

Les analyses pangénomiques dans l'exploration génétique de la déficience intellectuelle : de la recherche de gènes candidats du syndrome d'Aicardi, à la caractéristation du spectre mutationnel des gènes IL1RAPL1 et MBD5

Asma Ali Khan

▶ **To cite this version:**

Asma Ali Khan. Les analyses pangénomiques dans l'exploration génétique de la déficience intellectuelle : de la recherche de gènes candidats du syndrome d'Aicardi, à la caractéristation du spectre mutationnel des gènes IL1RAPL1 et MBD5. Médecine humaine et pathologie. Université de Lorraine, 2012. Français. NNT : 2012LORR0147 . tel-01749343

HAL Id: tel-01749343
https://hal.univ-lorraine.fr/tel-01749343

Submitted on 29 Mar 2018

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.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Ecole Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITÉ DE LORRAINE

Mention : « Sciences de la Vie et de la Santé »

par Asma Ali KHAN

Les analyses pangénomiques dans l'exploration génétique de la déficience intellectuelle : de la recherche de gènes candidats du syndrome d'Aicardi, à la caractérisation du spectre mutationnel des gènes *IL1RAPL1* et *MBD5*

Date de soutenance : 13 Novembre 2012

Membres du jury :
Rapporteurs :

Pr. Jean-Luc BRESSON	Génétique, Histologie - Biologie du Développement et de la reproduction, Université de Franche-Comté, Besançon
Dr. Elisabeth FLORI	Cytogénétique, Hôpitaux Universitaires de Strasbourg, Strasbourg

Examinateurs :

Pr. Bruno LEHEUP	Génétique Humaine, Université de Lorraine, Nancy
Pr. Philippe JONVEAUX	Génétique Humaine, Université de Lorraine, Nancy (Directeur de thèse)

Membres invités:

Dr. Marie-Dominique DEVIGNES	LORIA, CNRS, Université de Lorraine, Nancy
Dr. Christophe NEMOS	Génétique Humaine, Université de Lorraine, Nancy (Co-Directeur de thèse)

Laboratoire de génétique-EA4368-IFR 111 Déficiences mentales et anomalies de structure du génome, CHU de Nancy Rue du Morvan 54511 Vandoeuvre-les-Nancy

Remerciements

Mes remerciements les plus sincères vont à Monsieur le Professeur Philippe JONVEAUX, mon directeur de thèse : merci Monsieur de m’avoir accueillie dans votre équipe, et de m'avoir accordé votre confiance. Je vous suis également reconnaissante pour votre encadrement scientifique, votre gentillesse et votre disponibilité, ainsi que vos précieux conseils et votre aide tout au long de mon stage. Votre patience, votre motivation, votre immense savoir sur le génétique humaine ainsi que vos encouragements m’ont aidé dans la rédaction de cette thèse.

Je souhaite exprimer ma plus profonde reconnaissance à Madame le Docteur Elisabeth FLORI d’avoir accepté la tache d’être rapporteur de cette thèse.

Je remercie très sincèrement Monsieur le Professeur Jean-Luc BRESSON pour l'honneur qu'il me fait en acceptant d'évaluer ce travail, et d'être rapporteur de thèse.

J’exprime toute ma reconnaissance à Monsieur le Professeur Bruno LEHEUP qui m’a fait l’honneur de présider le jury de cette thèse, d’examiner mes travaux et pour son intérêt et son soutien chaleureux.

Je tiens à exprimer mes sincères remerciements à Madame le Docteur Marie-Dominique DEVIGNES et à Emmanuel BRESSO de l’équipe du LORIA pour leur collaboration et pour les études bioinformatiques.

Toute ma profonde gratitude à Madame le Docteur Céline BONNET pour son aide, ses conseils, son encadrement, sa disponibilité, son savoir-faire inestimable. Cette thèse n'aurait pas été possible sans son aide et son soutien, à la fois au niveau professionnel et au niveau personnel. Je lui suis très reconnaissante.

Mes remerciements très chaleureux vont aussi à Madame Marie-José GREGOIRE, pour son aide à la correction de ce manuscrit et pour ses encouragements.

Je tiens à remercier Madame Brigitte LEOTARD pour ses chaleureux conseils avisés, son soutien et ses encouragements.

J’exprime toute ma reconnaissance à Madame le Docteur Mylène BERI-DEXHEIMER pour sa gentillesse, sa disponibilité et ses conseils avisés.

Je tiens à remercier Monsieur le Docteur Christophe PHILIPPE, qui accorde toujours gentiment son temps pour répondre à toutes mes questions, j'ai également bénéficié de ses conseils.

Je remercie également Madame le Docteur Joanna SOKOLOWSKA-GILLOIS, Madame Fabienne FRANCOIS et Madame le Docteur Myriam BRONNER pour leur accueil chaleureux et leur bonne humeur.

Je tiens également à remercier Madame le Docteur Saliha YILMAZ qui a travaillé avant moi sur le syndrome d’Aicardi.

Mes remerciements s’adressent aux médecins et techniciennes du secteur Cytogénétique : Madame le Docteur Mylène VALDUGA, Monsieur le Docteur Marc MULLER, Stéphanie, Marie, Marie-Pierre, Sylvie et particulièrement Séverine pour la réalisation des analyses par CGH-array.

Je tiens également à exprimer mes remerciements amicaux aux techniciens du secteur oncogénique : Carole, Evelyne, Christelle, Claire, Didier qui m’ont toujours m’intégrée dans le planning très chargé du séquenceur ABI 3130.

Que soient également remerciés l’ensemble du personnel du laboratoire pour son accueil chaleureux et sa courtoisie :

-au secteur génétique constitutionnelle : Madame le Docteur Lætitia LAMBERT, Madame le Docteur Virginie ROTH, Peggy et Aïssa et particulièrement Aline pour son aide dans des manipulations génétiques moléculaires.

-au secteur hématologie : Nathalie pour son aide sur la RT-PCR, Alice, Emeline, Valérie, Julie et Christelle.

-au secteur prénatal et postnatal : Mireille, Sophie, Brigitte, Mélina, Mathieu, Elise, Mylène, Chantal.
-aux secrétaires : Séverine, Magalie, Clarisse et Claire.

Je garderai une place particulière pour Lila ALLOU et Minh T. HUYNH que j'ai côtoyés tout au long de mon stage et qui m'ont été d'une grande sympathie.

Cette thèse n'aurait pu être réalisée sans le soutien financier de la Higher Education Commission of Pakistan (HEC) et sans le suivi administratif de SFERE : par Madame Anna MANETA et Madame Somia ASHRAF qui sont toujours là pour résoudre les problèmes administratifs en France.

Mes remerciements les plus profonds vont à un grand visionnaire Monsieur le Professeur Atta-ur-Rahman, de la HEC du Pakistan pour avoir eu l'audace et la vision de concevoir notre programme de bourses en dépit des nombreuses critiques.

Je voudrais également remercier les gens du Pakistan pour le parrainage de nos études, en dépit de nos ressources limitées. J'espère être en mesure, un jour, de redonner une partie de celui-ci à la nation.

Mes parents méritent une mention spéciale pour leur soutien et leurs prières.

Table des matières

REMERCIEMENTS 2

TABLE DES FIGURES 7

LISTE DES TABLEAUX 9

LISTE DES ABREVIATIONS 10

PREAMBULE ET OBJECTIF DU TRAVAIL DE THESE 16

1. INTRODUCTION 18

1.1 La déficience intellectuelle 18
1.1.1 Classification d'une déficience intellectuelle 18
1.1.2 Prévalence de la DI 19
1.1.3 Etiologie de la DI 20
1.1.4 La déficience intellectuelle liée au chromosome X (DILX) 28
1.1.5 Déficience intellectuelle liée aux autosomes 41
1.1.6 Physiopathologie des DILX 42

1.2. Les stratégies pour la recherche de gène candidat dans la DI 48
1.2.1 Cartographie génétique 48
1.2.2 Le caryotype et l'hybridation in situ en fluorescence 49
1.2.3 L'hybridation génomique comparative sur microréseau (CGH-array) 49
1.2.4 Séquençage de nouvelle génération / séquençage haut débit 50
1.2.5 La recherche de gène candidat par l'analyse des données du Séquençage haut débit dans les maladies génétiques 59

2. PATIENTS ET METHODES 66

2.1 Patients 66
2.1.1 La cohorte des patients avec DI 66
2.1.2 La cohorte de patientes ayant un syndrome d'Aicardi 66

2.2 Matériel biologique 67

2.3 Extraction des acides nucléiques 68
2.3.1 ADN 68
2.3.2 ARN 68

2.4 Cytogénétique et cytogénétique moléculaire 69
2.4.1 Préparation des chromosomes métaphasiques 69
2.4.2 Caryotype en bandes GTG (bandes G, Trypsine, Giemsa) 69
2.4.3 L'hybridation in situ en fluorescence (FISH) 69
2.4.4 Inactivation du chromosome X 70
2.4.5 L'hybridation génomique comparative sur microréseau d'ADN 70

2.5 PCR quantitative sur ADN génomique ou sur cDNA 73
2.5.1 Le principe 73
2.5.2 Le protocole 73

2.5.3 Transcription inverse pour obtenir le cDNA 74

2.6 Le séquençage d'ADN selon la méthode Sanger 75

2.7 Séquençage à haut débit 77
2.7.1 Le séquençage de l'exome du chromosome X 77
2.7.2 Le séquençage de l'exome complet 77

3. RESULTATS 79

3.1 Réarrangements intragéniques du gène IL1RAPL1 et déficience intellectuelle 79
3.1.1 Données cliniques des patients 79
3.1.2 Résultats de CGH-arrray 80
3.1.3 Résultats moléculaires 81

3.2 Caractérisation du spectre mutationnel du gène MBD5 dans la déficience intellectuelle 85
3.2.1 Données cliniques des patients 85
3.2.2 Résultats de CGH-array 86
3.2.3 Résultats moléculaires 87

3.3 Les résultats sur le syndrome d'Aicardi 91
3.3.1 CGH-array 1M à façon du chromosome X 91
3.3.2 Le séquençage haut débit de l'exome du chromosome X 93
3.3.3 Le Séquençage haut débit de l'exome complet 100

4. DISCUSSION 141

4.1 Microremaniements du gène *IL1RAPL1* et déficience intellectuelle 142

4.2 Microremaniements du gène *MBD5* et déficience intellectuelle 145

4.3 Recherche du gène candidat du syndrome d'Aicardi 147

CONCLUSION ET PERSPECTIVES 150

REFERENCES INTERNET 152

BIBLIOGRAPHIE 154

LISTE DES PUBLICATIONS, LISTE DES POSTERS

ANNEXES

Table des figures

Figure 1 : Les gènes responsables de DILX syndromique 28
Figure 2 : Agénésie complète du corps calleux en coupe sagittale I.R.M. 30
Figure 3 : Les lacunes choriorétiniennes dans un fond d'œil 31
Figure 4 : Les gènes responsables de DILX-NS 40
Figure 5 : Une organisation schématique du réseau des protéines au niveau des synapses excitatrices dont les mutations ont été observées lors de la DI 45
Figure 6 : Stratégie d'amplification par PCR en émulsion (PCRem) 52
Figure 7 : Stratégie d'amplification par Bridge PCR 53
Figure 8 : Les trois plateformes du sequençage haut debit 54
Figure 9 : Schéma de la réaction du pyroséquençage 55
Figure 10 : La représentation de la technique du séquençage par Illumina genome analyzer 56
Figure 11 : Le principe du séquençage selon la technologie Ion Torrent 57
Figure 12 : La stratégie pour des patients atteints d'une DI syndromique sporadique 60
Figure 13 : La stratégie pour chercher le gène candidat dans les cas de DI sporadique non syndromique 62
Figure 14 : La stratégie pour les familles avec DI autosomique récessive 63
Figure 15 : Stratégies pour la recherche du gène candidat dans les maladies monogéniques à l'aide de microréseaux génomiques et le séquençage de l'exome 64
Figure 16 : Profils des remaniements du gène *IL1RAPL1* détectés par CGH-array 81
Figure 17 : Résultat de l'analyse par qPCR sur ADN génomique 82
Figure 18 : Résultats de l'analyse en FISH pour la duplication chez le patient 2 et sa mère 83
Figure 19 : Récapitulatif des remaniements du gène *IL1RAPL1* pour les 3 patients 84
Figure 20 : Les remaniements du gène *MBD5* chez des 4 patients détectés par CGHarray 86
Figure 21 : Le détail des remaniements du gène *MBD5* sur le locus 2q23.1 selon la base de données DGV (build36/hg 18) 87
Figure 22 : Résultat de l'analyse par RT-qPCR sur cDNA 88
Figure 23 : Résultat de l'analyse par RT-qPCR sur cDNA 88
Figure 24 : Transcrits aberrants du gène *MBD5*, caractérisés par RT PCR et séquençage pour le patient C 89
Figure 25 : Profil du séquençage Sanger sur l'ADN génomique et cDNA du patient E 90
Figure 26 : Profil du séquençage d'ADN génomique du patient E, du frère, de la mère et du père 90
Figure 27 : Interface du logiciel d'analyse des données de séquençage haut débit de l'exome du chromosome X 94
Figure 28 : Expression du gène *RNF175* selon la base de données Gene Cards® 104
Figure 29 : Profil du séquençage ciblé Sanger du gène *RNF175*. Il s'agit du variant non sens hérité du père (un témoin, et le trio ont été séquencés) 105
Figure 30 : Localisation des gènes *USP17L2* et *FAM66D* sur la locus 8p23.1 selon la base de données DGV 106
Figure 31 : Expression du gène *USP17L2* selon la base de données Gene Card® 106
Figure 32 : Résultats du séquençage ciblé Sanger de mutation dans le gène *USP17L2* 108
Figure 33 : Expression de gène *RBMXL3* dans les tissues neuronale, la moelle épinière et la rétine (en vert) selon la base de donnée Gene Card® 111
Figure 34 : Profil du séquençage ciblé Sanger a montré la varaition hérité de la mère détecté dans le gène *RBMXL3* 113
Figure 35 : Expression du gène *GABRE* selon la base de donnée Gene Card® 114
Figure 36 : Le profil du séquençage ciblé Sanger pour la variation dans le gène *GABRE* 115
Figure 37 : Expression du gène *NBPF10* selon la base de donnée Gene Card® 116
Figure 38 : Expression du gène *WDR52* selon la base de données Gene Card® 121
Figure 39 : Profil du séquençage Sanger pour la mutation du gène *WDR52* chez la patiente trio 5 et ses parents 122
Figure 40 : Un profil « sauvage » et non muté par séquençage Sanger pour la variation du gène *WDR52* chez le trio 4 123
Figure 41 : Profil du séquençage Sanger pour la variation dans le gène *MRGPRG* pour le trio5 126
Figure 42 : Profil du séquençage Sanger pour la variation dans le gène *MRGPRG* pour le trio 3 127
Figure 43 : Expression du gène *PPARGC1B* selon la base de données Gene Card® 128
Figure 44 : Profil du séquençage Sanger pour la variation du gène *PPARGC1B* chez le trio 3 130
Figure 45 : Expression du gene *DOCK11* selon la base de données GeneCard® 132

Figure 46 : Expression du gène *FAM104B* selon GeneCard® 133
Figure 47 : Expression du gène *ELF4* selon la base de données GeneCard® 135
Figure 48 : Expression du gène *POLA1* selon la base de données Gene Card® 137
Figure 49 : Expression de gène *FAM47A* selon la base de donnée GeneCard® 139
Figure 50 : Profil du séquençage Sanger pour la variation du gène *FAM47A* chez la patiente du trio 3 et ses parents 140

Liste des tableaux

Tableau 1: Causes principales de déficience intellectuelle et leur fréquence 20
Tableau 2 : Critères diagnostiques Aicardi élargis 33
Tableau 3 : Résultats détectés par la puce 1M à façon du chromosome X, et leur contrôle par la technique du qPCR 92
Tableau 4 : Représentation des résultats de séquençage haut débit de l'exome du chromosome X contrôlés par le séquençage Sanger 97
Tableau 5 : Légende des annotations des SNPs selon IntegraGen 101
Tableau 6 : Mutations non-sens *de novo* détectée par Séquençage haut débit de l'exome 103
Tableau 7 : Résultats du séquençage de l'exome pour le gène *RNF175* pour la patiente du trio 5 et ses parents (Extrait du tableau 6) 104
Tableau 8 : La mutation dans le gène *RNF175* chez le trio 5 105
Tableau 9 : Résultats du séquençage de l'exome pour le gène *RNF175* pour la patiente (trio 5) et ses parents (Extrait du tableau 6) 107
Tableau 10 : Mutations *de novo*, avec une profondeur de lecture supérieure à 20x et non décrites dans les bases de données de polymorphismes connus 109
Tableau 11 : Les deux gènes mutés localisés sur le chromosome X chez les deux patientes (trio 1 et 3) (Extrait du tableau 10) 111
Tableau 12 : La mutation dans le gène *RBMXL3* pour le trio 1 112
Tableau 13 : La variation dans le gène *GABRE* chez le trio 3 115
Tableau 14 : La mutation *de novo* du gène *NBPF10* mutés chez 2 patientes (trio 5) et (trio 2) 117
Tableau 15 : La mutation dans le gène *NBPF10* chez le trio 5 117
Tableau 16 : La mutation dans le gène *NBPF10* chez le trio 2 118
Tableau 17 : La mutation *de novo* du gène *POTEE* chez les patientes des trio 4 et trio 5 119
Tableau 18 : La mutation dans le gène *POTEE* pour le trio4 120
Tableau 19 : La mutation dans le gène *POTEE* pour le trio5 120
Tableau 20 : La mutation *de novo* du gène *WDR52* chez les patientes des trio 5 et trio 4 121
Tableau 21 : La mutation dans le gène *WDR52* pour le trio 5 122
Tableau 22 : La mutation dans le gène *WDR52* pour le trio 4 123
Tableau 23 : Mutations *de novo* du même gène, quelque soit la profondeur, chez au moins 2 patientes .. 124
Tableau 24 : Détail de la profondeur de lecture et du type de mutation chez les patientes, pour les 13 gènes mutés chez au moins 2 patientes 124
Tableau 25 : La mutation *de novo* du gène *MRGPRG* chez les patientes des trio 5 et trio 3 126
Tableau 26 : La mutation du gène *PPARGC1B* chez les patientes des trio 1 et trio 3 128
Tableau 27 : La mutation dans le gène *PPARGC1B* pour le trio 1 129
Tableau 28 : Le détail de la mutation dans le gène *PPARGC1B* pour le trio 3 129
Tableau 29 : Tous types de mutation chez 5 patientes 131
Tableau 30 : Les variations du gène *DOCK11* chez 3 patientes 132
Tableau 31 : Tout types de mutations des gènes mutés chez 4 patientes 134
Tableau 32 : Tous types de mutations des gènes chez 3 patientes 136
Tableau 33 : Détail de la mutation détectée dans le gène *POLA1* dans les trois familles 137
Tableau 34 : Tableau des gènes localisés sur le chromosome X, porteurs de variations chez au moins 2 patientes 138
Tableau 35 : Détail des variations dans le gène *FAM47A* chez les patientes des trio 3 et trio 4 139
Tableau 36 : Détail de la mutation du gène *RBMXL3* chez les trois filles 140
Tableau 37 : Critères pour mieux appréhender la distinction entre CNV bénin ou pathogène 141
Tableau 38 : Données de la littérature sur les remaniements du gène *IL1RAPL1* et les conséquences cliniques associées chez les garçons et les mères vectrices 144

Liste des abréviations

ADN : Acide désoxyribonucléique

ADNc : ADN complémentaire

ADNg : ADN génomique

AKT1 : Akt murine thymoma viral oncogene homolog 1

AMPA : Alpha amino 3-hydroxy 5-méthylisoazol 4-propionate

APS : Adénosine 5' - phosphosulfate

ARHGAP6 : Rho GTPase activating protein 6

ARHGEF6 : Rho Guanine nucleotide Exchange Factor 6

ARN : Acide ribonucléique

ARNm : ARNmessager

ARX : Aristaless-related homeobox, X-linked

AS : Syndrome d'Angelman

ATP : Adénosine triphosphate

BAC : Chromosome bactérien artificiel

BCOR : Bcl 6 corepressor

Bdnf : Brain derived neurotrophic factor

BSA : Albumen de sérum bovin

CC2D1A : Coiled-coil and C2 domain containing 1A

CCD : Charge coupled dispositif

CCDS : Base de données Consensus coding sequences

CDKL5 : Cyclin Dependent Kinase Like 5

CGH-array : Hybridation génomique comparative sur microréseau

cm : Centimètre

CNV : Copy number variation

CRBN : Cereblon

Ct : threshold cycle

CTPS2 : Cytidine triphosphate synthase II

Cy : Cyanine

DAPI : 4',6'-diamidino-2-phénylindole

dbSNP : Base de données Single Nucleotide Polymorphism

dCTP : Deoxcytosine triphosphate

DECIPHER : DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources

DGV : Database of Genomic Variants

DI : Déficience intellectuelle

DIAD : Déficiences intellectuelles autosomique dominante

DIAR : Déficience intellectuelle autosomique récessive

DILX : Déficience intellectuelle liée au chromosome X

DILX-NS : DILX non syndromiques ou non spécifiques

DLG3 : Discs large homolog 3

DLG4 : Discs Large homoloG 4

DMSO : Dimétylsulfoxide

dNTP : Désoxyribonucléoside triphosphate

DOCK11 : Dedicator of cytokinesis 11

DS : Déviation standard

DSM-IV : The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition

DTT : Dithiotreitol

EA : Equipe d'accueil

ECARUCA : European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations

EDTA : Acide éthylène diamine tétraacétique

EEG : Electroencéphalographique

ELF4 : E74-like factor 4

ExoSAP : Exonucléase I et une phosphatase alcaline

FAM104B : Family with sequence similarity 104, member B

FAM66D : Family with sequence similarity 66, member D

FISH : Hybridation in situ en fluorescence

FLNA : Filamin A

FMR1 : Fragile X Mental Retardation 1

FMR2 : Fragile X Mental Retardation 2

FMRP : Fragile X Mental Retardation Protein

FoSTeS : Fork stalling and template switching

FOXG1 : Forkhead box G1

FTSJ1 : FtsJ RNA methyltransferase homolog 1

GABA : Acide gamma-aminobutyrique

GAP : GTP ase activating protein

GluR1 : Glutamate recepteur1

GRIK2 : Glutamate receptor, ionotropic, kainate 2

HAT : Histone acéyltransférase

HCl : Acide chlorhydrique

HDAC : Histones désacétylases

HMR : Hyperphosphatasie et retard mental

HSD17B10 : 17 beta HydroxySteroid Dehydrogenase 10

HUMARA : Human Androgen Receptor A

HUWE1 : HECT, UBA, and WWE domains containing protein 1

IBD : Identiques par descendance

IL1RAPL1 : InterLeukin 1 Receptor Accessory Protein-Like 1

IRM : Imagerie par Résonance Magnétique

IRSp53/BAIAP2 : brain-specific angiogenesis inhibitor associated protein 2

JNK : c-Jun terminal kinase

kb : Kilobase

KCl : Chlorure de potassium

kg : Kilogramme

LCR : Low Copy Repeat

LORIA : Laboratoire Lorraine d'Informatique est ses Applications

MAGUK : Membrane-associated guanylate kinase

MAP kinases : Mitogen activated protein kinases

MASA : Mental retardation, Aphasia, Shuffling gait, Adducted thumbs

MASC : MAGUK-associated signaling complex

Mb : Megabase

MBD5 : Methyl-CpG binding domain protein 5

MECP2 : Méthyl CpG-binding protein 2

MEF2C : Myocyte enhancer factor 2C

mg : Milligramme

$MgCl_2$: Chlorure de magnésium

MID1 : Midline 1

min : Minute

miR : microARN

miRBase : Base de données microARN

ml : Millilitre

MLL2 : Myeloid/lymphoid or mixed-lineage leukemia 2

MP : Mate-paired librairie

MRGPRG : MAS-related GPR, member G

MUC16 : Mucin 16

NAHR : Recombinaison homologue non allélique

NBPF10 : Neuroblastoma breakpoint family, member 10

NC : Non couvert

NCBI : National Center for Biotechnology Information

ng : Nanogramme

NGRL : National Genetics Reference Laboratory

NGS : Séquençage de nouvelle génération

NHEJ : Non homologous end joining

NHS : Nance-Horan syndrome

NMDA : Ion channel forming N-methyl D-aspartic acid

NR1 : NMDA recepteur1

NR2A : NMDA recepteur 2A

NR2B : NMDA recepteur 2B

NRC : NMDA receptor complex

OMIM : Online Mendelian Inheritance in Man

OMS : Organisation mondiale de la santé

OPHN1 : Oligophrenin 1

ORC4 : Origin recognition complex, subunit 4

p : Bras court de chromosome

PAC : P1-derived Artificial Chromosome

PAK3 : p21 Protein Activated Kinase 3

PAKs : p21 activating kinase

pb : Paires de bases

PBS : Phosphate buffered saline

PCR : Réaction en chaîne par polymérisation

PCRem : PCR en émulsion

PE : Paired-end librairie

PGM : Personal Genome Machine

PIGV : Phosphatidylinositol glycan anchor biosynthesis, class V

PLP1 : Protéolipidique protein 1

POLA1 : Polymerase (DNA directed), alpha 1

POTEE : Prostate, ovary, testis, and placenta expressed gene ankyrin domain family

PPARGC1B : Peroxisome proliferator-activated receptor gamma, coactivator 1 beta

PPi : Pyrophosphate inorganique

PRSS12 : protease, serine, 12 (neurotrypsin, motopsin)

PSD : Postsynaptic density protein

PWS : Syndrome de Prader–Willi

q : Bras long de chromosome

QD : Quotient de Développement

QI : Quotient intellectuel

qPCR : PCR quantitative en temps réel

RBM8A : RNA binding motif protein 8a

RBMXL3 : RNA binding motif protein, X-linked-like 3

RefSeq : Base de données Reference Sequence

REPS2 : *RALBP1* associated Eps domain containing 2

RNF175 : Ring finger protein 175

rpm : Révolution par minute

RPPH1 : Ribonuclease P RNA component H1

RT : Reverse transcriptase

SAF : Syndrome d'alcoolisme foetal

SAi : Syndrome d'Aicardi

SAP102 : Synapse-associated protein 102

SDS : Sequence Detection System

SE : Single-end librairie

SETBP1 : SET binding protein 1

SHANK2 : SH3 and multiple ANKyrin repeat domains 2

SHANK3 : SH3 and multiple ANKyrin repeat domains 3

Sin3A : SIN3 transcription regulator homolog A

SLC6A8 : Solute carrier family 6 member 8

SNP : Single Nucleotide Polymorphism

SNV : Single Nucleotide Variant

SOLiD : Support oligonucléotid ligation détection

SOX3 : Sry box 3

SSC : Saline sodium citrate

STXBP1 : Syntaxin binding protein 1

SYNGAP1 : Synaptic Ras GTPase activating protein 1

TAR : Thrombocythémie et Absence de Radius

TE : Tris-HCl et EDTA

TGFB1 : Transforming growth factor, beta 1

TIR domaine : Toll/IL-1Receptor homologue domaine

Tm : Température de demi-dissociation

TRAPPC9 : Trafficking protein particle complex 9

TSPYL2 : Testis specific protein, Y-linked like2

TUSC3 : Tumor suppressor candidate 3

UBE3A : Ubiquitin protein ligase E3A

USP17L2 : Ubiquitin specific peptidase 17-like 2

VEGA : Annotation des gènes des vertébrés,

WDR52 : WD repeat domain 52

WHO : World health organization

ZNF674 : Zinc Finger Protein 674

λ *:* Longueur d'onde

Préambule et objectif du travail de thèse

La déficience intellectuelle (DI) est caractérisée par un fonctionnement intellectuel significativement en dessous de la moyenne associé avec la limitation des fonctions adaptatives, avant l'age de 18 ans.

