réalisé. Une IRM cérébrale a retrouvé un corps calleux mince. De plus, il souffre d'un glaucome bilatéral, de surdité asymétrique, d'une syndactylie des orteils 2 et 3 du pied droit, d'un angiome à la main droite, d'une hydrocèle légère et d'infections récurrentes des voies respiratoires supérieures.

### 6.1.2. Variations génomiques additionnelles modulant le phénotype

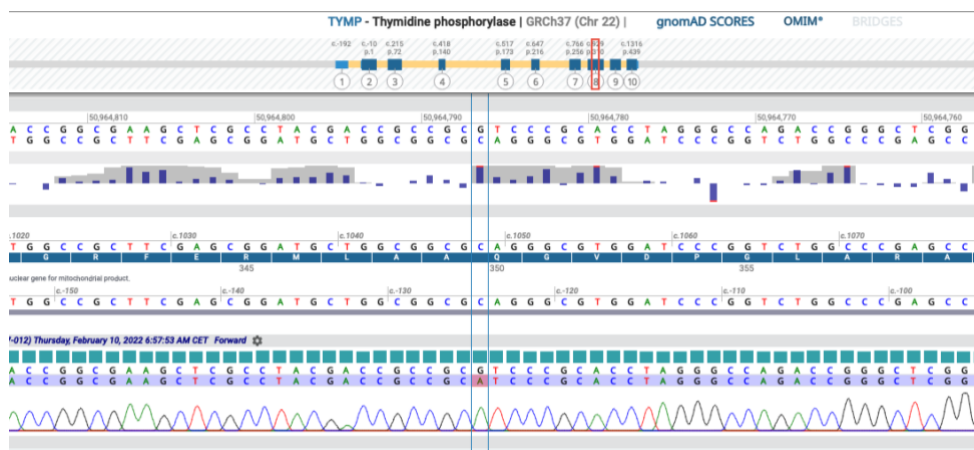
Nous avons recherché des variations additionnelles qui pourraient participer à la sévérité ou à des éléments du phénotype des 7 patients porteurs d'anneaux. Nous avons identifié des variations pour 5 d'entre eux qui se révèlent intéressantes (table 5).

Patient	CNVs additionnel	SNV/indel additionnels
P1	4q21.23q21.3(85620879_86990027)x3 pat ( <b>WDFY3</b> )	22:50964786:G:A ; c.1048C>T ; p.Gln350Ter ( <b>TYMP</b> )
P2	20q12(41361424_41626428)x1 dn ( <b>PTPRT</b> ), 22q11.21(19255856_20978977)x3 dn ( <b>TBX1</b> )	
P3	16p11.2(29595483_30198151)x3 mat ( <b>TBX6</b> )	
P4		1:205116700:C:A ; c.2776G>T ; p.Asp926Tyr, pat ( <b>DSTYK</b> )
P6		9:119461222:A:T ; c.1201A>T ; p.Lys401* ( <b>TRIM32</b> ) 11:16010716:C:T ; c.1793G>A ; p.Gly598Glu ( <b>SOX6</b> )

**Table 5.** Variations additionnelles (version du génome : GRCh37)

L'étude du WGS de P1 retrouve 2 variations additionnelles. La première est une mutation stop dans l'exon 8/10 du gène **TYMP**, en 22q13.33 en regard de la délétion 22qter. **TYMP** est associé à une encéphalopathie myo-neuro-gastrointestinale mitochondriale (syndrome MNGIE, OMIM #603041) autosomique récessive. Cette pathologie associe l'apparition progressive de troubles de la motricité gastro-intestinale, une neuropathie périphérique, une ophtalmoplégie externe chronique progressive et une leuco-encéphalopathie (Filosto et al., 2018; Laforce et al., 2009), qui pourrait s'inscrire dans la sévérité clinique de P1.

Nous avons vérifié et confirmé cette mutation par Sanger (figure 36). Malheureusement, l'ADN des parents n'était pas disponible pour réaliser une étude de ségrégation.



**Figure 36.** Profil Sanger de la mutation c.1048C>T de *TYMP*.

Par ailleurs, nous avons identifié une duplication de 1,4 Mb en 4q21.23q21.3 interrompant le gène **WDFY3** (OMIM \*617485), impliquée dans une forme autosomique dominante de microcéphalie primaire. L'analyse des CRAMs montre que cette duplication est directe, et l'analyse par FISH (sonde RP11-8N8) a montré qu'elle était héritée du père asymptotique a priori. Il est donc peu probable que cette duplication ait pu influencer le phénotype de ce patient.

Chez **P2**, une duplication classique de la région 22q11.21 englobant **TBX1** et une délétion de 289 kb de 20q12 englobant les exons 2 à 6 du gène **PTPRT** ont été identifiées. Les deux CNVs ne sont pas transmis par la mère (l'ADN du père n'étant pas disponible). L'étude de génotypage du patient et de la mère par les données de SNP array a montré que la délétion et la duplication ont eu lieu sur le chromosome maternel, et sont donc survenues *de novo*.

La microduplication 22q11 est bien décrite et associée à un syndrome très variable qui se caractérise par un risque de développer un TND ou un TSA. De plus, des anomalies ophtalmologiques telles que l'hypermétropie ou le strabisme, deux traits présentés par P2, sont présents dans 28% des individus porteurs de cette microduplication (Bartik et al., 2022). Les pertes auditives/surdité sont également décrites pour 11 à 18% des patients, de même que les anomalies urogénitales. La duplication 22q11.2 peut donc expliquer ces traits phénotypiques chez P2.

Par ailleurs, le gène **PTPRT** (OMIM \*608712) pourrait être impliqué dans des processus d'adhésion cellulaire du système nerveux central (J.-R. Lee, 2015). Il est



considéré comme un gène candidat aux TSA (SFARI 2 et dans la liste des *Candidate-NDD genes* de GeneTrek), et des délétions ont été rapportées chez des patients avec des troubles du comportement et des troubles déficit de l'attention/hyperactivité (Bernaciak et al., 2017). Une association positive aux troubles alimentaires est également rapportée (Kushima et al., 2022). La délétion de novo de *PTPRT* pourrait donc moduler le phénotype de ce patient. Des études fonctionnelles seraient nécessaires pour étayer cette hypothèse.

Nous avons identifié chez **P3** une duplication proximale 16p11.2 héritée de la mère. Il s'agit d'un CNV récurrent considéré comme un PIEV par le réseau français AchroPuce, c'est à dire un facteur de susceptibilité aux TND.

**P4** porte une mutation faux sens héritée du père (p.Asp926Tyr), dans le gène ***DSTYK*** (MIM \*612666), dont la fréquence allélique dans la base de donnée gnomAD est de 7.13e-4. Les prédictions sur l'effet de la protéine par les bases de données SIFT et Polyphen sont incertaines. Deux études récentes ont décrit un modèle de poisson-zèbre muté dans *Dstyk* avec de graves défauts vertébraux, entraînant un phénotype similaire à une scoliose congénitale (Bagwell et al., 2020; Sun et al., 2020). Ces variations pourraient expliquer la scoliose présentée par P4, d'autant plus que son père a également été traité plus jeune pour une scoliose.

Enfin, deux variations intéressantes ont été identifiées chez **P6**. La première est une mutation stop dans le gène ***TRIM32*** (OMIM \*602290) prédite pathogène. Ce gène est associé à deux syndromes récessifs : le syndrome de Bardet Biedl (#616988) et une dystrophie musculaire (#254110), mais il est aussi référencé dans la base SFARI avec un score 3 puisqu'il a été montré un enrichissement de délétions et variations faux-sens dans des cohortes de patients avec TND (Lionel et al., 2014).

La seconde variation est un faux-sens dans le gène ***SOX6*** (OMIM \*607257) qui est associé à une DI syndromique dans le syndrome de Tolchin-Le Caignec (OMIM #618971) (Schneider et al., 2022; Tolchin et al., 2020).

L'ADN des parents n'est pas disponible, aussi nous n'avons pu réaliser d'étude de ségrégation. Ces deux variations sont toutefois susceptibles d'exacerber le phénotype neurodéveloppemental du patient.

## 6.2. Variabilité clinique et génétique de patients PMS non porteurs d'anneau du chromosome 22

Dans le cadre de l'étude du PMS menée plus largement dans l'unité GHFC, A Vitrac, postdoctorante au laboratoire mène un projet dont l'un des principaux objectifs est d'étudier le rôle des variations additionnelles (*multiple hits*) dans l'hétérogénéité phénotypique du syndrome (travaux en cours, données non publiées).

Cette étude se base sur une cohorte de 90 patients PMS incluant plusieurs patients précédemment discutés dont les 7 patients porteurs d'un anneau du chromosome 22 et les patients porteurs de réarrangements complexes du chromosome 22 (figure 36). Le génome des 90 patients a été séquencé afin d'étudier les multiple-hits et les mettre en relation avec leur phénotype. Lorsque l'ADN des parents était disponible, ces derniers ont également été séquencés ce qui a permis l'analyse de variations *de novo* pour 64 individus.

Ma participation à ce projet s'est faite par la vérification visuelle et l'interprétation d'une grande partie des CNVs identifiés par WGS ainsi qu'à la collection des données cliniques de certains de ces patients.

### Présentation de la cohorte

La cohorte est constituée de 45 filles et de 45 garçons indiquant un sexe ratio de 1:1. L'origine ethnique des individus inclus dans la cohorte est majoritairement européenne (63%, figure 37). Parmi les 90 patients, 64 individus sont porteurs d'une délétion emportant *SHANK3* et 26 individus sont porteurs d'une variation ponctuelle délétère affectant *SHANK3*. Plusieurs traits cliniques associés au PMS ont été renseignés pour chacun des patients. Nous avons ainsi observé dans 98,6% des cas des troubles du langage, 97,7% de déficience intellectuelle, 95,6% de retard de développement, 68,2% de TSA, 61,7% de patients avec des particularités morphologiques, 47,7% de cas de reflux gastro-œsophagien ainsi que 27,7% de cas d'épilepsie (figure 38).

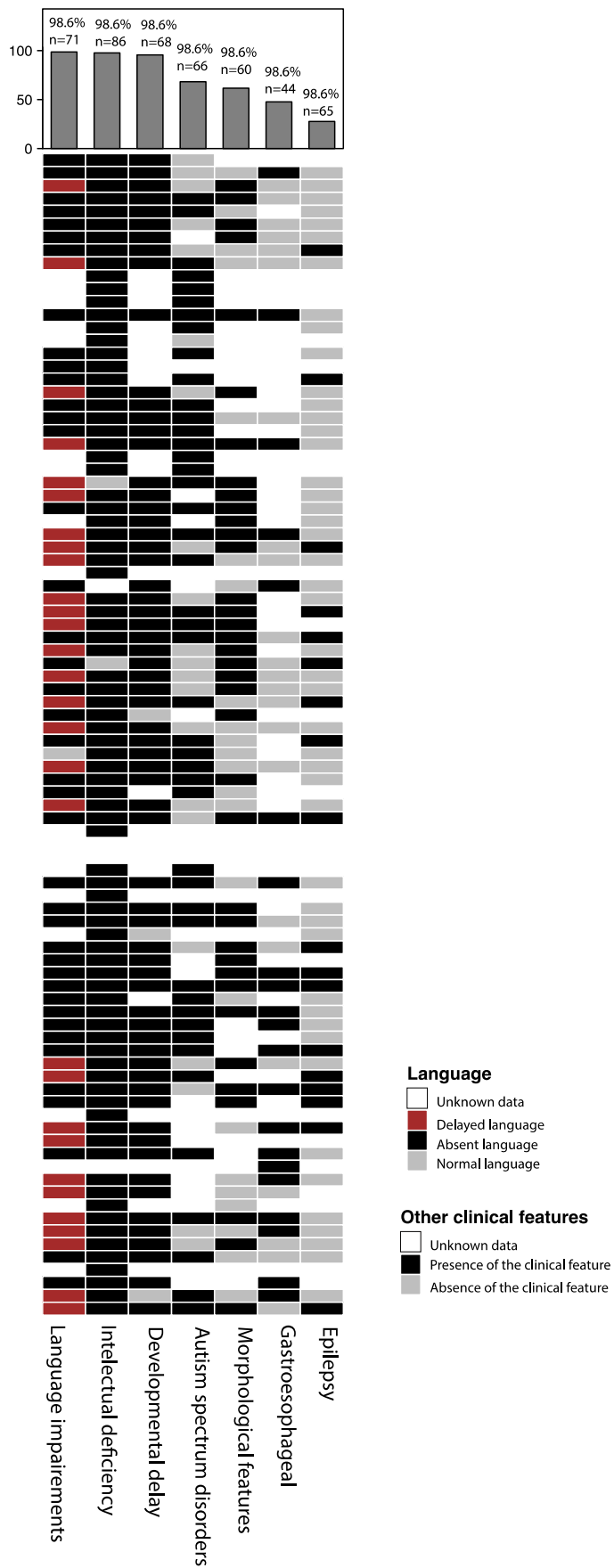


Participants with two parents	64					
Sex:						
F	45	Del	46	3	15	64
M	45	LoF	22	1	3	26
Ancestry (% of European)	63% (n=57)	<b>Total</b>	<b>68</b>	<b>4</b>	<b>18</b>	<b>90</b>

**Figure 37.** Présentation de la cohorte de patients avec PMS

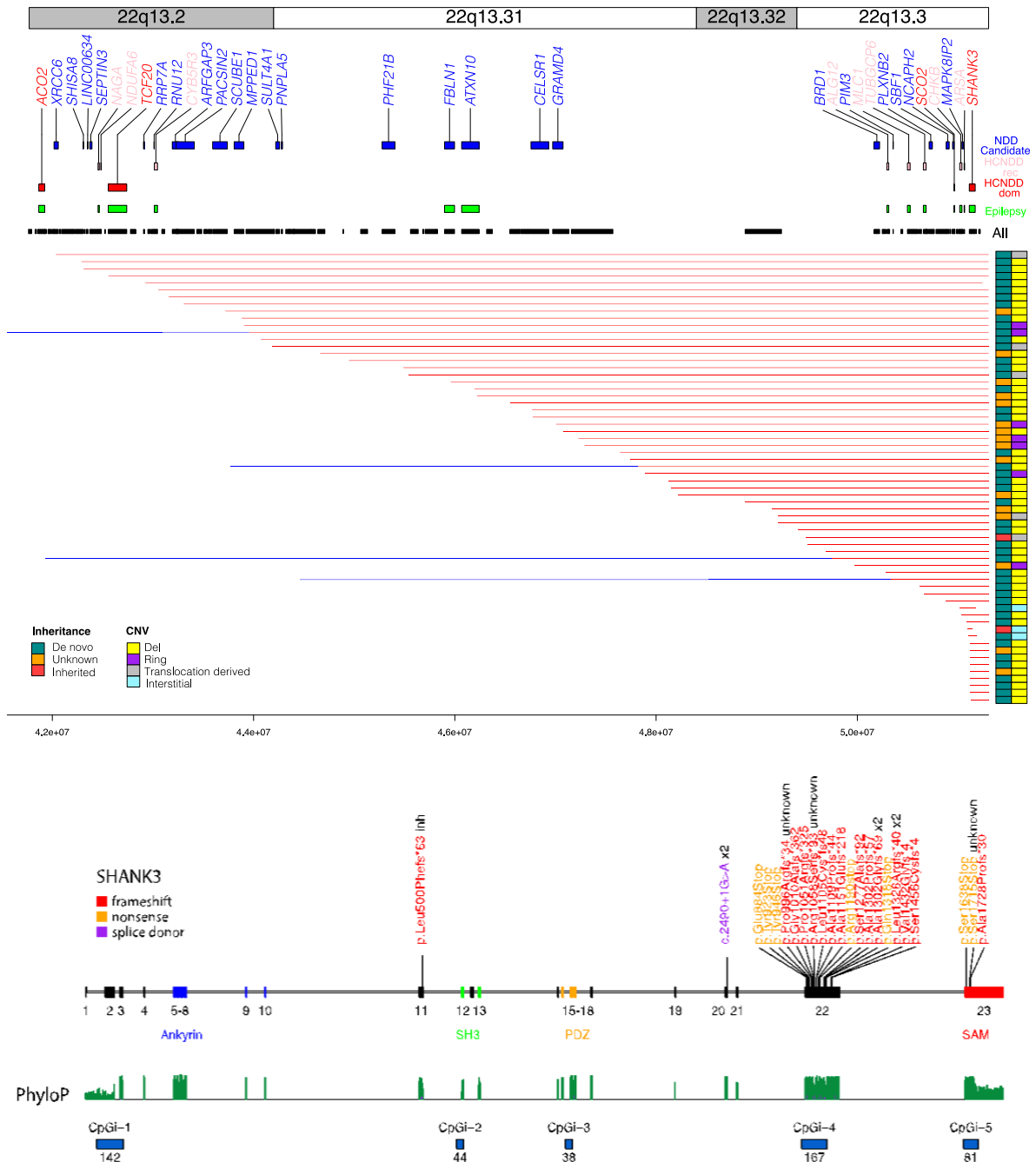
Del : délétion, LoF : mutations perte de fonction

Figure réalisée par A. Vitrac, Institut Pasteur



**Figure 38.** Profil clinique des patients avec PMS  
*Figure réalisée par A. Vitrac, Institut Pasteur*

La plupart des variations *SHANK3* sont apparus *de novo* (n=68) (figures 37 et 39). L'ensemble de ces variations sont représentées sur la figure 39. Les délétions ont une taille variable comprise entre 46 kb et 9,3 Mb, comme observé dans d'autres études. Pour 3 individus la délétion incluant *SHANK3* est interstitielle, et 5 délétions sont dérivées de translocation. Les variations ponctuelles sont majoritairement localisées sur l'exon 22 du gène.

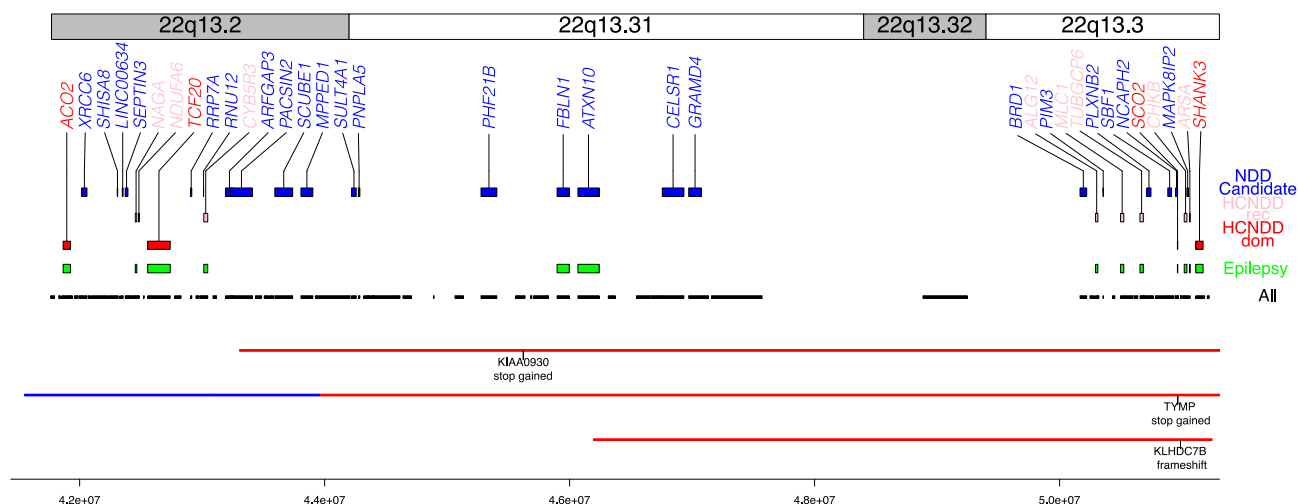


**Figure 39.** Illustration des délétions et variations ponctuelles *SHANK3* des patients avec PMS

Figure réalisée par A. Vitrac, Institut Pasteur

L'ensemble des CNVs et mutations identifiées dans l'ensemble du génome des patients et de leurs parents ont été vérifiés visuellement sur les fichiers BAM issus du séquençage complet du génome.

Les analyses de variations génétiques additionnelles sont effectuées à différentes échelles. Dans un premier temps il était important de voir si les patients avec PMS porteurs de délétions dans la région 22q13 portaient un second hit délétère sur le chromosome non délété. Nous avons ainsi recherché en ciblant la région délétée, les variations de type perte de fonction ou faux sens ayant un score mpc supérieur à deux, c'est à dire à haut risque d'avoir un effet délétère (Samocha et al., 2017). Lorsque le score mpc n'était pas disponible nous avons utilisé le score CADD phred et avons sélectionné les variations ayant un score supérieur à 28,3. Parmi les 64 patients portant une délétion dans la région 22q13, nous avons identifié un second hit sur le chromosome controlatéral chez 3 patients (figure 39). Un cas concerne le patient P1 de l'étude précédente et a déjà été discuté au [chapitre 2/4.2](#). Le deuxième cas caractérisé par une hypotonie axiale est porteur d'un variant non sens dans le gène *KIAA0930* non référencé dans les listes de gènes associés au neurodéveloppement. Une étude récente a néanmoins montré que ce gène pourrait être à l'origine d'atrophie musculaire dans certains cancers (Yamakawa et al., 2023). Le troisième cas est porteur d'une délétion d'un nucléotide dans le gène *KLHDC7B* composé d'un unique exon. Ce gène serait surexprimé chez les personnes atteintes de cancer du sein et pourrait jouer un rôle dans le contrôle de la prolifération cellulaire (Jeong et al., 2018). Ce gène pourrait également être impliqué dans des surdités liées à l'âge comme démontré par des études d'association et des modèles de souris dont le gène *KLHDC7B* a été invalidé (Cornejo-Sanchez et al., 2023). Peu de données cliniques sont disponibles pour ce patient et la littérature ne permet pas de conclure sur l'implication du variant retrouvé au sein de la séquence de *KLHDC7B* dans le phénotype du patient.