Les causes génétiques à l'origine de la DI sont clairement reconnues depuis de nombreuses années et leur caractérisation précise a progressé parallèlement au rythme de l'évolution des outils d'analyse du génome. A coté des anomalies chromosomiques telle la trisomie 21, l'identification des causes monogéniques de DI a porté initialement, avec succès, sur les DI liées au chromosome X. Cependant grâce aux progrès des biotechnologies, il est apparu une part croissante et prépondérante des causes génétiques de DI, sévère à modérée, en rapport avec des microaltérations chromosomiques, autosomes ou chromosome X, et de plus en plus de défauts de gènes autosomiques selon un modèle d'hérédité dominante ou récessive.

Le travail de thèse s'inscrit dans la caractérisation des anomalies génétiques associées à la survenue de la DI. L'objectif repose sur l'utilisation des technologies d'analyse pangénomique tant dans un but de recherche de nouveaux gènes candidats que pour la caractérisation du spectre mutationnel de gènes de la DI.

Nous présentons, dans une première partie introductive, des données générales sur la déficience intellectuelle, sa variabilité et les anomalies génétiques associées avec les voies de signalisation moléculaires et cellulaires impliquées. Nous abordons plus particulièrement dans ce contexte de DI, le syndrome d'Aicardi. Nous exposons, parmi les stratégies actuelles de recherche de gènes candidats, les différents outils d'exploration pangénomique utilisés, notamment l'hybridation génomique comparative sur microréseau d'ADN et la technique de séquençage nouvelle génération (NGS).

Dans une deuxième partie, nous détaillons le recrutement des patients explorés et les méthodes que nous avons utilisées pour la caractérisation des anomalies génétiques. Ce travail s'appuie sur le centre de référence maladies rares « Anomalies du développement » de l'inter région Est qui participe à l'évaluation génétique des patients ayant une DI. Par ailleurs, le laboratoire de recherche universitaire EA 4368, en partenariat avec l'association A.A.L syndrome d'Aicardi, contribue à rassembler plusieurs familles dans lesquelles une enfant est

atteint du syndrome d'Aicardi afin de participer à identifier la cause génétique de cette affection neurodéveloppementale.

Dans une troisième partie, nous faisons état des résultats obtenus à l'aide de la technique d'hybridation génomique comparative sur microréseau avec l'identification de remaniements intragéniques des gènes *IL1RAPL1* et *MBD5*. S'agissant du syndrome d'Aicardi, nous rapportons les résultats de la stratégie de recherche de gènes candidats tant par hybridation génomique comparative sur microréseau que par séquençage nouvelle génération. Cette dernière approche a été techniquement assurée par le laboratoire du GENOSCOPE (Evry, France) et de la société INTEGRAGEN (Evry, France) grâce au soutien financier du GIS Maladies Rares, de la Fondation Jérôme Lejeune et de la Fondation pour la Recherche Médicale. Les résultats ont été analysés et interprétés au laboratoire de génétique en étroite collaboration avec Emmanuel Bresso et le Dr Marie-Dominique Devignes (LORIA, CNRS, Vandoeuvre les Nancy).

La dernière partie de notre travail témoigne des avantages de ces stratégies d'analyse pangénomiques, dont l'analyse sur microréseau, mais souligne aussi la complexité, les limites et parfois incertitudes en terme d'interprétation des résultats, tout particulièrement pour le séquençage de nouvelle génération, reflet manifeste d'une « jeune» expérience dans l'utilisation de cet outil qui dans les très proches années deviendra tout à fait opérationnel.

1. INTRODUCTION

1.1 La déficience intellectuelle

La déficience intellectuelle (DI) est définie par l'existence avant l'age de 18 ans d'un «fonctionnement intellectuel significativement en dessous de la moyenne associé avec des limitations des fonctions adaptatives » (The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), classification DSM-IV-R) [1]. L'Organisation mondiale de la santé (OMS) la définit comme un développement incomplet ou insuffisant des capacités intellectuelles [2]. Pour Tredgold [3], elle correspond à un développement arrêté ou incomplet du cerveau. Ces troubles surviennent souvent pendant l'enfance.

Elle est caractérisée par des altérations fonctionnelles de l'encéphale de sévérité variable, avec ou sans autres anomalies neurologiques, du comportement, des troubles métaboliques, voire des anomalies malformatives [4-5].

La caractérisation quantitative de la DI repose sur la mesure du quotient intellectuel (QI), composé d'un **quotient verbal** et du **quotient non verbal** ou de la performance. L'autre aspect fondamental de la définition de DI dans le DSM-IV est la limitation des **fonctions adaptatives** dans des secteurs d'aptitude comme la communication, l'autonomie, l'apprentissage scolaire, la vie sociale, la responsabilité individuelle, le travail, les loisirs, la santé et la sécurité. La quantification de ces limitations adaptatives est donc complémentaire de l'étude du QI [1].

1.1.1 Classification d'une déficience intellectuelle

Selon le QI (test de Wechsler)

La classification la plus utilisée est celle du DSM-IV qui distingue les DI légères, modérées, sévères, et profondes, (QI déterminé selon le test de Wechsler) [6-7].

- DI légère : le QI se situe entre 50 et 70. Les personnes connaissant des difficultés scolaires mais capables de s'intégrer à la société de façon autonome à l'âge adulte.
- DI modérée : le QI se situe entre 35 à 50 et décrit des personnes connaissant dans l'enfance des retards de développement importants et une indépendance partielle. Ces individus, ont des problèmes d'intégration dans la société à l'âge adulte. Leur niveau intellectuel ne dépasse pas celui de la classe préparatoire.

- DI sévère : le QI se situe entre 20 et 35 et définit des personnes ayant besoin d'une assistance prolongée. Leur niveau intellectuel ne dépasse pas celui d'un enfant de 2 à 3 ans, leur langage est rudimentaire et les acquisitions sont incomplètes.
- DI profond : le QI est inférieur à 20 et décrit des personnes ayant peu de capacités à communiquer, une absence de langage, des troubles moteurs graves et une autonomie très limitée voire nulle. Un à deux pour cent des déficients sont concernés par ce retard.

Cette première classification est souvent simplifiée (classification simplifiée OMS-WHO, 1985) : Entre 70 et 50, il s'agit d'une déficience intellectuelle légère et d'une déficience intellectuelle sévère au-dessous de 50.

Les échelles de Wechsler comprennent des épreuves verbales et non verbales et sont les plus utilisées chez l'enfant de plus de 4 ans.

Les tests psychométriques

Il existe des tests psychométriques utilise pour l'exploration des capacités intellectuelles. Les résultats de ces tests doivent être interprétées en fonction du contexte de vie de l'enfant. En effet le développement des capacités cognitives, du langage, de la motricité et de la socialisation de l'enfant sont très dépendants de son développement affectif.

La plupart des tests de QI permettent d'évaluer l'orientation visuo-spatiale, l'organisation de séquences temporelles, la mémoire auditive, la compréhension et l'expression du langage, la faculté d'attention, et les fonctions d'intégration cognitive supérieure comme le raisonnement abstrait et symbolique et les facultés de conceptualisation.

Le test pour définir un Quotient de Développement (QD) est suggéré par Brunet-Lézine révisée par Josse en 1997 est la plus utilisée en France chez l'enfant de moins de 3 ans. Les tests de développement permettent également de définir l'âge mental des enfants de plus de 3 ans et leur QI.

La note de QI d'un sujet situe ses performances dans une distribution gaussienne.

1.1.2 Prévalence de la DI

La prévalence de la déficience intellectuelle est extrêmement difficile à évaluer en raison de la forte connotation socio-culturelle et de l'existence de biais dans l'évaluation psychométrique, en rapport avec le défaut d'actualisation régulière des cotations des tests psychométriques.

Dans la population générale, la prévalence de la DI est de l'ordre de 2 à 3 % [4, 8-9]. Une DI profonde, sévère ou modérée a une prévalence de 3,8 ‰ contre 2,5 % pour une DI légère : la DI légère est donc 10 à 12 fois plus fréquente que la DI sévère [10-14]. Il est bien rapporté que les garçons sont plus souvent atteints de DI que les filles (ratio entre 1,3 à 1,9 :1) metteant en avant la place des mutations dans des gènes liés au chromosome X [15-16].

1.1.3 Etiologie de la DI

Toutes les atteintes encéphaliques quelle qu'en soit la cause, sont susceptibles d'entraîner une diminution des capacités intellectuelles. Les causes de déficience intellectuelle sont multiples et variées, elles sont identifiées dans environ 50% des cas (tableau1) [17-18]. Sur le plan statistique, il existe une corrélation entre la profondeur du déficit intellectuel et l'existence d'une étiologie génétique : plus le déficit est profond, plus la probabilité de trouver une cause génétique est grande. Plus de 1000 maladies ont, par ailleurs, été associées à une DI [17].

Tableau 1: Causes principales de déficience intellectuelle et leur fréquence

Causes	Fréquences
Anomalies chromosomiques	4 – 28%
Syndromes dysmorphiques	3 – 7%
Maladies génétiques connues	4 – 14%
Anomalies morphologiques du système nerveux central	7 – 17%
Complications liées à la prématurité	2 – 10%
Retard mental lié à l'environnement familial	3 – 12%
Facteurs environnementaux et médicamenteux	5 – 13%
Causes endocriniennes	1 – 5%
Inconnues	30 – 50%

D'après Curry et al.,1997 [18].

La déficience intellectuelle peut être qualifiée de syndromique quand elle est associée à un phénotype clinique plus ou moins spécifique avec des signes cliniques, radiologiques, métaboliques ou biologiques. Dans les formes de déficience intellectuelle non syndromiques, le déficit cognitif est isolé sans aucune autre caractéristique clinique, biologique ou morphologique. Face à une complexité, la limite entre les formes syndromiques et non spécifiques n'est pas si aisée à définir, lorsque les signes cliniques sont plus particulièrement détaillés, des formes, jusqu'alors non syndromiques, peuvent apparaître syndromiques [19-20].

Environ 80% des déficiences intellectuelles provoque par des causes anténatales, et environ 10% chacune par des causes périnatales et postnatales. La majorité des causes anténatales a un déterminisme génétique hormis les embryopathies et les foetopathies toxiques ou infectieuses qui sont liées à des facteurs environnementaux. Les causes périnatales et postnatales sont souvent liées à une étiologie environnementale. Une étude menée par Stevenson en 2003 [21] a révélé ces données sur la population de Caroline du Sud de 10997 individus, âgés de 20 ans et dont 59% étaient des garçons.

1.1.3.1 Les causes environnementales

Plusieurs facteurs non génétiques de DI provoque par des lésions cérébrales peuvent agir avant la naissance ou durant la petite enfance. Il s'agit des infections ou maladies contractées par la mère durant la grossesse, surtout pendant les trois premiers mois sont la rougeole ou la rubéole. Les causes peuvent être post-natales : l'**infection** intra-utérines à cytomégalovirus et la méningite. La naissance très prématurée ou des troubles périnataux : l'**anoxie** et hémorragies cérébrales, les foetopathies et les embryopathies d'**origine toxique**, infectieuse ou métabolique peuvent expliquer le développement d'une déficience intellectuelle semblent être responsables de 3 à 8% des déficiences mentales [22] ; de plus une des causes les plus fréquentes de déficience intellectuelle est le syndrome d'alcoolisme foetal (SAF). La prévalence varie entre 0,6 et 3/1000 selon les communautés [23-24].

Les personnes atteintes de SAF présentent une atteinte du système nerveux central avec la présence d'un petit périmètre crânien, d'anomalies cérébrales (microcéphalie, anomalies structurelles), d'une déficience intellectuelle, d'anomalies neurologiques plus ou moins sévères. La plupart des individus ont des problèmes comportementaux et cognitifs tels que des troubles d'apprentissage, de la mémoire et d'attention, et ces caractères sont associés à une dysmorphie faciale. Parmi les **anomalies métaboliques** dans les origines de déficience

intellectuelle, la phénylcétonurie a été la première cause traitable qui provoque une DI dans 100% des cas avec microcéphalie et retard de croissance intra-utérin. l'anoxie cérébrale et l'ictère nucléaire (lié à un taux élevé de bilirubine) en sont les principales causes de DI liée à la souffrance fœtale.

Les causes postnatales sont responsables de 3 à 15% de déficience intellectuelle, Il peut s'agir de lésion traumatique, intoxication dont la moitié par **traumatisme** (enfants secoués), les autres affections cérébrales (tumeurs cérébrales, hydrocéphalies, encéphalites aiguës), la malnutrition et un environnement socioculturel défavorable ; tout causes pouvant atteindre développement du cerveau [25] .

1.1.3.2 Les causes génétiques

Les causes génétiques, comme facteurs étiologiques, distingue essentiellement les anomalies chromosomiques et les anomalies géniques. Les principales causes de la déficience intellectuelle sont exposées dans le tableau 1. Les causes génétique des déficiences intellectuelles seraient d'environ 50% [26-27]. Des mutations, préalablement décrites dans certains cas de déficience intellectuelle, impliqueraient plusieurs gènes (hétérogénéité génétique) comme le syndrome de Rett [28].

1.1.3.2.1 Les anomalies chromosomiques

Les anomalies chromosomiques (fréquence globale de 0,9% à la naissance d'individus vivants) touchent aussi bien les autosomes que les chromosomes sexuels et sont responsables de DI, le plus souvent en rapport avec un déséquilibre, gain ou perte de tout ou partie d'un chromosome. La DI est associée fréquemment un syndrome dysmorphique dont les tableaux cliniques ont été bien décrites et caractéristiques d'une anomalie chromosomique [29]. Ces anomalies chromosomiques sont identifiées grâce à la réalisation du caryotype dont le nombre de bandes par génome haploïde définit le seuil de résolution de l'analyse, et donc la précision dans l'identification et la caractérisation de l'anomalie. On distingue **les anomalies de nombre** (gain ou perte d'un chromosome entier) homogènes ou plus rarement en mosaïque si elle ne sont présentes que dans certaines cellules de l'organisme. La seule monosomie complète viable concerne le chromosome X ou syndrome de Turner. La présence d'un chromosome surnuméraire est responsable de l'apparition d'une trisomie; la plus fréquente est la trisomie 21 avec une prévalence de 0,12%. **Les anomalies de structure** se caractérisent par la cassure et le recollement de fragments chromosomiques. Ces anomalies peuvent être **équilibrées, telles les translocations (roberstonnienne, réciproque) et les inversions** sans

conséquence phénotypique sauf dans le rare cas de pathologie de point de cassure où le point de cassure interrompt la structure ou la régulation de l'expression d'un gène important pour le développement. Dans le cas d'anomalies **déséquilibrées telles les délétions et les duplications**, il y a perte ou excès de matériel chromosomique, avec des conséquences cliniques plus ou moins sévères, selon le nombre et la fonction des gènes en déséquilibre.

Les délétions peuvent être interstitielles ou terminales. La délétion du bras court du chromosome 5 ou syndrome du cri du chat est associée à une DI en rapport avec une perte spécifique de la sous-bande 5p15.2. Les microdélétions chromosomiques ne sont identifiables que par l'analyse dite « en haute résolution » du caryotype, et le plus souvent nécessitent des approches complémentaires de cytogénétique moléculaire telle l'hybridation in situ en fluorescence (FISH) [30]. Elles sont une cause importante de DI syndromique avec une présentation clinique caractéristique. Les délétions 15q11-13, 22q11.2 et 7q11.23 correspondant respectivement au syndrome de Prader-Willi ou d'Angelman (1/20.000 naissances), au syndrome de diGeorge (1/5.000 naissances) et au syndrome de Williams (1/20.000 naissances) [31].

1.1.3.2.2. Du déséquilibre chromosomique au déséquilibre génomique : la contribution des variations de nombre de copies dans la déficience intellectuelle

Des délétions ou duplications de la plupart des chromosomes ont été observés chez des patients atteints de déficience intellectuelle. Si les réarrangements chromosomiques de plus de 5 à 10 Mb peuvent être détectés par un caryotype, un grand nombre d'anomalies chromosomiques correspondent à des réarrangements submicroscopiques de taille inférieure à 5 à 10 Mb constituant ainsi une cause significative de DI détectable par l'hybridation génomique comparative sur microréseau d'ADN ou CGH-array (Hybridation génomique comparative sur microréseau) [32].

L'amélioration de cette technologie en microréseau a permis de mettre en évidence des petites délétions ou duplications de séquence génomique aboutissant à la variation du nombre de copies (CNV – Copy Number Variation) de ces petits fragments d'ADN. Elle a contribué à la découverte de nombreux CNV supplémentaires, responsables de troubles cognitifs mais aussi neuropsychiatriques tels que la schizophrénie et l'autisme [33-34].

À l'heure actuelle, plus de 55.000 CNV sont décrits. Les CNV bénins, **polymorphismes du nombre de copies** couvrent environ 12% de l'ensemble du génome dans la population normale [35]. Certains de ces polymorphismes sont fréquents alors que d'autres sont très rares soulignant tout l'intérêt de rapporter ces CNV rares considérés comme polymorphes

dans des bases de données afin de mieux appréhender les corrélations génotype-phénotype. Ces CNV sont répertoriés dans des bases de données telle DGV (Database of Genomic Variants, http://projects.tcag.ca/variation).

Plusieurs études pangénomiques des CNV effectuées chez les patients atteints de DI ont considérablement augmenté la capacité d'identifier les **syndromes microdélétionels ou résultant de microduplications**. A titre d'exemple, le syndrome de microdélétion en 17q21.31 a été identifié grâce à cette approche (prévalence estimée de 1 sur 16.000), chez des individus atteints de DI modérée à sévère avec une dysmorphie faciale évocatrice et un comportement jovial [36-37]. Les points de cassure de cette microdélétion récurrente de 500-650 kb sont situés dans des duplicons (LCR ou Low Copy Repeat) témoignant d'un mécanisme sous jacent de recombinaison homologue non allélique (NAHR) à l'origine de cette délétion. Cette microdéletion contient 6 gènes et récemment il vient d'être confirmé que le phénotype lié à cette microdéletion résulte de l'haplo insuffisance du gène *KANSL1* qui code une protéine du complexe histone acéyltransférase (HAT) [38].

Les microréarrangements subtélomériques ont été également impliqués dans l'étiologie de la DI. De nombreuses études chromosomiques ont montré qu'ils sont responsables de 3 à 6% de l'ensemble des cas DI [39-40]. Fait important, ces anomalies subtélomériques déséquilibrées sont dans la moitié des cas le résultat d'un déséquilibre d'un remaniement subtélomérique équilibré chez l'un des parents.

Globalement, ces CNV pathogènes ont été détectés dans 10-15% des individus avec DI à travers de nombreuses études [29, 41]. Ces CNV sont à la fois rares et apparus *de novo*. Toutefois, il est maintenant bien décrit que certains CNV, microdéletion ou microduplication, peuvent être hérités d'un parent cliniquement indemne soulignant une **pénétrance incomplète** tels les CNV en 1q21.1 [42-43] 15q13.3 [44] et 16p13.11 [45]. Des hypothèses sont évoquées et ont été confirmées récemment dans certaines situations : un phénomène de double hit où une seconde altération génomique contribuerait à l'apparition du phénotype chez l'enfant. Ainsi, le syndrome TAR (Thrombocythémie et Absence de Radius) est associé avec une microdélétion en 1q21.1, parfois héritée d'un parent indemne. Le phénotype TAR résulterait de la déficience du gène *RBM8A* inclus dans la délétion en présence de SNP (single nucleotide polymorphism) rares sur l'autre allèle, modifiant la transcription du gène [46]. Les réarrangements génomiques sont une source importante de **variation phénotypique et génétique**. Les syndromes génétiques sont bien caractérisés par certains des signes mais des variations phénotypiques et de sévérité sont connues depuis longtemps chez les patients

atteints de la même maladie génétique. Les études génétiques récentes sur les effets des variants de nombre de copies potentiellement pathogènes ont élargi le spectre phénotypique associée à des maladies tout à fait distinctes. Pour expliquer cette variabilité, Girirajan et al., en 2012 ont proposé récemment le modèle double hit où un deuxième CNV, dont l'effet composite avec un variant rare de grand effet contribue à l'hétérogénéité phénotypique [47]. Le deuxième CNV contribue à augmenter le nombre de gènes perturbés par un effet additif ou synergique. La combinaison de variants rares avec des perturbateurs de grand effet aboutit à des phénotypes différents, y compris dans la déficience intellectuelle, l'épilepsie, l'autisme ou la schizophrénie. Un premier CNV prédispose l'individu porteur à la survenue d'une maladie et un deuxième CNV vient moduler le tableau clinique et sa sévérité. Par exemple les CNV touchant les régions 16p11.2 [48] ou 1q21.1 [42] sont responsables de tableaux cliniques extrêmement variables et même parfois aucune signe clinique n'est observé chez le parent qui a transmis ce CNV à l'enfant atteint. Il est probable que cette variabilité soit liée à la présence d'un second CNV. Toutefois il est aujourd'hui difficile d'identifier ce second CNV avec certitude afin de guider le conseil génétique pour les familles porteuses.

Des bases de données telles que ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations, http://www.ecaruca.net) et DECIPHER (DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources, https://decipher.sanger.ac.uk/application/) récoltent en continue la description des CNV pathologènes et des phénotypes qui leurs sont associés.

Les microremaniements chromosomiques sont la conséquence de l'instabilité génomique liée aux caractéristiques de l'architecture du génome. La maladie génomique sont survenue a cause de ces microremaniements introduite par Lupski [49] : qui désigne une pathologie causée par une altération du génome (délétions, duplications, inversions ou translocations) conduisant à une perte, un gain ou une perturbation de la structure d'un gène. La taille peut varier de quelques centaines à plusieurs millions de paires de bases. Les points de cassure de ces remaniements chromosomiques se distribuassent sur l'ensemble du génome, plus particulièrement les régions péricentromériques et subtélomériques qui sont enrichies en ce type de séquences ; du fait que ces zones présentent une architecture génomique complexe comme des duplications segmentaires ou des séquences palindromiques [50].

Différents mécanismes impliqués dans l'origine de ces remaniements qui sont le plus souvent sporadiques avec 20 à 30% du génome humain consistent en des séquences répétitives. Les duplications segmentaires ou duplicons ou « low copy repeats » (LCR) sont des séquences

d’ADN de 10 à 300kb, dupliqués dans le génome et qui ont un degré remarquable d’identité de séquence (>95 %) souvent associées à des séquences Alu.

5 à 10% du génome humain peuvent correspondre à des gènes, des pseudogènes, des fragments de gènes, des « clusters » de gènes qui sont désigné comme des points chauds de recombinaison. On peut distinguer la région du génome soumise au remaniement chromosomique dont : (i) les remaniements récurrents avec les points de cassure sont fixes, (ii) des remaniements non récurrents pour lesquels les points de cassure sont variables mais il est possible de définir une région minimale critique dans laquelle se trouve(nt) le(s) gène(s) responsable(s) du phénotype. Plus récemment les progrès de la cytogénétique moléculaire ont conduit à proposer trois grands mécanismes pour expliquer l'apparition des désordres génomiques [51] : (i) la recombinaison homologue non allélique (ii) la ligature d'extrémités d'ADN non homologues (iii) l'interruption de la fourche de réplication et commutation de cible.

Réarrangements récurrents résultant d'une recombinaison homologue non allélique

La recombinaison homologue non allélique (NAHR) est une recombinaison homologue impliquant deux locus différents. Il est la première cause d'origine de remaniements génomiques. Elle survient pendant la méiose ou la mitose et nécessite deux répétitions segmentaires LCR ou duplicons des séquences homologues, qui agissent comme substrats moléculaires de recombinaison. Sa survenue au cours de la mitose est en particulier lors de la survenue de cancers. Une recombinaison homologue non allélique entre deux duplications segmentaires sur le même chromosome de même orientation va entraîner une délétion ou une duplication du segment intermédiaire. Si les séquences sont en orientation inversée il peut entraîner une inversion. Un remaniement produit des duplications segmentaires provoque par une translocation réciproque sur des chromosomes différents. La récurrence de syndromes microdélétionnels, et d'un syndrome de microduplication sont expliqué par ce mécanisme correspondant à la même région et souvent décrit parallèlement. Outre la microdélétion 17q21.31 citée préalablement, la microdélétion d'une taille de 3,7Mb responsable du syndrome de Smith-Magenis et la microduplication responsable du syndrome de Potocki-Lupski en 17p11.2 s'expliquent par la présence de LCR (SMS-REP) de 200kb dans cette région [52].

Ligature d'extrémités d'ADN non homologues

Contrairement aux cassures simple brin, qui peuvent être facilement réparées, la réparation d'une cassure double brin apparaissent soit au cours de processus physiologiques tels que la méïose, soit à la suite de stress génotoxiques (radiations ionisantes, ultraviolets, produits chimiques mutagènes) qui nécessite des mécanismes moléculaires plus complexes. Ce type de cassure a été mis en évidence avec une fréquence de 5 à 10% dans des cellules et sont extrêmement délétères. Ce mécanisme utilise deux systèmes de réparation d'une cassure double brin : la recombinaison homologue qui utilise une séquence d'ADN matrice identique à la séquence endommagée ou la ligature d'extrémités d'ADN non homologues (NHEJ : non homologous end joining) qui correspond à la resoudure des extrémités d'ADN cassées. Ce mécanisme de réparation existe chez la plupart des organismes, des bactéries aux mammifères. Dans ce mécanisme, après la détection d'une cassure double brin de l'ADN, les systèmes de réparation réalisent un « pontage moléculaire », les extrémités de la cassure sont souvent digérées avant d'être recollées afin de les rendre compatibles. La perte ou l'ajout de quelques nucléotides est l'une des caractéristiques de ce type de réparation, et est à l'origine de délétions ou de duplications. Il laisse une signature détectée au niveau des points de cassure des réarrangements au niveau de la cassure réparée. Un exemple de délétions ou duplications non récurrentes dans le gène *PLP1* (proteolipid protein 1) en Xq22 associées au syndrome de Pelizaeus-Merzbacher (OMIM 312080) expliqué par NHEJ pour lesquelles le séquençage des points de cassure montre l'ajout de nucléotides d'origine inconnue [53].

Interruption de la fourche de réplication et commutation de cible (FoSTeS)

Un mécanisme de réarrangement proposé n'est pas lié aux cassures double brin, mais basé sur les erreurs de la réplication de l'ADN explique les réarrangements non récurrents tels que des duplication/délétions séparées par des régions non dupliquées ou des duplications-triplications : il s'agit du « fork stalling and template switching » (FoSTeS) (interruption de la fourche de réplication et commutation de cible). En effet, certaines cassures seraient dues à des changements de brin matrice au cours de la réplication. Les sites de jonction présentent des microhomologies de 2 à 5 nucléotides qui sont situées à proximité de LCR, de séquences palindromiques ou de structures cruciformes. Lorsque la fourche de réplication entraîne une rupture par une brèche au niveau d'un brin d'ADN cible . Le brin en cours de synthèse pourrait alors décroché et se réhybrider au niveau d'une autre fourche du fait de microhomologies à l'extrémité 3' située à proximité (à une distance de 120 à 550kb). L'insertion d'une nouvelle séquence provenant d'un chromosome différent aboutit donc au croisement de deux fourches

de réplication. Selon sa position, le résultat final est une délétion ou une duplication. Il s'agit d'un mécanisme dépendant des erreurs du processus de réplication [54].

1.1.4 La déficience intellectuelle liée au chromosome X (DILX)

Les observations cliniques et génétiques ont montré que les DI liées à l'X (DILX) sont très hétérogènes. Les résultats des études génétiques ont démontré un excès de DI chez les garçons par rapport aux filles. Cette différence devient particulièrement marquée dans les DI légères [29, 55-56]. Cette inégalité de répartition est expliquée, en partie, par le caractère hémizygote des garçons pour les gènes situés sur le chromosome X [29]. Les études les plus récentes [57] ont montré que la contribution du chromosome X au DI pourrait être d'environ 10-15%. Le contenu des gènes sur l'X représente environ 4% du génome humain (environ 818 gènes du chromosome X codent des protéines sur les 19.586 contenu dans le génome humain) [58] ; (VEGA annotation des gènes des vertébrés, version 35; mars 2009, http://vega.sanger.ac.uk/index.html). À ce jour, 90 de ces 818 gènes annotés sur le chromosome X (11% du X) ont été impliqués dans 215 DILX (Figure 1) (Greenwood Genetics Center, http://www.ggc.org/xlmr.htm).

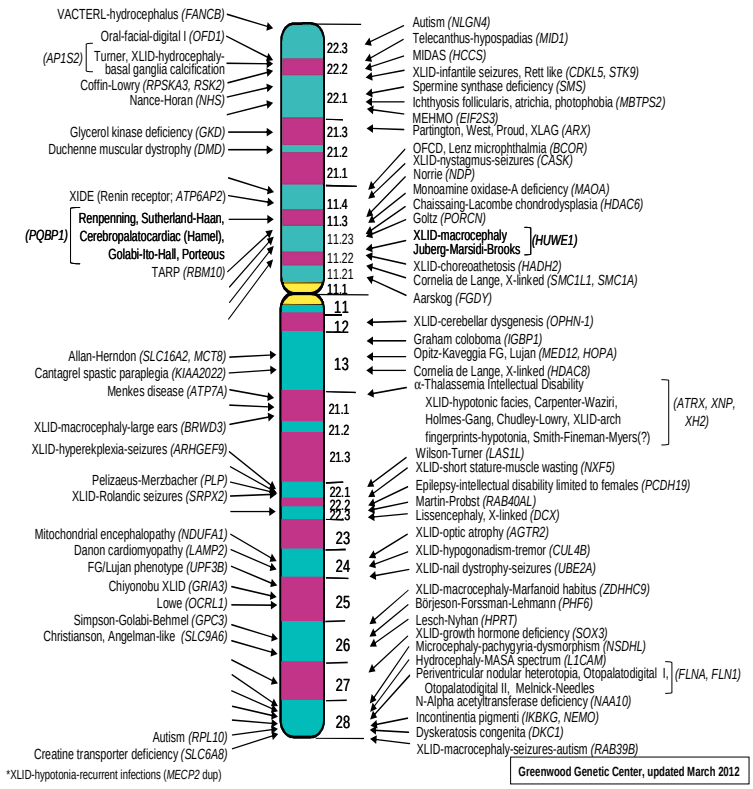


Figure 1 : Les gènes responsables de DILX syndromique

1.1.4.1 Déficience intellectuelle liée au chromosome X (DILX) syndromique

L'amélioration de la connaissance des mécanismes physiopathologiques des DILX a permis de faire évoluer les paramètres de la classification des DILX. La classification initiale des DILX repose sur des critères cliniques avec la distinction de deux sous-groupes, syndromiques et non syndromiques. Dans les **formes syndromiques**, la DI est associée à un phénotype clinique, radiologique ou biologique plus ou moins spécifique ; les signes cliniques associés, définissant les DILX syndromiques, sont de type malformatif (imperforation anale dans le syndrome FG, sténose de l'aqueduc dans le syndrome MASA (Mental retardation, Aphasia, Shuffling gait, Adducted thumbs), de type neurologique (syndrome de Rett), ou encore une expression exclusivement féminine (syndrome de Goltz, syndrome d'Aicardi).

1.1.4.2 Le Syndrome d'Aicardi

1.1.4.2.1 Historique

Le syndrome d'Aicardi (OMIM 304050) a été décrit pour la première fois en 1965 par le docteur Jean Aicardi suite à l'observation de filles présentant des spasmes en flexion [59]. En 1969, il a rapporté 15 enfants, uniquement des filles, avec une triade de signes définissant le syndrome d'Aicardi [60] : agénésie du corps calleux, spasmes infantiles, et lacunes choriorétiniennes. D'autres malformations de l'œil et du squelette sont aussi observées, inconstamment, chez ces filles [60-62].

1.1.4.2.2 Spectre clinique

C'est un syndrome rare dont la prévalence est estimée à 0,06 cas sur 100 000, avec un nombre revu récemment à 500 cas en Europe mais la prévalence réelle est probablement sous-estimée. L'accumulation des cas du syndrome d'Aicardi (SAi) publiés dans la littérature et l'apport des nouvelles techniques d'imagerie a permis d'appréhender l'étendue du spectre clinique. Une minorité des patientes ne présente pas la triade classique.

Les anomalies neurologiques

Premières manifestations de la maladie, les crises convulsives apparaissent souvent tôt, dans les 3 premiers mois et dominent la scène. Les spasmes infantiles sont souvent asymétriques ou unilatéraux ; ils apparaissent très précocement et sont souvent associés à des crises d'épilepsies. La DI est le plus souvent sévère mais dans de très rares cas une DI légère à modérée a été décrite [63]. A l'heure actuelle seulement un seul cas de syndrome d'Aicardi typique a été rapporté avec des fonctions cognitives normales [64]. Chez les patientes avec SAi, l'agénésie du corps calleux (Figure 2) n'est pas toujours complète [65-66]. D'autres malformations du cerveau accompagnent l'agénésie du corps calleux : une dysplasie corticale, périventriculaire de la substance grise, des hétérotopies, des formations kystiques, des anomalies vermiennes, et des papillomes du plexus choroïde.



Figure 2 : Agénésie complète du corps calleux en coupe sagittale I.R.M.

Le tracé électro-encéphalographique (EEG) le plus fréquent est caractérisé par des bouffées d'ondes lentes, étroites et de haute amplitude séparées d'intervalles d'ondes de basse amplitude. Ces crises épileptiques à type de spasmes et les anomalies EEG ont tendance à peu se modifier au cours de la vie des patientes [65-66].

Les anomalies oculaires

Les lacunes choriorétiniennes (Figure 3) seraient pathognomoniques mais les deux cas décrit de syndrome d'Aicardi ne présentant pas ces lacunes [66]. En fondoscopie, les lacunes apparaissent comme des tâches arrondies rosâtres ou blanchâtres avec une forme bien délimitée et des bords hyperpigmentés [67]. Quelques anomalies oculaires peuvent s'associer aux lacunes choriorétiniennes telles qu'un colobome irien ou rétinien et les microphtalmies sont les plus fréquemment rencontrées.



Figure 3 : Les lacunes choriorétiniennes dans un fond d'œil

Les anomalies extra neurologiques

Des anomalies des côtes et vertèbres (côtes manquantes, hémivertèbres, vertèbres soudées) sont présentes dans la moitié des cas conduisant à une scoliose parfois marquée [68]. une dysmorphie faciale caractéristiques existerait associée au SAi [62] ; dans cette étude l'analyse clinique de 40 patientes du SAi, présentent un maxillaire proéminent, des narines antéversées, une pointe du nez retroussée, une queue du sourcil épars. Des lésions dermatologiques (22,5%) et des malformations des mains (7,5%) sont également rapportés.

1.1.4.2.3 Pronostic et histoire naturelle

Le syndrome d'Aicardi débute par la survenue, avant l'âge de trois mois, de spasmes infantiles. L'évolution clinique est très variable entre des patientes. Certaines filles acquièrent la marche et prononcent quelques mots alors que d'autres n'acquièrent ni marche, ni langage. A ce jour, seuls 7 cas atteintes avec une DI moins sévère ont été rapportés dans la littérature, cependant aucun de ces 7 cas ne présentaient la triade caractéristique du SAi [64, 69-74].

Le lien entre la gravité des anomalies cérébrales et l'atteinte des fonctions cognitives chez les patientes Aicardi est difficile à établir. L'épilepsie ne semble pas être un facteur pronostique car un bon contrôle pharmacologique des crises n'est pas associé à un meilleur développement psychomoteur. Dans le plus part des cas l'absence de l'acquisition de la marche (79%) et langage (96%) a été décrit. L'étude de Menezes [68] a rapporté 75% des patientes ont une survie estimée à 6 ans et 40% à 15 ans. Au cours des dernières années, des enfants avec une évolution moins sévère ont été de plus en plus rapporté [75].

1.1.4.2.4 Proposition des critères pour le diagnostic clinique du syndrome d'Aicardi

A partir de la description initiale du syndrome en 1965, Jean Aicardi revisita les critères en 1999 [65] mettant en exergue un spectre clinique plus large et proposa en 2005 de nouveaux critères diagnostiques [66] avec des critères majeurs et mineurs (Tableau 2). Le diagnostic reste toujours basé sur la triade : agénésie du corps calleux, spasmes en flexion, et lacunes choriorétiniennes. Cependant, les patientes sans agénésie du corps calleux mais présentant d'autres critères majeurs peuvent être considérées comme ayant un syndrome d'Aicardi. Ainsi, la présence d'au moins deux critères de la triade associés à deux autres critères (majeurs ou mineurs) suggère fortement le diagnostic clinique de SAi.

Tableau 2 : Critères diagnostiques Aicardi élargis

Anomalies	Critères majeurs	Critères associés
Malformations complexes du SNC	**Agénésie partielle ou totale du corps calleux** • Dysplasie corticale • Hétérotopies périventriculaires et souscorticales • Formations kystiques épendymaires interhémisphériques ou autour du troisième ventricule • malformations corticales (souvent microgyrie) • Papillomes des plexus choroïdes	• Asymétrie des hémisphères cérébraux • Elargissement du Tectum
Anomalies ophtalmologiques	**Lacunes choriorétiniennes** • Colobome du disque et nerf optique, souvent unilatéral	Microphtalmie
	Spasmes infantiles	EEG: “dissociation corticale”, asynchronisme
Anomalies squelettiques		• Hémivertèbre, vertèbre en papillon • Côte manquante ou double • Scoliose • Anomalie des extrémités (campodactylie, pouce bas implanté, hypoplasie du 5ème doigt)
Dysmorphie faciale		• Maxillaire proéminent • Narines anteversées • Sourcils épars
Lésions cutanées		• Naevus multiples • Taches hypopigmentées • Papillomes verruqueux • Hémangiomes

(Aicardi 2005 [66] ; Sutton et al., 2005 [62]; Hopkinns et al., 2008 [76])

1.1.4.2.5 Mode d'hérédité de la maladie

Le syndrome d'Aicardi ne s'observe que chez les filles ou de façon exceptionnelle les garçons présentant par ailleurs un syndrome de Klinefelter (47, XXY) [77]. Le seul cas exceptionnel de deux soeurs atteintes du syndrome d'Aicardi [78] plaide pour une origine génétique de ce syndrome. Pour expliquer le cas rapporté de jumelles monozygotes discordantes pour le syndrome d'Aicardi, il est évoqué une mutation somatique précoce au cours du développement embryonnaire chez la jumelle atteinte [78] Cinq paires de jumelles dizygotes discordantes pour le syndrome d'Aicardi ont été décrites, ces cas ne semblent pas en faveur de facteurs périnatals à l'origine du syndrome d'Aicardi.

Le syndrome d'Aicardi survient de façon sporadique, excepté le cas décrit par Molina un cas des jumelles monozygotes. Ainsi, les auteurs suggéraient que la meilleure hypothèse pour expliquer cette récurrence familiale serait une mosaïque germinale chez l'un des deux parents [78]. L'hypothèse d'une néomutation dominante liée au chromosome X expliquerait l'atteinte exclusive des filles. Le SAi touche exclusivement des sujets de sexe féminin, l'hypothèse d'une mutation dominante liée à l'X a été proposée. Un effet létal de cette mutation chez les garçons hémizygotes pour le chromosome X est avancé, toutefois, l'absence de biais dans le sex ratio en faveur des filles dans les fratries des cas sporadiques de syndrome d'Aicardi va à l'encontre de ce modèle [79].

1.1.4.2.6 Inactivation de l'X et syndrome d'Aicardi

L'inactivation du chromosome X est un processus cellulaire normal mis en place très tôt au cours de l'embryogenèse (14 jours après la fécondation). Chez les mammifères, deux type de chromosomes X et Y sont responsables des différences entre hommes et femmes. Alors que les femmes portent deux chromosomes X, les mâles n'en ont qu'un seul auquel est associé un chromosome Y, de petite taille avec peu de gènes ; chez la femme et dont la fonction est d'assurer un dosage génique équivalent entre les deux sexes pour les gènes portés par l'X [80]. Chez les femmes avec le phénotype normales, sans antécédent familial de pathologie liée à l'X, Le choix de l'inactivation du chromosome X se fait au hasard dans chaque cellule somatiques touchant l'X paternel ou l'X maternel qui soit inactivé dans une proportion équivalente (50%-50%). La grande majorité des femmes (80 % à 90 %) a un profil d'inactivation aléatoire (50 %-50 % à 80%-20 %) dans la population féminine normale. Dans la population générale de par les phénomènes stochastiques, les femmes (10 à 20 % de la

population générale) présentant par pur hasard une inactivation fortement biaisée (90%-10% à 100%-0%) en faveur de l'un ou l'autre des chromosomes X.

Mais, quand il y a une maladie dominante liée à l'X et létale chez les garçons , le choix de l'inactivation de l'X ne se fait pas au hasard mais va se faire de façon préférentielle pour le chromosome qui ne porte pas l'anomalie chez la femme. Il s'agit d'un biais d'inactivation. C'est le cas chez les femmes porteuses de mutation dans ATR-X qui ne sont pas atteintes. Des exceptions existent comme par exemple le syndrome de Rett où la majorité des patientes ne présente pas de biais d'inactivation. Mais la majorité des maladies dominantes liées à l'X, comme l'incontinentia pigmenti, le syndrome oro-facio-digital de type I et le syndrome Goltz, est associée à une inactivation biaisée du chromosome X [81-83].

Dans une première étude, un biais d'inactivation a été retrouvé chez les 3 sur 7 patientes du SAi les plus sévèrement atteintes (déficience intellectuelle plus sévère et une épilepsie pharmaco résistante). Contrairement, deux des filles avec un DI légère avaient une inactivation aléatoire [75]. Néanmoins, le lien entre biais d'inactivation et phénotype clinique ne retrouve pas dans une deuxième étude. Ainsi, un profil d'inactivation aléatoire est observé chez les 10 patientes SAi étudiées (triade classique) [79].

Les mères asymptomatiques des trois patientes du SAi présentent une inactivation biaisée. Cette observation est en concordance avec une distribution binomiale de l'inactivation de l'X dans la population générale. Pourtant, chez un peu plus de 20 % des femmes saines présenteraient l'inactivation de l'X biaisée. Les patientes atteintes de SAi présentant un biais d'inactivation décrits précédemment par Neidich, vu le nombre restreint de patientes étudiées, pourraient faire partie de ce biais stochastique [79]. D'autres auteurs n'ont pas confirmé ce biais. Dernièrement, Eble [84] a réalisé des études d'inactivation de l'X dans le sang de 35 filles atteintes du syndrome d'Aicardi. L'auteur montre que la fraction des filles SAi avec un biais d'inactivation de l'X était plus importante comparée à la population générale. Par ailleurs, le degré du biais d'inactivation de l'X a été corrélé à la gravité du syndrome d'Aicardi chez les patientes. Ces résultats ne permettent pas d'affirmer ou d'infirmer l'hypothèse d'une transmission liée à l'X. En effet, le biais d'inactivation de l'X pourrait être tissu-dépendant et ne concerner que les tissus où ce syndrome s'exprime particulièrement (cerveau, oeil).

Si l'on admet que l'inactivation du chromosome X chez les filles SAi est aléatoire, alors en découle un certain nombre de possibilités et d'hypothèses.

- L'inactivation de chromosome X pourrait ne pas intervenir dans la pathologie.

- Il est ainsi possible que des cellules avec le chromosome X muté sont dans une très faible proportion et qui sont suffisante à conduire au phénotype du syndrome d'Aicardi. Alors que un biais complet en faveur de chromosome X muté conduise à une létalité (comme chez les garçons hémyzigotes),
- Il est aussi possible que le gène du syndrome d'Aicardi ne soit pas exprimé dans les cellules hématopoïétiques. Ainsi, une inactivation biaisée liée à la maladie ne serait pas observable dans le tissu sanguin que les tissus neuronale ou rétine où ce syndrome s'exprime particulièrement.
- Enfin, un défaut d'inactivation conduisant à une disomie fonctionnelle de la région du gène SAi pourrait expliquer l'atteinte des femmes sans biais d'inactivation et l'absence de cas de syndrome d'Aicardi chez les hommes.

1.1.4.2.7 Le syndrome d'Aicardi chez les garçons

Le syndrome d'Aicardi s'observe quasi exclusivement chez les filles. Toutefois, trois garçons ayant une formule chromosomique 47,XXY, ont été rapportés dans la littérature, avec des caractéristiques cliniques de SAi [66, 85-86]. Ces données plaidaient dès lors en faveur d'une affection génétique dominante liée au chromosome X. Il est à noter que des observations de garçons avec une formule 46,XY ont aussi été décrites mais le tableau clinique est très atypique, et pour l'un d'entre eux le diagnostic de SAi n'a pas été retenu par Jean Aicardi [87]. Récemment, Chappelow et al., en 2008 [88] ont rapporté l'observation d'un garçon de 5 ans avec la triade classique du SAi, en présence d'une formule 46,XY, observation allant, a priori, à l'encontre de l'hypothèse initiale du caractère létal de la mutation chez le garçon.

1.1.4.2.8 Recherche de gènes candidats pour le syndrome d'Aicardi

Le clonage positionnel vise à identifier la localisation du gène responsable de la maladie génétique étudiée s'effectue alors par analyse de liaison [89]. La première étape du clonage positionnel consiste donc en la localisation du gène dans une région chromosomique, par des études de liaison génétique dans les familles où ségrége la maladie et à l'aide de marqueurs polymorphes.

Dans les bases de données biologiques la carte intégrée (génétique, physique et transcriptionnelle) de la région candidate est souvent disponible. Si non, une analyse moléculaire de la région doit être réaliser. L'étape suivante consiste en la détermination de la séquence nucléotidique de la région candidate, permettant ainsi d'identifier les séquences

codantes contenues dans l'intervalle critique afin d'identifier un ou plusieurs gènes candidats. l'étape ultime est l'identification du gène candidat par la présence de mutations délétères chez les sujets malades. Cette localisation peut être facilitée par la découverte, chez les individus porteurs du caractère étudié de la maladie génétique, d'une anomalie chromosomique (translocation, délétion, inversion).

Les gènes candidat de la maladie génétique sont localisés par des études de liaison dans les familles où ségrège la maladie. Le développement des cartes génétiques base sur les marqueurs de polymorphisme de locus microsatellites couvrant la totalité du génome est très efficace pour la recherche des gènes candidats [90]. La faible densité de ces microsatellites dans certaine région du génome humain est un facteur limitant pour les cartes génétiques. Une approche alternative a récemment utilisé basé sur les SNP qui sont plus nombreux dans le génome. La base de données NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov) contient des informations sur le nombre, la cartographie et le pourcentage d'hétérozygotie des SNP.

Les analyses de liaison génétique permettent de déterminer la région, dans lequel il existe une forte probabilité que soit contenu le gène morbide.

Dans le cas particulier du syndrome d'Aicardi cette approche n'est pas possible puisqu'aucun cas de transmission familiale n'a été répertorié à ce jour.