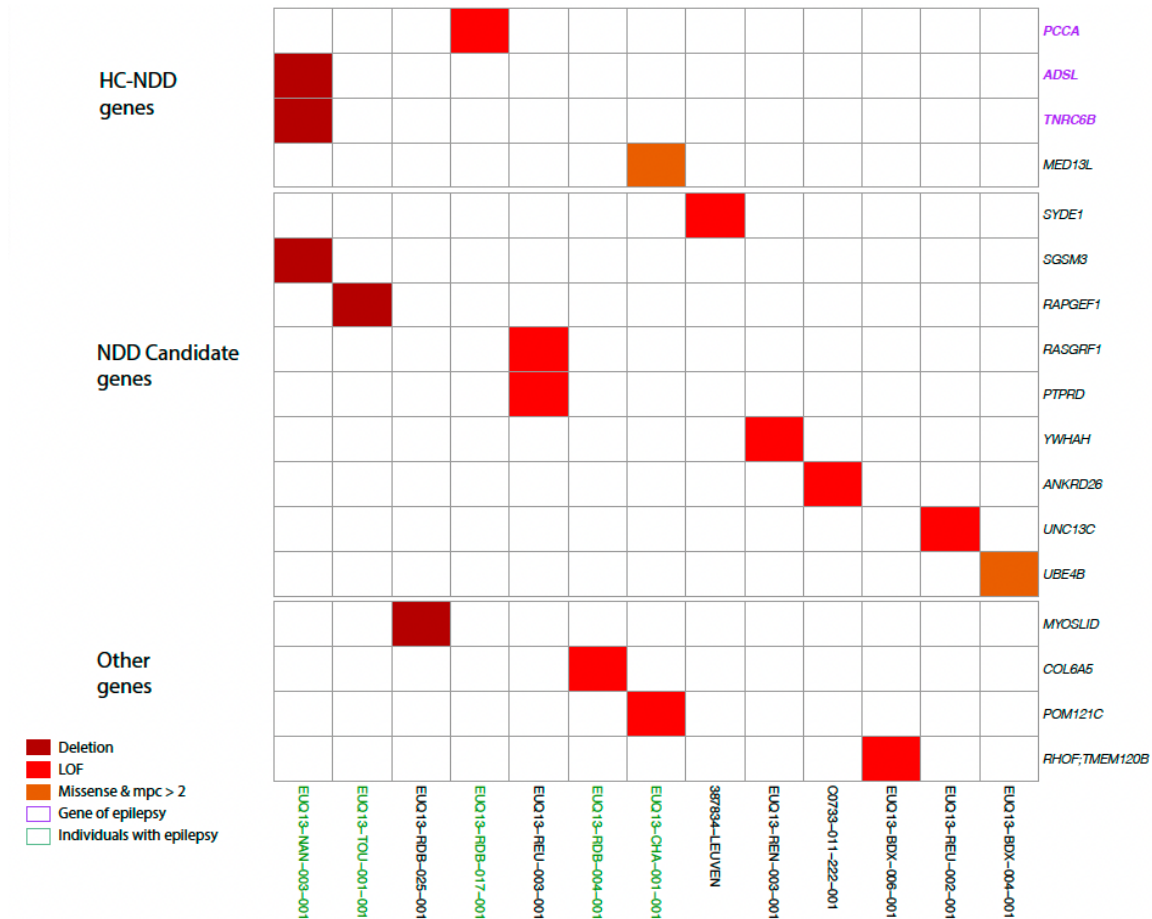


**Figure 40.** Multiple hits des patients PMS retrouvés dans la région 22q13 sur le chromosome controlatéral

Figure réalisée par A. Vitrac, Institut Pasteur

Nous nous sommes ensuite intéressés aux variations génétiques *de novo* délétères. Pour cette analyse nous avons sélectionné les délétions détectées par ClinSV, les variations de type perte de fonction ainsi que les variations faux sens prédites délétères selon les mêmes critères que décrit précédemment. Au total, 20% (13/64) des patients avec PMS sont porteurs de variations génétiques *de novo* additionnelles (Figure 41). Il est intéressant de noter que parmi 5 patients sujets à l'épilepsie, nous avons retrouvé des variations *de novo* et délétères affectant des gènes appartenant à la liste d'épilepsie d'HPO (<https://hpo.jax.org/app/>) dans 2 cas (EUQ13-NAN-003-001 et EUQ13-RDB-017-001) (Figure 41). Un autre de ces patients (EUQ13-CHA-001-001) est porteur d'un variant faux sens dans *MED13L* qui pourrait également être impliqué dans l'épilepsie selon une étude de 2018 (Smol et al., 2020).

Pour compléter cette analyse, nous avons également étudié les variations rares héritées et les variations rares de statut inconnu (dans le cas où nous ne disposons pas des données de SHD des deux parents) touchant les gènes HC NDD dominant et les NDD candidate.



**Figure 41.** Variations *de novo* additionnelles des patients PMS

Heatmap montrant les variations *de novo* additionnelles retrouvées dans le génome des patients PMS. Les délétions, LoF et variations faux-sens prédites délétères sont représentées respectivement en rouge foncé, rouge, et orange. Cinq patients sujets à l'épilepsie sont annotés en vert. Les gènes appartenant à la liste des gènes à risque pour l'épilepsie selon la classification HPO sont annotés en violet. Les gènes sont classés selon trois catégories HC NDD, NDD Candidate et ther (de haut en bas) *Figure réalisée par A. Vitrac, Institut Pasteur*

Pour cette analyse, seuls les variants passant les filtres de LOFTEE (Loss-Of-Function Transcript Effect Estimator) ont été retenus. Nous avons observé que 52% (n=47) des patients avec PMS portaient en plus du variant *SHANK3* identifié en tant que premier hit, des seconds hits dans un ou plusieurs gènes HC NDD dominant. Parmi 71 variants identifiés, 52 sont hérités d'un des deux parents et 16 sont de statut inconnu. Les variations rares héritées et de statut inconnu dans les gènes NDD candidats sont en cours d'analyse.

Les résultats préliminaires de ce projet suggèrent qu'au moins une partie de l'hétérogénéité clinique pourrait être expliquée par la présence de seconds hits.

# Chapitre 3 : Étude des familles de gènes des contactines

## 1. Objectifs

Les gènes des familles des contactines et des contactines *associated proteins* jouent un rôle essentiel dans le développement du système nerveux central et la myélinisation par leur rôle de protéines d'adhésion neurale. De fait, leur implication dans les TND a été étudiée avec des associations positives notamment avec des TSA ou des troubles neurosensoriels chez ces patients. D'autres études concluent au contraire à une absence d'association entre des variations de ces gènes et l'apparition de TND.

L'objectif de ce projet est de **préciser l'implication de CNVs touchant des gènes de la famille des contactines ou des contactines *associated proteins* dans l'apparition de troubles du neurodéveloppement**, à travers l'étude d'une cohorte rétrospective de notre laboratoire recueillant plus de 12 000 individus et l'étude de patients autistes inclus dans des cohortes Américaines (SPARK et SSC) pour lesquels les données d'ACPA sont disponibles dans mon laboratoire d'accueil.



## 2. Méthodes d'analyse de la cohorte de puce à ADN

### 2.1. Cohorte clinique française NDD

#### Cohorte clinique française NDD

Nous avons travaillé sur une cohorte rétrospective qui regroupe des données d'ACPA réalisées entre 2013 et 2017 chez des patients issus de 4 hôpitaux de l'assistance publique – hôpitaux de Paris (Trousseau, Jean Verdier, Robert Debré et Pitié Salpêtrière) dans un but diagnostique. Un consentement a été recueilli pour chaque patient ou son représentant légal précisant qu'ils ne s'opposaient pas à l'utilisation de leurs données dans le cadre de la recherche. Au total **12 013** individus ont ainsi été recrutés.

Les individus ont été classés en 7 groupes phénotypiques, d'après la catégorie déclarée à l'Agence de Biomédecine (ABM) jusqu'en 2017, que nous avons réadapté (table 6) : **DI syndromiques (DI Sd)**, **Malformation congénitale (MCA)**, **DI isolée (DI Iso)**, **TSA**, **Autre**, **Apparentés** et **Contrôles (Ctrls)**.

Codes ABM		Codes adaptés à notre étude	
1	Déficiência intellectuelle syndromique	1	Déficiência intellectuelle syndromique
2	Malformations sans retard psychomoteur	2	Malformations sans retard psychomoteur
3	Déficiência intellectuelle isolée	3	Déficiência intellectuelle isolée
4	Troubles envahissants du développement / Autisme / Épilepsie	4	Troubles envahissants du développement / Autisme
5	Caractérisation d'une anomalie découverte par caryotype ou par une autre technique	5	Autres (indication non 1, 2, 3 ou 4 : endocrinologie, fœtus, neurologie,..)
6	Fœtopathologie	6	Apparentés non 1, 2, 3 ou 4
7	Études familiales	7	Contrôles
8	Autres		

**Table 6.** Groupe phénotypique pour la cohorte

La catégorie **Autre** de notre étude regroupe tous les individus ne correspondant pas aux 6 autres catégories. Ce sont essentiellement des patients avec des pathologies endocrinologiques (troubles de la croissance, anomalies du développement sexuel, troubles de la fertilité), des troubles neurologiques (syndrome cérébelleux,

neuropathie, ...), des syndromes de Beckwith Wiedmann et de Silver-Russel, des myopathies, des surdités ou des fœtus.

Les individus Contrôles sont issus d'une étude sur les cardiomyopathies et n'ont pas été diagnostiqués pour un TND (Proust et al., 2015).

Comme décrit dans la littérature, nous observons une majorité d'hommes dans les phénotypes associés à un TND avec au maximum un sexe ratio de 4.52 pour les patients présentant un TSA (table 7).

	<b>Masculin</b>	<b>Féminin</b>	<b>Sex Ratio</b>
DI Sd	1813	1105	1,64
MCA	554	408	1,36
DI Iso	666	361	1,84
TSA	1529	338	4,52
Autres	651	867	0,75
Apparentés	956	1082	0,88
Ctrls	1200	483	2,48
<b>Total</b>	<b>7369</b>	<b>4644</b>	<b>1,59</b>

**Table 7.** Répartition homme/femme par catégorie et sex-ratio

## **2.2. Cohortes américaines d'individus autistes et leurs apparentés**

### **SPARK**

La cohorte de la SPARK est une cohorte américaine multicentrique qui recueille aujourd'hui plus de 100 000 individus avec TSA et leur famille (SPARK Consortium, 2018), (<https://www.sfari.org/resource/spark/>).

L'ADN a été extrait à partir de prélèvement salivaire.

### **SCC**

La cohorte SCC (*Simons Simplex Collection*) se compose de près de 2 500 familles avec un patient atteint de TSA, les deux parents et dans la majorité des cas un frère/sœur non atteint (quatuors) (Fischbach & Lord, 2010). Les patients ont été inclus aux États-Unis et le diagnostic de TSA posé suite aux évaluations de l'ADOS et de l'ADI-R. Les patients de moins de 4 ans ou plus de 18 ans ont été exclus, ainsi que ceux dont la sévérité ne permettait pas d'appliquer les outils de diagnostic d'autisme (déficit neurologique sévère, diagnostic génétique de trisomie 21 ou X-fragile). L'ADN a été extrait à partir d'un prélèvement sanguin.

Les données de ces deux cohortes sont disponibles au laboratoire de Génétique Humaine et Fonctions Cognitives.

### 2.3. ACPA et analyse des CNVs

Les ACPA de notre cohorte ont été réalisées à la plateforme P3S de la Pitié Salpêtrière par la technique de puce SNP Illumina avec 700 000 sondes (Illumina OmniExpress, Illumina Inc., San Diego, California) ou pour les plus anciens des puces SNP Human CytoSNP 12 comportant près de 300 000 sondes (Illumina).

Les données phénotypiques de chaque individu ont été contrôlées pendant mon projet de Master 2. L'application des critères de qualité et d'exclusion des échantillons (exclusion des duplicatas, concordance des données démographiques et des données génétiques telles que le sexe et les liens de parenté, qualité technique) ont également fait l'objet de mon travail de M2. Les critères de qualité techniques retenus étaient les suivants :

- Le **Call rate  $\geq 95\%$**  qui est un indice de la qualité du génotypage. Le Call rate est la fraction des SNPs qui ont bien été génotypés chez un individu.

- Deux paramètres quantitatifs : intensité de fluorescence (Log R Ratio, **LRR**) et valeur de la Fréquence de l'Allèle B ou **BAF**. Les valeurs retenues pour la qualité des CNVs sont une déviation standard de la **LRR  $< 0,35$**  et de la **BAF  $< 0,08$** . La qualité de la LRR et de la BAF est essentielle pour bien identifier les CNVs. Si les valeurs sont trop importantes (i.e. que les points sont trop éparpillés) il y a un risque trop important d'identifier des faux négatifs et des faux positifs.

- La valeur absolue du wave factor **|WF|  $< 0,05$** . La *wave factor* est une mesure complémentaire de la LRR, qui donne une moyenne des points de la LRR et représente sa déviation par rapport à la moyenne.

Les CNVs ont été déterminés à partir de deux algorithmes : QuantiSNP et PennCNV (Colella et al., 2007; Wang et al., 2007), qui marquent un CNV si au moins 3 sondes consécutives sont déviées dans le même sens et si la taille est  $\geq 1$  kb, avec un score de confiance  $> 15$  (comme décrit dans les matériels de (Huguet et al., 2018)). Les résultats des deux analyses sont fusionnés dans un fichier par CNVision (Sanders et al., 2011) (<http://www.cnvision.org/>) : pour chaque CNV, il est précisé s'il a été détecté par l'un ou les deux algorithmes et le cas échéant le pourcentage de chevauchement entre les deux. CIA (CNV *inheritance analysis*) est un programme mis au point par le laboratoire GHFC du Pr. Thomas Bourgeron qui va permettre de

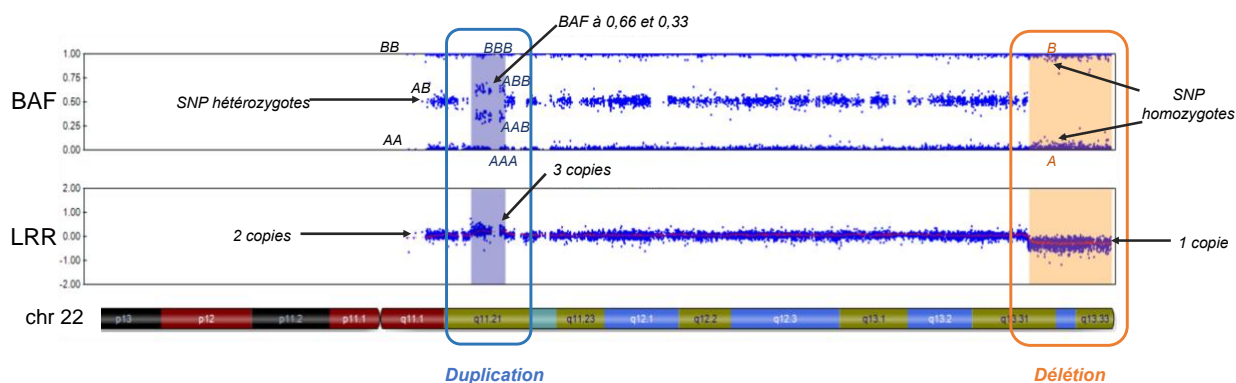
concaténer les CNVs a priori mal séparés selon certaines conditions : même variation (gain ou perte) < 500 kb, avec au moins 100 SNPs, séparés par un espace < 150 kb. L'annotation des CNV est ensuite réalisée grâce à AnnoVar (Wang et al., 2010).

Deux CNVs sont considérés identiques s'il y a un chevauchement réciproque d'au moins 80%. Nous avons éliminé les CNVs qui chevauchent à plus de 50% les duplications segmentaires.

Les coordonnées des CNVs sont fournies dans la version **GRCh37** du génome.

L'ensemble des CNVs touchant au moins un gène de la famille des contactines ou des contactines *associated proteins* ont ensuite été validés visuellement sur les profils de puce.

Les CNVs d'intérêt de la cohorte française ont tous été vérifiés visuellement sur les profils de puce. La LRR donne l'information de la quantité d'ADN présente et la BAF illustre le nombre d'allèles B, c'est à dire de SNP correspondant à l'allèle de référence dénommé B. On peut donc connaître le statut hétérozygote, homozygote ou plus complexe pour chaque sonde (figure 42). Un segment dupliqué aura ainsi une LRR augmentée et une BAF correspondant à des fréquences pour trois allèles : 0, 0,33, 0,66 ou 1.



**Figure 42.** Analyse d'un profil de puce SNP (Genome Studio)

## 2.4. Analyses statistiques

Les analyses statistiques et figures ont été réalisées grâce au logiciel R Studio (<http://www.rstudio.com/>). Les Odds Ratio (OR), intervalles de confiance (IC) et la P-value ont été calculés partir d'un test de Fisher.

Une méta-analyse a été réalisée en joignant les données de notre cohorte, la cohorte SPARK et la cohorte SSC. Les projets de puces utilisés pour la SSC sont des puces SNP Illumina Human1M-Duov3, Human1Mv1 ou HumanOmni2.5, tandis que la SPARK a utilisé des projets Infinium GSA-24 v1.0.

Pour l'analyse nous avons conservé pour chaque famille un probant autiste et un frère ou une sœur non autiste. Concernant notre cohorte de patients, nous avons conservé seulement des individus indépendants avec un TND c'est-à-dire les patients avec DI syndromique ou isolée et TSA.

Pour homogénéiser les 3 cohortes au maximum, nous avons conservé les CNVs correspondant aux critères suivants :

- taille > 50 kb
- au moins 20 sondes consécutives
- score de confiance > 30

Nous avons utilisé le package Metabin sur R avec un modèle à effet aléatoire (<https://cran.r-project.org/web/packages/meta/meta.pdf>), afin de prendre en compte la diversité de recrutement et d'analyse des différentes cohortes étudiées.

### 3. Les CNVs impactant les gènes des contactines

#### 3.1. Analyse globale des CNVs

Nous avons regardé la couverture de chaque gène sur les différents projets de puce utilisés dans les différentes cohortes (table 8). Le nombre de sondes couvrant les gènes sont proportionnels à leur taille, et nous voyons que pour les plus petits gènes, le nombre de sonde est inférieur à 10 voir nul pour les plus anciens projets de notre cohorte (*CNTNAP3*, *CNTNAP1* et *CNTN2*, puce CytoSNP pour n = 2957).

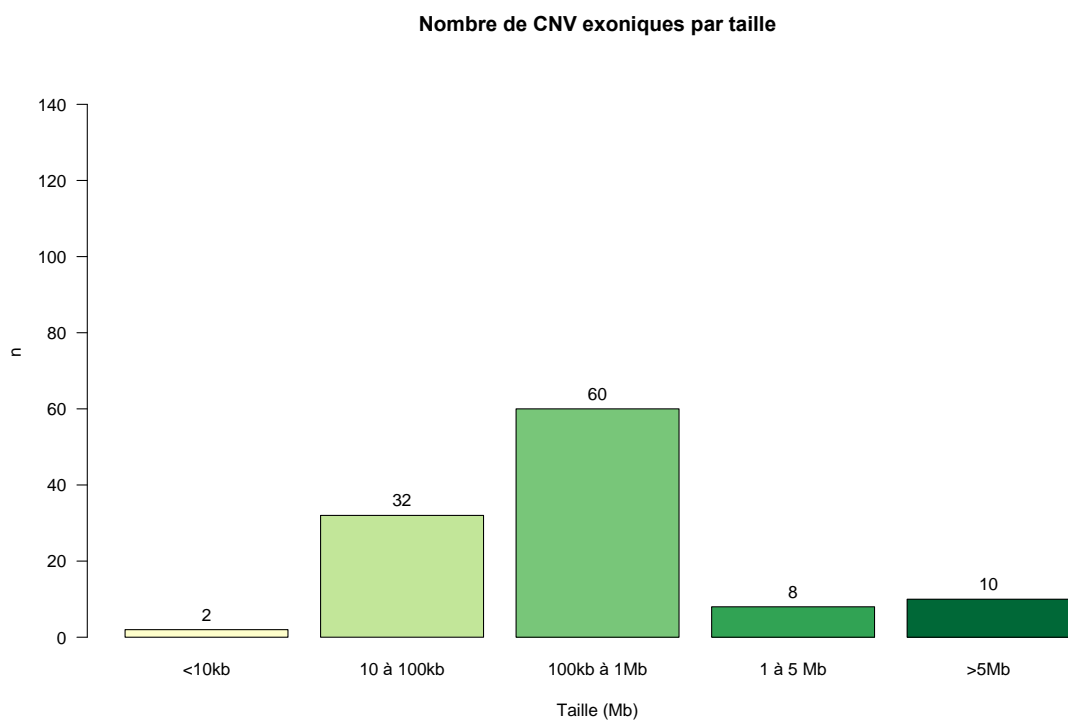
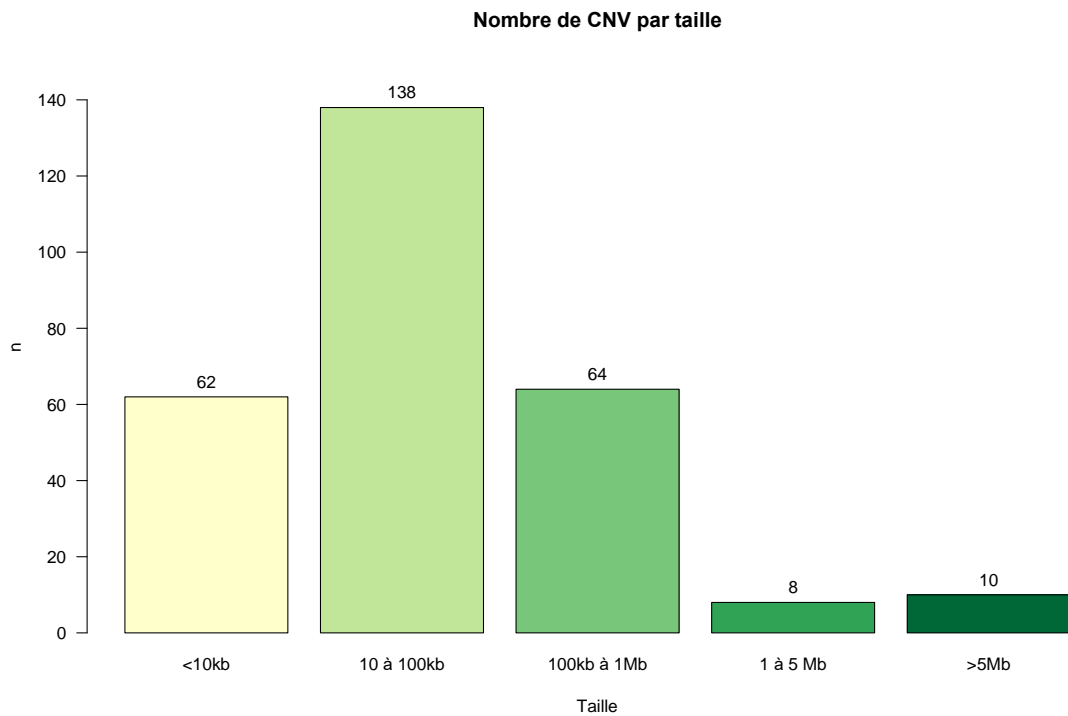
	SSC			SPARK	Notre cohorte	
	Human 1Mv1	Human 1M-Duov3	Human Omni2.5	Infinium GSA-24 v1.0	Human CytoSNP-12v2	Human OmniExpress-24
CNTN1	154	169	313	73	64	93
CNTN2	32	34	59	22	7	25
CNTN3	90	94	255	88	22	71
CNTN4	620	663	1462	361	157	463
CNTN5	711	746	1479	310	196	550
CNTN6	231	246	526	141	60	188
CNTNAP1	15	16	14	6	3	6
CNTNAP2	860	935	2160	573	192	605
CNTNAP3	37	36	36	11	0	5
<i>CNTNAP3B</i>	13	19	4	3	0	0
CNTNAP4	131	144	366	74	21	91
CNTNAP5	296	326	792	191	87	231

**Table 8.** Nombre de sondes couvrant les différents gènes sur les différents projets de puce utilisés dans les trois cohortes

Nous n'avons pas pris en considération le gène *CNTNAP3B* car non couvert par les deux technologies utilisées dans notre cohorte, ni *CNTNAP3C* qui n'apparaît qu'avec la version du génome GRCh38.