Anomalies chromosomiques associées au syndrome d'Aicardi

Les anomalies chromosomiques ont été décrites chez des trois patientes SAi avec une anomalie de la région p22 du chromosome X. La première patiente présentant le phénotype clinique du SAi avec une translocation chromosomique équilibrée de novo, t(X;3) et un point de cassure en Xp22.3 pouvant interrompre le locus Aicardi [91].

Cette patiente présentent les signes cliniques du syndrome SAi : les lacunes choriorétiniennes, une microphtalmie, une agénésie du corps calleux, des anomalies du squelette, un profil EEG anormal, une déficience intellectuelle, mais l'absence de spasmes infantiles. De plus anomalies ophtalmologiques atypiques, avec une hydrocéphalie, une anomalie de Dandy Walker et une asymétrie de la tête ont également rapporté chez cette patiente. L'analyse cytogénétiques, sur les lymphocytes aussi bien sur les fibroblastes, révèlent la formule 46,XX,t(X;3)(p22.3;q12). Le phénotype clinique de cette patiente expliqué par plusieurs hypothèses :

- la monosomie fonctionnelle 3q provoque les symptômes atypiques chez cette patiente

- la monosomie fonctionnelle 3q est moins désavantageuse que la perte complet de fonction du gène SAi au niveau cellulaire, d'où une inactivation préférentielle de l'X transloqué.
- Le gène du syndrome d'Aicardi serait soumis à l'inactivation de l'X.

Il y a deux point contradictoire ressortent de cette étude. Le premier problème concerne les études d'inactivation de chromosome X. Les études précédentes chez les patientes avec une translocation équilibrée, le chromosome X normal est le plus souvent inactivé. Dans le cas de l'inactivation de X normal il n'y aura aucune conséquence, alors que l'inactivation du chromosome X anormal entraînerait la monosomie fonctionnelle. Dans cette étude ils ont rapporte l'inactivation de l'X anormal avec une monosomie fonctionnelle 3q partielle. Si comme le présentent les auteurs, l'absence du gène SAi est plus désavantageuse que la monosomie, une inactivation biaisée devrait être observée pour le syndrome SAi, ce qui n'est pas le cas. De plus, tous les gènes transloqué de l'X, l'expression sera biallélique. En effet, il est possible d'expliquer le phénotype observé par une pathologie du point de cassure. Jean Aicardi aurait émis des réserves quant à la validité du diagnostic de syndrome d'Aicardi concernant ce cas précis (d'après OMIM 304050).

Deux autres patientes atteintes du syndrome d'Aicardi ont été décrit avec des délétions de Xp22 [75]. Le caryotype 46,XX chez six patientes et une patiente avec une délétion cytogénétique del(X)(p22pter) a été rapporté dans cette cohorte. Les détails du diagnostic de syndrome d'Aicardi n'étaient pas donnés. Chez cette patiente avec la délétion del(X)(p22pter) l'analyse moléculaire de la région Xp22 a montré l'absence des loci DXS278 et DXS85. En 1990, la même équipe a réalisé une recherche moléculaire des délétions de la région Xp22 chez 7 filles SAi, mais ne retrouve pas d'anomalies pour les 8 marqueurs polymorphes étudiés (incluant DXS278) [75].

Approche fonctionnelle

Cette approche est basée sur la fonction des gènes et non plus sur les anomalies structurales de l'ADN. Jusqu'à présent, le gène *FLNA* (Filamine A) a été séquencé. Une étude postmortem du cerveau de deux patientes atteintes SAi, a révélé l'accumulation de filamine dans les astrocytes. La protéine filamine A est impliquée dans hétérotrophies nodulaires bilatérales périventriculaires, une pathologie familiale liée à des mutations du gène *FLNA* qui touche spécifiquement les sujets de sexe féminin et létale chez le sujet de sexe masculin. Dans cette étude la présence de Filamine A dans les astrocytes des patientes SAi a conduit les auteurs à

proposer l'hypothèse que il serait aussi impliqué dans l'étiologie de SAi. Le séquençage des 48 exons du gène *FLNA*, incluant les jonctions exon-intron n'a révélé aucune mutation délétère chez les 10 filles SAi testées [92].

Approche pangénomique pour la recherche du gène candidat

Dans notre la laboratoire de recherche, Yilmaz et al., en 2007 [93] en partant de l'hypothèse qu'il n'existe pas de région candidate définie sur le chromosome X, ont utilisé la technique de d'hybridation génomique comparative sur microréseau de clones BAC (chromosome bactérien artificiel) pour cribler le chromosome X entier et détecter la présence de microdélétions ou de microduplications non détectables par analyse cytogénétique conventionnelle. Cette étude n'a pas trouvé de variation délétère du nombre de copie du génome de l'X, au seuil de résolution étudié. Wang et al., en 2009 [94] ont utilisé la même approche, CGH-array, mais pangénomique de haute résolution, à l'aide d'un microréseau de 244K oligonucléotides (Agilent), pour rechercher des variations de nombre de copies. Ils ont étudié des échantillons d'ADN de 38 filles atteintes du syndrome d'Aicardi. Il y avait cinq variations uniques détectées respectivement sur les chromosomes 7, 12, X et deux sur le chromosome 17. Toutes les variations étaient héritées d'un parent où ont été relevées comme des régions polymorphes. Cette étude a ainsi échoué à identifier une potentielle région candidate. Notre laboratoire, en utilisant cette même approche pangénomique, 244K, sur notre série de patientes avec SAi , a toutefois identifié chez une des enfants une anomalie microdélétionnelle *de novo*, en 1p36, soulignant que la monosomie 1p36 peut constituer une phénocopie du syndrome d'Aicardi [95].

1.1.4.3 Déficience intellectuelle liée au chromosome X (DILX) non-syndromique

Les **DILX non syndromiques (ou non spécifiques)** (DILX-NS) concernent des individus avec une DI apparemment isolée, sans aucune autre caractéristique clinique, biologique ou morphologique. Il est considéré que 2 tiers des DILX sont non syndromiques, mais cette proportion diminue au fur et à mesure de la découverte de nouveaux gènes et de la réévaluation du phénotype clinique sur une plus grande série de patients [96]. De plus, il apparaît aujourd'hui que des mutations dans certains gènes de DILX sont à la fois responsables de formes syndromiques et non syndromiques, ce qui rend les frontières entre les deux formes encore plus floues [19]. La prise en compte d'une anomalie morphologique cérébrale à l'IRM (Imagerie par Résonnance Magnétique) est actuellement considérée comme

un élément d'information important dans la classification. Des progrès importants ont été réalisés dans la recherche des gènes impliqués dans les DILX-NS, tout particulièrement grâce à la technique de séquençage haut débit de l'exome du chromosome X. Chez 53 des 208 familles (25%) dans lesquelles ségrégait une DILX-NS, une mutation faux-sens, une variation du nombre de copies ou un codon stop ont été identifiés [97]. Parmi les familles du consortium EuroMRX, 42% des mutations sont identifiées avec une femme conductrice obligatoire [98]. A l'heure actuelle, 38 gènes de DILX-NS sont reconnus dans ces familles et le plus souvent, chacun de ces gènes contribue à moins de 0,2 à 0,5% des cas (Figure 4). Ces données diffèrent donc de la fréquence observée pour les gènes *FMR1* (Fragile X Mental Retardation 1) et *FMR2* (Fragile X Mental Retardation 2) du syndrome de X fragile, *MECP2* (methyl CpG-binding protein 2) du syndrome de Rett, et *ARX* (aristaless-related homeobox, X-linked).

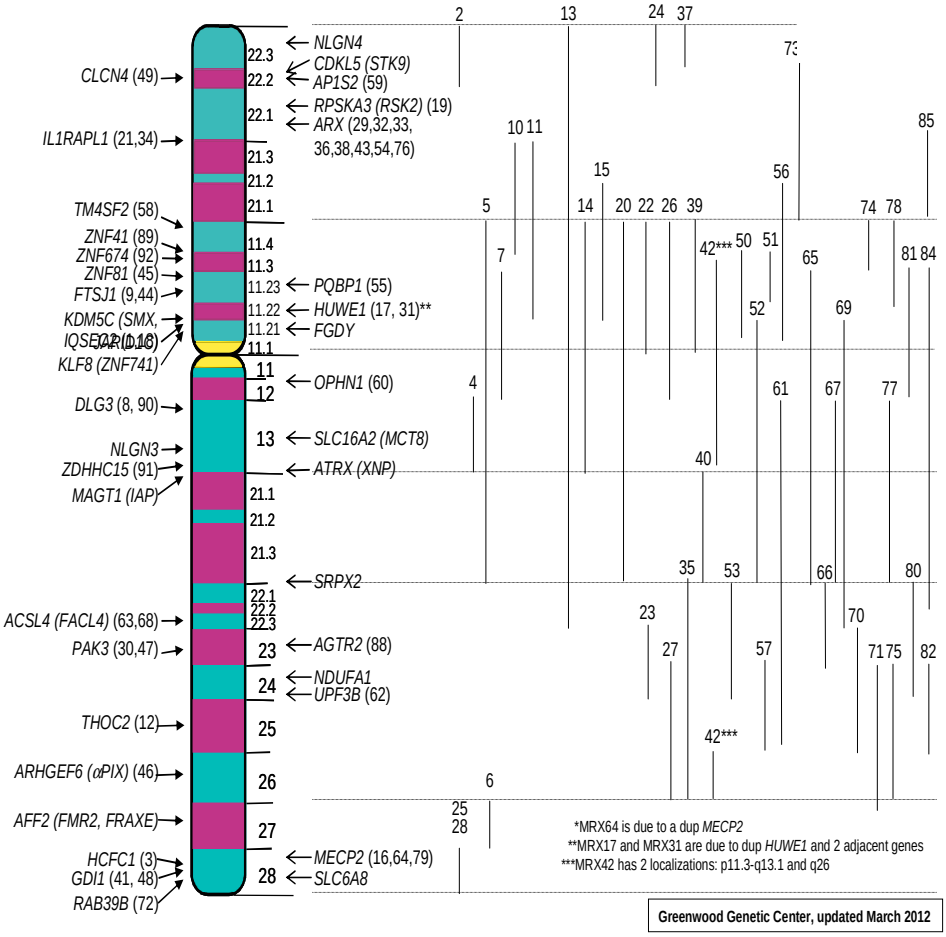


Figure 4 : Les gènes responsables de DILX-NS

Les nouvelles approches pangénomiques ont aussi contribué à mieux appréhender la fréquence de certaines microduplications dont certaines incluent des gènes de DI prouvant l'importance du dosage génique de ces gènes dans le développement des fonctions cognitives: (i) au locus Xp11.22, d'une taille variant entre 0,3 et 0,8 Mb englobant les gènes *HSD17B10* (17 beta HydroxySteroid Dehydrogenase 10) et *HUWE1* (HECT, UBA, and WWE domains containing protein 1**)** dans six familles différentes porteuses d'une DILX-NS [99]. (ii) en Xq26.2-q27, des duplications d'une taille allant de 3,9 à 7,5 Mb, concernant le gène *SOX3* (Sry bOX 3) sont associées à un hypopituitarisme lié à X chez 3 patients non apparentés [100]. (iii) en Xq21-q22 une duplication de taille variable contenant le gène protéolipidique protein 1 (*PLP1*) au syndrome de Pelizaeus Merzbacher (DILX syndromique) détectée dans plus de 50 familles [101] (iv) la duplication en Xq28, de taille variant entre 0,2 et 2,2 Mb et contenant le gène *MECP2* représentent une cause reconnue de DI chez les garçons [102-104].

1.1.5 Déficience intellectuelle liée aux autosomes

1.1.5.1 Hérédité autosomique dominante

Les déficiences intellectuelles répondant à une hérédité autosomique dominante (DIAD), dans leur expression modérée à sévère, sont le plus souvent en rapport avec la survenue d'une mutation *de novo*, les individus atteints ayant un coefficient d'adaptation ne favorisant pas la procréation. Pour certaines affections telles la neurofibromatose (incidence ~ 1/4000) la sclérose tubéreuse de Bourneville, (~ 1/6000), et la dystrophie myotonique (~ 1/8000), une DI peut être observée de façon inconstante au sein d'une même famille [57].

Parmi les gènes de DIAD non syndromique, des études ciblées sur des voies de signalisation spécifiques ou par criblage systématique utilisant les techniques de séquençage haut débit ont permis d'identifier des mutations, à ce jour pour un petit nombre d'individus, dans certains gènes dont: (i) le gène ***SYNGAP1*** (synaptic Ras GTPase activating protein 1) [105] contrôlant la plasticité synaptique par l'insertion des récepteurs de l'alpha-amino-3-hydroxy-5-méthyl-4-isoxazole (AMPA) dans la membrane synaptique et régulant également le cytosquelette d'actine ; (ii) le gène ***STXBP1*** (syntaxin binding protein 1) [106] exercant une fonction essentielle de régulation de la fusion des vésicules synaptiques et la libération de neurotransmetteurs dans la fente synaptique ; (iii) le gène ***FOXG1*** (forkhead box G1) codant un facteur de transcription dont l'altération est aussi associée à une anomalie morphologique du développement cérébral [107] ; (iv) le gène ***MEF2C*** (myocyte enhancer factor 2C) qui régule négativement le nombre et la fonction des synapses [108]. Très vraisemblablement,

l'extension des capacités de séquençage haut débit donnera une nouvelle vision des gènes impliqués dans la survenue de DIAD isolée, en appréciant le niveau d'hétérogénéité génétique et la fréquence respective des mutations.

1.1.5.2 Hérédité autosomique récessive

A l'heure actuelle, tout comme les DIAD, l'élucidation moléculaire de la déficience intellectuelle autosomique récessive (DIAR) a pris du retard par rapport à DILX. La recherche sur les DIAR a été ralentie par les familles de petite taille en Europe ou aux Etats unis, où la plupart de ces recherches a lieu. La plupart des patients atteints de DIAR sont des cas isolés grevant les analyses de liaison génétique. Parfois, de petites délétions ou des réarrangements chromosomiques équilibrés qui altèrent une copie du gène peuvent démasquer des mutations récessives de l'autre allèle. La cartographie d'homozygotie dans les grandes familles consanguines est une méthode de choix pour localiser les gènes des DIAR. Ces familles sont préférentiellement observées dans les pays avec de fréquents mariages consanguins [109]. Plusieurs gènes de DIAR non syndromique ont été récemment identifiés, participant à des voies de signalisation cellulaire variable, tels *PRSS12* [110], *CRBN* [111], *CC2D1A* [112], *GRIK2* [113], *TUSC3* [114], et *TRAPPC9* [115]. La part de ces gènes dans l'étiologie de la DI, en terme de fréquence, reste encore à déterminer, manifestement avec le soutien des outils d'exploration pan génomique disponibles à titre diagnostique dans les très proches années.

1.1.6 Physiopathologie des DILX

Le grand nombre des gènes impliqués dans les DILX code des protéines qui, à première vue, semblent participer à la diversité des fonctions biologiques très variés exercées par leurs protéines. Les gènes peuvent être implique dans des fonctions très générales (régulation de la transcription et remodelage de la chromatine), ou bien plus spécifiques (régulation du cytosquelette d'actine et donc de la morphogenèse neuronale). Parmi elles, la structure et la fonction de la synapse fait l'unanimité [17, 20, 116-118].

D'autre part, la dérégulation de la transcription et le remodelage de la chromatine semblent fréquemment en cause dans les DILX. Tous ces trois vois biologique sont essentiel pour le bon fonctionnement cognitif.

La majorité des produits des gènes impliqués dans les DI, si l'on met à part les gènes intervenant dans la transcription et le remodelage de la chromatine, est localisée dans l'espace

pré ou post synaptique. La description de l'ensemble des gènes de DILX et des pathologies qui résultent de leurs mutations distingue les catégories de DILX selon le type de dysfonctionnement cellulaire : (i) la réorganisation du cytosquelette d'actine et le contrôle de la croissance des neurites, *OPHN1* (Oligophrenin 1), *PAK3* (p21 Protein Activated Kinase 3), *ARHGEF6* (Rho Guanine nucleotide Exchange Factor 6), (ii) la traduction *FTSJ1* (FtsJ RNA methyltransferase homolog 1), la dégradation protéique *MID1* (Midline 1), le métabolisme énergétique *SLC6A8* (solute carrier family 6 member 8), (iii) les molécules impliquées dans la structure et les fonctions des synapses, *DLG4* (Discs Large homoloG 4), *FMR1*, *IL1RAPL1* (InterLeukin 1 Receptor Accessory Protein-Like 1) , *SHANK3* (SH3 and multiple ANKyrin repeat domains 3), (iv) le remodelage de la chromatine et le contrôle de la transcription, *MECP2*, *CDKL5* (Cyclin Dependent Kinase Like 5), *BCOR* (Bcl6 CORepressor), *ARX*, *FMR1 et 2*.

1.1.6.1 Gènes impliqués dans la réorganisation du cytosquelette d'actine et le contrôle de la croissance des neurites

Trois gènes (*OPHN1*, *PAK3*, *ARHGEF6*), associés à la DI, codent des protéines qui interviennent dans la voie de signalisation des RhoGTPases, de la famille Ras-like GTPases qui agit dans la dynamique du cytosquelette d'actine en réponse à des stimuli extracellulaires [119]. L'étude fonctionnelle a montré les mutations dans RhoGTPases conduit la perte de fonction et des anomalies de la migration neuronale et de la croissance dendritique et axonale [120].

OPHN1, code l'Oligophrenine 1, fortement exprimée dans le cerveau fœtal et adulte, dans les cellules neuronales et gliales. L'activité GTPase des membres de la famille Rho, (RhoA, Rac et Cdc42) stimule grâce au domaine Rho-GAP de l'Oligophrenine 1 , qui est impliqué dans l'organisation du cytosquelette et la dynamique des cônes de croissance [121].

PAK3 code une protéine kinase de la famille des PAKs (p21 activating kinase), qui a un rôle d'effecteur en aval des Rho-GTPases [122]. L'activation des cascades de MAP kinases (mitogen activated protein) est également initié par *PAK3* et il implique dans la dynamique du cytosquelette d'actine [123]. La mutation ponctuelle d'un codon stop dans le domaine kinase est responsable de DI, sans modification du site de liaison-activation à Rac et Cdc42.

ARHGEF6 (Rho Guanine nucleotide Exchange Factor 6) code une protéine qui possède des homologies avec les RhoGEFs. Elle recrute des protéines PAK du cytosquelette, structures

riches en actine et cette formation du cytosquelette est contrôlée par l'activité de Cdc42 et Rac1.

1.1.6.2 Gènes impliqués dans le métabolisme énergétique

SLC6A8 code une protéine transporteur de la créatine qui est dépendant des ions sodium et chlore. Il recapture de certains neurotransmetteurs (dopamine, GABA (acide gamma-aminobutyrique), sérotonine et noradrénaline) et acides aminés (glycine, proline, taurine). Une maladie métabolique caractérisée par une DI liée à l'X, survenu par des mutations dans le gène *SLC6A8.* Un déficit de la créatine cérébrale avec un comportement autistique et parfois une épilepsie sont également associés avec ce gène [124-125]. Le mécanisme impliqué dans la déficience intellectuelle serait en rapport avec une perturbation de l'apport énergétique pour le développement du cerveau [124].

1.1.6.3 L'activité synaptique dans la déficience intellectuelle

Le glutamate est un des principaux neurotransmetteurs du système nerveux central. Il est distribuer dans la fente synaptique modulent la libération de neurotransmetteur en agissant comme auto-récepteurs pré-synaptiques. Ainsi, la transduction du signal par les mGluRs dans la membrane post synaptique est déclenchée. Le NMDA (ion channel forming Nmethyl D aspartic acid) et les récepteurs métabotropiques du glutamate (mGluR) couplés aux protéines G hétérotrimériques sont les deux principaux récepteurs du glutamate.
Les 39 des 1124 gènes codant des protéines du PSD (postsynaptic density) et 7 des 186 gènes correspondant aux protéines du NRC/MASC (NMDA receptor complex/MAGUK-associated signaling complex) sont localisés sur le chromosome X. NMDA impliqués dans le phénomène de potentialisation à long terme, processus de mémorisation et d'apprentissage ont été fréquemment impliqués dans des DILX. Tous les autres gènes localisés sur le chromosome X liés au complexe PSD sont donc des candidats potentiels pour les DILX. La figure 5 montre l'implication des différents gènes de ces complexes dans les DILX dans l'espèce humaine et la souris.

1.1.6.3.1 Interleukin-1 receptor accessory protein- like 1 (*IL1RAPL1*)

Ce gène est impliqué dans une déficience intellectuelle non syndromique et parfois avec troubles autistiques. *IL1RAPL1* appartient de la famille des récepteurs de Toll/IL-1. Le niveau d'expression de la protéine est plus élevé dans l'hippocampe, le gyrus denté et le cortex entorhinal, suggérant un rôle spécialisé dans les processus physiologiques de la mémoire et la capacité de l'apprentissage [126-130].

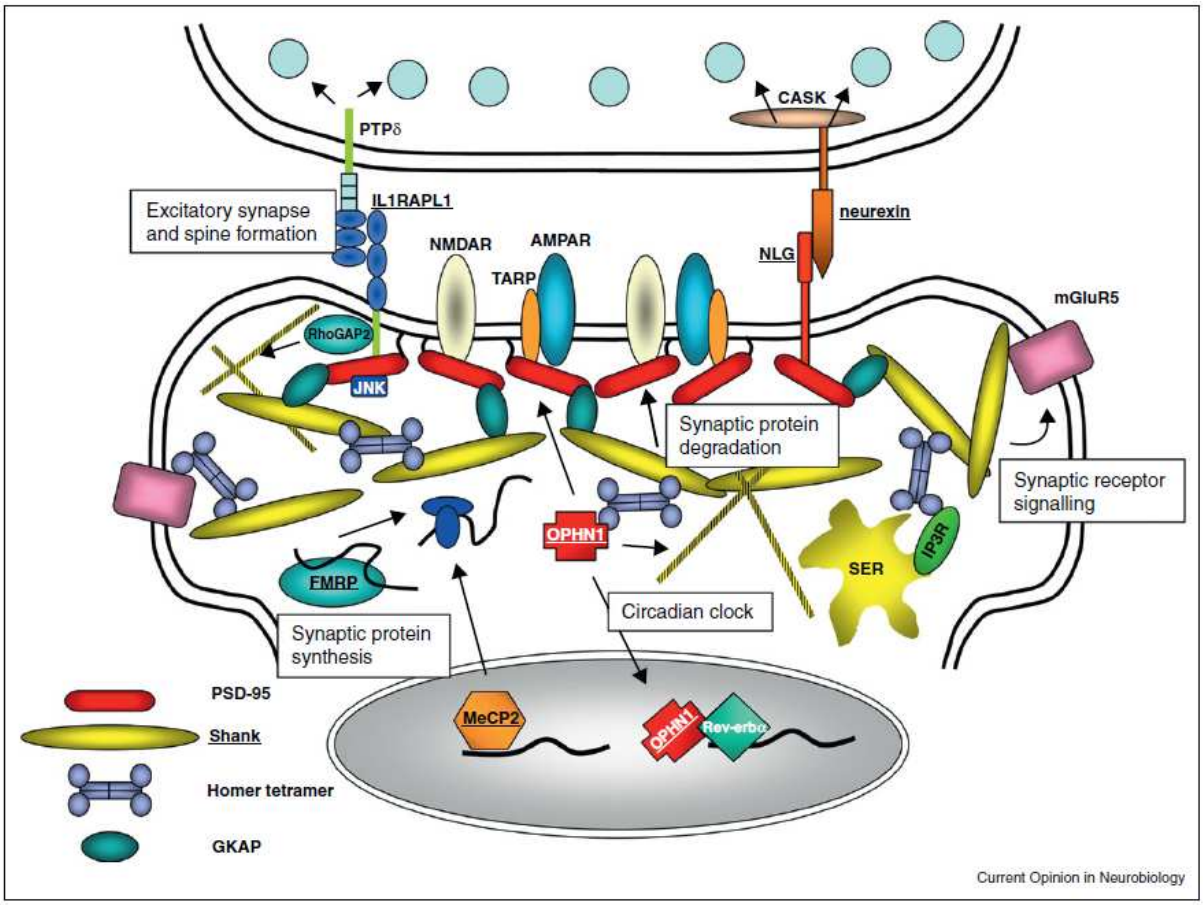


Figure 5 : Une organisation schématique du réseau des protéines au niveau des synapses excitatrices dont les mutations ont été observées lors de la DI

D'après Verpelli et al., 2012 [131]

IL1RAPL1 se lie à la protéine de densité post-synaptique 95 (PSD-95) et régule sa phosphorylation et son association synaptique par l'activation de JNK (c-Jun terminal kinase) [132]. *IL1RAPL1* est considéré comme un médiateur trans-synaptique de signalisation qui régule la formation de synapses excitatrices et d'épines dendritiques. Des mutations perte de

fonction de ces protéines entraîneraient ainsi la réduction du nombre de synapses excitatrices associée à une déficience cognitive.