Les deux parents étaient disponibles pour 659 patients, la mère seule dans 228 cas et le père seul dans un cas. Nous disposons des parents pour 9 des patients porteurs d'un CNV d'intérêt. Lorsque les deux parents étaient disponibles (n = 8), 100% des CNVs étaient hérités (5 paternels, 3 maternels). Pour le 9<sup>ème</sup> patient chez qui la mère seule était présente, le CNV n'était pas d'origine maternelle.

La taille des CNVs varie de 1,3 kb à 37 Mb, avec une médiane à 41,6 kb (figure 41). La majorité des CNVs identifiés ont une taille comprise entre 100 kb et 1Mb.

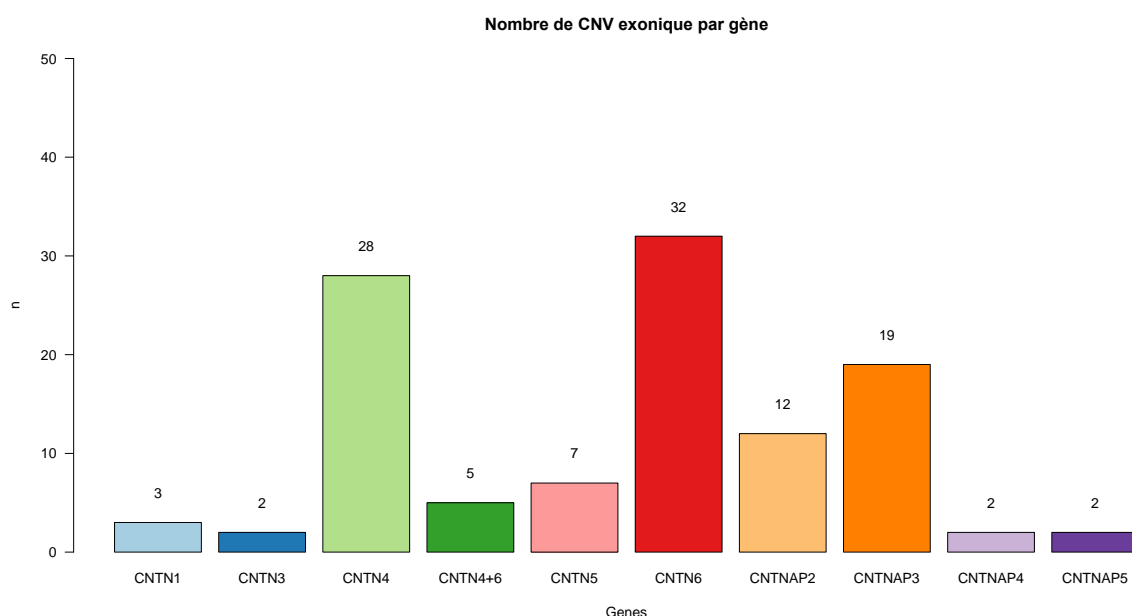


**Figure 41.** Répartition des tailles des CNVs identifiés  
Haut : tout CNVs confondus. Bas : CNVs exoniques seulement

Ainsi nous observons un excès de CNVs entre 10 et 100 kb qui sont introniques puisque non retrouvés après le filtre sur les CNVs exoniques. Cet excès correspond

très probablement à deux des trois CNVs récurrents que nous avons identifié dans la cohorte. Le premier de 41 kb touche l'intron 3 de *CNTN5* (chr11:99525126-99566782), et est récurrent dans la base de donnée DVG, nous le considérons donc comme un polymorphisme. Le deuxième concerne l'intron 3 du gène *CNTNAP2* (chr7:146571183-146586456) et varie entre 13 et 20 kb chez nos patients. Cette délétion n'est pas présente dans la base de données DGV mais rapportée dans la base de données gnomAD. Le troisième CNV récurrent est localisé dans l'intron 1 de *CNTN5* (chr11:99086508-99096479, de 9,9kb) et n'est pas rapporté dans la base de données DGV (des délétions plus grandes le sont en revanche, englobant ce CNV). Ils sont donc à considérer comme des CNVs probablement bénins.

Nous avons analysé la répartition des CNVs par gène étudié (figure 43). Aucun CNV ne touche les gènes *CNTN2* ni *CNTNAP1*. Ce sont les deux plus petits gènes de ces deux familles, faisant respectivement 35,3 et 17,4 kb, et les moins bien couverts par les profils de puce (25 et 6 sondes respectivement, table 8), ce qui peut contribuer à l'absence de CNVs identifiés.



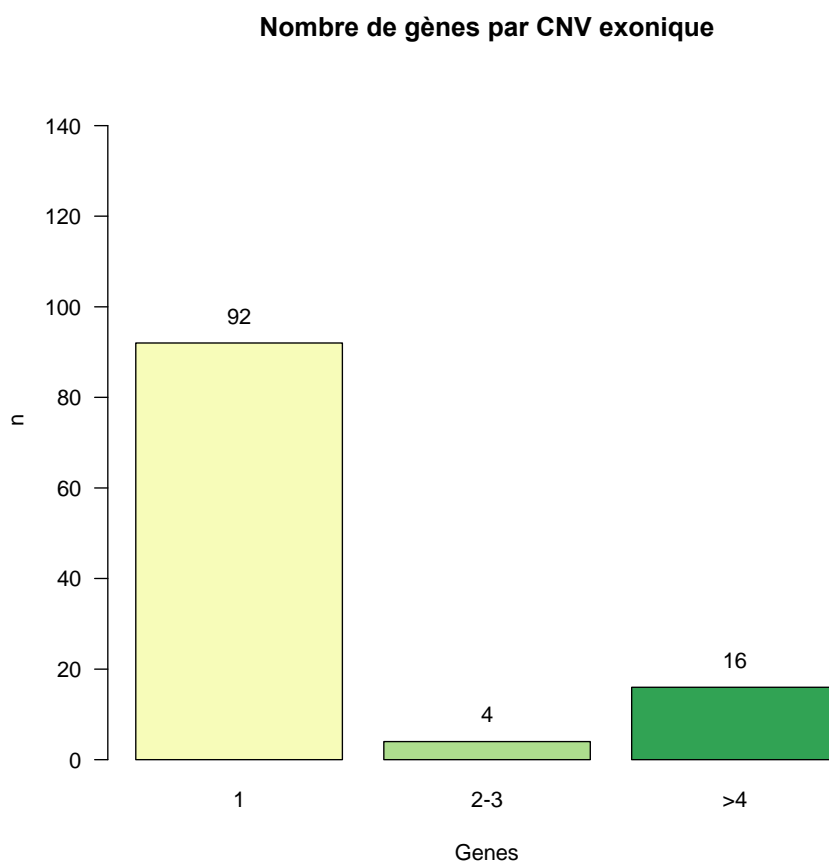
**Figure 43.** Répartition des CNV identifiés selon le gène

Nous observons un excès de CNVs dans la région intronique de *CNTN5* comme cité plus haut (86 individus). Il s'agit du deuxième plus grand gène de ces familles, d'une taille de 1 337,9 kb, juste après le gène *CNTN2AP* (2304,6 kb).

Ils ne sont plus que 7 patients porteurs d'un CNV exonique de *CNTN5*.



Le nombre de gènes impliqués dans les CNVs identifiés varie de 1, pour la plus grande majorité des cas à plus de 4 pour 16 individus (figure 44). La majorité des CNVs impliquant une contactine ou une contactine *associated protein* ne touche donc pas d'autre gènes dans cette cohorte.



**Figure 44.** Nombre de gènes impliqués dans les CNVs identifiés.

Les 16 CNVs impliquant au moins 4 gènes ont une taille allant de 4,7 Mb (6 gènes) à 16,9 Mb (66 gènes) pour les délétions (7 patients) et de 197 kb (4 gènes) à 37,5 Mb (194 gènes) pour les duplications. Ces patients sont atteints de DI syndromique (9 cas) ou isolée (2 cas), et 3 cas sont dans le groupe Autre.

Les 7 délétions sont classées pathogènes, en raison de la perte d'autres gènes que les contactines. Quatre d'entre-elles sont des délétions terminales englobant la région 3p26, la plus petite bornant la région chr3:61495-4818792. Elles délètent d'autres gènes haploinsuffisants, dont *ITPR1* (OMIM \*147265) responsable du syndrome de Gillespie (OMIM #206700) associant aniridie, ataxie cérébelleuse et DI dans une forme autosomique dominante. Un patient du groupe « Autre » porte une délétion en chr7:142913569-149639889, qui touchent *CNTNAP2* et délète 3 gènes impliqués dans des pathologies dominantes, dont *EZH2* (OMIM \*601573) responsable

d'un syndrome de Weaver associant une forme de gigantisme avec retard psychomoteur (OMIM #277590). Un autre patient avec une DI syndromique porte une délétion de 7,4 Mb en chr11:97164437-104591614 qui implique *CNTN5* mais aussi *YAP1* (OMIM \*606608) responsable d'une dysmorphie caractérisée par un colobome uvéal parfois associé à des fentes labiales ou palatines, un déficit auditif ou une DI. La septième délétion, interstitielle chez un patient avec une DI syndromique, en chr2:115472604-132380909 délète, en plus de *CNTNAP5* le gène *GLI2* (OMIM \*165230) associé au syndrome de Culler-Jones (OMIM #615849), des anomalies antéhypophysaires avec polydactylie et dysmorphie, et à une holoprosencéphalie de type 9 dans des formes plus sévères.

La plus petite duplication est portée par un apparenté et implique en plus de *CNTN4* les gènes *CRBN*, *TRNT1* et *IL5RA* (chr3:3000908-3198513). Elle n'est pas rapportée comme un polymorphisme et serait classée comme un VSI.

Quatre patients portent des duplications en 9p, deux proximales de 863 kb et 1,7 Mb, péricentromériques, chez deux patients avec une DI syndromiques et interprétées comme VSI. Les deux autres beaucoup plus grandes sont en rapport avec une trisomie et une tétrasomie 9p respectivement. Un patient est porteur d'une trisomie 12, et le dernier d'une trisomie 7q35qter.

Ainsi pour ces patients, le remaniement en lui-même et les autres gènes impliqués seront responsables du phénotype et le poids des contactines sera sans doute faible pour la plupart de ces 16 patients.

### 3.2. Analyse des données en fonction du phénotype

Après les contrôles qualités décrits dans la [section 3/2.3](#), un total de **12 013** individus a été inclus dans l'étude de cette cohorte avec un résultat d'ACPA, comprenant 5 812 patients avec un trouble du neurodéveloppement (c'est à dire les patients avec DI syndromique, DI isolée et TSA), 1027 avec malformations congénitales multiples sans DI (MCA), 1 518 avec une autre pathologie, 2 038 apparentés et 1 683 individus contrôles (table 9).

Un total de 282 CNVs impliquant une contactine ou contactine *associated-protein* a été identifié (189 délétions, 93 duplications). Lorsque les CNVs introniques sont exclus, les chiffres sont de 43 délétions et 69 duplications.

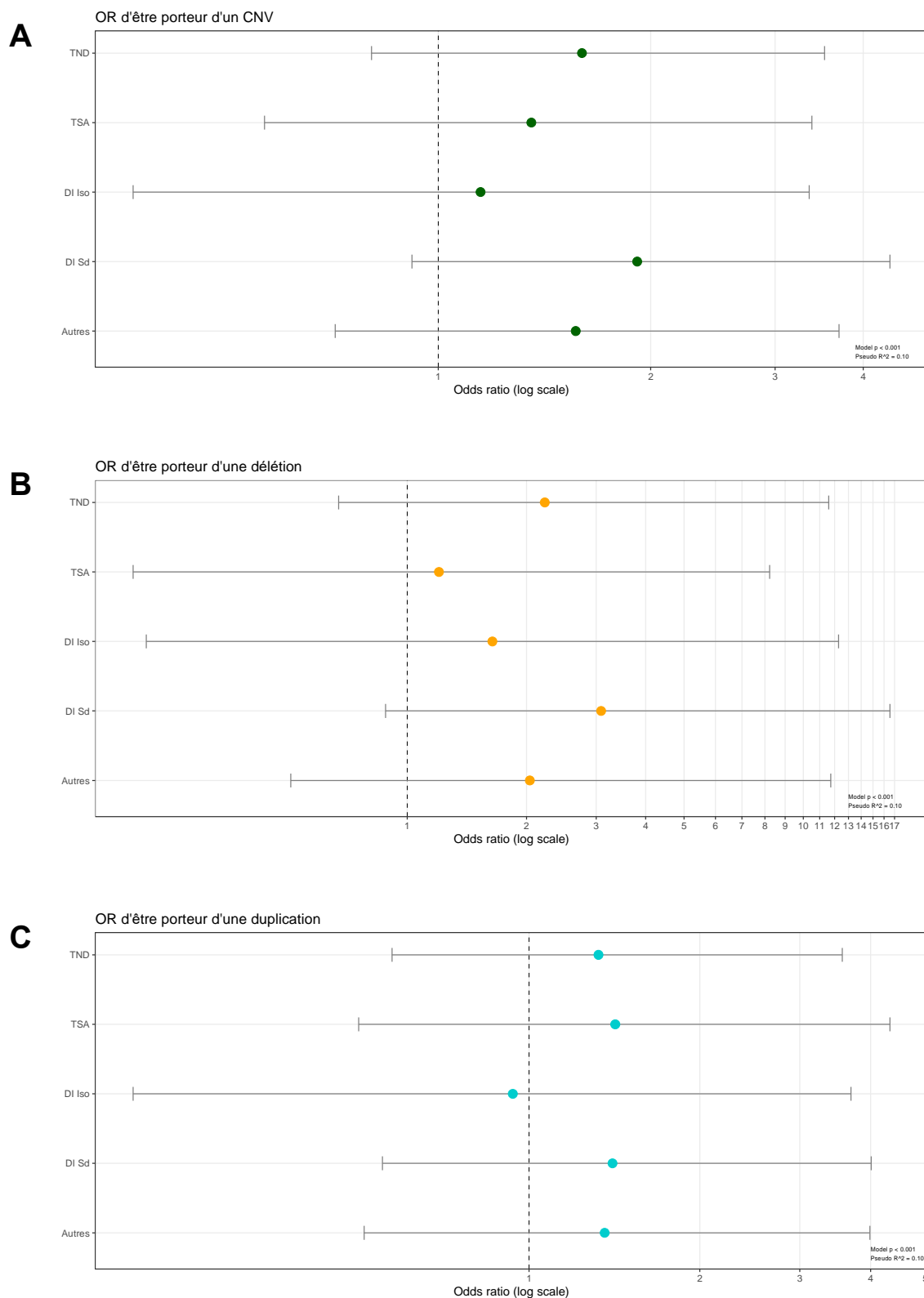
Phénotype	DI Syndromique		DI Isolée		TSA		MCA		Autres		Apparentés		Contrôles	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
<b>n</b>	2918		1027		1867		962		1518		2038		1683	
<b>Nb CNVs</b>	<b>73</b>	<b>2,5%</b>	<b>22</b>	<b>2,1%</b>	<b>48</b>	<b>2,6%</b>	<b>19</b>	<b>2,0%</b>	<b>34</b>	<b>2,2%</b>	<b>50</b>	<b>2,5%</b>	<b>36</b>	<b>2,1%</b>
<b>Délétion</b>	51	1,7%	16	1,6%	33	1,8%	13	1,4%	23	1,5%	30	1,5%	23	1,4%
<b>exonique</b>	<b>16</b>	<b>0,5%</b>	<b>3</b>	<b>0,3%</b>	<b>4</b>	<b>0,2%</b>	<b>2</b>	<b>0,2%</b>	<b>7</b>	<b>0,5%</b>	<b>8</b>	<b>0,4%</b>	<b>3</b>	<b>0,2%</b>
<b>OR (VS contrôles)</b>	3,09		1,64		1,2		1,17		2,59		2,21		-	
<b>P-value</b>	0,09		0,68		1		1		0,21		0,36		-	
<b>Duplication</b>	22	0,67	6	0,49	15	0,92	6	0,73	11	0,74	20	1	13	0,77
<b>exonique</b>	<b>17</b>	<b>0,54</b>	<b>4</b>	<b>0,29</b>	<b>11</b>	<b>0,71</b>	<b>4</b>	<b>0,42</b>	<b>10</b>	<b>0,68</b>	<b>16</b>	<b>0,8</b>	<b>7</b>	<b>0,42</b>
<b>OR (VS contrôles)</b>	1,4		0,93		1,42		1		1,59		1,89		-	
<b>P-value</b>	0,52		1		0,49		1		0,47		0,21		-	

**Table 9.** Répartition des CNVs selon le type et le groupe phénotypique

Il n'y a pas de risque significatif d'être porteurs d'un CNV touchant une contactine parmi les différents groupes comparés aux contrôles, aussi bien en total que pour les délétions ou les duplications étudiées séparément (table 9).

Nous avons calculé le risque d'être porteur d'un CNV parmi les patients avec un TND, c'est à dire en regroupant les groupe DI syndromique, DI isolée et TSA (n = 55 porteurs d'un CNV sur 5 812 patients) (figure 45.A) qui ne retrouve pas de différence significative par rapport aux contrôles d'un point de vue global (Odd Ratio OR = 1.60, IC95 : 0.80-3.52, p = 0,23). Il n'y a pas non plus de risque quand nous nous intéressons

aux délétions seules (OR = 2,22,  $p = 0,24$ ) (figure 45.B) ou aux duplications seules (OR = 1,33,  $p = 0,57$ ) (figure 45.C).

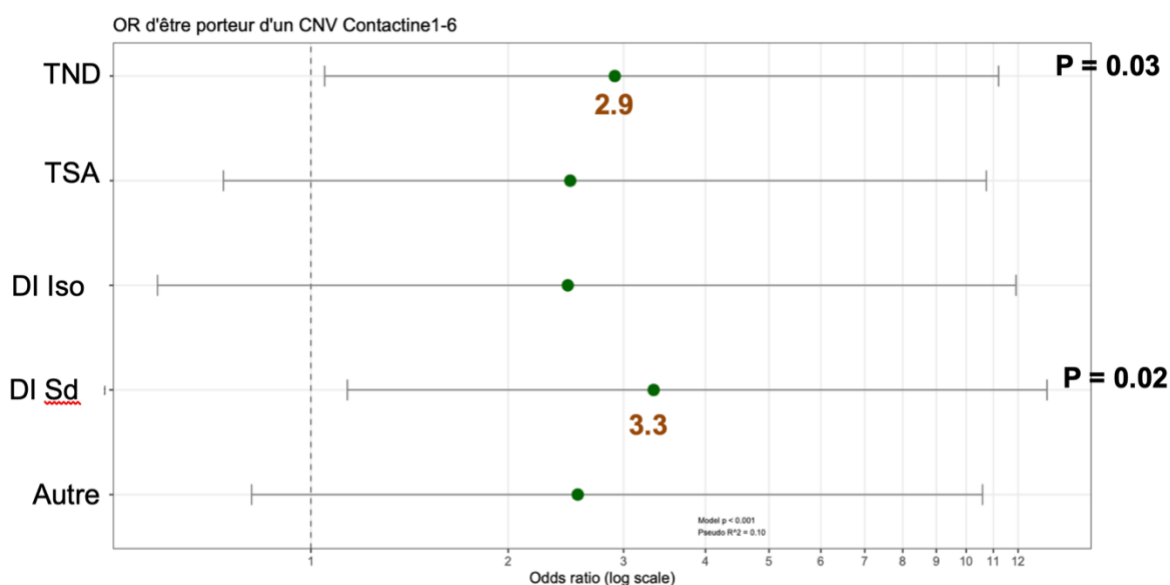


**Figure 45.** Odd Ratio d'être porteur d'un CNV selon les groupes phénotypiques  
**A.** Tous types de CNVs étudiés **B.** Cas des délétions seules **C.** Cas des duplications seules.  
 « Autres » comprend les patients non atteints de TND (MCA + Autre). Le groupe « TND » comprend les patients des groupes TSA + DI isolée + DI syndromique.

Pour aucun groupe, que ce soit de façon globale ou par type de CNV, le risque d'être porteur d'un CNV touchant une contactine n'est pas significatif par rapport aux contrôles. Bien que non significatif, nous observons tout de même une tendance qui ressort pour les délétions chez le groupe TND et les patients avec DI syndromique (OR 3.09, IC95 : 0.88-16.55,  $p = 0.09$ ) (figure 45.B).

Nous ne trouvons pas d'association significative en excluant les 16 patients porteurs d'un remaniement contenant plus de 4 gènes discutés précédemment.

Les analyses ont été ensuite réalisées en se concentrant d'une part sur les CNVs touchants les gènes des *CNTN1* à *CNTN6*, et d'autre part les gènes *CNTNAP1* à *CNTNAP5*. Nous trouvons une association significative pour les groupes TND (OR 2,9 ;  $p=0,03$ ) et déficience syndromique (OR 3,3 ;  $p=0,02$ ) porteurs d'un CNV touchant une contactine 1 à 6 (figure 46).



**Figure 46.** Odd Ratio d'être porteur d'un CNV touchant un gènes des contactines 1 à 6 selon les groupes phénotypiques

« *Autres* » comprend les patients non atteints de TND (MCA + Autre). Le groupe « *TND* » comprend les patients des groupes TSA + DI isolée + DI syndromique.

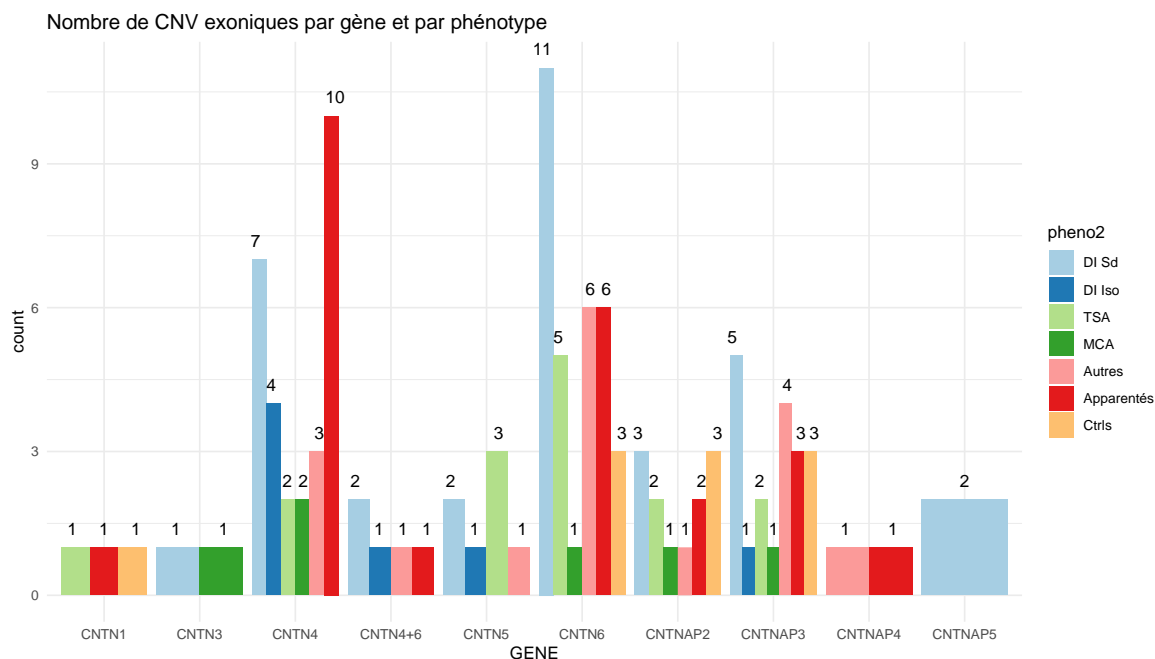
L'analyse ne trouve pas d'association significative pour les patients porteurs d'un CNVs dans les gènes *CNTNAP1* à *CNTNAP5*.

En rapportant les gènes au phénotype, nous observons qu'aucun apparenté ou contrôle n'est porteur d'un CNV d'une région codante des gènes *CNTN3*, *CNTN5* ou *CNTNAP5* (figure 47 et 48).

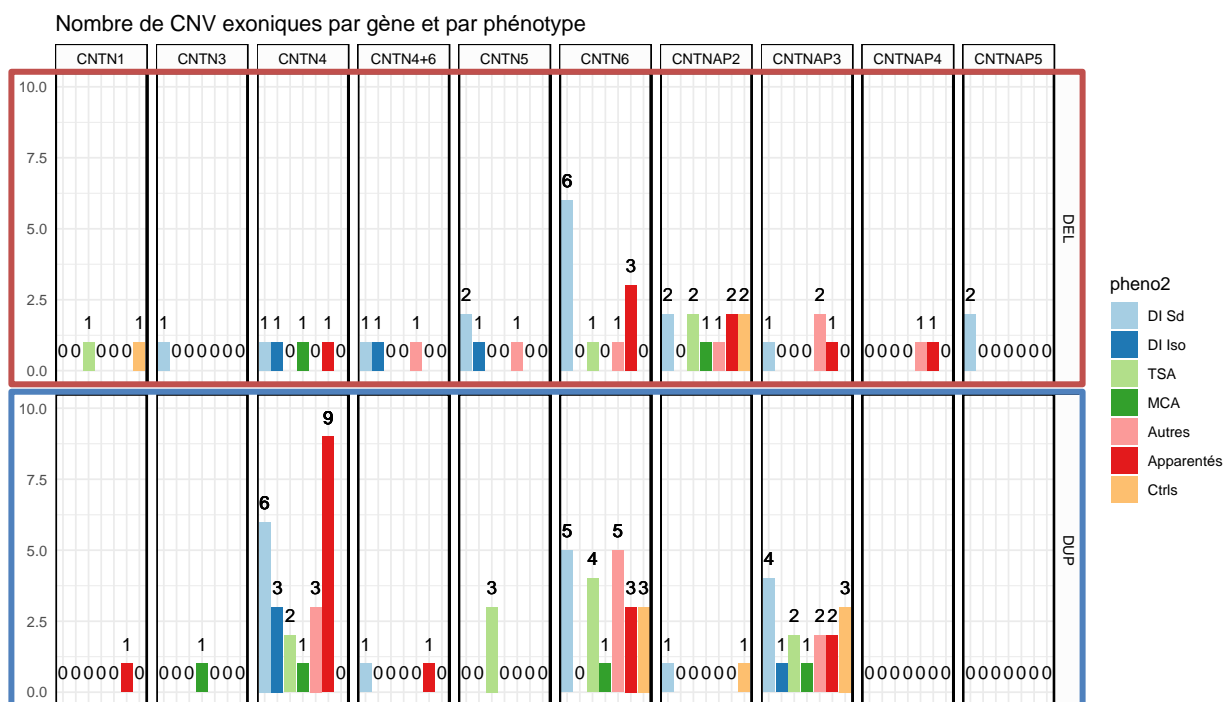
Cinq individus (4 patients et un apparenté) sont porteurs d'une délétion ( $n = 3$ ) ou d'une duplication ( $n = 2$ ) de la région 3p26.3 impliquant les deux gènes *CNTN4* et *CNTN6*

(figures 47 et 48). Les trois délétions sont terminales et font 4,7 Mb pour un patient du groupe MCA, 5,9 Mb pour un patient du groupe DI isolée, et 10,6 Mb pour un patient du groupe « Autre ». Les deux duplications font 61 kb (interstitielle) pour un parent, et 1,3 Mb (duplication terminale) chez un patient avec DI syndromique.

Dans le groupe contrôle, des délétions sont identifiées seulement sur la *CNTN1* (1 cas) et *CNTNAP3* (2 cas).



**Figure 47.** Nombre de CNVs identifiés par gène et selon le phénotype



**Figure 48.** Nombre de délétions identifiées par gène et selon le phénotype  
Encadré orange : délétions ; Encadré bleu : duplications

Les délétions les plus fréquentes touchent le gène **CNTN6** et sont représentées par une majorité de patients avec DI syndromique (0,2% des DI syndromiques) et les apparentés (0,1%). Mais comme décrit précédemment, 4 de ces délétions sont terminales et de grande taille.

Les duplications impliquant **CNTN4** et **CNTN6** sont, elles, les plus fréquentes (figure 48). Deux tiers des duplications CNTN6 concernent des patients avec déficience syndromique tandis que un tiers touchent des apparentés. Les duplications de **CNTN4** seule sont également retrouvées pour un tiers chez des apparentés (0,44% des apparentés). Des duplications de ces deux gènes sont décrites dans les bases de données DGV et gnomAD et sont également plus fréquemment retrouvées chez des patients avec un TSA, ce qui a amené à les considérer au moins comme un facteur de susceptibilité aux TND (Guo et al., 2017).

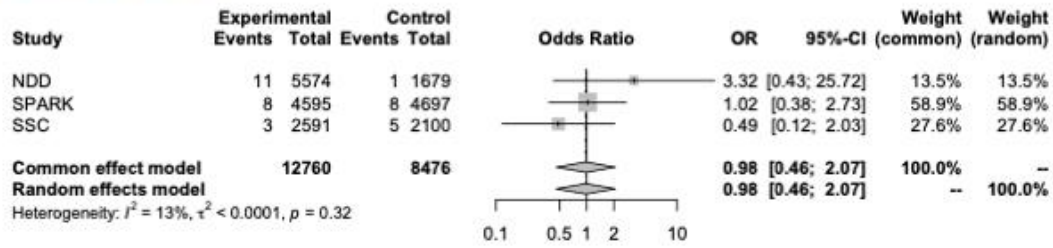
On observe également des duplications de **CNTNAP3** dans chacun des groupes. Des duplications de **CNTNAP3** sont décrites dans la base de données DGV et considérées comme un polymorphisme. **CNTNAP3** est localisée sur le bras court du chromosome 9 en région péricentromérique, ce qui explique sa faible couverture par les projets de puce, et d'autre part ce polymorphisme.

### 3.3. Méta-analyse

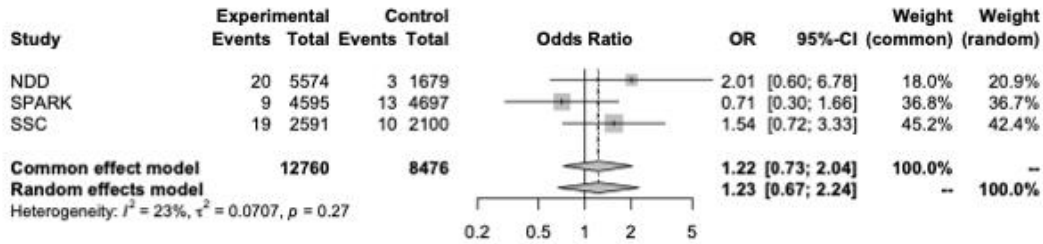
Nous avons réalisé une méta-analyse avec les patients de notre cohorte, la cohorte SPARK (4595 probants et 4697 apparentés non atteints) et la cohorte SSC (2591 probants et 2100 apparentés non atteints). Après la sélection d'individus indépendants pour notre cohorte, nous comptons 5574 patients avec un TND et 1679 contrôles, parmi lesquels 146 sont porteurs d'un CNV répondants aux critères stricts de qualité.

Aucune association significative avec des TND n'est mise en évidence, que ce soit pour les délétions ou duplications des contactines (figure 49), les délétions ou duplications des contactines *associated proteins* (figure 50), ou pour les délétions ou duplications de l'ensemble des gènes (figure 51).

### Délétions *CNTN1-6*

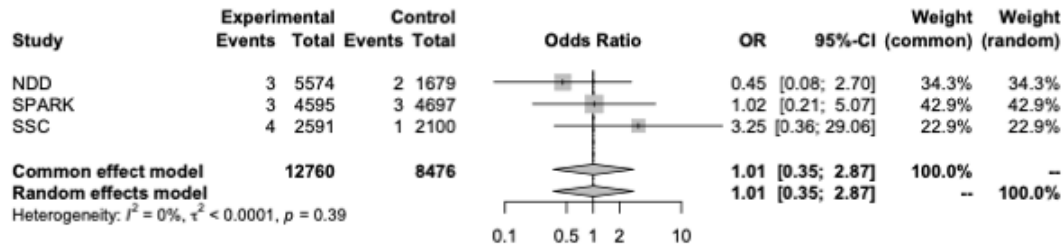


### Duplication *CNTN1-6*

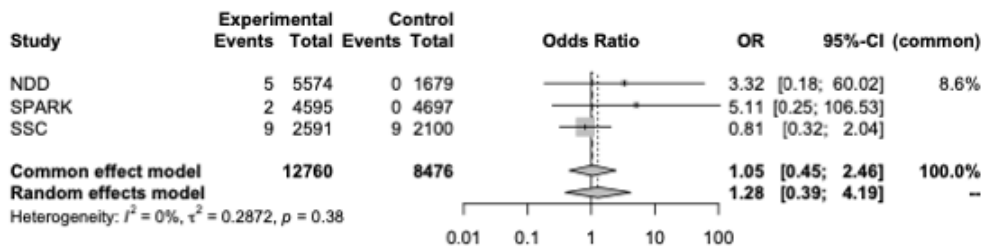


**Figure 49.** Méta-analyse des délétions et duplications des deux familles de gènes pour notre cohorte, la SPARK et la SSC

### Délétions *CNTNAPI-5*



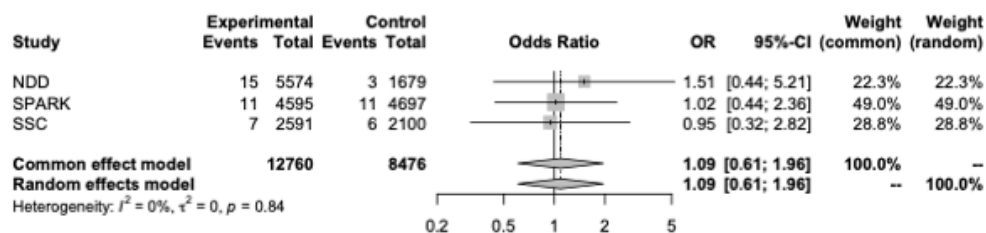
### Duplication *CNTNAPI-5*



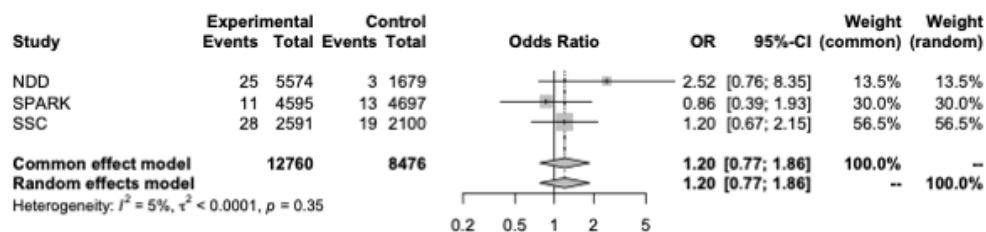
**Figure 50.** Méta-analyse des délétions et duplications des gènes Contactines pour notre cohorte, la SPARK et la SSC



## Délétions



## Duplication



**Figure 51.** Méta-analyse des délétions et duplications des gènes *Contactin associated proteins* pour notre cohorte, la SPARK et la SSC

Au final, nos résultats ne permettent pas de conclure à une association significative entre la présence d'un CNVs, qu'il s'agisse de délétion ou de duplication, touchant un gène des familles des contactines et l'apparition d'un TND.

Dans notre cohorte, nous avons identifié 13 patients avec grand CNV en rapport pour certains avec des trisomies partielles, qui vont expliquer le phénotype du patient. Le rôle des contactines dans ces 13 cas est négligeable.

La méta-analyse ne conclut pas non plus à une association significative. Nos résultats restent toutefois difficiles d'interprétation en raison d'un important biais : les individus des cohortes SPARK et SSC ne sont pas indépendants. En effet, si nous avons exclus les parents, le groupe d'individus non atteints correspond aux frères/sœurs des probants. Par ailleurs, le ratio famille simplex (un seul individu atteint) et multiplex est également très différent dans les trois cohortes. Nous n'avons pas non plus matché les cohortes sur l'âge des individus.

## Chapitre 4 : DISCUSSION

Aujourd'hui, le diagnostic génétique d'un trouble du neurodéveloppement passe par la réalisation d'une ACPA puis d'un WES, plus ou moins couplée à une recherche d'X-fragile, pour proposer un diagnostic le plus complet explorant des CNVs et des SNV/indel (Souche et al., 2022). A partir de 2018, le plan France Médecine Génomique 2025 ouvre l'accès au WGS à des patients répondant à des critères de « pré-indications » dont la liste évolue régulièrement, et pourrait se placer un jour en diagnostic de première intention (Nisar et al., 2021; Scocchia et al., 2019; van der Sanden et al., 2023). En effet, à ce jour le WGS permet au minimum d'identifier en une seule technique les variations mises en évidence à la fois par le WES et l'ACPA, avec une meilleure résolution pour l'identification des CNVs. Toutefois, le WGS est limité par le contenu des génomes de référence actuellement utilisés pour l'alignement des séquences, GRCh37 et 38, qui ne couvrent pas certaines régions du génome riches en séquences répétées, comme les bras courts des chromosomes acrocentriques (BCCA).

La récente publication d'une séquence « complète » du génome, *Telome to Telomere* (T2T) (Nurk et al., 2022), apporte une information plus complète, notamment sur la séquence des BCCA. Ainsi nous avons pu cibler une analyse des points de cassure de patients avec un anneau du chromosome 22, dont la mécanique va entraîner la fusion des extrémités des bras courts et bras longs avec dans tous nos cas une perte de matériel génétique au moins en 22qter. L'étude des CRAM a mis en évidence des *reads* qui étaient non alignés sur les références GRCh37 ou 38. Un BLAT sur la référence T2T-CHM13 a mis en évidence un alignement sur plusieurs BCCA, cohérent avec l'anneau connu. Les BCCA partageant de nombreuses régions pseudo homologues, ADN satellite et rDNA, il est encore difficile à ce jour de déterminer spécifiquement de quel chromosome il s'agit. La méthode a pu être validée en ré-analysant l'ensemble des points de cassure terminaux des patients PMS porteurs de délétion terminale, qui a conduit à l'identification d'un anneau supplémentaire. Aidée et soutenue par les bioinformaticiens de l'équipe de recherche qui m'a accueillie, nous avons pu mettre au point un outil Python visant à détecter de façon automatique ces anomalies d'alignement. Cet algorithme nommé *gollum* et décrit dans notre article, a été validé pour le chromosome 22. L'outil est disponible pour l'ensemble de la

communauté médicale et scientifique et devrait pouvoir s'appliquer à l'ensemble des chromosomes acrocentriques dans de prochaines études.

L'analyse des données de WGS associé à T2T-CHM13 est donc prometteuse et ouvre un nouveau champ de diagnostic chromosomique jusque-là limité aux techniques conventionnelles. Le caryotype garde toutefois une place cruciale dans la détermination des remaniements de structure. La ré-analyse des points de cassure terminaux a en effet identifié sur notre cohorte deux patients porteurs d'un remaniement autre qu'un anneau impliquant un BCCA et la région terminale d'un chromosome. Dans un premier cas il s'agissait d'un dérivé 13 d'une translocation entre le bras court d'un chromosome 13 et le bras long d'un chromosome 22. Le deuxième patient portait un dérivé d'inversion du 22 entraînant une duplication 22qter insérée dans le bras court 22p. Ces deux cas reflètent donc une limite du WGS, qui resterait, s'il était proposé en première intention, un outil de dépistage de ce type de remaniements et permettrait alors de cibler des patients nécessitant la réalisation d'un caryotype.

L'intérêt de savoir si des délétions terminales sont liées à la formation d'un anneau, dans le cas du chromosome 22 et du PMS, réside dans le risque de tumeurs lié à la neurofibromatose de type 2, par la perte somatique de l'anneau instable. En effet, le risque de NF2 s'associe à la formation de tumeurs bénignes du système nerveux central, notamment des schwannomes, qui peuvent se manifester par une perte d'audition ou des troubles neurologiques selon leur position. Il devient donc important de pouvoir les dépister le plus tôt possible, et proposer un bilan auditif et une IRM cérébrale à ces patients pourrait aider à ce dépistage. De plus, les orienter vers une structure de soin adaptée, des centres experts de cette pathologie, augmente l'espérance de vie de ces patients.

Aujourd'hui le caryotype n'est plus pratiqué de façon systématique et le nombre d'anneaux du chromosome 22 est probablement sous-estimé. Le WGS couplé à notre outil et aligné sur T2T permettra d'identifier plus facilement ces patients et assurer une prise en charge adaptée.

L'analyse des WGS offre un aperçu global des variations génétiques portées par un patient. Aussi nous avons pu identifier chez 5 de nos 7 patients des variations additionnelles aux délétions 22qter qui vont influencer le phénotype de ces patients. C'est notamment le cas du patient P1 qui porte une mutation hémizygote responsable d'un phénotype neurodéveloppemental sévère, ou de P4 qui porte un variant qui pourrait expliquer sa scoliose en plus du PMS. Ces premières données restent

descriptives, mais nous montrent que l'impact de variations additionnelles dans la complexité des phénotypes est peut-être moins rare qu'attendu. Ainsi il semble essentiel d'avoir des données génomiques complètes pour la compréhension des pathologies présentées par les patients, et cet axe fait l'objet d'un travail de recherche à part entière dans notre laboratoire auquel j'ai pu participer.

De nombreuses causes génétiques sont aujourd'hui rattachées à l'apparition de TND, dont l'étiologie est mise en évidence dans près de la moitié des cas. Une grande partie des gènes impliqués dans ces pathologies touchent entre autre les voies synaptiques. Par ailleurs, on identifie de plus en plus de gènes et/ou variations dont la pénétrance est incomplète, retrouvée dans des populations saines, et l'expressivité variable, ou encore de variations qui vont s'exprimer en fonction du fond génétique d'un individu (Bourgeron, 2015; Parenti et al., 2020). Nous nous sommes ainsi intéressés au poids de CNVs identifiés par puce SNP touchant les gènes de deux familles de protéines d'adhésion neurales, les contactines et contactines *associated proteins* sur une cohorte de 12 013 individus.

Notre étude sur les gènes des contactines et les *contactin associated proteins* n'a pas montré de résultat significatif entre les CNVs impliquant ces gènes et la présence d'un TND, mais toutefois une association significative chez les patients porteurs d'un TND et dans le groupe des DI syndromique porteur d'un CNV touchant une contactine 1 à 6.

Plusieurs limites à notre étude sont identifiées. Tout d'abord, la sélection de nos patients : il s'agit de patients explorés à l'hôpital en génétique pour diagnostic de malformations et/ou de TND. Cette population est probablement plus sévèrement atteinte que les individus habituellement explorés dans les études autour des contactines, qui incluent le plus souvent des patients porteurs de TSA ou DI isolés.

Par ailleurs, concernant le phénotype de nos patients, nous n'avons récupéré que la catégorie phénotypique dans laquelle les patients sont classés selon le retour fait par les laboratoires à l'Agence de Biomédecine (bien que quelques grands syndromes aient été précisés dans la catégorie « Autre » : les pathologies endocrinologiques, surdité,...). Ces catégories regroupaient entre 2013 et 2017 (années d'inclusions de la cohorte) 8 groupes que nous avons simplifié en 7 pour mieux cibler les TND, mais ces données restent approximatives. Or les symptômes souvent décrits dans la littérature qui résultent d'une atteinte hétérozygote d'un gène des contactines sont le plus souvent des TSA isolés ou encore des atteintes

neurosensorielles, et notamment des hyperacousies. Nos données cliniques ne sont pas assez fines pour mettre en évidence des associations avec des phénotypes discrets. De façon générale, des données phénotypiques complètes font souvent défaut dans les grandes cohortes.

Une autre limite à notre étude est l'exploration de données d'ACPA sans données de séquençage. Seuls les CNVS d'au moins quelques kb ont donc été discutés. Nous n'avons pas étudié les individus porteurs de variations de séquences, SNV ou Indel. De plus, aucun patient dans notre cohorte ne porte de variation homozygote, mais nous ne connaissons pas l'état de l'allèle controlatéral qui pourrait porter une variation délétère identifiable en SHD et qui pourrait être responsable de phénotype sévère comme l'encéphalopathie liée à *CNTNAP1* (Garel et al., 2022) ou *CNTNAP2* causant des épilepsie syndromiques avec DI sévère dans un syndrome de Pitt-Hopkins-like (OMIM #610042) (Mittal et al., 2021; Smogavec et al., 2016). De plus, nous ne discutons pas dans notre étude de la présence ou non d'autres CNVs identifiés et pouvant être en lien avec le phénotype des patients. Dans ce sens, nous avons déjà vu que pour au moins 13 patients, le remaniement impliquant une contactine est pathogène de par sa taille ou sa nature et explique à lui seul le phénotype. Ce point sera donc abordé dans une prochaine étape de ce projet.