1.1.6.3.2 FMRP : la protéine du retard mental lié à l'X fragile

Le syndrome de l'X fragile est la principale cause de DI héréditaire chez l'homme. Les patients atteints de ce syndrome ont des épines dendritiques plus longues et plus minces que la normale [133]. Le gène *FMR1* code la protéine FMRP (Fragile X Mental Retardation Protein) appartenant à la famille des ribonucléoprotéines se liant aux ARN; elle régule le transport et la traduction d'une partie des ARN messagers neuronaux aux synapses [133-135]. Son expression et sa localisation dans les dendrites augmentées après la stimulation synaptique, suggèrent un lien direct entre FMRP et la plasticité synaptique [136-138]. Lorsque la FMRP n'est pas exprimée, la dérégulation de la traduction des ARNm cibles semble participer à l'altération de l'épine dendritique et des modifications synaptiques qui caractérisent le syndrome de l'X fragile [134]. Les niveaux d'expression des protéines scaffold postsynaptiques Shank1, Shank 3, et IRSp53/BAIAP2 (brain-specific angiogenesis inhibitor associated protein 2), ainsi que des sous-unités NR1 (NMDA recepteur1) et NR2B (NMDA recepteur 2B) du récepteur NMDA et GluR1 (glutamate recepteur1), sous-unité du récepteur AMPA (Alpha amino 3-hydroxy 5-méthylisoazol 4-propionate) sont augmentés dans le cortex et l'hippocampe des souris invalidées pour *FMR1*. FMRP est également un régulateur négatif de la transcription de la sous-unité NR2A (NMDA recepteur 2A) de NMDA. Cette régulation est influencée par le microARN miR-125b. Ces résultats suggèrent que l'absence de FMRP modifie la plasticité synaptique dans le syndrome de l'X fragile en modifiant la composition des sous-unités du récepteur NMDA [139].

1.1.6.3.3 Les protéines d'échafaudage synaptique ou protéines scaffold

PSD-95 est la protéine scaffold la plus abondante de la densité post synaptique (PSD), protéine assurant un rôle de régulateur-clé de la signalisation synaptique et de l'apprentissage et codée par le gène *DLG4* (discs large homolog 4). Le gène *DLG3* (discs large homolog 3) humain, qui code la protéine post synaptique SAP102 (synapse-associated protein 102), est clairement associé à la DI [140-141]. Les mutations identifiées introduisent des codons stop prématurés empêchant la capacité de SAP102 tronquée à interagir avec le récepteur de NMDA et d'autres protéines de sa voie de signalisation [142].

1.1.6.3.4 Famille des protéines SHANK / ProSAP

La monosomie 22q13.3 ou syndrome de Phelan-McDermid est caractérisée par une hypotonie néonatale, un retard global de développement, une absence ou un retard sévère en matière d'acquisition du langage ainsi qu'un comportement de type autistique. Le gène ProSAP2/*SHANK3* (SH3 and multiple ANKyrin repeat domains 3), code une protéine structurale post synaptique au niveau des synapses excitatrices. La délétion de ce gène est considérée comme responsable de ce syndrome. Des mutations de novo dans *SHANK3* mais aussi du gène de la même famille, *SHANK2,* ont été identifiées chez les patients atteints de troubles autistiques et de DI [143-145].

1.1.6.4 Régulation de la transcription et remodelage de la chromatine

Certains gènes régulant l'expression d'autres gènes peuvent être responsables lorsqu'ils sont mutés de DILX. Parmi les mécanismes impliqués dans la répression génique, les modifications épigénétiques jouent un rôle fondamental. Les mutations du gène methyl-CpG binding protein 2 (*MECP2*) sont caractéristique ; elles sont responsables du syndrome de Rett, une maladie génétique quasi exclusivement diagnostiqué chez les filles ; MECP2 est une protéine à domaine MBD (methyl binding domain) qui se lie aux dinucléotides CpG methylés, permettant ainsi le recrutement de corépresseurs (tel Sin3A- SIN3 transcription regulator homolog A) et d'histones désacétylases (HDAC1 et HDAC2).
L'organisation structurale de ce complexe protéique de la chromatine étant primordiale pour le contrôle de l'expression des gènes cibles, intervenant dans la maturation post mitotique des neurones. Par ailleurs, l'analyse in vivo sur des modèles de souris invalidées pour *MECP2* ont permis de mettre en évidence ; l'expression de certain nombre des gènes est affectée lorsque la protéine Mecp2 dans la formation de boucles chromatiniennes silencieuses est absente tels DLX5/DLX6 (distal less homeobox 5/6) et Bdnf (brain derived neurotrophic factor).

Les troubles liés au développement neuronal sont souvent causés par la dérégulation de l'expression de gènes soumis à empreinte génomique parentale, par exemple, au locus 15q11-13, le syndrome de Prader–Willi (PWS) et le syndrome d'Angelman (AS). Des manifestations cliniques similaires parmi les syndromes mentionnés ci-dessus suggère que les réseaux communs et des voies moléculaires peuvent expliquer une partie des défauts neurologiques. Les gènes soumis à empreinte peuvent influencer le fonctionnement du cerveau en affectant les processus de développement des neurones [146].

1.2. Les stratégies pour la recherche de gène candidat dans la DI

1.2.1 Cartographie génétique

Au cours de la méïose, la ségrégation indépendante des chromosomes aide à déterminer si des gènes sont situés sur le même chromosome ou sur des chromosomes différents. Les deux loci qui ségrégent ensemble sont liés donc situés sur le même chromosome et ils seront hérités ensemble.

On appelle ce phénomène « liaison génétique » des gènes qui sont situés sur le même chromosome à être transmis ensemble. Lors du crossing over (événement d'échange réciproque) au cours de la méïose ; il existe également une possibilité de recombinaison entre les chromatides de chromosomes homologues peut séparer des gènes liés.

La fréquence de la recombinaison méiotique a été utilisé pour calculer la distance génétique entre les gènes. Plus deux gènes sont proches, moins est de probabilité d'une recombinaison entre eux, donc plus leur liaison génétique est élevée. La cartographie du génome humain a été développé par des différentes marqueurs génétiques. Ces marqueurs ont un locus précis et unique au niveau du génome ; dont ils ont été choisis en fonction non seulement de leur spécificité mais aussi en fonction de leur polymorphisme avec des allèles différents. Ainsi, les gènes déterminant des pathologies peuvent être cartographiés en établissant leur co-ségrégation avec des marqueurs polymorphes de positions connues dans le génome, dont les allèles se transmettent spécifiquement avec la maladie.

Ainsi la localisation du gène candidat peut être déterminer par un intervalle de liaison. Il est nécessaire de pouvoir tester un grand nombre de marqueurs polymorphes régulièrement répartis sur l'ensemble des chromosomes. Ces marqueurs de polymorphismes correspondent à des variations de séquences entre les différents allèles d'un même locus dans une population. Il en existe différents types tels les polymorphismes de longueur des fragments de restriction, les microsatellites, et les marqueurs de type SNP. Ainsi, une pathologie autosomique récessive a pu être identifiées partagée par les sujets porteurs des régions d'homozygotie.

Toutefois l'utilisation et le succès de l'analyse de liaison dans le clonage de gène candidat repose sur la notion de familles informatives avec au mieux plusieurs cas de sujets atteints et de sujets non atteints. Pour les affections survenant de novo, de façon sporadique, isolée dans une famille, ce type d'analyse n'est pas adapté. Toutefois, dans certains cas de maladies

autosomiques récessives dans les familles consanguines la cartographie d'homozygotie peut contribuer à identifier des gènes candidats.

1.2.2 Le caryotype et l'hybridation in situ en fluorescence

Il donne lieu à une approche globale de l'ensemble des chromosomes, visualisant les anomalies de structure sans perte ni gain de matériel telles que les inversions et les translocations réciproques mais aussi des aneusomies segmentaires, duplications ou délétions à condition que la taille en soit suffisante (supérieure à 10Mb). Il peut donc jouer un rôle important dans la découverte d'un gène à l'origine d'une maladie, notamment en élucidant les points de cassure d'un remaniement équilibré qui peuvent interrompre la structure ou la régulation de l'expression d'un gène. Pour cela, les techniques d'hybridation in situ en fluorescence ont largement facilité la marche sur le chromosome pour la quête de gènes candidats. Cette technique est basée sur le principe d'hybridation de 2 brins d'ADN de séquence complémentaire. FISH utilise ce processus pour former une molécule d'ADN double brin et localise des sondes sur des chromosomes de la région à étudier (sur un chromosome métaphasique ou un noyau en interphase), marquée par une molécule fluorescente [147]. La lecture se fait par microscopie à épifluorescence. Les sondes utilisées sont le plus souvent des clones de BAC (Chromosome Artificiel de Bactérie) dont la cartographie est actuellement référencée dans les bases de données génomiques. Ces sondes contribuent à délimiter avec précision la région des points de cassures afin de rechercher, connaissant la position exacte des clones BAC et leur contenu en gènes, les gènes candidats potentiellement remaniés par l'anomalie chromosomique.

1.2.3 L'hybridation génomique comparative sur microréseau (CGH-array)

Les conséquences phénotypiques d'une anomalie chromosomique sont dans la majorité des cas le résultat d'un déséquilibre touchant un ou plusieurs gènes sensibles au dosage génique. L'hybridation génomique comparative sur microréseau d'ADN (CGH-array) a récemment modifié l'approche diagnostique des anomalies chromosomiques. Cette technique détecte des déséquilibres génomiques de petite taille de quelques dizaines de kilobases à quelques mégabases, jusqu'alors non identifiés par l'étude du caryotype. Le développement de la

CGH-array, notamment des puces à oligonucléotides, permet d'identifier les remaniements déséquilibres génomiques, tels que les duplications et délétions, de quelques kilobases. Le principe de cette technique repose sur la cohybridation deux ADN équimoléculaire et compétitive, un de témoin et l'autre de patient, marqués par un fluorophore différent (cyanine 5 et cyanine 3) sur des sondes de séquence connue fixées sur une lame de verre.
En comparant le ratio des intensités relatifs de fluorescence mesurées par un scanner, on peut identifier les gains et les pertes le long des chromosomes. Les sondes sont réparties sur l'ensemble du génome. La résolution de la technique varie considérablement est défini par la nature, le nombre, la distance génomique moyenne entre les sondes et leur répartition, les sondes à oligonucléotides (de 25 à 85 nucléotides) permettent d'obtenir une résolution allant jusqu'à 6kb actuellement (244 000 oligonucléotides). Cette technologie a été la plus utilisée pour détecter les CNV de quelques dizaines de paires de bases seulement et au diagnostic étiologique moléculaires de nombreuses pathologies, notamment dans le domaine de la cytogénétique constitutionnelle pour des nombreuses pathologies [148-150].
Grâce à cette technique plusieurs nouveaux syndromes microdélétionnels et des gènes candidats ont été identifiés, en délimitant des régions minimales critiques. Des réarrangements complexes sont bien caractérisées après la détection par CGH-array en particulier des translocations apparemment équilibrées, révélant ainsi des déséquilibres submicroscopiques. Plusieurs études ont démontré l'importance de variation de nombre de copies dans la déficience intellectuelle, d'autisme, et les retards de développement [48, 151-152].

1.2.4 Séquençage de nouvelle génération / séquençage haut débit

Au cours des cinq dernières années, il y a eu un changement fondamental de l'application du séquençage par systèmes de Sanger automatisés pour l'analyse du génome. La méthode de Sanger avait dominé l'industrie depuis deux décennies et a conduit à un certain nombre de réalisations fondamentales, y compris le séquençage du génome humain. Malgré de nombreuses améliorations dans la technique, les limites du séquençage Sanger ont montré un besoin d'une nouvelle technologie pour améliorer le séquençage de nombreux génomes humains. La méthode automatisée Sanger est considérée comme le séquençage de «première génération», et les nouvelles méthodes sont appelées « séquençage de nouvelle génération » (NGS). D'une façon générale, le processus de séquençage peut être regroupé en préparation de la librairie, séquençage, imagerie, et l'analyse des données. Nous sommes à l'orée d'une

période charnière entre les technologie à haut débit dites de 2ème génération qui requièrent une étape d'amplification des molécules d'ADN en amont du décodage, de celles dites de 3ème génération permettant le décryptage direct d'une seule molécule matrice d'ADN.
L'avancement majeur offert par la NGS est la capacité de produire un énorme volume de données à moindre coût. La capacité de séquencer le génome entier de nombreux organismes a offert la possibilité d'études comparatives et évolutives à grande échelle jusqu'alors inimaginables. La plus large application de la NGS est le re-séquençage du génome humain afin d'améliorer notre compréhension de la façon dont les différences génétiques influent sur les maladies génétiques [153].

1.2.4.1 Préparation des banques ou librairies

L'échantillon d'ADN est fragmenté, et la ligation d'adaptateurs aux deux extrémités permet l'utilisation ultérieure d'amorces universelles pour fixer et amplifier chacun de ces fragments d'ADN. Trois types de préparation de librairie sont classiquement décrits : single-end (SE), paired-end (PE), et mate-paired (MP) librairies [154]. Les librairies SE sont créées par la sonication de l'ADN génomique en fragments qui ont moins de 1 kb [154]. Le protocole de la préparation de la librairie paired-end (PE) est proche du protocole utilisé pour la librairie SE, mais les amorces de séquençage différentes sont ligaturées à chaque extrémité. Ainsi les lectures PE sont générées par le séquençage d'un fragment d'ADN à partir des deux extrémités d'une manière séquentielle. Après une analyse de la première lecture avec amorce du séquençage S1, les modèles sont régénérés et la seconde lecture est séquencée par l'utilisation de amorce du séquençage S2. Cette technologie permet la création de paires de lecture qui se situent entre 200 à 500 pb (paires de bases) [154]. Dans des protocoles Mate-pair librairie (MP), l'ADN, fragmenté en fragments de 2 à 5 kb, est ensuite marqué à l'extrémité, circularisé, segmenté et donc à nouveau linéarisé. Les fragments contenant le marqueur et donc les deux extrémités du fragment d'ADN choisis sont séquencés tel que décrit dans le séquençage PE [154]. Après l'étape de la préparation de la librairie initiale, une étape d'amplification d'ADN assure une intensité du signal suffisante pour la détection des nucléotides (PCR en émulsion ou l'amplification par Bridge PCR).

PCR en émulsion : l'ADN simple brin s'hybride sur les billes qui portent l'oligonucléotide cible. Les billes font partie de l'émulsion (eau et huile) contenant les réactifs

nécessaires pour l'amplification par PCR (Réaction en chaîne par polymérisation). Après l'amplification, les billes avec les brins d'ADN amplifiés sont lavées et les billes sont purifiées et immobilisées pour le séquençage (Figure 6). [154-155].

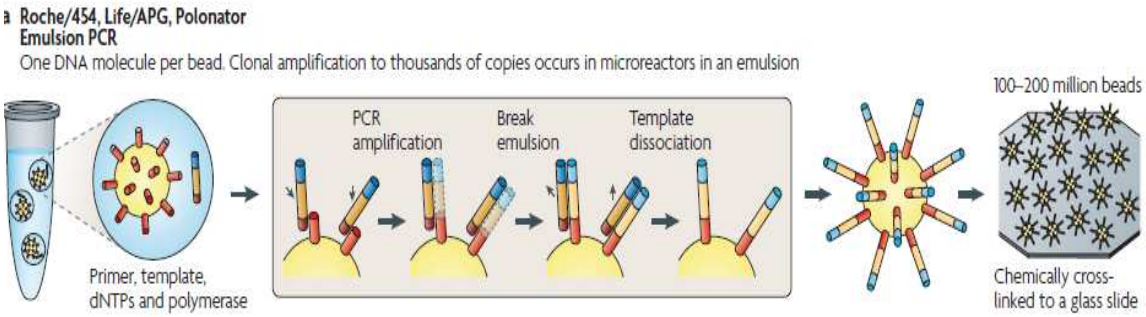


Figure 6 : Stratégie d'amplification par PCR en émulsion (PCRem)

Un mélange réactionnel constitué d'une émulsion huile-aqueuse est créé pour encapsuler les complexes bille-ADN dans des gouttelettes aqueuses. L'amplification par PCR est réalisée au sein de ces gouttelettes pour créer des billes contenant plusieurs milliers de copie du même fragment. Les billes PCRem sont chimiquement liées à une lame de verre ou déposées dans les puits de la plaque PicoTiter Plate. D'après Metzker 2010 [154].

Bridge PCR : dans cette approche, les deux amorces de PCR sens et anti-sens sont hybridées sur un substrat ou un support solide par une molécule linker. Après l'amplification tous les amplicons sont immobilisés et regroupés sur un seul emplacement physique. (Figure 7)**.** Après dénaturation, une étape de lavage élimine tous les fragments d'origine. Les autres liés de manière covalente avec des amorces inverses complémentaires de l'extrémité permettent la création des ponts et la synthèse du brin complémentaire. Ensuite l'ADN est dénaturé pour obtenir l'ADN simple brin. De ce fait, des clusters ADN sont produits sur toute la surface solide. [156-157].

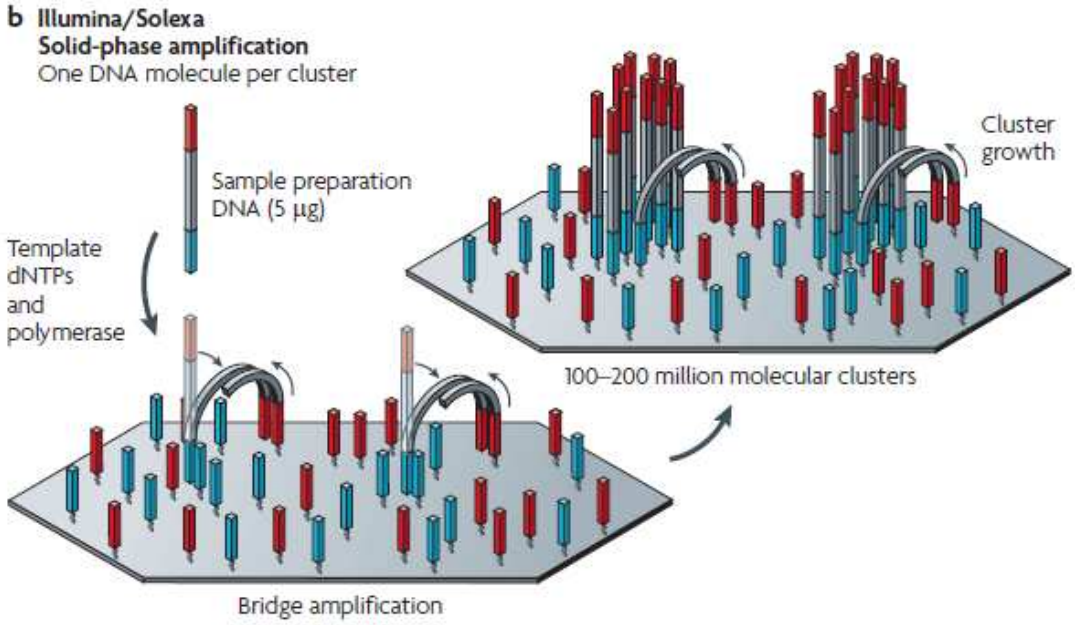


Figure 7 : Stratégie d'amplification par Bridge PCR

La PCR en pont ou Bridge PCR est faite en deux étapes: hybridation des amorces initiales et élongation simple brin d'un fragment et l'amplification en pont des fragments d'ADN immobilisé avec des amorces adjacentes pour former des clusters. D'après Metzker 2010 [154].

Les trois plateformes de NGS les plus utilisées actuellement sont celles des sociétés commerciales Roche, Illumina/Solexa, et Life Technologies (Figure 8)

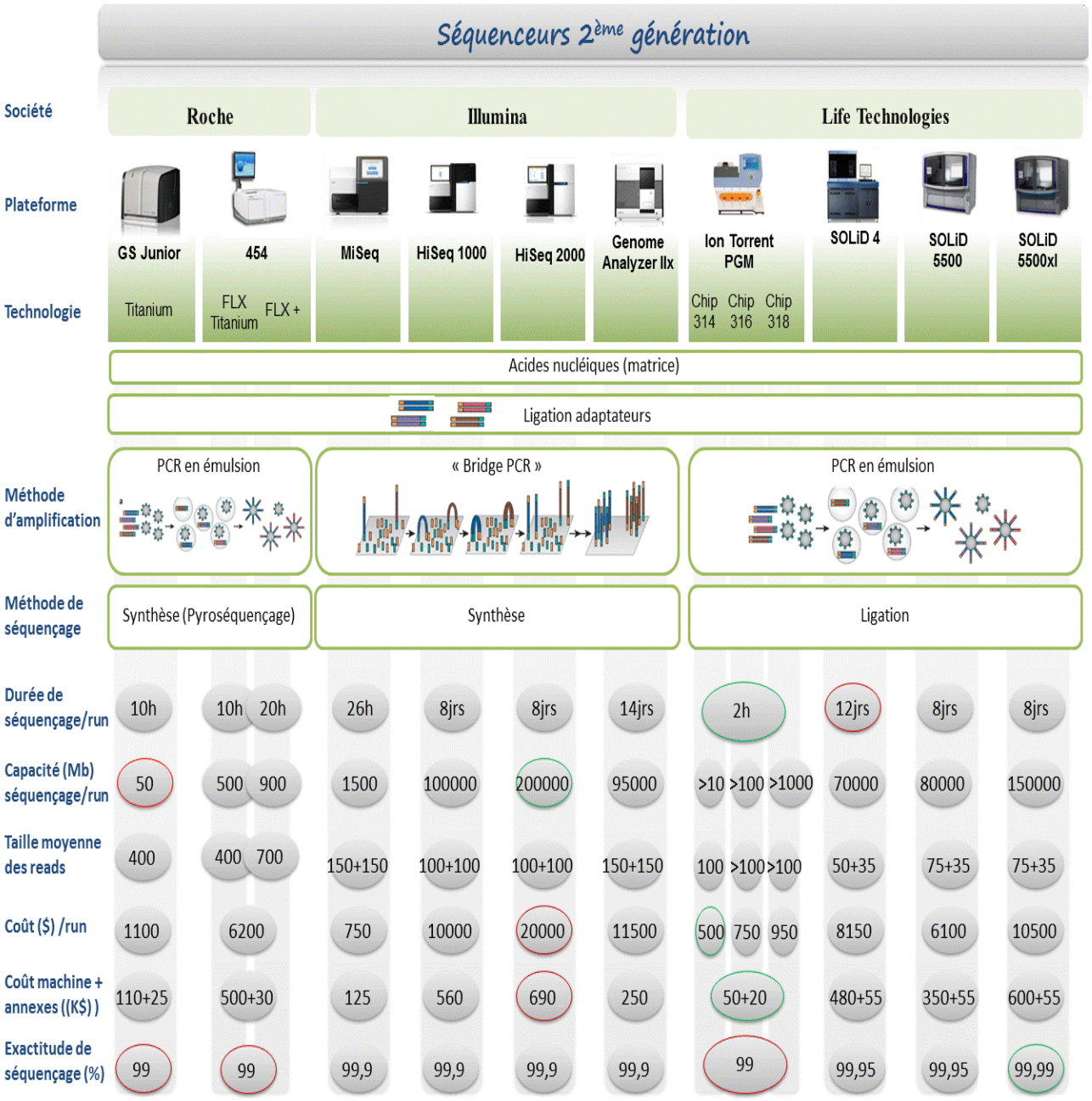


Figure 8 : Les trois plateformes du sequençage haut debit

D'après Biorigami (http://www.biorigami.com)

Les techniques d'amplification des fragments d'ADN cibles, la méthode du séquençage, la génération des données avec la taille moyenne des lectures, le coût par run et la fidelité de séquence. Les points postifs sont signalés en couleur verte et les points faibles en rouge.

1.2.4.2 454 Genome Sequencer FLX instrument (Roche Applied Science)

Cette plateforme est basée sur le principe du **pyroséquençage** (Figure 9). Les nucléotides (sous forme désoxyribonucleotide triphosphate) sont ajoutés l'un après l'autre. Si le nucléotide ajouté est complémentaire du nucléotide du brin matrice, il est incorporé dans le brin en cours de synthèse et un pyrophosphate inorganique (PPi) est libéré. L'ATP sulfurylase transforme le pyrophophate libéré en ATP en présence d'un substrat : l'adénosine 5' - phosphosulfate (APS). L'ATP formé est utilisé par une luciférase qui transforme la luciférine en oxyluciférine qui génère un signal lumineux dans le visible proportionnel à la quantité d'ATP. L'apyrase dégrade les nucléotides non incorporés et l'excès d'ATP. Le capteur CCD du séquenceur capte le signal lumineux et le traduit par un pic sur le pyrogramme™. A noter que la précision de lecture est plus faible dans les régions avec des bases répétées (homopolymères).