De la même façon, l'étude de cohortes ayant bénéficié d'un WES ou WGS sera le prochain point exploré par ce projet, afin de proposer une analyse complète des variations touchant les contactines dans les TND.

A l'avenir, il deviendra essentiel de disposer de données cliniques précises dans de grandes cohortes pour réussir à décrire des associations avec des traits phénotypiques de plus en plus précis. C'est un point sur lequel le laboratoire qui m'a accueilli s'implique, notamment dans le recensement des patients atteints de PMS pour qui nous recueillons une évaluation clinique la plus complète possible.

# Bibliographie

1000 Genomes Project Consortium, Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., Hurles, M. E., & McVean, G. A. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061-1073. <https://doi.org/10.1038/nature09534>

Abe-Hatano, C., Iida, A., Kosugi, S., Momozawa, Y., Terao, C., Ishikawa, K., Okubo, M., Hachiya, Y., Nishida, H., Nakamura, K., Miyata, R., Murakami, C., Takahashi, K., Hoshino, K., Sakamoto, H., Ohta, S., Kubota, M., Takeshita, E., Ishiyama, A., ... Goto, Y.-I. (2021). Whole genome sequencing of 45 Japanese patients with intellectual disability. *American Journal of Medical Genetics Part A*, *185*(5), 1468-1480. <https://doi.org/10.1002/ajmg.a.62138>

Aboukais, R., Zairi, F., Bonne, N.-X., Baroncini, M., Schapira, S., Vincent, C., & Lejeune, J.-P. (2015). Causes of mortality in neurofibromatosis type 2. *British Journal of Neurosurgery*, *29*(1), 37-40. <https://doi.org/10.3109/02688697.2014.952266>

Abyzov, A., Urban, A. E., Snyder, M., & Gerstein, M. (2011). CNVnator : An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research*, *21*(6), 974-984. <https://doi.org/10.1101/gr.114876.110>

Alibutud, R., Hansali, S., Cao, X., Zhou, A., Mahaganapathy, V., Azaro, M., Gwin, C., Wilson, S., Buyske, S., Bartlett, C. W., Flax, J. F., Brzustowicz, L. M., & Xing, J. (2023). Structural Variations Contribute to the Genetic Etiology of Autism Spectrum Disorder and Language Impairments. *International Journal of Molecular Sciences*, *24*(17), 13248. <https://doi.org/10.3390/ijms241713248>

Álvarez-Mora, M. I., Sánchez, A., Rodríguez-Revenge, L., Corominas, J., Rabionet, R., Puig, S., & Madrigal, I. (2022). Diagnostic yield of next-generation sequencing in 87 families with neurodevelopmental disorders. *Orphanet Journal of Rare Diseases*, *17*(1), 60. <https://doi.org/10.1186/s13023-022-02213-z>

Anderson, S. (2020). Café au Lait Macules and Associated Genetic Syndromes. *Journal of Pediatric Health Care*, *34*(1), 71-81. <https://doi.org/10.1016/j.pedhc.2019.05.001>

Antonarakis, S. E. (2022). Short arms of human acrocentric chromosomes and the completion of the human genome sequence. *Genome Research*, *32*(4), 599-607. <https://doi.org/10.1101/gr.275350.121>

Ardern-Holmes, S., Fisher, G., & North, K. (2017). Neurofibromatosis Type 2 : Presentation, Major Complications, and Management, With a Focus on the Pediatric Age Group. *Journal of Child Neurology*, *32*(1), 9-22. <https://doi.org/10.1177/0883073816666736>

Audebert, C., Hot, D., Lemoine, Y., & Caboche, S. (2014). Le séquençage haut-débit—Vers un diagnostic basé sur la séquence complète du génome de l'agent infectieux. *médecine/sciences*, *30*(12), Article 12. <https://doi.org/10.1051/medsci/20143012018>

Bagwell, J., Norman, J., Ellis, K., Peskin, B., Hwang, J., Ge, X., Nguyen, S. V., McMenamin, S. K., Stainier, D. Y., & Bagnat, M. (2020). Notochord vacuoles absorb compressive bone growth during zebrafish spine formation. *eLife*, *9*, e51221. <https://doi.org/10.7554/eLife.51221>

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder : Evidence from a British twin study.

*Psychological Medicine*, 25(1), 63-77. <https://doi.org/10.1017/s0033291700028099>

Bailey, S. F., Alonso Morales, L. A., & Kassen, R. (2021). Effects of Synonymous Mutations beyond Codon Bias : The Evidence for Adaptive Synonymous Substitutions from Microbial Evolution Experiments. *Genome Biology and Evolution*, 13(9), evab141. <https://doi.org/10.1093/gbe/evab141>

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson Rosenberg, C., White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L.-C., Harrington, R., Lopez, M., Fitzgerald, R. T., Hewitt, A., ... Dowling, N. F. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002)*, 67(6), 1-23. <https://doi.org/10.15585/mmwr.ss6706a1>

Bartik, L. E., Hughes, S. S., Tracy, M., Feldt, M. M., Zhang, L., Arganbright, J., & Kaye, A. (2022). 22q11.2 duplications : Expanding the clinical presentation. *American Journal of Medical Genetics Part A*, 188(3), 779-787. <https://doi.org/10.1002/ajmg.a.62577>

Baser, M. E., Friedman, J. M., Aeschliman, D., Joe, H., Wallace, A. J., Ramsden, R. T., & Evans, D. G. R. (2002). Predictors of the Risk of Mortality in Neurofibromatosis 2. *American Journal of Human Genetics*, 71(4), 715-723.

Battaglia, A., Bianchini, E., & Carey, J. C. (1999). Diagnostic yield of the comprehensive assessment of developmental delay/mental retardation in an institute of child neuropsychiatry. *American Journal of Medical Genetics*, 82(1), 60-66.

Bay, H., Haghghatfard, A., Karimipour, M., Seyedena, S. Y., & Hashemi, M. (2023). Expression alteration of Neuroligin family gene in attention deficit and hyperactivity disorder and autism spectrum disorder. *Research in Developmental Disabilities*, 139, 104558. <https://doi.org/10.1016/j.ridd.2023.104558>

Bernaciak, J., Wiśniowiecka-Kowalnik, B., Castañeda, J., Kutkowska-Każmierczak, A., & Nowakowska, B. (2017). A novel de novo 20q13.11q13.12 microdeletion in a boy with neurodevelopmental disorders—Case report. *Developmental Period Medicine*, 21(2), 91-94.

Bhatia, N. S., Lim, J. Y., Bonnard, C., Kuan, J.-L., Brett, M., Wei, H., Cham, B., Chin, H., Bosso-Lefevre, C., Dharuman, P., Escande-Beillard, N., Devasia, A. G., Goh, C. Y. J., Kam, S., Liew, W. K.-M., Liew, W. K., Lin, G., Jain, K., Ng, A. Y.-J., ... Jamuar, S. S. (2021). Singapore Undiagnosed Disease Program : Genomic Analysis aids Diagnosis and Clinical Management. *Archives of Disease in Childhood*, 106(1), 31-37. <https://doi.org/10.1136/archdischild-2020-319180>

Bonaglia, M. C., Giorda, R., Beri, S., De Agostini, C., Novara, F., Fichera, M., Grillo, L., Galesi, O., Vetro, A., Ciccone, R., Bonati, M. T., Giglio, S., Guerrini, R., Osimani, S., Marelli, S., Zucca, C., Grasso, R., Borgatti, R., Mani, E., ... Zuffardi, O. (2011). Molecular mechanisms generating and stabilizing terminal 22q13 deletions in 44 subjects with Phelan/McDermid syndrome. *PLoS Genetics*, 7(7), e1002173. <https://doi.org/10.1371/journal.pgen.1002173>

Bose, P., Hermetz, K. E., Conneely, K. N., & Rudd, M. K. (2014). Tandem Repeats and G-Rich Sequences Are Enriched at Human CNV Breakpoints. *PLoS ONE*, 9(7), e101607. <https://doi.org/10.1371/journal.pone.0101607>

Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews. Neuroscience*, 16(9), 551-563. <https://doi.org/10.1038/nrn3992>

Brotelle, T., & Bay, J.-O. (2016). La voie de signalisation PI3K-AKT-mTOR :

Description, développement thérapeutique, résistances, marqueurs prédictifs/pronostiques et applications thérapeutiques en cancérologie. *Bulletin du Cancer*, 103(1), 18-29. <https://doi.org/10.1016/j.bulcan.2015.09.011>

Burnside, R. D. (2015). 22q11.21 Deletion Syndromes: A Review of Proximal, Central, and Distal Deletions and Their Associated Features. *Cytogenetic and Genome Research*, 146(2), 89-99. <https://doi.org/10.1159/000438708>

Bursted, B., Zamariolli, M., Bellucco, F. T., & Melaragno, M. I. (2022). Mechanisms of structural chromosomal rearrangement formation. *Molecular Cytogenetics*, 15(1), 23. <https://doi.org/10.1186/s13039-022-00600-6>

Byers, H. M., Adam, M. P., LaCroix, A., Leary, S. E. S., Cole, B., Dobyns, W. B., & Mefford, H. C. (2017). Description of a new oncogenic mechanism for atypical teratoid rhabdoid tumors in patients with ring chromosome 22. *American Journal of Medical Genetics. Part A*, 173(1), 245-249. <https://doi.org/10.1002/ajmg.a.37993>

Caba, L., Rusu, C., Plăiașu, G., Grămescu, M., Bujoran, C., Ochiană, D., Voloșciuc, M., Popescu, R., Braha, E., Pânzaru, M., Butnariu, L., Sireteanu, A., Covic, M., & Gorduza, E. (2012). Ring Autosomes : Some Unexpected Findings. *Balkan Journal of Medical Genetics : BJMG*, 15(2), 35-46. <https://doi.org/10.2478/bjmg-2013-0005>

Cargile, C. B., Goh, D. L.-M., Goodman, B. K., Chen, X.-N., Korenberg, J. R., Semenza, G. L., & Thomas, G. H. (2002). Molecular cytogenetic characterization of a subtle interstitial del(3)(p25.3p26.2) in a patient with deletion 3p syndrome. *American Journal of Medical Genetics*, 109(2), 133-138. <https://doi.org/10.1002/ajmg.10323>

Carvalho, C. M. B., & Lupski, J. R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nature Reviews Genetics*, 17(4), 224-238. <https://doi.org/10.1038/nrg.2015.25>

Caumes, R., Smol, T., Thuillier, C., Balerdi, M., Lestienne-Roche, C., Manouvrier-Hanu, S., & Ghoumid, J. (2020). Phenotypic spectrum of SHANK2-related neurodevelopmental disorder. *European Journal of Medical Genetics*, 63(12), 104072. <https://doi.org/10.1016/j.ejmg.2020.104072>

Charng, W.-L., Karaca, E., Coban Akdemir, Z., Gambin, T., Atik, M. M., Gu, S., Posey, J. E., Jhangiani, S. N., Muzny, D. M., Doddapaneni, H., Hu, J., Boerwinkle, E., Gibbs, R. A., Rosenfeld, J. A., Cui, H., Xia, F., Manickam, K., Yang, Y., Faqeih, E. A., ... Lupski, J. R. (2016). Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. *BMC Medical Genomics*, 9(1), 42. <https://doi.org/10.1186/s12920-016-0208-3>

Chatterjee, M., Schild, D., & Teunissen, C. E. (2019). Contactins in the central nervous system : Role in health and disease. *Neural Regeneration Research*, 14(2), 206-216. <https://doi.org/10.4103/1673-5374.244776>

Cheng, S. S. W., Chan, K. Y. K., Leung, K. K. P., Au, P. K. C., Tam, W.-K., Li, S. K. M., Luk, H.-M., Kan, A. S. Y., Chung, B. H. Y., Lo, I. F. M., & Tang, M. H. Y. (2019). Experience of chromosomal microarray applied in prenatal and postnatal settings in Hong Kong. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, 181(2), 196-207. <https://doi.org/10.1002/ajmg.c.31697>

Cho, E. H., Park, J. B., & Kim, J. K. (2014). Atypical teratoid rhabdoid brain tumor in an infant with ring chromosome 22. *Korean Journal of Pediatrics*, 57(7), 333-336. <https://doi.org/10.3345/kjp.2014.57.7.333>

Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D., &



Kingsmore, S. F. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genomic Medicine*, 3, 16. <https://doi.org/10.1038/s41525-018-0053-8>

Coe, B. P., Girirajan, S., & Eichler, E. E. (2012). The genetic variability and commonality of neurodevelopmental disease. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 160C(2), 118-129. <https://doi.org/10.1002/ajmg.c.31327>

Coe, B. P., Witherspoon, K., Rosenfeld, J. A., van Bon, B. W. M., Vulto-van Silfhout, A. T., Bosco, P., Friend, K. L., Baker, C., Buono, S., Vissers, L. E. L. M., Schuurs-Hoeijmakers, J. H., Hoischen, A., Pfundt, R., Krumm, N., Carvill, G. L., Li, D., Amaral, D., Brown, N., Lockhart, P. J., ... Eichler, E. E. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nature Genetics*, 46(10), 1063-1071. <https://doi.org/10.1038/ng.3092>

Cornejo-Sanchez, D. M., Li, G., Fabiha, T., Wang, R., Acharya, A., Everard, J. L., Kadlubowska, M. K., Huang, Y., Schrauwen, I., Wang, G. T., DeWan, A. T., & Leal, S. M. (2023). Rare-variant association analysis reveals known and new age-related hearing loss genes. *European Journal of Human Genetics: EJHG*, 31(6), 638-647. <https://doi.org/10.1038/s41431-023-01302-2>

Cottrell, C. E., Bir, N., Varga, E., Alvarez, C. E., Bouyain, S., Zernzach, R., Thrush, D. L., Evans, J., Trimarchi, M., Butter, E. M., Cunningham, D., Gastier-Foster, J. M., McBride, K. L., & Herman, G. E. (2011). Contactin 4 as an autism susceptibility locus. *Autism Research: Official Journal of the International Society for Autism Research*, 4(3), 189-199. <https://doi.org/10.1002/aur.184>

Cummings, K., Watkins, A., Jones, C., Dias, R., & Welham, A. (2022). Behavioural and psychological features of PTEN mutations : A systematic review of the literature and meta-analysis of the prevalence of autism spectrum disorder characteristics. *Journal of Neurodevelopmental Disorders*, 14(1), 1. <https://doi.org/10.1186/s11689-021-09406-w>

Cuoco, C., Ronchetto, P., Gimelli, S., Béna, F., Divizia, M. T., Lerone, M., Mirabelli-Badenier, M., Mascaretti, M., & Gimelli, G. (2011). Microarray based analysis of an inherited terminal 3p26.3 deletion, containing only the CHL1 gene, from a normal father to his two affected children. *Orphanet Journal of Rare Diseases*, 6, 12. <https://doi.org/10.1186/1750-1172-6-12>

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>

De Rubeis, S., Siper, P. M., Durkin, A., Weissman, J., Muratet, F., Halpern, D., Trelles, M. del P., Frank, Y., Lozano, R., Wang, A. T., Holder, J. L., Betancur, C., Buxbaum, J. D., & Kolevzon, A. (2018). Delineation of the genetic and clinical spectrum of Phelan-McDermid syndrome caused by SHANK3 point mutations. *Molecular Autism*, 9(1), 31. <https://doi.org/10.1186/s13229-018-0205-9>

de Ravel, T. J. L., Devriendt, K., Fryns, J.-P., & Vermeesch, J. R. (2007). What's new in karyotyping? The move towards array comparative genomic hybridisation (CGH). *European Journal of Pediatrics*, 166(7), 637-643. <https://doi.org/10.1007/s00431-007-0463-6>

Des Portes, V. (2020). Troubles du neurodéveloppement : Aspects cliniques. *Contraste*, 51(1), 21-53. <https://doi.org/10.3917/cont.051.0021>

Djurovic, S., Gustafsson, O., Mattingsdal, M., Athanasiu, L., Bjella, T., Tesli, M.,

Agartz, I., Lorentzen, S., Melle, I., Morken, G., & Andreassen, O. A. (2010). A genome-wide association study of bipolar disorder in Norwegian individuals, followed by replication in Icelandic sample. *Journal of Affective Disorders*, *126*(1-2), 312-316. <https://doi.org/10.1016/j.jad.2010.04.007>

Dolzhenko, E., Bennett, M. F., Richmond, P. A., Trost, B., Chen, S., van Vugt, J. J. F. A., Nguyen, C., Narzisi, G., Gainullin, V. G., Gross, A. M., Lajoie, B. R., Taft, R. J., Wasserman, W. W., Scherer, S. W., Veldink, J. H., Bentley, D. R., Yuen, R. K. C., Bahlo, M., & Eberle, M. A. (2020). ExpansionHunter Denovo : A computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biology*, *21*(1), 102. <https://doi.org/10.1186/s13059-020-02017-z>

Durand, C. M., Betancur, C., Boeckers, T. M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I. C., Anckarsäter, H., Sponheim, E., Goubran-Botros, H., Delorme, R., Chabane, N., Mouren-Simeoni, M.-C., de Mas, P., Bieth, E., Rogé, B., Héron, D., ... Bourgeron, T. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genetics*, *39*(1), 25-27. <https://doi.org/10.1038/ng1933>

Durand, C. M., Perroy, J., Loll, F., Perrais, D., Fagni, L., Bourgeron, T., Montcouquiol, M., & Sans, N. (2012). SHANK3 mutations identified in autism lead to modification of dendritic spine morphology via an actin-dependent mechanism. *Molecular Psychiatry*, *17*(1), Article 1. <https://doi.org/10.1038/mp.2011.57>

Evans, D. G., Hartley, C. L., Smith, P. T., King, A. T., Bowers, N. L., Tobi, S., Wallace, A. J., Perry, M., Anup, R., Lloyd, S. K. W., Rutherford, S. A., Hammerbeck-Ward, C., Pathmanaban, O. N., Stapleton, E., Freeman, S. R., Kellett, M., Halliday, D., Parry, A., Gair, J. J., ... Smith, M. J. (2020). Incidence of mosaicism in 1055 de novo NF2 cases : Much higher than previous estimates with high utility of next-generation sequencing. *Genetics in Medicine*, *22*(1), 53-59. <https://doi.org/10.1038/s41436-019-0598-7>

Evans, D. G., Huson, S. M., Donnai, D., Neary, W., Blair, V., Teare, D., Newton, V., Strachan, T., Ramsden, R., & Harris, R. (1992). A genetic study of type 2 neurofibromatosis in the United Kingdom. I. Prevalence, mutation rate, fitness, and confirmation of maternal transmission effect on severity. *Journal of Medical Genetics*, *29*(12), 841-846. <https://doi.org/10.1136/jmg.29.12.841>

Evans, Dg. R. (2009). Neurofibromatosis type 2 (NF2): A clinical and molecular review. *Orphanet Journal of Rare Diseases*, *4*(1), 16. <https://doi.org/10.1186/1750-1172-4-16>

Fernandez, T., Morgan, T., Davis, N., Klin, A., Morris, A., Farhi, A., Lifton, R. P., & State, M. W. (2004). Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. *American Journal of Human Genetics*, *74*(6), 1286-1293. <https://doi.org/10.1086/421474>

Filosto, M., Cotti Piccinelli, S., Caria, F., Gallo Cassarino, S., Baldelli, E., Galvagni, A., Volonghi, I., Scarpelli, M., & Padovani, A. (2018). Mitochondrial Neurogastrointestinal Encephalomyopathy (MNGIE-MTDPS1). *Journal of Clinical Medicine*, *7*(11), Article 11. <https://doi.org/10.3390/jcm7110389>

Fischbach, G. D., & Lord, C. (2010). The Simons Simplex Collection : A resource for identification of autism genetic risk factors. *Neuron*, *68*(2), 192-195. <https://doi.org/10.1016/j.neuron.2010.10.006>

Frazier, T. W. (2019). Autism Spectrum Disorder Associated with Germline Heterozygous PTEN Mutations. *Cold Spring Harbor Perspectives in Medicine*, *9*(10), a037002.

<https://doi.org/10.1101/cshperspect.a037002>

Froimchuk, E., Jang, Y., & Ge, K. (2017). Histone H3 lysine 4 methyltransferase KMT2D. *Gene*, 627, 337-342. <https://doi.org/10.1016/j.gene.2017.06.056>

Furley, A. J., Morton, S. B., Manalo, D., Karagogeos, D., Dodd, J., & Jessell, T. M. (1990). The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neurite outgrowth-promoting activity. *Cell*, 61(1), 157-170. [https://doi.org/10.1016/0092-8674\(90\)90223-2](https://doi.org/10.1016/0092-8674(90)90223-2)

Gandawijaya, J., Bamford, R. A., Burbach, J. P. H., & Oguro-Ando, A. (2020). Cell Adhesion Molecules Involved in Neurodevelopmental Pathways Implicated in 3p-Deletion Syndrome and Autism Spectrum Disorder. *Frontiers in Cellular Neuroscience*, 14, 611379. <https://doi.org/10.3389/fncel.2020.611379>

Gardner, R. J. M., Sutherland, G. R., & Shaffer, L. G. (2012). *Chromosome Abnormalities and Genetic Counseling*. Oxford University Press, USA.