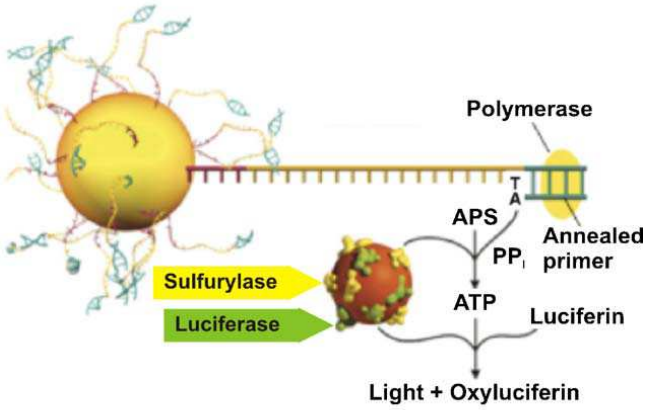


Figure 9 : Schéma de la réaction du pyroséquençage

D'après Roche life science (http://www.454.com)

1.2.4.3 Illumina (Solexa) Genome Analyzer

Le séquençage est basé sur l'amplification, l'accrochage liaison sur puce et l'utilisation de terminateurs de chaîne réversibles marqués par des fluorochromes (Figure 10). Une diminution de la fiabilité du séquençage aux extrémités des lectures est rapportée.

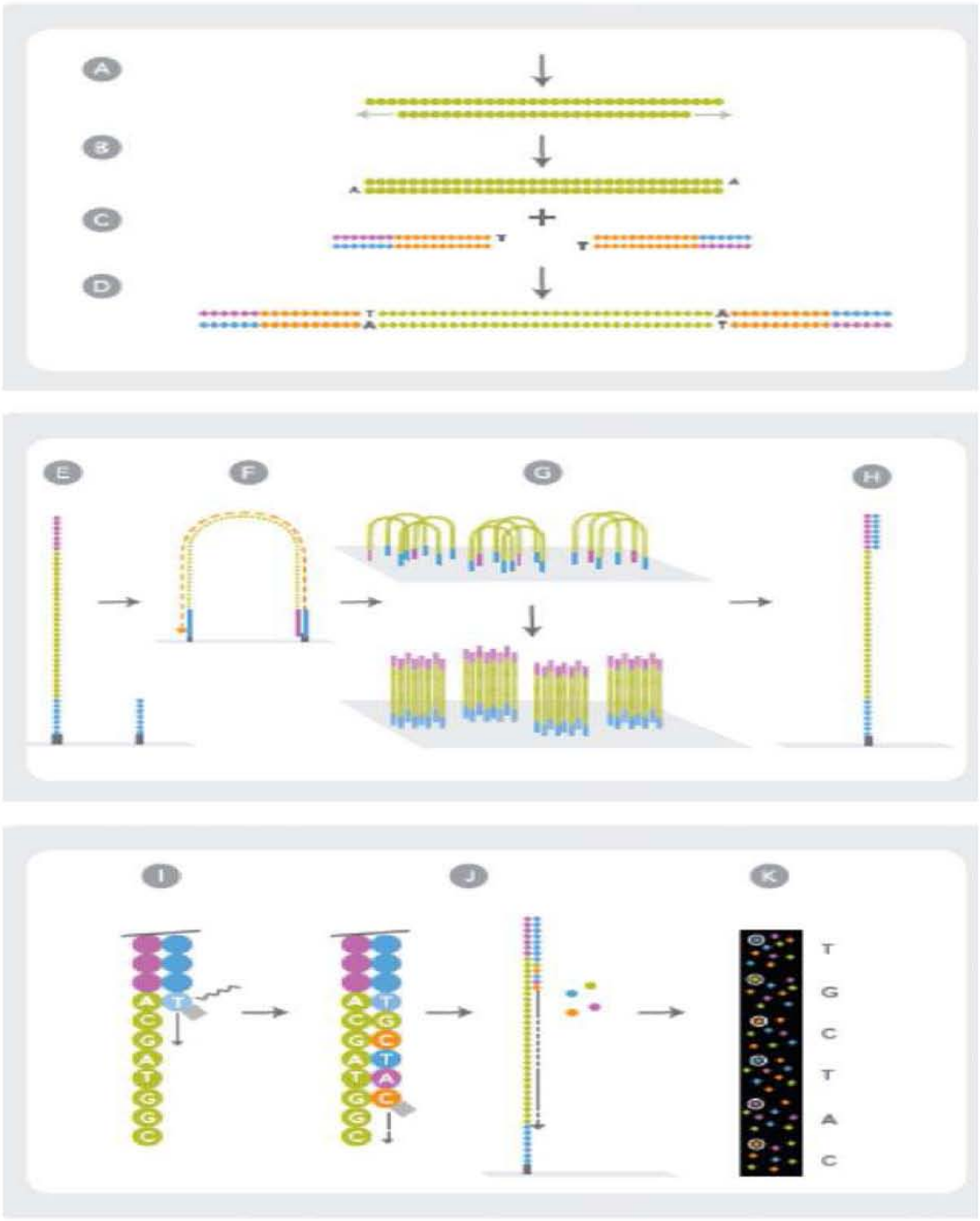


Figure 10 : La représentation de la technique du séquençage par Illumina genome analyzer

A. l'ADN fragmenté ; B. Réparation des extrémités ; C et D. ligature des adaptateurs; E. les fragments sont hybridés sur un support/lame ; F et G. le PCR en pont; H. hybridation des amorces du sequencage ; I et J l'élongation ; K. base calling.

D'après Illumina (http://www.illumina.com)

1.2.4.4 Les systèmes de Life Technologies (SOLID et Ion Torrent PGM)

Le system SOLiD™ (Supported Oligonucleotide Ligation and Detection) est basé sur l'amplification par émulsion et l'hybridation-ligature chimique. Il utilise une ligation avec une ADN ligase. Les amorces de séquençage universelles, l'ADN ligase et un pool d' oligonucleotides (interrogeant les deux premières bases) marqués sont ajoutés. La séquence d'ADN est déterminée par la base du code de couleur, représentant la couleur des deux premières bases. En raison de l'encodage à deux couleurs chaque base est déterminée de manière indépendante deux fois. Le séquençage est réalisé par des cycles répétés d'hybridation par un mélange des amorces du séquençage et des sondes marquées par fluorescence. Cette étape est suivie par la détection des signaux fluorescents sur les sondes qui codent pour les bases interrogées. Parce que chaque base est déterminée par un marqueur fluorescent différent, le taux d'erreur est réduit dans cette technique.

La technologie Ion torrent (PGM Personal Genome Machine) est basée sur des puces semi-conductrices remplies de puits. Un proton est relargué quand un nucléotide est incorporé par la polymérase dans l'ADN ; cela résulte en un changement de pH local qui est détecté par cette technologie sensible à la variation d'ions (Figure 11). Cette technologie ne nécessite pas de caméra, pas de scanner, pas de cascade enzymatique, pas de fluorophore ou chemiluminescence. Il est annoncé un débit de 1Go pb/expérience. Temps total d'expérience (construction de la banque/données intégrées) : 2 jours et un tour (run) : 3.5 heures.

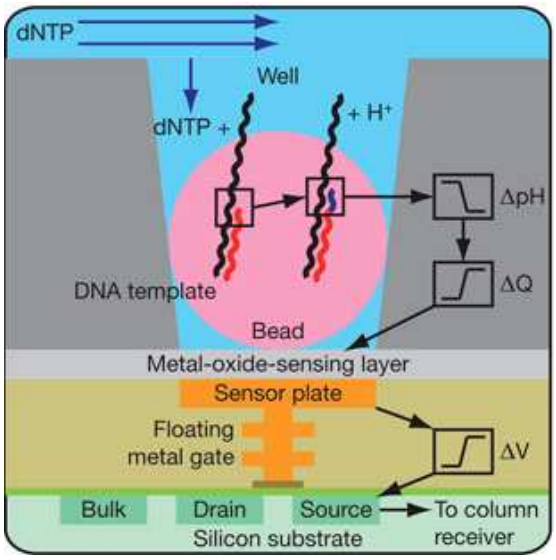


Figure 11 : Le principe du séquençage selon la technologie Ion Torrent

(D'après Rothberg et al., 2011 [158])

1.2.4.5 Alignement et interprétation des variants

C'est un processus qui aligne tous les fragments d'ADN après le séquençage par rapport un génome de référence. En règle générale, tous les programmes d'alignement de courte lecture utilisent une procédure en deux étapes pour cartographier une séquence. Tout d'abord, est identifié un petit locus du génome de référence où la lecture est le plus susceptible de s'aligner. Puis, un algorithme d'alignement plus précis est utilisé pour déterminer la position exacte de la séquence lue [159].

L'analyse des variants par rapport au génome de référence doit notamment tenir compte de :

- la profondeur de séquençage : le nombre de fois que la séquence a été lue afin de diminuer les faux positif.
- l'homogénéité de couverture des régions d'intérêt
- la reproductibilité de la méthode
- le nombre d'échantillons à traiter

Cette analyse utilisera des filtres selon les stratégies de recherche, notamment les bases de données de variants connus et rapportés dans une population témoin (dbSNP (Base de données Single Nucleotide Polymorphism, http://www.ncbi.nlm.nih.gov/projects/SNP/); 1000 génomes (http://www.1000genomes.org/); bases de données privées). L'interprétation sera complétée par des approches bioinformatiques utilisant des bases de données permettant de prioriser des gènes candidats selon différents critères, l'affection étudiée, le profil d'expression tissulaire du gène, la voie de signalisation cellulaire, éléments importants à prendre en compte pour valider l'implication d'un gène dans une maladie.

1.2.5 La recherche de gène candidat par l'analyse des données du Séquençage haut débit dans les maladies génétiques

La mise en place progressive des technologies de séquençage de nouvelle génération a révolutionné l'approche de recherche de gènes candidats dans les maladies génétiques.
L'exome, représentant 1% de l'ensemble du génome humain, a été la cible d'analyse des techniques de séquençage à haut débit. Cette approche a été privilégiée par rapport à celle du séquençage du génome complet de par l'expérience acquise jusqu'alors sur les mutations à l'origine des maladies génétiques, mutations essentiellement localisées dans les parties codantes des gènes. La technique NGS a été utilisée dans des affections génétiques sporadiques ou héritées, selon différents modes, dominant ou récessif.

1.2.5.1 Séquençage de l'exome dans la déficience intellectuelle syndromique, sporadique

Si l'on considère une affection syndromique sporadique, et dans le cas présent responsable de DI, l'expertise clinique permet de sélectionner une population cliniquement « homogène » avec l'hypothèse d'une mutation survenue *de novo* dans un gène candidat. Une hétérogénéité génétique de cette affection demeure bien évidemment possible. Cette approche consiste à séquencer plusieurs patients présentant le même phénotype, chaque patient devant, a priori, partagé un variant causal dans le même gène.

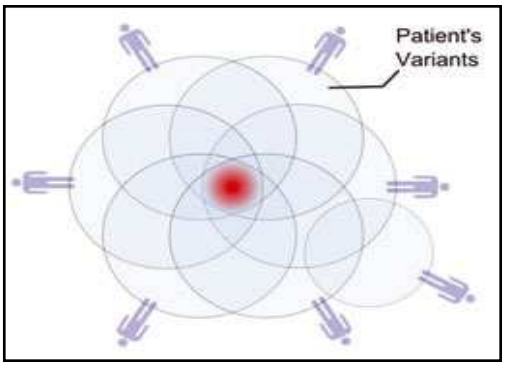


Figure 12 : La stratégie pour des patients atteints d'une DI syndromique sporadique

Ils sont groupés par rapport aux signes cliniques communs. Le gène responsable est muté chez la plupart d'entre eux.

(D'après Topper et al., 2011 [160])

Cette stratégie a été utilisée pour identifier des mutations dans le gène *MLL2* impliqué dans le **syndrome de Kabuki** (déficience intellectuelle, dysmorphie faciale, et cardiopathie) [161] et des mutations *de novo* dans le gène *SETBP1* lors du **syndrome de Schinzel-Giedion** (déficience intellectuelle sévère, dysmorphie faciale, et de multiples anomalies congénitales) [162].

L'analyse de trios (l'exome du patient et celui de ses parents) est une approche particulièrement utilisée dans l'interprétation des grands volumes de données du séquençage (Figure 13). Cette stratégie est basée sur l'hypothèse qu'une mutation *de novo* a une probabilité élevée d'être causale dans la survenue de la maladie. Toutefois, l'accumulation régulièrement progressive des données de séquençage, ont remis en question cette vision simplifiée et la présence d'une mutation *de novo*, même une mutation non sens, peut être observée dans une population témoin normale. Pour valider l'hypothèse de gène candidat, une confirmation sur un grand nombre de patients partageant le même syndrome doit être réalisée.

Ng et al., [161] ont concentré leur analyse initiale sur l'exome de 10 individus atteints du syndrome de Kabuki**.** Ils ont d'abord filtré les variants observés grâce à dbSNP, le projet 1000 génomes, et des bases de données internes. Un seul gène *MUC16*, a été observé comme muté chez les 10 patients. Toutefois, les auteurs ont fait valoir qu'il s'agissait d'une fausse piste, car *MUC16* est l'un des grands gènes (190 kb) du génome humain, et par conséquent, avec une probabilité plus élevée de variations eu égard à sa taille. Dans une nouvelle analyse, ils ont donné plus de poids aux mutations troncantes (mutations non-sens et les mutations frameshift). Cela a conduit à l'identification de mutations dans le gène *MLL2* chez 7 des 10

patients, permettant ainsi de le reconnaître gène candidat responsable pour le syndrome de Kabuki.

S'agissant du syndrome de Schinzel-Giedion, Hoischen et al., [162] ont séquencé l'exome de quatre patients. Ils ont concentré leurs analyses sur les variations de structure nucléotidiques (SNV Single Nucleotide Variant) non synonymes. Pour les 4 patients, seuls deux gènes ont été observés avec des changements faux-sens différents. Un des deux gènes était localisé dans une région présentant une forte homologie avec d'autres loci génomiques, faisant suggérer des artefacts en rapport avec un alignement incorrecte des séquences. Le gène candidat restant était *SETBP1*, sur le chromosome 18q12.3. Tous les quatre patients ont eu des mutations faux-sens de *SETBP1*, et analyse de la séquence ciblée chez les parents a indiqué que ces variants étaient en effet des mutations *de novo*.

1.2.5.2 Séquençage de l'exome dans la déficience intellectuelle non-syndromique, sporadique

L'identification de gènes candidats pour les cas de DI cliniquement isolée, non syndromique est plus difficile dans le sens où il n'existe pas de moyens cliniques pour distinguer les patients les uns des autres hormis la DI. Les populations étudiées sont donc très vraisemblablement hétérogènes et la stratégie de validation de gène candidat (récurrence de mutation dans le même gène chez d'autres patients) se heurtera à cette hétérogénéité. Qui plus est, le caractère sporadique ne permet pas d'utiliser les données familiales pour conforter la ségrégation de la mutation avec le phénotype. Les réserves quant au caractère *de novo* de la mutation et son implication dans la survenue de la DI sont les mêmes que celles notées précédemment [163-164].

Vissers et al., [165] ont appliqué la stratégie d'analyse de l'exome complet des trios pour 10 patients ayant une **déficience intellectuelle non syndromique, sporadique**. Deux patients avaient chacun une mutation *de novo* au sein d'un gène préalablement identifié comme étant responsable d'une DI non syndromique. Pour 4 autres patients, un variant *de novo* a été retenu dans un potentiel gène candidat ayant un rôle dans le développement cérébral.

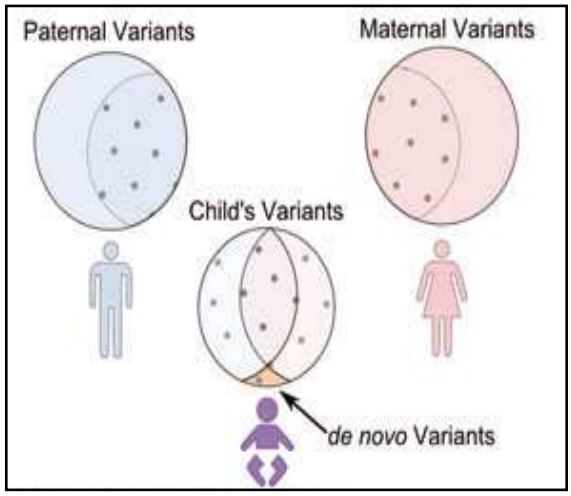


Figure 13 : La stratégie pour chercher le gène candidat dans les cas de DI sporadique non syndromique

(D'après Topper et al., 2011 [160])

1.2.5.3 Séquençage de l'exome pour une DI survenant dans des familles avec union consanguine

La DI peut répondre aussi à un mode d'hérédité autosomique récessif, et dans ce contexte la description de familles avec union consanguine contribue à l'identification de nouveaux gènes candidats. Les études préalables de cartographie d'homozygotie et déséquilibre de liaison permettent de réduire la région critique au sein de laquelle le gène candidat est localisé. (Figure 14)

Récemment, Najmabadi et al., [166] ont recherché le gène candidat de DI autosomique récessive chez 136 familles consanguines. Ils ont séquencé des exons dans les régions cibles pour chaque famille, au lieu de séquencer l'exome complet. Ils ont trouvé des mutations pour 26 familles (23 gènes), gènes rapportés à la déficience intellectuelle, en fournissant un diagnostic définitif.

Pour les familles restantes, ils ont identifié 50 nouveaux gènes candidats, chacun avec une mutation homozygote.

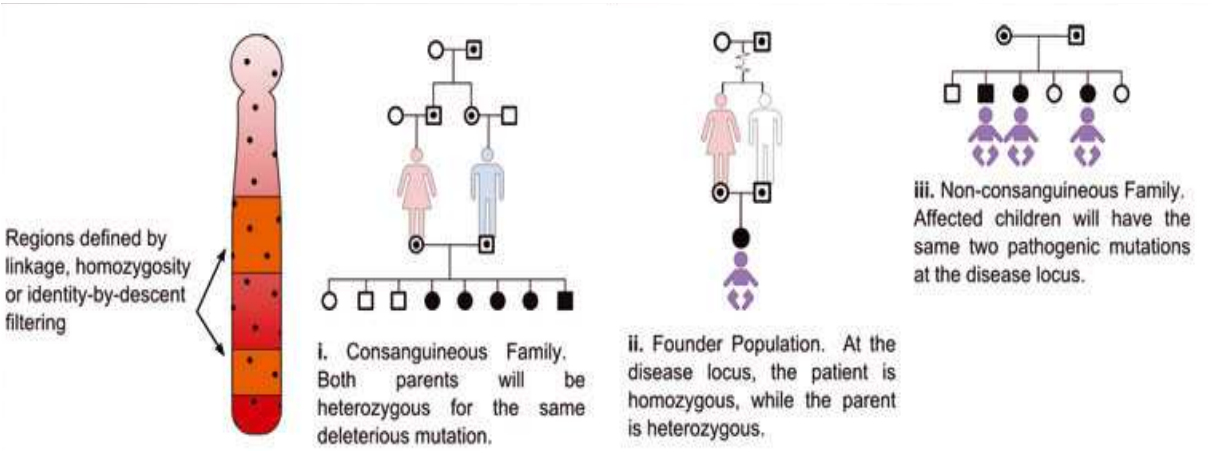


Figure 14 : La stratégie pour les familles avec DI autosomique récessive

(D'après Topper et al., 2011 [160])

1.2.5.4 Séquençage de l'exome dans les familles non-consanguines avec déficience intellectuelle

Des individus apparentés ont une tendance plus importante d'avoir des génotypes similaires que des individus non apparentés car ils ont une probabilité plus élevée d'avoir des allèles IBD ou Identiques par descendance. Krawitz et al., [167] ont étudié une famille non consanguine avec trois enfants avec un syndrome d'hyperphosphatasie et retard mental (HMR), avec une hérédité autosomique récessive. Après l'acquisition de données de séquençage d'exome pour les trois enfants avec HMR, ils ont identifié des régions qui sont compatibles avec une identité par descendance. Cela a réduit leur espace de recherche à ~ 20% d'exome. Ils ont ensuite cherché dans ces régions des gènes porteurs de variants rares dans les trois enfants atteints, ciblant sur deux gènes candidats. La fonction de l'un de ces gènes, *PIGV*, était très compatible avec l'étiologie de la maladie. L'analyse de ce gène dans trois autres familles, non apparentées, a identifié des mutations homozygotes ou hétérozygotes composites confirmant ainsi l'implication de ce gène.

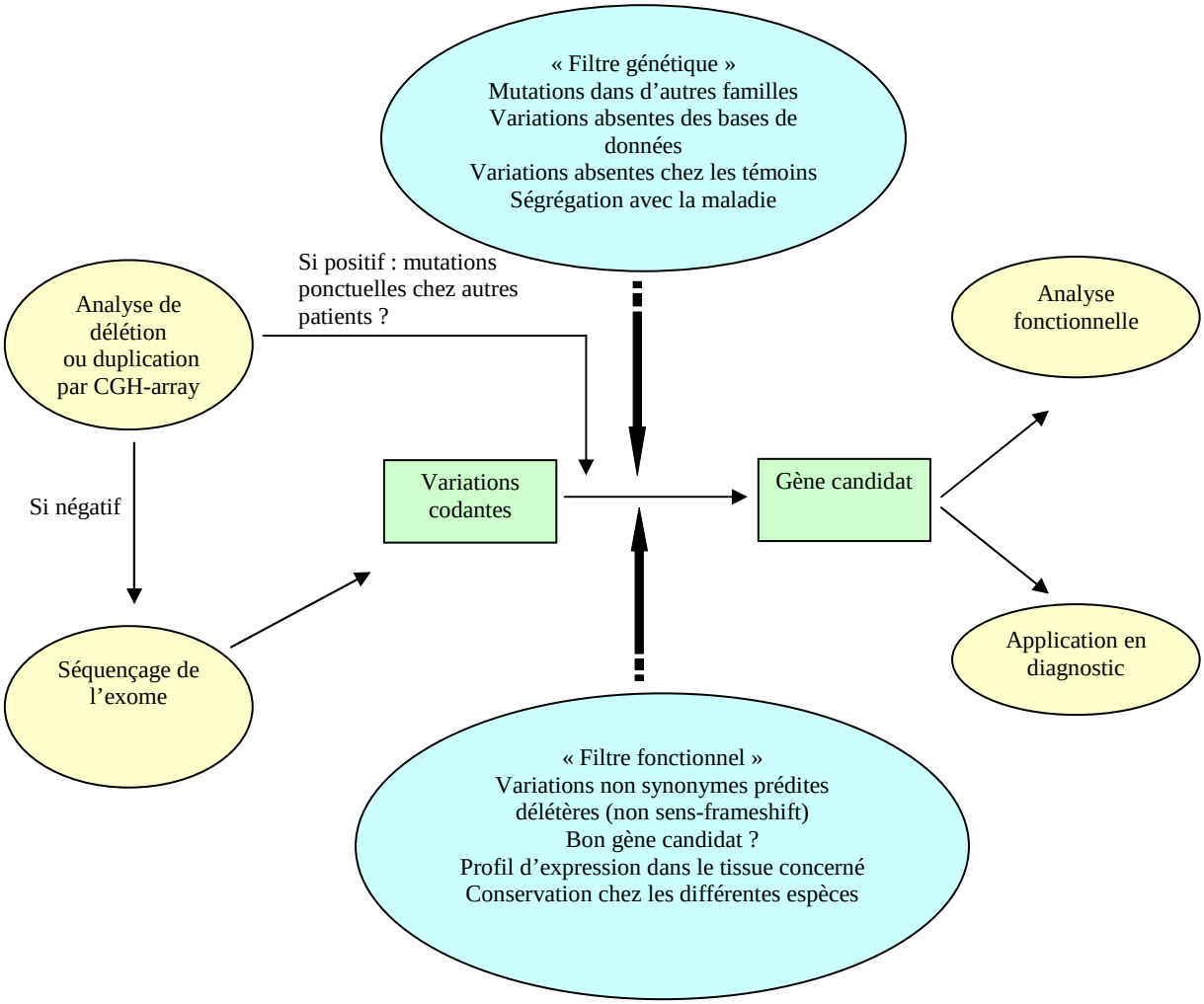


Figure 15 : Stratégies pour la recherche du gène candidat dans les maladies monogéniques à l'aide de microréseaux génomiques et le séquençage de l'exome

(D'après Kuhlenbäumer et al., 2011 [168])

1.2.5.5 Avantages et limites du séquençage de l'exome

Les capacités actuelles offertes de séquençage d'exome conduisent à proposer le plus souvent d'emblée une analyse sur l'exome complet sans se limiter à l'exploration d'une région définie. Des gènes dont la fonction s'exerce dans des voies biologiques ou processus cellulaires inattendus pour le phénotype peuvent être découverts par cette approche.

Le nombre d'échantillons requis pour une étude est plus réduit, souvent quelques individus.

Malgré le grand potentiel du séquençage d'exome pour identifier les mutations responsables de maladies rares, il y a encore de nombreux défis. Les données d'exomes générées par les technologies actuelles sont incomplètes : la couverture partielle pour certaines régions explorées, une profondeur de lecture faible. Cela peut se produire en raison de biais dans la

capture (biais par des divergences dans le ratio de guanine à la cytosine (GC) et la structure secondaire de séquence), le séquençage et pendant le processus d'alignement (biais produit par les séquences homologues et dans les régions de séquences répétées).

2. PATIENTS ET METHODES

2.1 Patients

2.1.1 La cohorte des patients avec DI

Ce travail s'appuie sur le centre de référence maladies rares « Anomalies du développement » de l'inter région Est qui participe à l'évaluation génétique des patients ayant une DI. Dans le cadre du diagnostic étiologique de DI, une analyse sur microréseau d'ADN est réalisée chez plus 600 patients /an au sein de la plateforme de CGH-array de l'inter-région Est –laboratoire de génétique du CHU de Nancy) regroupant les services de génétique des CHU de Besançon, Dijon, Nancy, Reims et Strasbourg. Les patients bénéficient au préalable d'une expertise clinique permettant d'orienter vers une cause éventuelle de DI via les données cliniques (antécédents familiaux, signes associés de dysmorphie, d'anomalies malformatives, de troubles du comportement), neuroradiologiques (imagerie cérébrale), métaboliques et génétiques.

2.1.2 La cohorte de patientes ayant un syndrome d'Aicardi

Vingt-deux patientes ont été inclues dans cette étude grâce à l'Association AAL Syndrome d'Aicardi (www.aicardi.info) et les neuropédiatres et généticiens qui ont validé le diagnostic de syndrome d'Aicardi selon les critères actuellement reconnus [66] en tenant compte notamment des données d'imagerie cérébrale et ophtalmologiques. Les ADN parentaux sont également disponibles pour toutes ces patientes. L'âge des patientes était compris entre 6 et 47 ans (moyenne 26,5 ans et 13 ans médiane). Treize patientes présentaient la triade caractéristique (agénésie complète du corps calleux, lacunes choriorétiniennes et spasmes infantiles), 10 avaient les critères cliniques élargis. Le caryotype constitutionnel lymphocytaire était normal au seuil de résolution de 550 bandes pour chacune d'entre elles.

2.2 Matériel biologique

Le matériel génétique, acides nucléiques ADN et/ou ARN, a été extrait à partir de prélèvements sanguins. Pour chaque patient prélevé, un consentement libre et éclairé est signé par les parents pour leur enfant et aussi pour eux-mêmes.

2.3 Extraction des acides nucléiques

2.3.1 ADN

L'extraction de l'ADN génomique est réalisée, à partir d'un prélèvement sanguin à l'aide du kit Nucleon BACC3TM (Amersham, GE healthcare, GmbH) selon les instructions du fournisseur. Après précipitation par l'éthanol, l'ADN est resuspendu dans du tampon TE (Tris-HCl 10mM pH8, EDTA 1mM pH8).

2.3.2 ARN

Les prélèvements sanguins sont effectués sur tubes PAXGene^{TM} (PreanalytiX, GmbH) afin d'assurer la stabilité des ARN. Les ARN sont extraits avec le kit d'extraction PAXGene blood RNA isolation kit (PreAnalytiX, GmbH) en accord avec le protocole du fournisseur.

Le contrôle de la qualité et la quantification d'ADN et d'ARN ont été réalisés par spectrophotométrie (ND-100 Nanodrop^{TM}) nécessitant au minimum 1 µl d'ADN. Les rapports d'absorbance (λ= 260/230 et λ= 260/280) ont été étudiés. Ces ratios donnent une indication de la présence des protéines, des polyphénols et des glucides dans l'échantillon. Une migration électrophorétique sur gel d'agarose 2% des échantillons d'ADN extrait (200 ng) a été réalisée avec le marqueur de taille λ d'ADN standard.

2.4 Cytogénétique et cytogénétique moléculaire

2.4.1 Préparation des chromosomes métaphasiques

A partir d'un prélèvement sanguin, les lymphocytes T sont cultivés en présence de phytohémaglutinine pendant 72 heures à 37°C. Les cellules sont bloquées en métaphase par 45µl de colchicine (20mg/ml) pendant 45min à 37°C en présence de 200µl de bromure d'éthidium (0,7mg/ml). Après centrifugation (1200 rpm, 5min), un choc hypotonique est appliqué (KCl à 0,0075M) pendant 17min à 37°C. La préfixation est obtenue par ajout de 1ml de fixateur (3 volumes d'éthanol absolu pour 1 volume d'acide acétique). Après centrifugation (1200rpm, 5min), suivent 3 lavages dans 12ml de fixateur (1200rpm, 5min) et la remise en suspension du culot dans du fixateur. Après étalement, les lames sont traitées pour la réalisation du caryotype. Dans ce cas, elles sont vieillies 24 à 48 heures à l'abri de la lumière.

2.4.2 Caryotype en bandes GTG (bandes G, Trypsine, Giemsa)

Les lames sont vieillies une nuit à 65°C. Elles sont ensuite incubées dans une solution de trypsine à 37°C pendant 10 secondes, rincées dans 2 bains de PBS (phosphate buffered saline) et colorées dans une solution contenant 9ml de Giemsa (Biolyon, France), 9ml d'acide citrique (3M) et 9ml de méthanol (1M). Les lames sont ensuite rincées et analysées.

2.4.3 L'hybridation in situ en fluorescence (FISH)

2.4.3.1 Préparation des sondes : culture bactérienne, extraction d'ADN et marquage des sondes

Les BAC utilisés comme sondes pour cartographier les duplications (Xp22.2 et Xp21.3) ont été sélectionnés à partir du site UCSC genome browser (http://genome.ucsc.edu/) et fournis par le Wellcome Trust Sanger Institute. (http://www.sanger.ac.uk).

Après culture bactérienne d'une nuit en milieu liquide LB Broth Miller à 37°C en présence de chloramphénicol (20µg/ml) et la centrifugation (20min à 40000 rpm), les sondes BAC sont obtenues à partir des culots bactériens par extraction de l'ADN du BAC avec le kit NucleoBond PC 100^{TM}(Macherey Nagel, GmbH) selon les recommandations du fabricant.

Le marquage par incorporation de fluorophore (fluorescéine isothyocyanate ou rhodamine ou Aqua) est réalisé par nick translation à l'aide du kit nick translation^{TM}(Vysis Abbot Laboratories, USA) selon les recommandations du fournisseur.

2.4.3.2 Hybridation et lavages

Les cellules obtenues après culture de lymphocytes T sont étalées sur une lame et ensuite prétraitées par une solution de HCl à 0,01M, 0,01% de pepsine à 37°C, lavées 2 fois 5min dans du PBS 1X et fixées par une solution de formaldéhyde à 3% pendant 2min à température ambiante et déshydratées dans 3 bains d'éthanol de concentration croissante (70, 85 et 100%). Après dénaturation 3min à 72°C et hybridation une nuit à 37°C avec la sonde marquée, les lames sont lavées 2min dans une solution de SSC (saline sodium citrate) 0,4X/0,3% NP40 à 72°C puis 15 secondes dans une solution de SSC 2X/0,1%NP40 à température ambiante.
Les cellules sont contre-colorées au DAPI. Les signaux sont visualisés par microscopie à épifluorescence (Axioplan 2 imaging, Zeiss).

2.4.4 Inactivation du chromosome X

L'étude du profil de l'inactivation du chromosome X chez les mères des patients avec les remaniements du gène *IL1RAPL1* a été réalisée au locus HUMARA (Human Androgen Receptor A) en Xq13 comme décrit précédemment [169]. Après la digestion par l'enzyme *HpaII* les échantillons d'ADN sont amplifiés par les amorces spécifiques sur le locus HUMARA. Les produits fluorescents de PCR ont été analysés sur un Genetic Analyser 3130 (Applied Biosystems, USA), et des pics de chaque produit de PCR ont été obtenus par le logiciel Gene Mapper v4.0 (Applied Biosystems, USA). Inactivation de l'X a été considéré comme biaisé si le ratio était supérieur à 80:20.

2.4.5 L'hybridation génomique comparative sur microréseau d'ADN

L'étude par hybridation comparative génomique a été réalisée sur des microréseaux (ou puces) d'oligonucléotides, fournis par Agilent Human Genome CGH microarray^{TM} (Agilent Technologies, USA), d'une densité variable, pangénomique 105K, 244K, 1M, ou avec une couverture spécifique du chromosome X de 1M. Les oligonucléotides sont synthétisés *in situ* grâce à la méthode *Agilent's 60-mer SurePrint technology*. Chaque puce a fait l'objet d'un contrôle qualité avant commercialisation. ***Les lames 244K*** comportent 236 381 oligomères

représentant à la fois des séquences codantes et non codantes avec une résolution théorique moyenne de 8,9 kb (7,4 kb dans les régions géniques). L'annotation des oligonucléotides est basée sur le version de génome humaine génome [UCSC hg18 -build NCBI 36- Mars 2006]. Pour ***la puce conçue à façon de 1M pour le chromosome X***, elle contient 974 016 oligonucléotides au total, avec 161 714 oligonucléotides pangénomiques et 789 109 oligonucléotides sur le chromosome X. Ce microréseau couvre tous les gènes connus sur le chromosome X dont ceux codant les microARN. La distance moyenne entre les oligonucléotides est de 204 pb avec couverture de la séquence codante et non codante du chromosome X [UCSC hg19 (build NCBI 37), Février 2009].

2.4.5.1 Protocole

Trois microgrammes d'ADN de la patiente ont été hybridés contre l'ADN d'un témoin de même sexe sur une lame Human Genome CGH microarray™ (Agilent Technologies). La qualité d'ADN pour chaque échantillon a été contrôlée comme expliqué précédemment.

L'hybridation est mise en œuvre selon les recommandations du fabricant. Tout d'abord, les ADN témoins et patientes sont digérés par les enzymes AluI (10U/µl) et RsaI (10U/µl) (Promega, USA) pendant 2 heures à 37°C, puis 20 minutes à 65°C pour inactiver les enzymes. Chaque tube contient : 2 µl eau sans nucléase; 2,6 µl 10X tampon C ; 0,2 µl BSA acétylé (10 µg/ µl) ; et 5 unités de chaque enzyme AluI et RsaI. Le contrôle de la digestion est réalisé par électrophorèse sur gel d'agarose à 2% avec un marqueur de taille 100pb (observation d'une traînée entre 200 et 500pb). Les ADN sont marqués par la technique de random priming pendant 2 heures à 37°C au bain marie (Enzo Life Sciences USA). Pour chaque marquage, nous avons utilisé 5 µl de random primers, 3 µl de cyanines 5 et 3 (à 1mM), 10 µl de tampon 5X, 5 µl de dNTP 10X et 1 µl de fragment exo-kleenow. L'ADN de la patiente est marqué par la cyanine 5-dCTP (deoxcytidine triphosphate) et celui du témoin par la cyanine 3-dCTP. Les produits de marquage ont ensuite été purifiés sur des colonnes de filtration Microcon YM30 (Millipore, USA). Le contrôle du taux d'incorporation des cyanines par mesure de l'activité spécifique (1000X concentration d'ADN (ng/µl)/324,5 X concentration en cyanine (pmol/µl)) et le dosage des ADN est réalisé par spectrophotométrie (ND-100 Nanodrop™). L'activité spécifique doit être inférieure à 45 pour une hybridation ultérieure de qualité.

La préhybridation est réalisée à l'aide du kit de pré-hybridation fourni par Agilent. L'ADN génomique marqué et purifié est ainsi mélangé à 50 µl d'ADN Cot-1, 52 µl d'agent de bloquage 10X et 260 µl de tampon d'hybridation 2X, dénaturé 3 minutes à 95°C et placé au

bain marie à 37°C pendant 30 minutes. L'hybridation a lieu à 65°C avec une agitation réglée à 20 rpm pendant 40 heures suivie de lavages à l'aide des solutions fournies par Agilent. Toutes les étapes de marquage, purification, hybridation et lavage des lames sont effectués dans un caisson antiozone. Après avoir analysé les lames à l'aide du scanner Agilent, les données de fluorescence sont ensuite analysées avec le logiciel Feature Extraction^{TM}v9.1. La visualisation des déséquilibres génomiques utilise le logiciel CGH analytics^{TM}v3.4 (Agilent, USA). L'algorithme d'analyse utilisé est ADM2, avec un seuil fixé à 6 et un minimum de 3 sondes consécutives déviantes pour considérer une déviation comme significative.

2.5 PCR quantitative sur ADN génomique ou sur cDNA

2.5.1 Le principe

Les variations du nombre de copie génomique, détectées par hybridation comparative sur microréseaux, sont vérifiées par une technique alternative, PCR quantitative en temps réel (qPCR). La technique consiste en une PCR classique en présence d'un agent fluorescent le plus souvent le SYBR® green™ (Applied Biosystem, USA) qui s'intercale exclusivement dans l'ADN double brin. La fluorescence ainsi détectée à chaque cycle de PCR est proportionnelle à la quantité d'ADN amplifié. Le principe de la PCR en temps réel repose sur la possibilité de suivre la quantité d'ADN présente dans la réaction. Le cycle seuil (ou threshold cycle Ct), est le cycle d'amplification à partir duquel la fluorescence est significativement différente de la fluorescence liée au bruit de fond. Les couples d'amorces d'amplification, spécifiques de chaque région génomique en déséquilibre relevée par CGH-array, ont été sélectionnés grâce au logiciel Primer-Express V2.0 (Applied Biosystem, USA) afin d'obtenir les meilleurs designs pour une efficacité et spécificité d'amplification au locus étudié. L'absence de polymorphisme de type SNP dans les amorces est également contrôlée sur le site SNP Check3 NGRL (National Genetics Reference Laboratory, https://ngrl.manchester.ac.uk/SNPCheckV3/snpcheck.htm). Nous avons choisi un couple d'amorces (synthèse réalisée par Eurogentec Inc. S.A) localisé en amont de la zone déviante et un couple d'amorce localisé en aval de la zone déviante.

La méthode nécessite l'établissement d'une droite standard par couple d'amorces à étudier avec un ADN témoin, permettant de contrôler l'efficacité de PCR pour chaque couple d'amorces et de déterminer la gamme de mesure permettant de quantifier tous les échantillons pour un couple d'amorce donné. La méthode de détermination des ΔΔCt est utilisée pour la quantification de la région cible, exprimée par rapport à un gène de référence sur le même échantillon d'ADN. La quantité de région cible et celle de gène de référence est fonction de l'efficacité de la PCR et du Ct enregistré.

2.5.2 Le protocole

La réaction de PCR est réalisée à l'aide du réactif *power* SYBR Master Mix™(Applied Biosystems, USA) sur un appareil de PCR en temps réel 7500 standard Real-Time PCR

System (Applied Biosystem, USA). Une courbe d'efficacité est réalisée pour chaque couple d'amorces afin d'évaluer sa spécificité vis-à-vis du génome humain. Pour la réalisation de cette courbe d'efficacité, des dilutions séquentielles en duplicat d'un ADN témoin sont utilisées de 100 à 0,1 ng/µl, les amorces sens et antisens sont diluées à 10 µM. L'amplification est réalisée dans les conditions suivantes : dans un volume final de 20 µl, le mélange réactionnel contient 0,4uM d'amorces sens et anti-sens, 10 µl de 2X *power* SYBR Master Mix^{TM}(Applied Biosystems, USA) et 5 µl de solution d'ADN correspondant à des dilutions successives de la cible de 100, 50, 10, 1 et 0,1ng pour réaliser une courbe d'étalonnage. La PCR, réalisée sur ABI PRISM 7500 Sequence Detection System (SDS), comprend un cycle de dénaturation à 95°C pendant 10min, suivi de 40 cycles à 95°C pendant 15sec et 60°C pendant 1min. La pente de la droite d'étalonnage traduit l'efficacité de la PCR. (entre -2.8 à -3.2). La spécificité des amorces et l'absence de dimères d'amorces sont vérifiées par une étape de dissociation mise en œuvre à la fin de l'amplification et qui consiste en une dénaturation (15 secondes à 95°C) suivie d'un refroidissement rapide à 60°C et d'une ascension progressive et lente en température jusqu'à 95°C. La dérivée de la courbe de fusion (quantité de fluorescence en fonction de la température) montre la présence d'un pic unique pour chaque amplification si celle-ci est spécifique. Puis les plaques tests sont réalisées avec tous les couples d'amorces. L'amplification est testée en triplicat pour chaque échantillon d'ADN dilué à 2ng/µl. Le mix de qPCR contient le «*power* SYBR Master Mix^{TM}», chaque couple d'amorces, et l'ADN (patiente/ témoin/parentaux). Le gène de référence *RPPH1* (ribonuclease P RNA component H1) est utilisé pour normaliser les résultats. Les résultats de la réaction de qPCR sont analysés grâce au logiciel ABI Prism 7500 SDS version 1.4^{TM}(Applied Biosystems, USA).

2.5.3 Transcription inverse pour obtenir le cDNA

La transcriptase inverse a été utilisée pour obtenir le cDNA exploitable en qPCR.

La réaction de transcription inverse est réalisée à partir de 1µg d'ARN totaux dans un volume de 20µl. Le mélange réactionnel contient, 4µl de tampon de la Superscript II, 0,01M de dithiotreitol (DTT), 100ng/µl d'hexanucléotides, 1 U/µl d'inhibiteur de RNase, 0,5mM de dNTP et 10 U/µl de Superscript II reverse transcriptase (Invitrogen, USA). Elle est réalisée selon le protocole fourni par le fabricant.

2.6 Le séquençage d'ADN selon la méthode Sanger

Pour analyser la séquence d'ADN génomique à la recherche de variations nucléotidiques, les amorces sont choisies grâce au site de Primer 3 v. 0.4.0 (http://frodo.wi.mit.edu/). Les amorces introniques localisées en bordure des exons ont été choisies à l'aide du logiciel ExonPrimer2. L'amplification d'une séquence ciblée du génome a été vérifiée à l'aide de la base de données Genome Browser UCSC. L'absence de polymorphisme de type SNP dans les amorces est également contrôlée sur le site SNP Check3 NGRL (National Genetics Reference Laboratory). Les amorces sont sélectionnées avec une température de demi-dissociation (Tm) comprises entre 55°C et 65°C et une taille entre 20 pb à 25 pb. Le séquençage est effectué sur un appareil ABI PRISM 3100 (Applied Biosystems, USA). La mise au point pour chaque couple d'amorces se fait donc en se bornant aux seuils de température entre 55°C et 65°C. Les autres paramètres sur lesquels nous avons agi sont les concentrations de $MgCl_2$ (0.8 mM à 2.0 mM) et de DMSO (5 à 10%).

Afin d'amplifier la séquence cible, la composition de chaque mélange réactionnel contient 100ng d'ADN génomique totaux, pour un volume final de 25 µl avec 2,5 mM de Tampon ; 0,2mM dNTP's ; 0,4µM d'amorces sens et antisens ; $MgCl_2$ (0.8 mM à 2.0 mM) ; 0,08U de Taq Thermoprime^{TM} ou Gold (selon la spécificité d'amplicon) et qsp H_2O. Le contrôle de la qualité, la taille, et de la spécificité des amplicons obtenus a été réalisé par dépôt de 5µl de la réaction de PCR sur un gel d'agarose 2 %. Les produits de PCR sont purifiés par l'ExoSAP (GE healthcare, GmbH) afin d'éliminer les dNTP non incorporés au cours de l'amplification et l'excès d'amorces (15min à 37°C suivi d'une dénaturation enzymatique de 15min à 80°C). Le mélange enzymatique ExoSAP-IT contient deux enzymes, une exonucléase I et une phosphatase alcaline (SAP). La première enzyme dégrade les amorces résiduelles et la phosphatase hydrolyse les dNTPs restants du mélange de PCR qui interféreraient avec la réaction de séquençage. Nous avons appliqué la technique de séquençage "dye terminator sequencing" décrite par Smith et al., en 1986 [170] qui utilise des didéoxyribonucléotides dont chacun est marqué par un fluorophore spécifique. Les fragments d'ADN synthétisés portent ce fluorophore terminal. On les appelle des terminateurs d'élongation ou "BigDye Terminators" ou "Dye-labeled terminator". Il n'y a qu'une seule réaction de séquençage en présence des 4 didésoxyribonucléotides. Le séquençage des produits de PCR a été réalisé en utilisant le kit BigDye Terminator v3.1 (Applied Biosystems, USA). Le mélange réactionnel (5ml) contient 2,5µl produits PCR purifiés par exosap, 1µl d'amorces 5mM sens ou antisens,

0,5µl de tampon 5X, 1µl BigDye. Trente cycles de PCR ont été réalisés avec une dénaturation à 95°C pendant 10sec, l'hybridation à 50°C pendant 5sec et l'élongation à 60°C pendant 4min. Les produits de PCR sont purifiés sur une résine Sephadex ® G-50 superfin (GE healthcare, GmbH).

Cette étape a pour but de purifier par gel-filtration les produits de réaction de séquence avant injection sur séquenceur capillaire. La résine Sephadex® G50 utilisée permet de désaler les échantillons et d'éliminer les nucléotides non incorporés et les amorces de PCR en excès. La résine été déposée dans chaque puits (45µl) du chargeur de colonne. 300 µl d'eau ultra pure ont été ajoutés dans chaque puits contenant de la résine. Les produits de séquençage à purifier sont dilués dans un volume d'eau de 20 à 35 µl ultra pure (par rapport à qualite de l'amplicon à sequencer) et déposés délicatement dans les puits. La plaque MultiScreen est centrifugée 4 min à 910 g. Les filtrats contiennent les produits purifiés dans l'eau. La séparation par électrophorèse capillaire et la détection des pics de fluorescence ont été réalisées par l'automate ABI PRISM 3130 analyser (Applied Biosystems, USA). Les profils de toutes les séquences ont été étudiés grâce au logiciel sequence scanner v 1.0 (Applied Biosystems). L'analyse des variations de séquences est réalisée à l'aide du logiciel Alamut 1.5^{TM}(Interactive Biosoftware, France) qui compile les informations de différentes bases de données bioinformatiques.

2.7 Séquençage à haut débit

La qualité de chaque échantillon ADN a été contrôlée par nanodrop ND-100 Nanodrop^{TM} avec un ratio A260/A280 entre 1,8 à 2,0.

La laboratoire de génétique de l'EA 4368 ne possédant l'équipement et l'expérience nécessaire pour le séquençage haut débit a confié les ADN à deux équipes différentes, dans un premier temps celle du Centre national de séquençage-GENOSCOPE (Evry, France) pour le séquençage de l'exome du chromosome X puis à celle de la société INTEGRAGEN (France) pour le séquençage de l'exome complet.

2.7.1 Le séquençage de l'exome du chromosome X

Le séquençage de l'exome du chromosome X a été effectué en utilisant la plateforme « Roche 454 Genome Sequencer FLX » par Centre national de séquençage-GENOSCOPE (Evry, France). Le détail de la technique a été décrit dans l'introduction (partie 1.2.4.2, page 55) 454 Genome Sequencer FLX instrument (Roche Applied Science) »

2.7.2 Le séquençage de l'exome complet

La société INTEGRAGEN (France) a fait le séquençage de l'exome complet sur la plateforme Illumina HiSeq2000. Le principe de cette technique a été détaillé dans l'introduction (partie 1.2.4.3, page 56) Illumina /Solexa Genome Analyzer.

Analyse bio-informatique

- L'analyse bioinformatique des données de séquençage est basée sur le pipeline fourni par Illumina (CASAVA1.8). CASAVA1.8 est une suite de scripts comprenant l'alignement des séquences sur le génome complet (build37), un comptage allèlique et une détection des variants (SNPs et Indels). L'algorithme d'alignement utilisé est ELANDv2e (alignement réduisant les mésappariements artéfactuels).
- L'annotation des variations génomiques est réalisée en tenant compte des informations issues de l'annotation des gènes (RefSeq), des polymorphismes référencés (dbsnp132, 1000Genomes) suivie d'une caractérisation de la mutation (exonique, intronique, non-

sens, faux-sens, synonyme). Pour chacune des positions, sont également déterminées les fréquences exomiques (Homozygote et Hétérozygote) en considérent les expériences antérieures de tous les exomes séquencés par la société INTEGRAGEN Les résultats sont rendus par échantillon, sous forme de fichiers textes tabulés.

3. RESULTATS

Nous présentons nos résultats en 3 parties. Les deux premières parties relatent notre expérience sur la diversité de microremaniements intragéniques, à partir de la cohorte de patients recrutés au sein du centre de référence maladies rares de l'inter région Est et concernent les gènes *ILRAPL1* et *MBD5*. La dernière partie rapporte les résultats de notre stratégie de recherche de gènes candidats du syndrome d'Aicardi.