Garel, P., Lesca, G., Ville, D., Poulat, A.-L., Chatron, N., Sanlaville, D., Des Portes, V., Arzimanoglou, A., & Lion-François, L. (2022). CNTNAP1-encephalopathy: Six novel patients surviving the neonatal period. *European Journal of Paediatric Neurology: EJPN: Official Journal of the European Paediatric Neurology Society*, 37, 98-104. <https://doi.org/10.1016/j.ejpn.2022.01.015>

Girirajan, S., Dennis, M. Y., Baker, C., Malig, M., Coe, B. P., Campbell, C. D., Mark, K., Vu, T. H., Alkan, C., Cheng, Z., Biesecker, L. G., Bernier, R., & Eichler, E. E. (2013). Refinement and Discovery of New Hotspots of Copy-Number Variation Associated with Autism Spectrum Disorder. *The American Journal of Human Genetics*, 92(2), 221-237. <https://doi.org/10.1016/j.ajhg.2012.12.016>

Goh, G., & Choi, M. (2012). Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. *Genomics & Informatics*, 10(4), 214-219. <https://doi.org/10.5808/GI.2012.10.4.214>

Gross, A. M., Ajay, S. S., Rajan, V., Brown, C., Bluske, K., Burns, N. J., Chawla, A., Coffey, A. J., Malhotra, A., Scocchia, A., Thorpe, E., Dzidic, N., Hovanes, K., Sahoo, T., Dolzhenko, E., Lajoie, B., Khouzam, A., Chowdhury, S., Belmont, J., ... Taft, R. J. (2019). Copy-number variants in clinical genome sequencing: Deployment and interpretation for rare and undiagnosed disease. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 21(5), 1121-1130. <https://doi.org/10.1038/s41436-018-0295-y>

Guarracino, A., Buonaiuto, S., De Lima, L. G., Potapova, T., Rhie, A., Koren, S., Rubinstein, B., Fischer, C., Human Pangenome Reference Consortium, Abel, H. J., Antonacci-Fulton, L. L., Asri, M., Baid, G., Baker, C. A., Belyaeva, A., Billis, K., Bourque, G., Carroll, A., Chaisson, M. J. P., ... Garrison, E. (2023). Recombination between heterologous human acrocentric chromosomes. *Nature*, 617(7960), 335-343. <https://doi.org/10.1038/s41586-023-05976-y>

Guilmatre, A., Huguet, G., Delorme, R., & Bourgeron, T. (2014). The emerging role of SHANK genes in neuropsychiatric disorders. *Developmental Neurobiology*, 74(2), 113-122. <https://doi.org/10.1002/dneu.22128>

Guo, H., Peng, Y., Hu, Z., Li, Y., Xun, G., Ou, J., Sun, L., Xiong, Z., Liu, Y., Wang, T., Chen, J., Xia, L., Bai, T., Shen, Y., Tian, Q., Hu, Y., Shen, L., Zhao, R., Zhang, X., ... Xia, K. (2017). Genome-wide copy number variation analysis in a Chinese autism spectrum disorder cohort. *Scientific Reports*, 7, 44155. <https://doi.org/10.1038/srep44155>

Gupta, P., Mirman, I., Shahar, S., & Dubey, D. (2023). Growing Spectrum of

Autoimmune Nodopathies. *Current Neurology and Neuroscience Reports*, 23(5), 201-212. <https://doi.org/10.1007/s11910-023-01264-4>

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., Miller, J., Fedele, A., Collins, J., Smith, K., Lotspeich, L., Croen, L. A., Ozonoff, S., Lajonchere, C., Grether, J. K., & Risch, N. (2011). Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. *Archives of General Psychiatry*, 68(11), 1095-1102. <https://doi.org/10.1001/archgenpsychiatry.2011.76>

Hamerton, J. L., Canning, N., Ray, M., & Smith, S. (1975). A cytogenetic survey of 14,069 newborn infants. *Clinical Genetics*, 8(4), 223-243. <https://doi.org/10.1111/j.1399-0004.1975.tb01498.x>

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B. L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., ... Hubbard, T. J. (2012). GENCODE : The reference human genome annotation for The ENCODE Project. *Genome Research*, 22(9), 1760-1774. <https://doi.org/10.1101/gr.135350.111>

Hochstenbach, R., Buizer-Voskamp, J. E., Vorstman, J. A. S., & Ophoff, R. A. (2011). Genome Arrays for the Detection of Copy Number Variations in Idiopathic Mental Retardation, Idiopathic Generalized Epilepsy and Neuropsychiatric Disorders : Lessons for Diagnostic Workflow and Research. *Cytogenetic and Genome Research*, 135(3-4), 174-202. <https://doi.org/10.1159/000332928>

Hochstenbach, R., van Binsbergen, E., Engelen, J., Nieuwint, A., Polstra, A., Poddighe, P., Ruivenkamp, C., Sikkema-Raddatz, B., Smeets, D., & Poot, M. (2009). Array analysis and karyotyping : Workflow consequences based on a retrospective study of 36,325 patients with idiopathic developmental delay in the Netherlands. *European Journal of Medical Genetics*, 52(4), 161-169. <https://doi.org/10.1016/j.ejmg.2009.03.015>

Huang, A. Y., Yu, D., Davis, L. K., Sul, J. H., Tsetsos, F., Ramensky, V., Zelaya, I., Ramos, E. M., Osiecki, L., Chen, J. A., McGrath, L. M., Illmann, C., Sandor, P., Barr, C. L., Grados, M., Singer, H. S., Nöthen, M. M., Hebebrand, J., King, R. A., ... Gilles de la Tourette Syndrome GWAS Replication Initiative (GGRI). (2017). Rare Copy Number Variants in NRXN1 and CNTN6 Increase Risk for Tourette Syndrome. *Neuron*, 94(6), 1101-1111.e7. <https://doi.org/10.1016/j.neuron.2017.06.010>

Huang, Y., Zhao, Y., Ren, Y., Yi, Y., Li, X., Gao, Z., Zhan, X., Yu, J., Wang, D., Liang, S., & Wu, L. (2019). Identifying Genomic Variations in Monozygotic Twins Discordant for Autism Spectrum Disorder Using Whole-Genome Sequencing. *Molecular Therapy - Nucleic Acids*, 14, 204-211. <https://doi.org/10.1016/j.omtn.2018.11.015>

Huguet, G., Schramm, C., Douard, E., Jiang, L., Labbe, A., Tihy, F., Mathonnet, G., Nizard, S., Lemyre, E., Mathieu, A., Poline, J.-B., Loth, E., Toro, R., Schumann, G., Conrod, P., Pausova, Z., Greenwood, C., Paus, T., Bourgeron, T., ... IMAGEN Consortium. (2018). Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. *JAMA Psychiatry*, 75(5), 447-457. <https://doi.org/10.1001/jamapsychiatry.2018.0039>

Ibañez, K., Polke, J., Hagelstrom, R. T., Dolzhenko, E., Pasko, D., Thomas, E. R. A., Daugherty, L. C., Kasperaviciute, D., Smith, K. R., McDonagh, E. M., Rueda, A., Polychronopoulos, D., Chan, G., Angus-Leppan, H., Bhatia, K. P., Davison, J. E., Festenstein, R., Fratta, P., Giunti, P., ... Zarowiecki, M. (2022). Whole genome sequencing for the diagnosis of neurological repeat expansion disorders in the UK : A retrospective diagnostic accuracy and

prospective clinical validation study. *The Lancet Neurology*, 21(3), 234-245. [https://doi.org/10.1016/S1474-4422\(21\)00462-2](https://doi.org/10.1016/S1474-4422(21)00462-2)

International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., ... Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164), 851-861. <https://doi.org/10.1038/nature06258>

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931-945. <https://doi.org/10.1038/nature03001>

Iossifov, I., Levy, D., Allen, J., Ye, K., Ronemus, M., Lee, Y., Yamrom, B., & Wigler, M. (2015). Low load for disruptive mutations in autism genes and their biased transmission. *Proceedings of the National Academy of Sciences*, 112(41), E5600-E5607. <https://doi.org/10.1073/pnas.1516376112>

Jamain, S., Quach, H., Betancur, C., Råstam, M., Colineaux, C., Gillberg, I. C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C., Bourgeron, T., & Paris Autism Research International Sibpair Study. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genetics*, 34(1), 27-29. <https://doi.org/10.1038/ng1136>

Jang, W., Kim, Y., Han, E., Park, J., Chae, H., Kwon, A., Choi, H., Kim, J., Son, J. O., Lee, S. J., Hong, B. Y., Jang, D. H., Han, J. Y., Lee, J. H., Kim, S. Y., Lee, I. G., Sung, I. K., Moon, Y., Kim, M., & Park, J. H. (2019). Chromosomal Microarray Analysis as a First-Tier Clinical Diagnostic Test in Patients With Developmental Delay/Intellectual Disability, Autism Spectrum Disorders, and Multiple Congenital Anomalies : A Prospective Multicenter Study in Korea. *Annals of Laboratory Medicine*, 39(3), 299-310. <https://doi.org/10.3343/alm.2019.39.3.299>

Jeong, G., Bae, H., Jeong, D., Ham, J., Park, S., Kim, H. W., Kang, H.-S., & Kim, S. J. (2018). A Kelch domain-containing KLHDC7B and a long non-coding RNA ST8SIA6-AS1 act oppositely on breast cancer cell proliferation via the interferon signaling pathway. *Scientific Reports*, 8(1), 12922. <https://doi.org/10.1038/s41598-018-31306-8>

Jung, J., Ohk, J., Kim, H., Holt, C. E., Park, H. J., & Jung, H. (2023). mRNA transport, translation, and decay in adult mammalian central nervous system axons. *Neuron*, 111(5), 650-668.e4. <https://doi.org/10.1016/j.neuron.2022.12.015>

Jurka, J. (2000). Repbase Update : A database and an electronic journal of repetitive elements. *Trends in Genetics*, 16(9), 418-420. [https://doi.org/10.1016/S0168-9525\(00\)02093-X](https://doi.org/10.1016/S0168-9525(00)02093-X)

Kaneko-Goto, T., Yoshihara, S.-I., Miyazaki, H., & Yoshihara, Y. (2008). BIG-2 mediates olfactory axon convergence to target glomeruli. *Neuron*, 57(6), 834-846. <https://doi.org/10.1016/j.neuron.2008.01.023>

Kaur, S., & Christodoulou, J. (1993). MECP2 Disorders. In M. P. Adam, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, & A. Amemiya (Éds.), *GeneReviews*®. University of Washington, Seattle. <http://www.ncbi.nlm.nih.gov/books/NBK1497/>

Kochinke, K., Zweier, C., Nijhof, B., Fenckova, M., Cizek, P., Honti, F., Keerthikumar, S., Oortveld, M. A. W., Kleefstra, T., Kramer, J. M., Webber, C., Huynen, M. A., & Schenck, A. (2016). Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *American Journal of Human Genetics*, 98(1), 149-164.

<https://doi.org/10.1016/j.ajhg.2015.11.024>

Kolevzon, A., Angarita, B., Bush, L., Wang, A. T., Frank, Y., Yang, A., Rapaport, R., Saland, J., Srivastava, S., Farrell, C., Edelmann, L. J., & Buxbaum, J. D. (2014). Phelan-McDermid syndrome: A review of the literature and practice parameters for medical assessment and monitoring. *Journal of Neurodevelopmental Disorders*, 6(1), 39. <https://doi.org/10.1186/1866-1955-6-39>

Kong, A., Masson, G., Frigge, M. L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P. I., Ingason, A., Steinberg, S., Rafnar, T., Sulem, P., Mouy, M., Jonsson, F., Thorsteinsdottir, U., Gudbjartsson, D. F., Stefansson, H., & Stefansson, K. (2008). Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature Genetics*, 40(9), Article 9. <https://doi.org/10.1038/ng.216>

Kosztolányi, G. (1987). Does « ring syndrome » exist? An analysis of 207 case reports on patients with a ring autosome. *Human Genetics*, 75(2), 174-179. <https://doi.org/10.1007/BF00591082>

Kosztolányi, G. (2009). The genetics and clinical characteristics of constitutional ring chromosomes. *Journal of the Association of Genetic Technologists*, 35(2), 44-48.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4), 693-705. <https://doi.org/10.1016/j.cell.2007.02.005>

Koza, S. A., Tabet, A. C., Bonaglia, M. C., Andres, S., Anderlid, B.-M., Aten, E., Stiefsohn, D., European Phelan-McDermid syndrome consortium, Evans, D. G., van Ravenswaaij-Arts, C. M. A., & Kant, S. G. (2023). Consensus recommendations on counselling in Phelan-McDermid syndrome, with special attention to recurrence risk and to ring chromosome 22. *European Journal of Medical Genetics*, 66(7), 104773. <https://doi.org/10.1016/j.ejmg.2023.104773>

Kurtas, N., Arrigoni, F., Errichiello, E., Zucca, C., Maghini, C., D'Angelo, M. G., Beri, S., Giorda, R., Bertuzzo, S., Delledonne, M., Xumerle, L., Rossato, M., Zuffardi, O., & Bonaglia, M. C. (2018). Chromothripsis and ring chromosome 22 : A paradigm of genomic complexity in the Phelan-McDermid syndrome (22q13 deletion syndrome). *Journal of Medical Genetics*, 55(4), 269-277. <https://doi.org/10.1136/jmedgenet-2017-105125>

Kushima, I., Imaeda, M., Tanaka, S., Kato, H., Oya-Ito, T., Nakatochi, M., Aleksic, B., & Ozaki, N. (2022). Contribution of copy number variations to the risk of severe eating disorders. *Psychiatry and Clinical Neurosciences*, 76(9), 423-428. <https://doi.org/10.1111/pcn.13430>

Laforce, R., Valdmánis, P. N., Dupré, N., Rouleau, G. A., Turgeon, A. F., & Savard, M. (2009). A novel TYMP mutation in a French Canadian patient with mitochondrial neurogastrointestinal encephalomyopathy. *Clinical Neurology and Neurosurgery*, 111(8), 691-694. <https://doi.org/10.1016/j.clineuro.2009.05.005>

Lai, M.-C., Lombardo, M. V., & Baron-Cohen, S. (2014). Autism. *Lancet (London, England)*, 383(9920), 896-910. [https://doi.org/10.1016/S0140-6736\(13\)61539-1](https://doi.org/10.1016/S0140-6736(13)61539-1)

Landlust, A. M., Koza, S. A., Carbin, M., Walinga, M., Robert, S., Cooke, J., Vyshka, K., European Phelan-McDermid syndrome consortium, van Balkom, I. D. C., & van Ravenswaaij-Arts, C. (2023). Parental perspectives on Phelan-McDermid syndrome : Results of a worldwide survey. *European Journal of Medical Genetics*, 66(7), 104771. <https://doi.org/10.1016/j.ejmg.2023.104771>

Layer, R. M., Chiang, C., Quinlan, A. R., & Hall, I. M. (2014). LUMPY : A probabilistic framework for structural variant discovery. *Genome Biology*, 15(6), R84.

<https://doi.org/10.1186/gb-2014-15-6-r84>

Le Couteur, A., Haden, G., Hammal, D., & McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments : The ADI-R and the ADOS. *Journal of Autism and Developmental Disorders*, 38(2), 362-372. <https://doi.org/10.1007/s10803-007-0403-3>

Leblond, C. S., Heinrich, J., Delorme, R., Proepper, C., Betancur, C., Huguet, G., Konyukh, M., Chaste, P., Ey, E., Rastam, M., Anckarsäter, H., Nygren, G., Gillberg, I. C., Melke, J., Toro, R., Regnault, B., Fauchereau, F., Mercati, O., Lemièrè, N., ... Bourgeron, T. (2012). Genetic and Functional Analyses of SHANK2 Mutations Suggest a Multiple Hit Model of Autism Spectrum Disorders. *PLoS Genetics*, 8(2), e1002521. <https://doi.org/10.1371/journal.pgen.1002521>

Leblond, C. S., Le, T.-L., Malesys, S., Cliquet, F., Tabet, A.-C., Delorme, R., Rolland, T., & Bourgeron, T. (2021). Operative list of genes associated with autism and neurodevelopmental disorders based on database review. *Molecular and Cellular Neurosciences*, 113, 103623. <https://doi.org/10.1016/j.mcn.2021.103623>

Leblond, C. S., Nava, C., Polge, A., Gauthier, J., Huguet, G., Lumbroso, S., Giuliano, F., Stordeur, C., Depienne, C., Mouzat, K., Pinto, D., Howe, J., Lemièrè, N., Durand, C. M., Guibert, J., Ey, E., Toro, R., Peyre, H., Mathieu, A., ... Bourgeron, T. (2014). Meta-analysis of SHANK Mutations in Autism Spectrum Disorders : A gradient of severity in cognitive impairments. *PLoS Genetics*, 10(9), e1004580. <https://doi.org/10.1371/journal.pgen.1004580>

Lee, C., Iafrate, A. J., & Brothman, A. R. (2007). Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nature Genetics*, 39(7 Suppl), S48-54. <https://doi.org/10.1038/ng2092>

Lee, J.-R. (2015). Protein tyrosine phosphatase PTPRT as a regulator of synaptic formation and neuronal development. *BMB Reports*, 48(5), 249-255. <https://doi.org/10.5483/bmbrep.2015.48.5.037>

Lengyel, A., Pinti, Év., Nebral, K., Pikó, H., Ujfalusi, A., Haas, O. A., Fekete, G., & Haltrich, I. (2021). Chromosome 2q14.3 microdeletion encompassing CNTNAP5 gene in a patient carrying a complex chromosomal rearrangement. *Journal of Genetics*, 100, 66.

Levy, T., Foss-Feig, J. H., Betancur, C., Siper, P. M., Trelles-Thorne, M. del P., Halpern, D., Frank, Y., Lozano, R., Layton, C., Britvan, B., Bernstein, J. A., Buxbaum, J. D., Berry-Kravis, E., Powell, C. M., Srivastava, S., Sahin, M., Soorya, L., Thurm, A., Kolevzon, A., & the Developmental Synaptopathies Consortium. (2022). Strong evidence for genotype–phenotype correlations in Phelan-McDermid syndrome : Results from the developmental synaptopathies consortium. *Human Molecular Genetics*, 31(4), 625-637. <https://doi.org/10.1093/hmg/ddab280>

Li, H., Takeda, Y., Niki, H., Ogawa, J., Kobayashi, S., Kai, N., Akasaka, K., Asano, M., Sudo, K., Iwakura, Y., & Watanabe, K. (2003). Aberrant responses to acoustic stimuli in mice deficient for neural recognition molecule NB-2. *European Journal of Neuroscience*, 17(5), 929-936. <https://doi.org/10.1046/j.1460-9568.2003.02514.x>

Li, P., Dupont, B., Hu, Q., Crimi, M., Shen, Y., Lebedev, I., & Liehr, T. (2022). The past, present, and future for constitutional ring chromosomes : A report of the international consortium for human ring chromosomes. *HGG Advances*, 3(4), 100139. <https://doi.org/10.1016/j.xhgg.2022.100139>

Lindstrand, A., Eisfeldt, J., Pettersson, M., Carvalho, C. M. B., Kvarnung, M., Grigelioniene, G., Anderlid, B.-M., Bjerin, O., Gustavsson, P., Hammarsjö, A., Georgii-

Hemming, P., Iwarsson, E., Johansson-Soller, M., Lagerstedt-Robinson, K., Lieden, A., Magnusson, M., Martin, M., Malmgren, H., Nordenskjöld, M., ... Nilsson, D. (2019). From cytogenetics to cytogenomics : Whole-genome sequencing as a first-line test comprehensively captures the diverse spectrum of disease-causing genetic variation underlying intellectual disability. *Genome Medicine*, 11(1), 68. <https://doi.org/10.1186/s13073-019-0675-1>

Lionel, A. C., Tammimies, K., Vaags, A. K., Rosenfeld, J. A., Ahn, J. W., Merico, D., Noor, A., Runke, C. K., Pillalamarri, V. K., Carter, M. T., Gazzellone, M. J., Thiruvahindrapuram, B., Fagerberg, C., Laulund, L. W., Pellecchia, G., Lamoureux, S., Deshpande, C., Clayton-Smith, J., White, A. C., ... Scherer, S. W. (2014). Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. *Human Molecular Genetics*, 23(10), 2752-2768. <https://doi.org/10.1093/hmg/ddt669>

Lipton, J. O., & Sahin, M. (2014). The Neurology of mTOR. *Neuron*, 84(2), 275-291. <https://doi.org/10.1016/j.neuron.2014.09.034>

Liu, B., Conroy, J. M., Morrison, C. D., Odunsi, A. O., Qin, M., Wei, L., Trump, D. L., Johnson, C. S., Liu, S., & Wang, J. (2015). Structural variation discovery in the cancer genome using next generation sequencing : Computational solutions and perspectives. *Oncotarget*, 6(8), 5477-5489. <https://doi.org/10.18632/oncotarget.3491>

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised : A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659-685. <https://doi.org/10.1007/BF02172145>

Lubetzki, C., Charles, P., Denisenko-Nehrbass, N., Barbin, G., & Girault, J.-A. (2003). Les jonctions paranodales des fibres myélinisées : Site d'ancrage, lieu d'interactions. *médecine/sciences*, 19(1), Article 1. <https://doi.org/10.1051/medsci/200319122>

Lubetzki, C., Sol-Foulon, N., & Desmazières, A. (2020). Nodes of Ranvier during development and repair in the CNS. *Nature Reviews Neurology*, 16(8), Article 8. <https://doi.org/10.1038/s41582-020-0375-x>

Lyons-Warren, A. M., Cheung, S. W., & Holder, J. L. (2017). Clinical Reasoning : A common cause for Phelan-McDermid syndrome and neurofibromatosis type 2: One ring to bind them. *Neurology*, 89(17), e205-e209. <https://doi.org/10.1212/WNL.0000000000004573>

Mack, H., & Swisshelm, K. (2013). Robertsonian Translocations. In S. Maloy & K. Hughes (Éds.), *Brenner's Encyclopedia of Genetics (Second Edition)* (p. 301-305). Academic Press. <https://doi.org/10.1016/B978-0-12-374984-0.01357-7>

Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A., & Jabado, N. (2011). What can exome sequencing do for you? *Journal of Medical Genetics*, 48(9), 580-589. <https://doi.org/10.1136/jmedgenet-2011-100223>

Manning, M., Hudgins, L., & Professional Practice and Guidelines Committee. (2010). Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 12(11), 742-745. <https://doi.org/10.1097/GIM.0b013e3181f8baad>

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., Thiruvahindrapuram, B., Fiebig, A., Schreiber, S., Friedman, J., Ketelaars, C. E. J., Vos, Y. J., Ficocioglu, C., Kirkpatrick, S., Nicolson, R., ... Scherer, S. W. (2008). Structural Variation of Chromosomes in Autism Spectrum Disorder. *The American*



*Journal of Human Genetics*, 82(2), 477-488. <https://doi.org/10.1016/j.ajhg.2007.12.009>

Martin, C. L., & Ledbetter, D. H. (2017). Chromosomal Microarray Testing for Children With Unexplained Neurodevelopmental Disorders. *JAMA*, 317(24), 2545-2546. <https://doi.org/10.1001/jama.2017.7272>

Mastrososa, F. K., Miller, D. E., & Eichler, E. E. (2023). Applications of long-read sequencing to Mendelian genetics. *Genome Medicine*, 15(1), 42. <https://doi.org/10.1186/s13073-023-01194-3>

McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., Fulton, R., Kucaba, T. A., Wagner-McPherson, C., Barbazuk, W. B., Gregory, S. G., Humphray, S. J., French, L., Evans, R. S., Bethel, G., ... Max-Planck-Institute for Molecular Genetics: (2001). A physical map of the human genome. *Nature*, 409(6822), Article 6822. <https://doi.org/10.1038/35057157>

Mefford, H. C., Batshaw, M. L., & Hoffman, E. P. (2012). Genomics, Intellectual Disability, and Autism. *New England Journal of Medicine*, 366(8), 733-743. <https://doi.org/10.1056/NEJMra1114194>

Mendelsohn, N. J., & Schaefer, G. B. (2008). Genetic evaluation of autism. *Seminars in Pediatric Neurology*, 15(1), 27-31. <https://doi.org/10.1016/j.spen.2008.01.005>

Meng, L., Pammi, M., Saronwala, A., Magoulas, P., Ghazi, A. R., Vetrini, F., Zhang, J., He, W., Dharmadhikari, A. V., Qu, C., Ward, P., Braxton, A., Narayanan, S., Ge, X., Tokita, M. J., Santiago-Sim, T., Dai, H., Chiang, T., Smith, H., ... Lalani, S. R. (2017). Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management. *JAMA Pediatrics*, 171(12), e173438. <https://doi.org/10.1001/jamapediatrics.2017.3438>

Mercati, O., Huguet, G., Danckaert, A., André-Leroux, G., Maruani, A., Bellinzoni, M., Rolland, T., Gouder, L., Mathieu, A., Buratti, J., Amsellem, F., Benabou, M., Van-Gils, J., Beggiano, A., Konyukh, M., Bourgeois, J.-P., Gazzellone, M. J., Yuen, R. K. C., Walker, S., ... Bourgeron, T. (2017). CNTN6 mutations are risk factors for abnormal auditory sensory perception in autism spectrum disorders. *Molecular Psychiatry*, 22(4), 625-633. <https://doi.org/10.1038/mp.2016.61>

Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G. A., Schneider, V. A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., ... Phillippy, A. M. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, 585(7823), 79-84. <https://doi.org/10.1038/s41586-020-2547-7>

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., Church, D. M., Crolla, J. A., Eichler, E. E., Epstein, C. J., Faucett, W. A., Feuk, L., Friedman, J. M., Hamosh, A., Jackson, L., Kaminsky, E. B., Kok, K., Krantz, I. D., Kuhn, R. M., ... Ledbetter, D. H. (2010). Consensus statement: Chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American Journal of Human Genetics*, 86(5), 749-764. <https://doi.org/10.1016/j.ajhg.2010.04.006>

Mills, R. E., Luttig, C. T., Larkins, C. E., Beauchamp, A., Tsui, C., Pittard, W. S., & Devine, S. E. (2006). An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research*, 16(9), 1182-1190. <https://doi.org/10.1101/gr.4565806>

Mitchel, M. W., Myers, S. M., Heidlebaugh, A. R., Taylor, C. M., Rea, H., Neuhaus, E., Kurtz-Nelson, E. C., Earl, R., Bernier, R., Ledbetter, D. H., Martin, C. L., & Eichler, E. E.

(1993). CHD8-Related Neurodevelopmental Disorder with Overgrowth. In M. P. Adam, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, & A. Amemiya (Éds.), *GeneReviews®*. University of Washington, Seattle. <http://www.ncbi.nlm.nih.gov/books/NBK585456/>

Mittal, R., Kumar, A., Ladda, R., Mainali, G., & Aliu, E. (2021). Pitt Hopkins-Like Syndrome 1 with Novel CNTNAP2 Mutation in Siblings. *Child Neurology Open*, 8, 2329048X211055330. <https://doi.org/10.1177/2329048X211055330>

Monteiro, P., & Feng, G. (2017). SHANK proteins : Roles at the synapse and in autism spectrum disorder. *Nature Reviews Neuroscience*, 18(3), 147-157. <https://doi.org/10.1038/nrn.2016.183>

Morris-Rosendahl, D. J., & Crocq, M.-A. (2020). Neurodevelopmental disorders-the history and future of a diagnostic concept. *Dialogues in Clinical Neuroscience*, 22(1), 65-72. <https://doi.org/10.31887/DCNS.2020.22.1/macrocq>

Mullaney, J. M., Mills, R. E., Pittard, W. S., & Devine, S. E. (2010). Small insertions and deletions (INDELs) in human genomes. *Human Molecular Genetics*, 19(R2), R131-R136. <https://doi.org/10.1093/hmg/ddq400>

Murai, K. K., Misner, D., & Ranscht, B. (2002). Contactin supports synaptic plasticity associated with hippocampal long-term depression but not potentiation. *Current Biology: CB*, 12(3), 181-190. [https://doi.org/10.1016/s0960-9822\(02\)00680-2](https://doi.org/10.1016/s0960-9822(02)00680-2)

Murdoch, J. D., Gupta, A. R., Sanders, S. J., Walker, M. F., Keaney, J., Fernandez, T. V., Murtha, M. T., Anyanwu, S., Ober, G. T., Raubeson, M. J., DiLullo, N. M., Villa, N., Waqar, Z., Sullivan, C., Gonzalez, L., Willsey, A. J., Choe, S.-Y., Neale, B. M., Daly, M. J., & State, M. W. (2015). No evidence for association of autism with rare heterozygous point mutations in Contactin-Associated Protein-Like 2 (CNTNAP2), or in Other Contactin-Associated Proteins or Contactins. *PLoS Genetics*, 11(1), e1004852. <https://doi.org/10.1371/journal.pgen.1004852>

Narita, A., Nagai, M., Mizuno, S., Ogishima, S., Tamiya, G., Ueki, M., Sakurai, R., Makino, S., Obara, T., Ishikuro, M., Yamanaka, C., Matsubara, H., Kuniyoshi, Y., Murakami, K., Ueno, F., Noda, A., Kobayashi, T., Kobayashi, M., Usuzaki, T., ... Kuriyama, S. (2020). Clustering by phenotype and genome-wide association study in autism. *Translational Psychiatry*, 10(1), 290. <https://doi.org/10.1038/s41398-020-00951-x>

Ng, P. C., Levy, S., Huang, J., Stockwell, T. B., Walenz, B. P., Li, K., Axelrod, N., Busam, D. A., Strausberg, R. L., & Venter, J. C. (2008). Genetic Variation in an Individual Human Exome. *PLoS Genetics*, 4(8), e1000160. <https://doi.org/10.1371/journal.pgen.1000160>

Nisar, H., Wajid, B., Shahid, S., Anwar, F., Wajid, I., Khatoon, A., Sattar, M. U., & Sadaf, S. (2021). Whole-genome sequencing as a first-tier diagnostic framework for rare genetic diseases. *Experimental Biology and Medicine*, 246(24), 2610. <https://doi.org/10.1177/15353702211040046>

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bizikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., ... Phillippy, A. M. (2022). The complete sequence of a human genome. *Science (New York, N. Y.)*, 376(6588), 44-53. <https://doi.org/10.1126/science.abj6987>

Ogawa, J., Lee, S., Itoh, K., Nagata, S., Machida, T., Takeda, Y., & Watanabe, K. (2001). Neural recognition molecule NB-2 of the contactin/F3 subgroup in rat : Specificity in neurite outgrowth-promoting activity and restricted expression in the brain regions. *Journal of*

*Neuroscience Research*, 65(2), 100-110. Scopus. <https://doi.org/10.1002/jnr.1133>

Oguro-Ando, A., Zuko, A., Kleijer, K. T. E., & Burbach, J. P. H. (2017). A current view on contactin-4, -5, and -6: Implications in neurodevelopmental disorders. *Molecular and Cellular Neurosciences*, 81, 72-83. <https://doi.org/10.1016/j.mcn.2016.12.004>

Parenti, I., Rabaneda, L. G., Schoen, H., & Novarino, G. (2020). Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends in Neurosciences*, 43(8), 608-621. <https://doi.org/10.1016/j.tins.2020.05.004>

Phelan, K., & McDermid, H. E. (2011). The 22q13.3 Deletion Syndrome (Phelan-McDermid Syndrome). *Molecular Syndromology*, 2(3-5), 186-201. <https://doi.org/10.1159/000334260>

Phelan, M. C. (2008). Deletion 22q13.3 syndrome. *Orphanet Journal of Rare Diseases*, 3(1), 14. <https://doi.org/10.1186/1750-1172-3-14>

Poliak, S., & Peles, E. (2003). The local differentiation of myelinated axons at nodes of Ranvier. *Nature Reviews. Neuroscience*, 4(12), 968-980. <https://doi.org/10.1038/nrn1253>

Poot, M., & Hochstenbach, R. (2021). Prevalence and Phenotypic Impact of Robertsonian Translocations. *Molecular Syndromology*, 12(1), 1-11. <https://doi.org/10.1159/000512676>

Pristyazhnyuk, I. E., & Menzorov, A. G. (2018). Ring chromosomes: From formation to clinical potential. *Protoplasma*, 255(2), 439-449. <https://doi.org/10.1007/s00709-017-1165-1>

Proust, C., Empana, J.-P., Boutouyrie, P., Alivon, M., Challande, P., Danchin, N., Escriou, G., Esslinger, U., Laurent, S., Li, Z., Pannier, B., Regnault, V., Thomas, F., Jouven, X., Cambien, F., & Lacolley, P. (2015). Contribution of Rare and Common Genetic Variants to Plasma Lipid Levels and Carotid Stiffness and Geometry: A Substudy of the Paris Prospective Study 3. *Circulation. Cardiovascular Genetics*, 8(4), 628-636. <https://doi.org/10.1161/CIRCGENETICS.114.000979>

Pruitt, K. D., & Maglott, D. R. (2001). RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Research*, 29(1), 137-140. <https://doi.org/10.1093/nar/29.1.137>

Querol, L., Delmont, E., & Lleixà, C. (2023). The autoimmune vulnerability of the node of Ranvier. *Journal of the Peripheral Nervous System: JPNS*, 28 Suppl 3, S12-S22. <https://doi.org/10.1111/jns.12570>

Radley, J. A., O'Sullivan, R. B. G., Turton, S. E., Cox, H., Vogt, J., Morton, J., Jones, E., Smithson, S., Lachlan, K., Rankin, J., Clayton-Smith, J., Willoughby, J., Elmslie, F. F., Sansbury, F. H., Cooper, N., Study, D. D. D. (DDD), & Balasubramanian, M. (2019). Deep phenotyping of 14 new patients with IQSEC2 variants, including monozygotic twins of discordant phenotype. *Clinical Genetics*, 95(4), 496-506. <https://doi.org/10.1111/cge.13507>

Rafi, S. K., & Butler, M. G. (2020). The 15q11.2 BP1-BP2 Microdeletion (Burnside-Butler) Syndrome: In Silico Analyses of the Four Coding Genes Reveal Functional Associations with Neurodevelopmental Phenotypes. *International Journal of Molecular Sciences*, 21(9), 3296. <https://doi.org/10.3390/ijms21093296>

Redaelli, S., Maitz, S., Crosti, F., Sala, E., Villa, N., Spaccini, L., Selicorni, A., Rigoldi, M., Conconi, D., Dalprà, L., Roversi, G., & Bentivegna, A. (2019). Refining the Phenotype of Recurrent Rearrangements of Chromosome 16. *International Journal of Molecular Sciences*, 20(5), 1095. <https://doi.org/10.3390/ijms20051095>

Reddy, K. S. (2005). Cytogenetic abnormalities and fragile-X syndrome in Autism

Spectrum Disorder. *BMC Medical Genetics*, 6, 3. <https://doi.org/10.1186/1471-2350-6-3>

Rentzsch, P., Schubach, M., Shendure, J., & Kircher, M. (2021). CADD-Splice—Improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Medicine*, 13(1), 31. <https://doi.org/10.1186/s13073-021-00835-9>

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD : Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1), D886-D894. <https://doi.org/10.1093/nar/gky1016>

Repnikova, E. A., Lyalin, D. A., McDonald, K., Astbury, C., Hansen-Kiss, E., Cooley, L. D., Pfau, R., Herman, G. E., Pyatt, R. E., & Hickey, S. E. (2020). CNTN6 copy number variations : Uncertain clinical significance in individuals with neurodevelopmental disorders. *European Journal of Medical Genetics*, 63(1), 103636. <https://doi.org/10.1016/j.ejmg.2019.02.008>

Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K. G., McKnight, D., Bai, R., Suchy, S., Friedman, B., Tahiliani, J., Pineda-Alvarez, D., Richard, G., Brandt, T., Haverfield, E., ... Bale, S. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 18(7), 696-704. <https://doi.org/10.1038/gim.2015.148>

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., & ACMG Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants : A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 17(5), 405-424. <https://doi.org/10.1038/gim.2015.30>

Rivera, H., & Domínguez, M. G. (2010). Variegated aneuploidy and ring chromosome syndromes overlap. *American Journal of Medical Genetics Part A*, 152A(1), 228-229. <https://doi.org/10.1002/ajmg.a.33195>

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), Article 1. <https://doi.org/10.1038/nbt.1754>

Rolland, T., Cliquet, F., Anney, R. J. L., Moreau, C., Traut, N., Mathieu, A., Huguet, G., Duan, J., Warriar, V., Portalier, S., Dry, L., Leblond, C. S., Douard, E., Amsellem, F., Malesys, S., Maruani, A., Toro, R., Børglum, A. D., Grove, J., ... Bourgeron, T. (2023). Phenotypic effects of genetic variants associated with autism. *Nature Medicine*, 29(7), Article 7. <https://doi.org/10.1038/s41591-023-02408-2>

Rosenberg, R. E., Law, J. K., Yenokyan, G., McGready, J., Kaufmann, W. E., & Law, P. A. (2009). Characteristics and Concordance of Autism Spectrum Disorders Among 277 Twin Pairs. *Archives of Pediatrics & Adolescent Medicine*, 163(10), 907-914. <https://doi.org/10.1001/archpediatrics.2009.98>

Rosenfeld, J. A., Coe, B. P., Eichler, E. E., Cuckle, H., & Shaffer, L. G. (2013). Estimates of penetrance for recurrent pathogenic copy-number variations. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 15(6), 478-481. <https://doi.org/10.1038/gim.2012.164>

Rossi, E., Messa, J., & Zuffardi, O. (2008). Ring syndrome : Still true? *Journal of Medical Genetics*, 45(11), 766-768. <https://doi.org/10.1136/jmg.2008.060442>

Roussignol, G., Ango, F., Romorini, S., Tu, J. C., Sala, C., Worley, P. F., Bockaert, J., & Fagni, L. (2005). Shank expression is sufficient to induce functional dendritic spine synapses in aspiny neurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(14), 3560-3570. <https://doi.org/10.1523/JNEUROSCI.4354-04.2005>

Rowe, L. R., Lee, J.-Y., Rector, L., Kaminsky, E. B., Brothman, A. R., Martin, C. L., & South, S. T. (2009). U-type exchange is the most frequent mechanism for inverted duplication with terminal deletion rearrangements. *Journal of Medical Genetics*, 46(10), 694-702. <https://doi.org/10.1136/jmg.2008.065052>

Rudenko, G. (2017). Dynamic Control of Synaptic Adhesion and Organizing Molecules in Synaptic Plasticity. *Neural Plasticity*, 2017, 6526151. <https://doi.org/10.1155/2017/6526151>

Runheim, H., Pettersson, M., Hammarsjö, A., Nordgren, A., Henriksson, M., Lindstrand, A., Levin, L.-Å., & Soller, M. J. (2023). The cost-effectiveness of whole genome sequencing in neurodevelopmental disorders. *Scientific Reports*, 13(1), 6904. <https://doi.org/10.1038/s41598-023-33787-8>

Samocha, K. E., Kosmicki, J. A., Karczewski, K. J., O'Donnell-Luria, A. H., Pierce-Hoffman, E., MacArthur, D. G., Neale, B. M., & Daly, M. J. (2017). *Regional missense constraint improves variant deleteriousness prediction* [Preprint]. Genomics. <https://doi.org/10.1101/148353>

Sánchez-Luquez, K. Y., Carpena, M. X., Karam, S. M., & Tovo-Rodrigues, L. (2022). The contribution of whole-exome sequencing to intellectual disability diagnosis and knowledge of underlying molecular mechanisms: A systematic review and meta-analysis. *Mutation Research/Reviews in Mutation Research*, 790, 108428. <https://doi.org/10.1016/j.mrrev.2022.108428>

Sanders, S. J., Ercan-Sencicek, A. G., Hus, V., Luo, R., Murtha, M. T., Moreno-DeLuca, D., Chu, S. H., Moreau, M. P., Gupta, A. R., Thomson, S. A., Mason, C. E., Bilguvar, K., Celestino-Soper, P. B. S., Choi, M., Crawford, E. L., Davis, L., Davis Wright, N. R., Dhodapkar, R. M., DiCola, M., ... State, M. W. (2011). Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron*, 70(5), 863-885. <https://doi.org/10.1016/j.neuron.2011.05.002>

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Hultman, C., Larsson, H., & Reichenberg, A. (2017). The Heritability of Autism Spectrum Disorder. *JAMA*, 318(12), 1182-1184. <https://doi.org/10.1001/jama.2017.12141>

Sarkar, A., Panati, K., & Narala, V. R. (2022). Code inside the codon: The role of synonymous mutations in regulating splicing machinery and its impact on disease. *Mutation Research. Reviews in Mutation Research*, 790, 108444. <https://doi.org/10.1016/j.mrrev.2022.108444>

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., Cordier, M.-P., Coubes, C., Demeer, B., Chaussonot, A., Demurger, F., Devillard, F., Doco-Fenzy, M., ... Sanlaville, D. (2019). Whole genome paired-end sequencing elucidates functional and phenotypic consequences of balanced chromosomal rearrangement in patients with developmental disorders. *Journal of Medical Genetics*, 56(8), 526-535. <https://doi.org/10.1136/jmedgenet-2018-105778>

Schmilovich, Z., Huguet, G., He, Q., Musa-Johnson, A., Douard, E., Loum, M. A., Liao, C., Ross, J. P., Dionne-Laporte, A., Spiegelman, D., Jean-Louis, M., Saci, Z., Hayward,

C., Banaschewski, T., Bokde, A., Desrivieres, S., Lemaitre, H., Schumann, G., Xiong, L., ... Rouleau, G. A. (2022). Copy-number variants in the contactin-5 gene are a potential risk factor for autism spectrum disorder. *Research in Autism Spectrum Disorders*, 99, 102055. <https://doi.org/10.1016/j.rasd.2022.102055>

Schneider, S. A., Mueller, C., Biskup, S., Fietzek, U. M., & Schroeder, A. S. (2022). Neurodevelopmental disorder with dystonia due to SOX6 mutations. *Molecular Genetics & Genomic Medicine*, 10(12), e2051. <https://doi.org/10.1002/mgg3.2051>

Schön, M., Lapunzina, P., Nevado, J., Mattina, T., Gunnarsson, C., Hadzsiev, K., Verpelli, C., Bourgeron, T., Jesse, S., Van Ravenswaaij-Arts, C. M. A., & Hennekam, R. C. (2023). Definition and clinical variability of SHANK3-related Phelan-McDermid syndrome. *European Journal of Medical Genetics*, 66(7), 104754. <https://doi.org/10.1016/j.ejmg.2023.104754>

Scocchia, A., Wigby, K. M., Masser-Frye, D., Del Campo, M., Galarreta, C. I., Thorpe, E., McEachern, J., Robinson, K., Gross, A., ICSL Interpretation and Reporting Team, Ajay, S. S., Rajan, V., Perry, D. L., Belmont, J. W., Bentley, D. R., Jones, M. C., & Taft, R. J. (2019). Clinical whole genome sequencing as a first-tier test at a resource-limited dysmorphology clinic in Mexico. *NPJ Genomic Medicine*, 4, 5. <https://doi.org/10.1038/s41525-018-0076-1>

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.-H., Hicks, J., Spence, S. J., Lee, A. T., Puura, K., Lehtimäki, T., ... Wigler, M. (2007). Strong Association of De Novo Copy Number Mutations with Autism. *Science*, 316(5823), 445-449. <https://doi.org/10.1126/science.1138659>

Sehn, J. K. (2015). Chapter 9—Insertions and Deletions (Indels). In S. Kulkarni & J. Pfeifer (Éds.), *Clinical Genomics* (p. 129-150). Academic Press. <https://doi.org/10.1016/B978-0-12-404748-8.00009-5>

Shen, Y.-M., Chan, B. S. M., Liu, J.-B., Zhou, Y.-Y., Cui, X.-L., He, Y.-Q., Fang, Y., Xiang, Y.-T., & Luo, X.-R. (2018). The prevalence of psychiatric disorders among students aged 6~16 years old in central Hunan, China. *BMC Psychiatry*, 18(1), 243. <https://doi.org/10.1186/s12888-018-1823-7>

Sheng, M., & Kim, E. (2000). The Shank family of scaffold proteins. *Journal of Cell Science*, 113 ( Pt 11), 1851-1856. <https://doi.org/10.1242/jcs.113.11.1851>

Shimoda, Y., & Watanabe, K. (2009). Contactins. *Cell Adhesion & Migration*, 3(1), 64-70.

Singh, V., van Pelt, E. D., Stoop, M. P., Stingl, C., Ketelslegers, I. A., Neuteboom, R. F., Catsman-Berrevoets, C. E., Luijck, T. M., & Hintzen, R. Q. (2015). Gray matter-related proteins are associated with childhood-onset multiple sclerosis. *Neurology(R) Neuroimmunology & Neuroinflammation*, 2(5), e155. <https://doi.org/10.1212/NXI.0000000000000155>

Smajlagic, D., Lavrichenko, K., Berland, S., Helgeland, Ø., Knudsen, G. P., Vaudel, M., Haavik, J., Knappskog, P. M., Njølstad, P. R., Houge, G., & Johansson, S. (2020). Population prevalence and inheritance pattern of recurrent CNVs associated with neurodevelopmental disorders in 12,252 newborns and their parents. *European Journal of Human Genetics*, 1-11. <https://doi.org/10.1038/s41431-020-00707-7>

Smogavec, M., Cleall, A., Hoyer, J., Lederer, D., Nassogne, M.-C., Palmer, E. E., Deprez, M., Benoit, V., Maystadt, I., Noakes, C., Leal, A., Shaw, M., Gecz, J., Raymond, L., Reis, A., Shears, D., Brockmann, K., & Zweier, C. (2016). Eight further individuals with

intellectual disability and epilepsy carrying bi-allelic CNTNAP2 aberrations allow delineation of the mutational and phenotypic spectrum. *Journal of Medical Genetics*, 53(12), 820-827. <https://doi.org/10.1136/jmedgenet-2016-103880>

Smol, T., Thuillier, C., Boudry-Labis, E., Dieux-Coeslier, A., Duban-Bedu, B., Caumes, R., Bouquillon, S., Manouvrier-Hanu, S., Roche-Lestienne, C., & Ghoumid, J. (2020). Neurodevelopmental phenotype associated with CHD8-SUPT16H duplication. *Neurogenetics*, 21(1), 67-72. <https://doi.org/10.1007/s10048-019-00599-w>

Souche, E., Beltran, S., Brosens, E., Belmont, J. W., Fossum, M., Riess, O., Gilissen, C., Ardeshirdavani, A., Houge, G., van Gijn, M., Clayton-Smith, J., Synofzik, M., de Leeuw, N., Deans, Z. C., Dincer, Y., Eck, S. H., van der Crabben, S., Balasubramanian, M., Graessner, H., ... Weiss, M. M. (2022). Recommendations for whole genome sequencing in diagnostics for rare diseases. *European Journal of Human Genetics*, 30(9), Article 9. <https://doi.org/10.1038/s41431-022-01113-x>

SPARK Consortium. (2018). SPARK : A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron*, 97(3), 488-493. <https://doi.org/10.1016/j.neuron.2018.01.015>

Srivastava, S., Love-Nichols, J. A., Dies, K. A., Ledbetter, D. H., Martin, C. L., Chung, W. K., Firth, H. V., Frazier, T., Hansen, R. L., Prock, L., Brunner, H., Hoang, N., Scherer, S. W., Sahin, M., & Miller, D. T. (2019). Meta-analysis and multidisciplinary consensus statement : Exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genetics in Medicine*, 21(11), 2413-2421. <https://doi.org/10.1038/s41436-019-0554-6>

Stankiewicz, P., & Lupski, J. R. (2010). Structural Variation in the Human Genome and its Role in Disease. *Annual Review of Medicine*, 61(1), 437-455. <https://doi.org/10.1146/annurev-med-100708-204735>

Stark, Z., Tan, T. Y., Chong, B., Brett, G. R., Yap, P., Walsh, M., Yeung, A., Peters, H., Mordaunt, D., Cowie, S., Amor, D. J., Savarirayan, R., McGillivray, G., Downie, L., Ekert, P. G., Theda, C., James, P. A., Yapfitee-Lee, J., Ryan, M. M., ... White, S. M. (2016). A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 18(11), 1090-1096. <https://doi.org/10.1038/gim.2016.1>

Stefanski, A., Calle-López, Y., Leu, C., Pérez-Palma, E., Pestana-Knight, E., & Lal, D. (2021). Clinical sequencing yield in epilepsy, autism spectrum disorder, and intellectual disability: A systematic review and meta-analysis. *Epilepsia*, 62(1), 143-151. <https://doi.org/10.1111/epi.16755>

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Walters, G. B., Jonsdottir, G. A., Doyle, O. M., Tost, H., Grimm, O., Kristjansdottir, S., Snorrason, H., Davidsdottir, S. R., Gudmundsson, L. J., Jonsson, G. F., Stefansdottir, B., Helgadottir, I., ... Stefansson, K. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*, 505(7483), 361-366. <https://doi.org/10.1038/nature12818>

Stoeckli, E. T., Kuhn, T. B., Duc, C. O., Ruegg, M. A., & Sonderegger, P. (1991). The axonally secreted protein axonin-1 is a potent substratum for neurite growth. *The Journal of Cell Biology*, 112(3), 449-455. <https://doi.org/10.1083/jcb.112.3.449>

Straub, L., Bateman, B. T., Hernandez-Diaz, S., York, C., Lester, B., Wisner, K. L., McDougle, C. J., Pennell, P. B., Gray, K. J., Zhu, Y., Suarez, E. A., Mogun, H., & Huybrechts, K. F. (2022). Neurodevelopmental Disorders Among Publicly or Privately Insured Children in

the United States. *JAMA Psychiatry*, 79(3), 232-242.  
<https://doi.org/10.1001/jamapsychiatry.2021.3815>

Südhof, T. C. (2008). Neuroligins and neuroligins link synaptic function to cognitive disease. *Nature*, 455(7215), 903-911. <https://doi.org/10.1038/nature07456>

Sun, X., Zhou, Y., Zhang, R., Wang, Z., Xu, M., Zhang, D., Huang, J., Luo, F., Li, F., Ni, Z., Zhou, S., Chen, H., Chen, S., Chen, L., Du, X., Chen, B., Huang, H., Liu, P., Yin, L., ... Chen, L. (2020). Dstyk mutation leads to congenital scoliosis-like vertebral malformations in zebrafish via dysregulated mTORC1/TFEB pathway. *Nature Communications*, 11(1), 479. <https://doi.org/10.1038/s41467-019-14169-z>

Tabet, A.-C., Rolland, T., Ducloy, M., Lévy, J., Buratti, J., Mathieu, A., Haye, D., Perrin, L., Dupont, C., Passemard, S., Capri, Y., Verloes, A., Drunat, S., Keren, B., Mignot, C., Marey, I., Jacquette, A., Whalen, S., Pipiras, E., ... Bourgeron, T. (2017). A framework to identify contributing genes in patients with Phelan-McDermid syndrome. *Npj Genomic Medicine*, 2(1), 32. <https://doi.org/10.1038/s41525-017-0035-2>

Tärnlungeanu, D. C., & Novarino, G. (2018). Genomics in neurodevelopmental disorders: An avenue to personalized medicine. *Experimental & Molecular Medicine*, 50(8), 1-7. <https://doi.org/10.1038/s12276-018-0129-7>

Thapar, A., Cooper, M., & Rutter, M. (2017). Neurodevelopmental disorders. *The Lancet. Psychiatry*, 4(4), 339-346. [https://doi.org/10.1016/S2215-0366\(16\)30376-5](https://doi.org/10.1016/S2215-0366(16)30376-5)

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijdsdijk, F. (2016). Heritability of autism spectrum disorders: A meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 57(5), 585-595. <https://doi.org/10.1111/jcpp.12499>

Tolchin, D., Yeager, J. P., Prasad, P., Dorrani, N., Russi, A. S., Martinez-Agosto, J. A., Haseeb, A., Angelozzi, M., Santen, G. W. E., Ruivenkamp, C., Mercimek-Andrews, S., Depienne, C., Kuechler, A., Mikat, B., Ludecke, H.-J., Bilan, F., Le Guyader, G., Gilbert-Dussardier, B., Keren, B., ... Le Caignec, C. (2020). De Novo SOX6 Variants Cause a Neurodevelopmental Syndrome Associated with ADHD, Craniosynostosis, and Osteochondromas. *The American Journal of Human Genetics*, 106(6), 830-845. <https://doi.org/10.1016/j.ajhg.2020.04.015>

Toma, C., Pierce, K. D., Shaw, A. D., Heath, A., Mitchell, P. B., Schofield, P. R., & Fullerton, J. M. (2018). Comprehensive cross-disorder analyses of CNTNAP2 suggest it is unlikely to be a primary risk gene for psychiatric disorders. *PLoS Genetics*, 14(12), e1007535. <https://doi.org/10.1371/journal.pgen.1007535>

Traut, W., Weichenhan, D., Himmelbauer, H., & Winking, H. (2006). New members of the neuroligin superfamily: Multiple rodent homologues of the human CASPR5 gene. *Mammalian Genome*, 17(7), 723-731. <https://doi.org/10.1007/s00335-005-0157-1>

Tsilchorozidou, T., Menko, F. H., Laloo, F., Kidd, A., De Silva, R., Thomas, H., Smith, P., Malcolmson, A., Dore, J., Madan, K., Brown, A., Yovos, J. G., Tsaligopoulos, M., Vogiatzis, N., Baser, M. E., Wallace, A. J., & Evans, D. G. R. (2004). Constitutional rearrangements of chromosome 22 as a cause of neurofibromatosis 2. *Journal of Medical Genetics*, 41(7), 529-534. <https://doi.org/10.1136/jmg.2003.016774>

Tsuboyama, M., & Iqbal, M. A. (2021). CHL1 deletion is associated with cognitive and language disabilities—Case report and review of literature. *Molecular Genetics & Genomic Medicine*, 9(7), e1725. <https://doi.org/10.1002/mgg3.1725>

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—New capabilities and interfaces. *Nucleic Acids Research*,



40(15), e115. <https://doi.org/10.1093/nar/gks596>

van der Sanden, B. P. G. H., Schobers, G., Corominas Galbany, J., Koolen, D. A., Sinnema, M., van Reeuwijk, J., Stumpel, C. T. R. M., Kleefstra, T., de Vries, B. B. A., Ruitenkamp-Versteeg, M., Leijsten, N., Kwint, M., Derks, R., Swinkels, H., den Ouden, A., Pfundt, R., Rinne, T., de Leeuw, N., Stegmann, A. P., ... Vissers, L. E. L. M. (2023). The performance of genome sequencing as a first-tier test for neurodevelopmental disorders. *European Journal of Human Genetics*, 31(1), Article 1. <https://doi.org/10.1038/s41431-022-01185-9>

Vasudevan, P., & Suri, M. (2017). A clinical approach to developmental delay and intellectual disability. *Clinical Medicine (London, England)*, 17(6), 558-561. <https://doi.org/10.7861/clinmedicine.17-6-558>

Velinov, M. (2019). Genomic Copy Number Variations in the Autism Clinic—Work in Progress. *Frontiers in Cellular Neuroscience*, 13. <https://www.frontiersin.org/articles/10.3389/fncel.2019.00057>

Vieira, M. M., Jeong, J., & Roche, K. W. (2021). The role of NMDA receptor and neuroligin rare variants in synaptic dysfunction underlying neurodevelopmental disorders. *Current Opinion in Neurobiology*, 69, 93-104. <https://doi.org/10.1016/j.conb.2021.03.001>

Vissers, L. E. L. M., van Nimwegen, K. J. M., Schieving, J. H., Kamsteeg, E.-J., Kleefstra, T., Yntema, H. G., Pfundt, R., van der Wilt, G. J., Krabbenborg, L., Brunner, H. G., van der Burg, S., Grutters, J., Veltman, J. A., & Willemsen, M. A. A. P. (2017). A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 19(9), 1055-1063. <https://doi.org/10.1038/gim.2017.1>

Vitrac, A., Leblond, C. S., Rolland, T., Cliquet, F., Mathieu, A., Maruani, A., Delorme, R., Schön, M., Grabrucker, A. M., van Ravenswaaij-Arts, C., Phelan, K., Tabet, A.-C., & Bourgeron, T. (2023). Dissecting the 22q13 region to explore the genetic and phenotypic diversity of patients with Phelan-McDermid syndrome. *European Journal of Medical Genetics*, 66(5), 104732. <https://doi.org/10.1016/j.ejmg.2023.104732>

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), e164. <https://doi.org/10.1093/nar/gkq603>

Warburton, P. E., & Sebra, R. P. (2023). Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. *Annual Review of Genomics and Human Genetics*, 24, 109-132. <https://doi.org/10.1146/annurev-genom-101722-103045>

Weischenfeldt, J., Symmons, O., Spitz, F., & Korbel, J. O. (2013). Phenotypic impact of genomic structural variation: Insights from and for human disease. *Nature Reviews Genetics*, 14(2), Article 2. <https://doi.org/10.1038/nrg3373>

Wheeler, D. L., Chappey, C., Lash, A. E., Leipe, D. D., Madden, T. L., Schuler, G. D., Tatusova, T. A., & Rapp, B. A. (2000). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 28(1), 10-14. <https://doi.org/10.1093/nar/28.1.10>

Willfors, C., Carlsson, T., Anderlid, B.-M., Nordgren, A., Kostrzewa, E., Berggren, S., Ronald, A., Kuja-Halkola, R., Tammimies, K., & Bölte, S. (2017). Medical history of discordant twins and environmental etiologies of autism. *Translational Psychiatry*, 7(1), Article 1. <https://doi.org/10.1038/tp.2016.269>

Wiszniewski, W., Gawlinski, P., Gambin, T., Bekiesinska-Figatowska, M., Obersztyn,

E., Antczak-Marach, D., Akdemir, Z. H. C., Harel, T., Karaca, E., Jurek, M., Sobecka, K., Nowakowska, B., Kruk, M., Terczynska, I., Goszczanska-Ciuchta, A., Rudzka-Dybala, M., Jamroz, E., Pyrkosz, A., Jakubiuk-Tomaszuk, A., ... Lupski, J. R. (2018). Comprehensive genomic analysis of patients with disorders of cerebral cortical development. *European Journal of Human Genetics: EJHG*, 26(8), 1121-1131. <https://doi.org/10.1038/s41431-018-0137-z>

Wolkenstein, P. (2001). La neurofibromatose 1. *médecine/sciences*, 17(11), Article 11. <https://doi.org/10.1051/medsci/200117111158>

Wright, M., De Geus, E., Ando, J., Luciano, M., Posthuma, D., Ono, Y., Hansell, N., Van Baal, C., Hiraishi, K., Hasegawa, T., Smith, G., Geffen, G., Geffen, L., Kanba, S., Miyake, A., Martin, N., & Boomsma, D. (2001). Genetics of cognition : Outline of a collaborative twin study. *Twin Research: The Official Journal of the International Society for Twin Studies*, 4(1), 48-56. <https://doi.org/10.1375/1369052012146>

Xu, J., & Chen, Z. (2003). Advances in molecular cytogenetics for the evaluation of mental retardation. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 117C(1), 15-24. <https://doi.org/10.1002/ajmg.c.10016>

Yamakawa, T., Zhang, G., Najjar, L. B., Li, C., & Itakura, K. (2023). The uncharacterized transcript KIAA0930 confers a cachexic phenotype on cancer cells. *Oncotarget*, 14, 723-737. <https://doi.org/10.18632/oncotarget.28476>

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST : A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134. <https://doi.org/10.1186/1471-2105-13-134>

Yoshihara, Y., Kawasaki, M., Tamada, A., Nagata, S., Kagamiyama, H., & Mori, K. (1995). Overlapping and differential expression of BIG-2, BIG-1, TAG-1, and F3 : Four members of an axon-associated cell adhesion molecule subgroup of the immunoglobulin superfamily. *Journal of Neurobiology*, 28(1), 51-69. <https://doi.org/10.1002/neu.480280106>

Yumoto, T., Kimura, M., Nagatomo, R., Sato, T., Utsunomiya, S., Aoki, N., Kitaura, M., Takahashi, K., Takemoto, H., Watanabe, H., Okano, H., Yoshida, F., Nao, Y., & Tomita, T. (2020). Autism-associated variants of neuroligin 4X impair synaptogenic activity by various molecular mechanisms. *Molecular Autism*, 11(1), 68. <https://doi.org/10.1186/s13229-020-00373-y>

Zablotsky, B., Black, L. I., Maenner, M. J., Schieve, L. A., Danielson, M. L., Bitsko, R. H., Blumberg, S. J., Kogan, M. D., & Boyle, C. A. (2019). Prevalence and Trends of Developmental Disabilities among Children in the United States : 2009-2017. *Pediatrics*, 144(4), e20190811. <https://doi.org/10.1542/peds.2019-0811>

Zablotsky, B., Bramlett, M. D., Visser, S. N., Danielson, M. L., & Blumberg, S. J. (2018). Latent Class Analysis of ADHD Neurodevelopmental and Mental Health Comorbidities. *Journal of Developmental & Behavioral Pediatrics*, 39(1), 10. <https://doi.org/10.1097/DBP.0000000000000508>

Zeidan, J., Fombonne, E., Scora, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism : A systematic review update. *Autism Research: Official Journal of the International Society for Autism Research*, 15(5), 778-790. <https://doi.org/10.1002/aur.2696>

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics*, 10, 451-481. <https://doi.org/10.1146/annurev.genom.9.081307.164217>

Zhang, S. Q., Fleischer, J., Al-Kateb, H., Mito, Y., Amarillo, I., & Shinawi, M. (2020).

Intragenic CNTN4 copy number variants associated with a spectrum of neurobehavioral phenotypes. *European Journal of Medical Genetics*, 63(3), 103736. <https://doi.org/10.1016/j.ejmg.2019.103736>

Zhao, W.-W., Wu, M., Chen, F., Jiang, S., Su, H., Liang, J., Deng, C., Hu, C., & Yu, S. (2015). Robertsonian Translocations: An Overview of 872 Robertsonian Translocations Identified in a Diagnostic Laboratory in China. *PLOS ONE*, 10(5), e0122647. <https://doi.org/10.1371/journal.pone.0122647>

Ziats, C. A., Jain, L., McLarney, B., Vandenboom, E., DuPont, B. R., Rogers, C., Sarasua, S., Nevado, J., Cordisco, E. L., Phelan, K., & Boccutto, L. (2020). Neurofibromatosis type 2 in Phelan-McDermid syndrome: Institutional experience and review of the literature. *European Journal of Medical Genetics*, 63(11), 104042. <https://doi.org/10.1016/j.ejmg.2020.104042>

Zirn, B., Arning, L., Bartels, I., Shoukier, M., Hoffjan, S., Neubauer, B., & Hahn, A. (2012). Ring chromosome 22 and neurofibromatosis type II: Proof of two-hit model for the loss of the NF2 gene in the development of meningioma. *Clinical Genetics*, 81(1), 82-87. <https://doi.org/10.1111/j.1399-0004.2010.01598.x>

Zou, Y., Zhang, W.-F., Liu, H.-Y., Li, X., Zhang, X., Ma, X.-F., Sun, Y., Jiang, S.-Y., Ma, Q.-H., & Xu, D.-E. (2017). Structure and function of the contactin-associated protein family in myelinated axons and their relationship with nerve diseases. *Neural Regeneration Research*, 12(9), 1551-1558. <https://doi.org/10.4103/1673-5374.215268>

Zuffardi, O., Bonaglia, M., Ciccone, R., & Giorda, R. (2009). Inverted duplications deletions: Underdiagnosed rearrangements?? *Clinical Genetics*, 75(6), 505-513. <https://doi.org/10.1111/j.1399-0004.2009.01187.x>

Zuko, A., Kleijer, K. T. E., Oguro-Ando, A., Kas, M. J. H., van Daalen, E., van der Zwaag, B., & Burbach, J. P. H. (2013). Contactins in the neurobiology of autism. *European Journal of Pharmacology*, 719(1), 63-74. <https://doi.org/10.1016/j.ejphar.2013.07.016>

# Table des illustrations

- Figure 1.** Différentes techniques de diagnostic en génétique
- Figure 2.** Sélection des gènes pour les *High Confidence* et *Candidate NDD genes*
- Figure 3.** Schématisation d'une synapse
- Figure 4.** Organisation de la chromatine et des nucléosomes
- Figure 5.** Schéma de la protéine mTOR et des complexes mTORC1 et mTORC2
- Figure 6.** Relation entre fond génétique et variation génétique dans l'expression des TSA
- Figure 7.** Schéma des domaines protéiques de *SHANK3*
- Figure 8.** Protéines SHANK et leurs interactions dans la synapse glutamatergique
- Figure 9.** Score de QI rapporté à l'ADI-R chez des patients autiste selon le gène
- Figure 10.** Répartition des délétions 22qter chez des patients PMS
- Figure 11.** Répartition des mutations ponctuelles et indel sur la protéine *SHANK3*
- Figure 12.** *CNTN1* et *CNTN2* au niveau des régions paranodales et juxtaparanodales
- Figure 13.** Schéma et image d'une translocation robertsonienne
- Figure 14.** Création de microremaniements par NAHR
- Figure 15.** Mécanisme de réparation par NHEJ
- Figure 16.** Mécanismes de réparation par FoSTeS et MMBIR
- Figure 17.** Formation d'un chromosome dicentrique aboutissant à une inv dup del
- Figure 18.** Schéma de différentes situations aboutissant à la formation d'un anneau
- Figure 19.** Critères d'évaluation de la pathogénicité d'une variation génomique selon l'ACMG
- Figure 20.** Les différentes étapes d'une technique de séquençage haut débit (technique Illumina)
- Figure 21.** Séquençage en paired-end
- Figure 22.** Représentation des reads fractionnés et discordants avant et après alignement
- Figure 23.** Cohorte des patients avec PMS
- Figure 24.** Principe d'analyse des points de cassure en 22qter
- Figure 25.** Représentation des analyses génétiques réalisées
- Figure 26.** Différentes visualisations et schématisation du chromosome 22 de P1
- Figure 27.** Caryotype du patient P1 avec anneau du chromosome 22
- Figure 28.** Première hypothèse de formation de l'anneau complexe de P1

- Figure 29.** Test d'amplification de la séquence chimérique
- Figure 30.** Proposition finale de mécanisme de formation de l'anneau complexe chez le patient P1
- Figure 31.** Image de caryotype et FISH montrant le dérivé 13 et la duplication 22qter en mosaïque
- Figure 32.** Visualisation du CRAM sur IGV au niveau du point de cassure 22qter
- Figure 33.** Point de cassure 22qter avec dérivé d'inversion
- Figure 34.** Région 22qter avec représentation des séquences répétées aux points de cassure terminaux
- Figure 35.** Représentation heatmap du phénotype des 7 patients
- Figure 36.** Profil Sanger de la mutation c.1048C>T de *TYMP*.
- Figure 37.** Présentation de la cohorte de patients avec PMS
- Figure 38.** Profil clinique des patients avec PMS
- Figure 39.** Illustration des délétions et variations ponctuelles SHANK3 des patients avec PMS
- Figure 40.** Multiple hits des patients PMS retrouvés dans la région 22q13 sur le chromosome controlatéral
- Figure 41.** Répartition des tailles des CNVs identifiés
- Figure 42.** Analyse d'un profil de puce SNP (Genome Studio)
- Figure 43.** Répartition des CNV identifiés selon le gène
- Figure 44.** Nombre de gènes impliqués dans les CNVs identifiés
- Figure 45.** Odd Ratio d'être porteur d'un CNV selon les groupes phénotypiques
- Figure 46.** Odd Ratio d'être porteur d'un CNV touchant un gènes des contactines 1 à 6 selon les groupes phénotypiques
- Figure 47.** Nombre de CNVs identifiés par gène et selon le phénotype
- Figure 48.** Nombre de délétions identifiées par gène et selon le phénotype
- Figure 49** Méta-analyse des délétions et duplications des deux familles de gènes pour notre cohorte, la SPARK et la SSC
- Figure 50** Méta-analyse des délétions et duplications des gènes Contactines pour notre cohorte, la SPARK et la SSC
- Figure 51.** Méta-analyse des délétions et duplications des gènes *Contactin associated proteins* pour notre cohorte, la SPARK et la SSC

# Tables

- Table 1.** Caractéristiques des gènes des contactines et contactines *associated-proteins*
- Table 2.** Amorces utilisées pour séquençage Sanger
- Table 3.** Couples d'amorces utilisées pour la séquence chimérique
- Table 4.** Nombre et type d'éléments répétés au niveau des points de cassure 22q
- Table 5.** Variations additionnelles (version du génome : GRCh37)
- Table 6.** Groupe phénotypique pour la cohorte
- Table 7.** Répartition homme/femme par catégorie et sex-ratio
- Table 8.** Nombre de sondes couvrant les différents gènes sur les différents projets de puce utilisés dans les trois cohortes
- Table 9.** Répartition des CNVs selon le type et le groupe phénotypique