3.1 Réarrangements intragéniques du gène IL1RAPL1 et déficience intellectuelle

3.1.1 Données cliniques des patients

Le **patient 1** est un garçon de 18 ans, né à la suite d'une grossesse non incidentée d'un couple non apparenté, sans antécédents particuliers. Les mensurations à la naissance étaient normales. A un mois, il est opéré d'une hernie inguinale droite. A six mois, une hypotonie globale est clairement notée, avec un mauvais contact visuel sans anomalie à l'examen ophtalmologique. Le développement psychomoteur est retardé. Il a acquis la position assise à 13 mois et la marche à 3 ans. Il ne faisait aucune phrase jusqu'à l'âge de 4 ans. Sa déficience intellectuelle a été estimée comme légère à modérée. La compréhension était relativement bien conservée. Une scoliose sévère a nécessité une intervention chirurgicale à 14 ans. A l'âge de 17 ans, ses mensurations étaient en dessous du 5^{ième} percentile (poids 45 kg, taille 146 cm). L'examen clinique révélait par ailleurs un visage long et triangulaire, une lèvre inférieure épaisse et un léger ptosis. L'IRM cérébrale effectuée à 1 an était sans particularités. Le caryotype sanguin montrait une formule masculine normale au seuil de résolution de 500 bandes. La recherche de maladie X fragile (gène *FMR1*) a été négative.

Le **patient 2** est un garçon de 4 ans né d'un couple non apparenté. La naissance a lieu à 38 semaines de grossesse avec des mensurations néonatales normales. La marche est acquise à 19 mois, il existe des difficultés d'apprentissage dans les gestes fins. A l'âge de 2 ans et 2 mois, au niveau du langage, seuls quelques mots sont prononcés avec un manque d'association. Le poids était de 13,6 kg (50ème percentile), la taille 90 cm (90ème percentile) et

le périmètre crânien de 48,5 cm (20ème percentile). Il est noté un front haut, un léger ptosis, un bout de nez arrondi, et au niveau des extrémités une clinodactylie des 5ème doigts et une syndactylie 2/3 des orteils bilatérale. A 3 ans et 9 mois, le décalage dans l'acquisition du langage est rattrapé. Son intelligence n'a pas été formellement évaluée, c'est un enfant calme et affectueux et les difficultés actuelles s'inscrivent essentiellement au niveau de l'attention et de concentration. Il existe un retard d'acquisition de la propreté. L'IRM cérébrale a révélé une agénésie complète du corps calleux. La recherche d'X fragile (gène *FMR1*) est négative.

Le **patient 3** est un garçon de 9 ans né de parents non apparentés d'origine turque, sans antécédent familial particulier. La grossesse est menée à son terme sans incident. Il est eutrophique à la naissance. A 5 mois, un strabisme est noté et corrigé par le port de lunettes. L'hypotonie et le retard psychomoteur sont relevés vers 1 an. La position assise est acquise à 9 mois, la marche à 2 ans et les premiers mots rares, mal prononcés, apparaissent au delà de 3 ans. Il a présenté une crise convulsive dans un contexte fébrile à 4 mois puis à 3 ans mais dans un contexte apyrétique. A 9 ans, il est orienté vers une classe d'apprentissage et de soutien spécialisée. Des progrès sont observés dans le domaine du langage et de l'écriture avec toutefois des troubles de la compréhension concomitant à des troubles importants de l'attention. A cet âge, la taille est de 140 cm (+2 DS), le poids 28 kg (+0,5DS), et le périmètre crânien 49,5 cm (-2,5 DS). Il existe un retard global du développement considéré comme modéré. L'examen clinique note aussi une rétraction bitemporale, un visage long, un grand nez, de longues fentes palpébrales, un palais étroit, un menton pointu, une clinodactylie des 5ème orteils. L'EEG est normal et l'IRM cérébrale révèle un discret retard de myélinisation.

3.1.2 Résultats de CGH-arrray

Une anomalie intragénique d'*IL1RAPL1* (Transcript: IL1RAPL1-001 ENST00000378993) a été identifiée chez les trois patients (Figure 16).

Pour le **patient 1**, une duplication intragénique d'une taille minimale de 65 291 pb comprenant les exons 7 à 11 : *arr Xp21.3(29,818,162-29,883,453)x2*

Pour le **patient 2**, trois régions dupliquées sont détectées : *arr Xp22.2(16,228,874-16,581,022)x2,Xp21.3(28,684,119-28,779,104)x2, Xp21.3(28,951,913-29,246,487)x2.* La première duplication localisée en Xp22.2 a une taille minimale de 352 148 pb et comprend une partie du gène *CTPS2* (Cytidine triphosphate synthase II). Les deux autres en Xp21.3 se

situent dans le gène *ILRAPL1*, l'une de 94 985 pb comprenant l'exon 2 et l'autre de 294 574 pb incluant l'exon 3.

Pour le **patient 3**, une délétion intragénique de 67 695pb comprenant l'exon 3 : *arr Xp21.3(29,166,541-29,234,236)x0*.

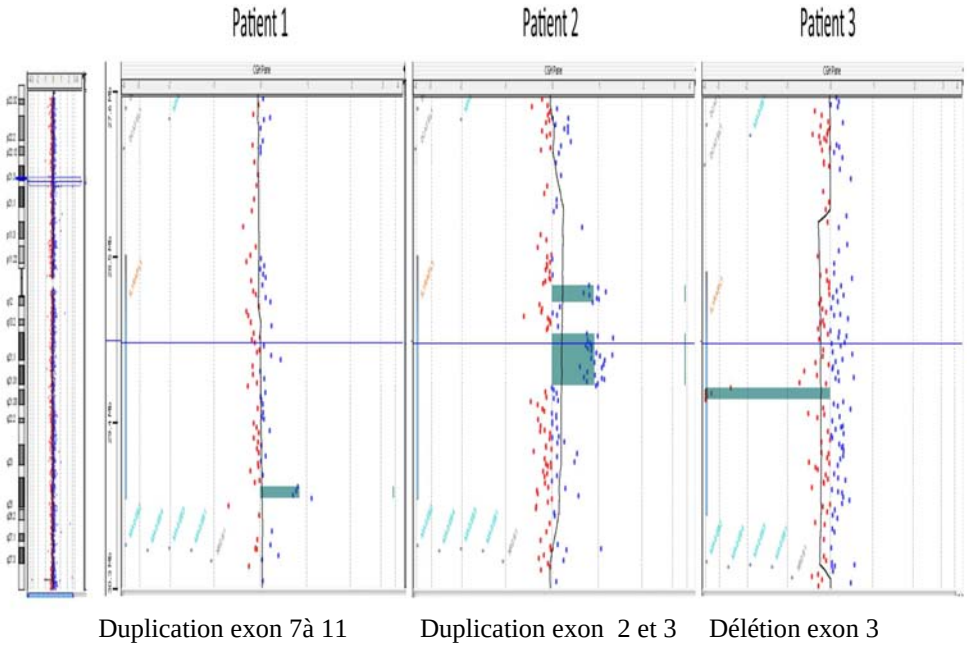


Figure 16 : Profils des remaniements du gène *IL1RAPL1* détectés par CGH-array (UCSC build 36/hg18)

3.1.3 Résultats moléculaires

qPCR

La qPCR sur l'ADN génomique a d'une part confirmé les anomalies intragéniques chez les 3 patients et d'autre part, comparée aux résultats sur les ADN parentaux respectifs, la transmission maternelle de l'anomalie dans les trois cas (Figure 17).

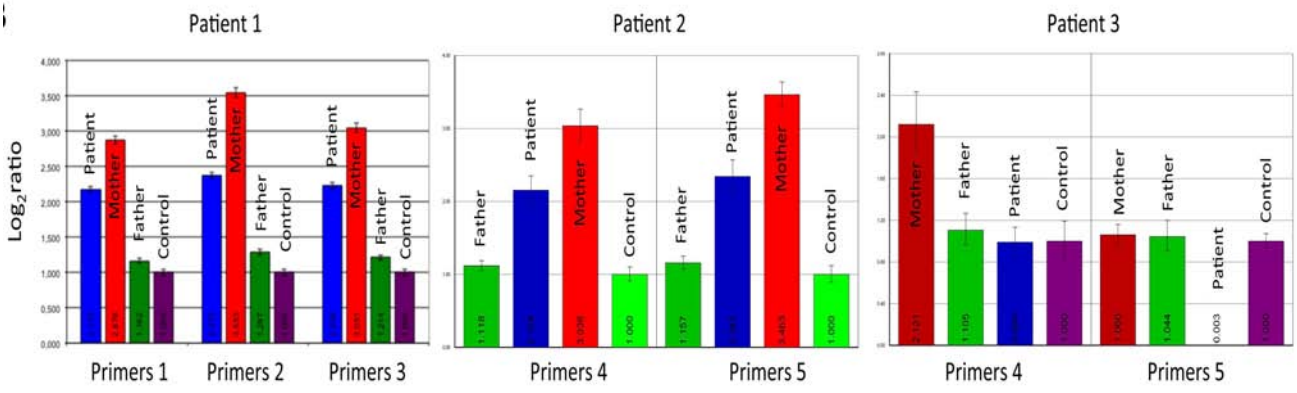


Figure 17 : Résultat de l'analyse par qPCR sur ADN génomique

Pour les trois patients (en bleu), leur père (en vert) et leur mère (en rouge) respectifs, avec l'ADN témoin (en mauve)

Inactivation du chromosome X

Le profil d'inactivation du chromosome X chez les mères de ces trois patients est analysé en comparaison avec un témoin. Le ratio a été déterminé sur le locus HUMARA du chromosome X. Les mères des patients 1 (67/33%) et 2 (79/21%) présentent une inactivation aléatoire et la mère du patient 3, une inactivation biaisée (90/10%).

Analyse en FISH

Pour le patient 2, une analyse en FISH a été réalisée à l'aide de sondes BAC, une en position distale dupliquée Xp22.2 (RP11-160F21 en aqua bleu) et deux en position proximale dupliquée Xp21.3 (CTD-2010F8 rouge et RP11-262B9 en vert), et une sonde témoin en Xqter. Au seuil de résolution obtenu, outre la duplication en tandem en Xp22.2, le remaniement apparait complexe avec une duplication intragénique en tandem de la sonde proximale Xp21.3 (rouge), et une duplication, non en tandem, mais avec insertion en situation distale de la sonde proximale Xp21.3 (verte) au delà de la sonde Xp22.2 (Figure 18)

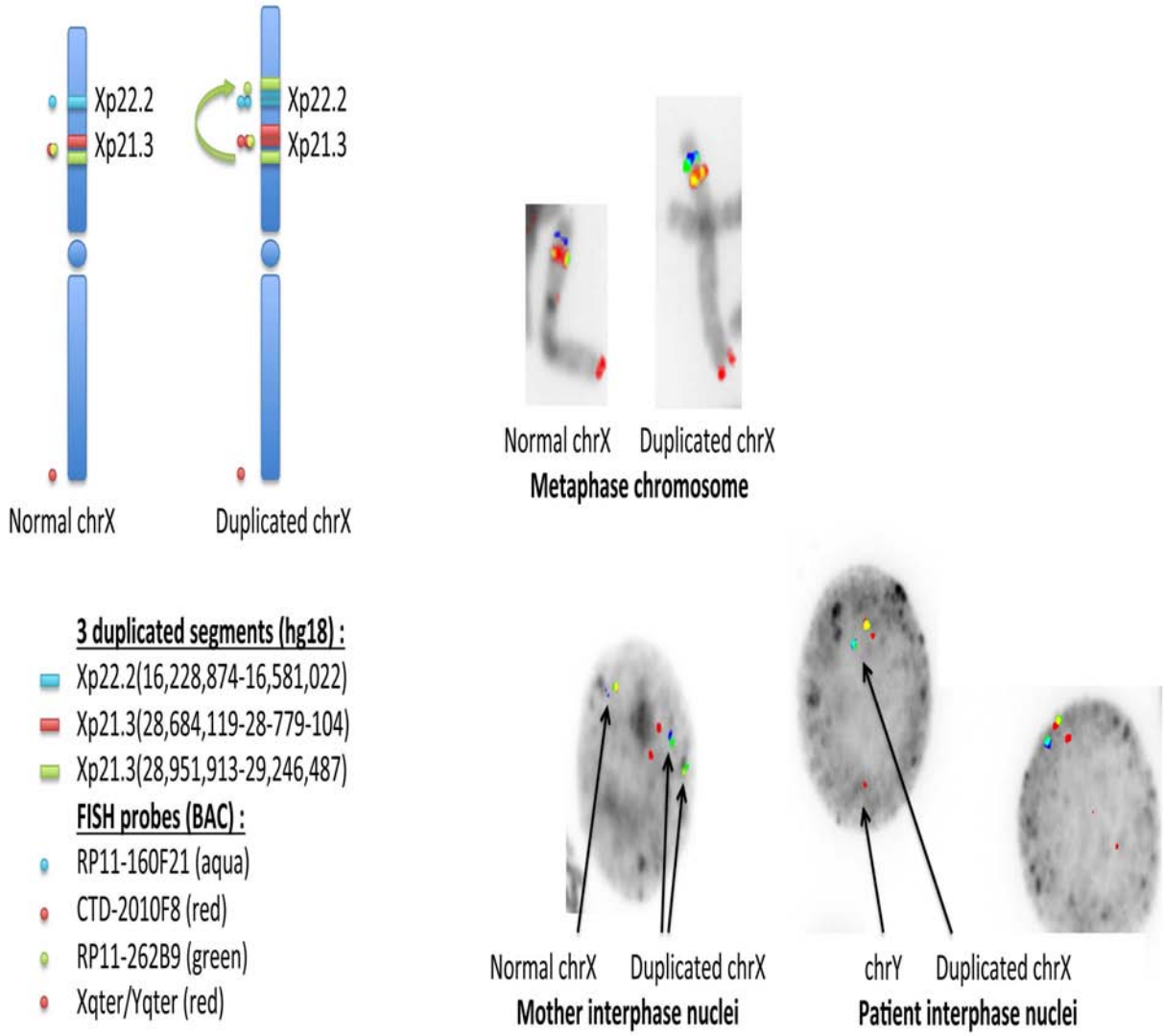


Figure 18 : Résultats de l'analyse en FISH pour la duplication chez le patient 2 et sa mère

Panel droit : Résultats de l'analyse en FISH sur chromosome métaphasique et sur noyau chez le patient et sa mère à l'aide de sondes BAC

Panel gauche : pictogramme des chromosomes X normal et dupliqué avec le positionnement de sondes BAC (cf aussi la cartographie physique des sondes de la Figure 19)



Figure 19 : Récapitulatif des remaniements du gène *IL1RAPL1* pour les 3 patients

Le détail des remaniements du gène *IL1RAPL1* pour les 3 patients, sur les locus Xp22.2 et Xp21.2 (en bleu les duplications, en rouge la délétion), avec le positionnement des sondes BAC utilisées

3.2 Caractérisation du spectre mutationnel du gène MBD5 dans la déficience intellectuelle

3.2.1 Données cliniques des patients

Les **patientes A et B** sont des jumelles monozygotes. Leur mère, traitée pour épilepsie, a arrêté son traitement pendant la grossesse. Le père et les deux autres enfants ne présentent pas de troubles cognitifs. La grossesse n'a pas été incidentée. Les jumelles présentent un retard de développement. La patiente A a acquis la station assise à 16 mois, et a marché à 2 ans et 6 mois, sa sœur B, à 17 mois et 3 ans, respectivement. Elles présentent un retard du langage, des stéréotypies et des troubles autistiques. A 3 ans et 6 mois, il est noté une taille, pour toutes les deux, à -3DS avec un périmètre crânien et un poids dans les normes pour l'âge. Une antéversion des narines est notée. L'IRM cérébrale de ces jumelles est normale

Le **patient C** est un garçon de 4 ans né de parents non consanguins. Les mensurations à la naissance sont normales. Il présente un retard de développement global. Il a commencé à marcher à 22 mois et les étapes d'acquisition du langage ont été retardées. A 4ans, les mensurations sont dans la norme, et sa DI, bien que non formellement évaluée, a été estimée de légère à modérée.

La **patiente D** est une fille de 4 ans née de parents non consanguins. Les mensurations à la naissance sont normales. Les étapes initiales du développement ont été rapportées comme normales. La marche est acquise à 19 mois. Les premiers mots sont prononcés à 13 mois, suivi d'une régression des acquis du langage, concomitamment avec une régression des interactions sociales. Elle présente un déficit de l'attention et une hyperactivité. A 3 ans et 11 mois, l'examen clinique révèle des paramètres de croissance normaux pour l'âge un visage arrondi, une antéversion des narines et une bouche aux coins tombants.

Le **patient E** est un garçon de 10 ans né de parents non consanguins. Les mensurations à la naissance étaient normales. Des difficultés alimentaires ont été observées précocement. Des épisodes d'opisthotonos ont été décrits les premiers mois de vie. A 14 mois, la tenue assise est instable. Le contact visuel est rapporté comme normal. A l'âge de 2 ans et 7 mois, il n'y avait aucun langage verbal, l'hypotonie était sévère, sans acquisition de la marche. A 10 ans et 3 mois, la taille est à + 2DS. La tenue debout est possible de façon indépendante pendant un court moment mais sans marche acquise. Le langage est absent. La personnalité est gaie avec

de fréquents sourires. Le visage est allongé, hypotonique avec une bouche entrouverte, les lobules des oreilles sont larges. Il a présenté des crises épileptiques tonico-cloniques généralisées à 4 ans ayant conduit à un traitement par du valproate, et à 8 ans des clonies des jambes, de la langue et de la bouche. L'EEG révèle des pointes et des pointes-ondes dans les régions frontales et temporales gauches. L'IRM cérébrale est normale. L'étude de la méthylation dans le cadre d'un syndrome d'Angelman est normale de même que l'analyse du gène *UBE3A* (A.Moncla, CHU Marseille)

3.2.2 Résultats de CGH-array

L'analyse par CGH-array a identifié les remaniements suivants dans le gène *MBD5*. Tous les résultats sont basés sur la version du génome UCSC build 36/hg18 (Figure 20) :

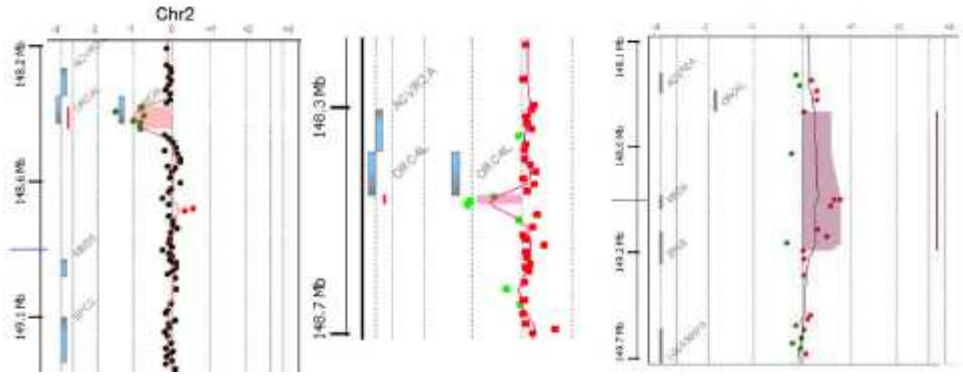
1. Chez les **jumelles A et B** une délétion interstitielle au locus 2q23.1 a été détectée : arr 2q23.1(148,447,496-148,515,776)x1

La taille minimale de cette délétion est 68.280 pb. La délétion implique la fin du gène *ORC4* et les deux premiers exons non codant de l'isoforme 1 du gène *MBD5*.

2. La **patiente D** a une délétion interstitielle au locus 2q23.1 :

arr 2q23.1(148,496,551-148,515,776)x1. La taille minimale de cette anomalie est 19.225 pb. Cette délétion touche la fin du gène *ORC4* et les deux premiers exons non codants de l'isoforme 1 du gène *MBD5*.

Chez le **patient C**, une duplication interstitielle au locus 2q23.1 est détectée : arr 2q23.1(148,944,718-148,979,574)x3. La taille minimale de cette anomalie est 34.856 pb. Cette duplication touche 4 exons (5 à 8) de l'isoforme 1 du gène *MBD5*.



Patiente A et B Patiente D Patient C

Figure 20 : Les remaniements du gène *MBD5* chez des 4 patients détectés par CGHarray

Dans cette région génomique aucune variation de nombre de copie n'a été décrite dans la base de donnée DGV (Figure 21).

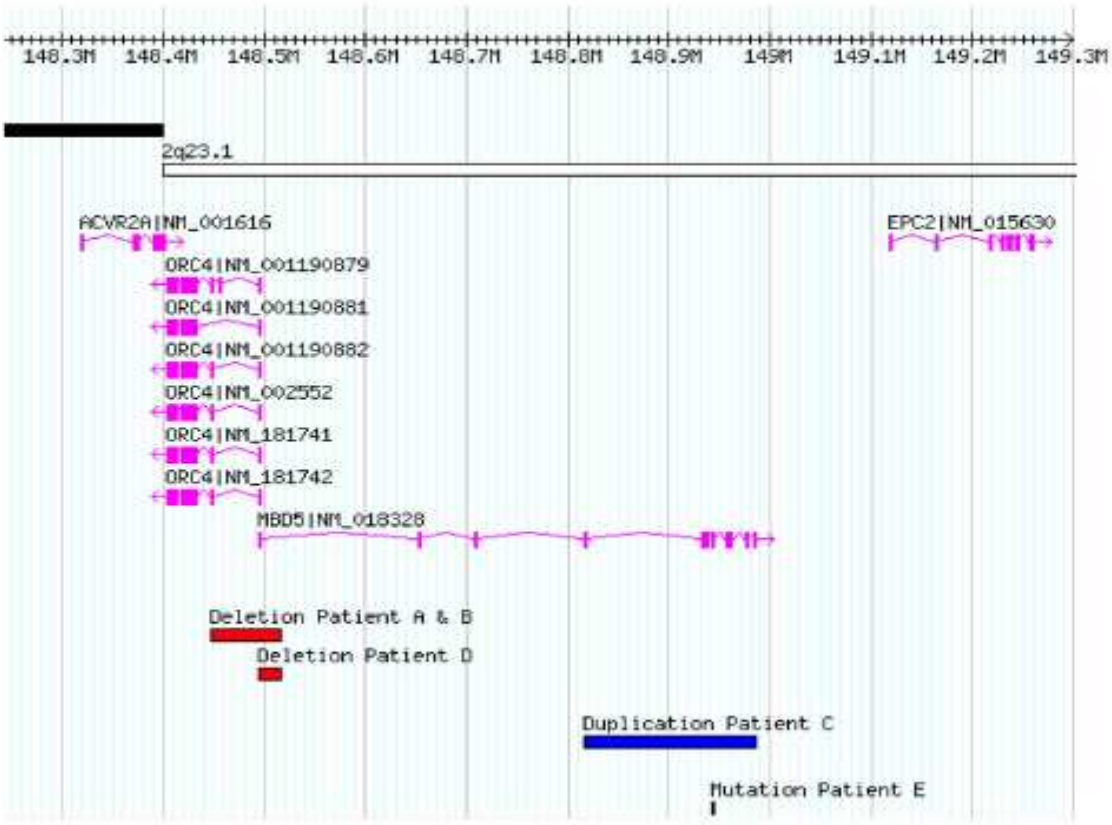


Figure 21 : Le détail des remaniements du gène *MBD5* sur le locus 2q23.1 selon la base de données DGV (build36/hg 18)

3.2.3 Résultats moléculaires

qPCR

Toutes les variations chez les patientes A, B, et D sont contrôlées par qPCR sur l'ADN génomique des patientes et de leurs parents : ces variations sont des délétions *de novo*. Les ADN parentaux ne sont pas disponibles pour le patient C. La duplication a été confirmée par comparaison avec de l'ADN témoin.

L'expression d'ARNm *MBD5* a été vérifiée chez tous les patients par RT-qPCR. Chez les patientes A, B (Figure 22) et D (Figure 23) l'expression de *MBD5* est diminuée. Chez le patient C, l'augmentation de l'expression est due à la duplication. Le séquençage des amplicons de la région dupliquée montre des transcrits anormaux (Figure 24)

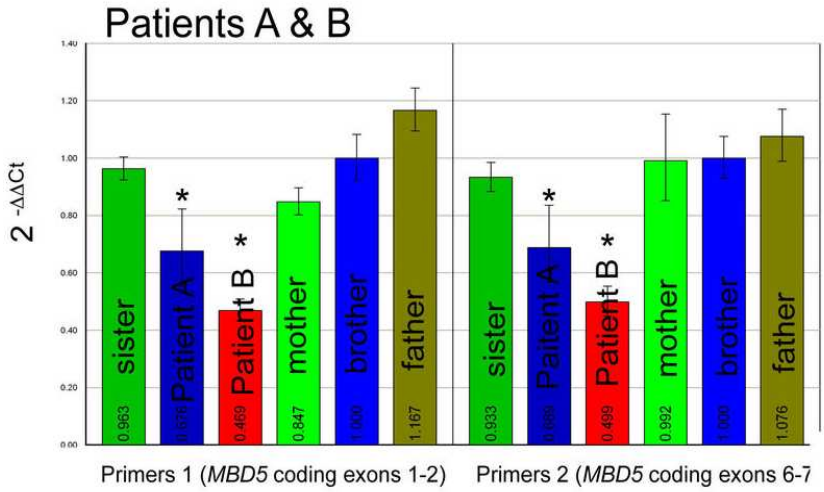


Figure 22 : Résultat de l'analyse par RT-qPCR sur cDNA

Pour les jumelles (en bleu et en rouge) et leur père, la mère, la sœur, et le frère respectifs

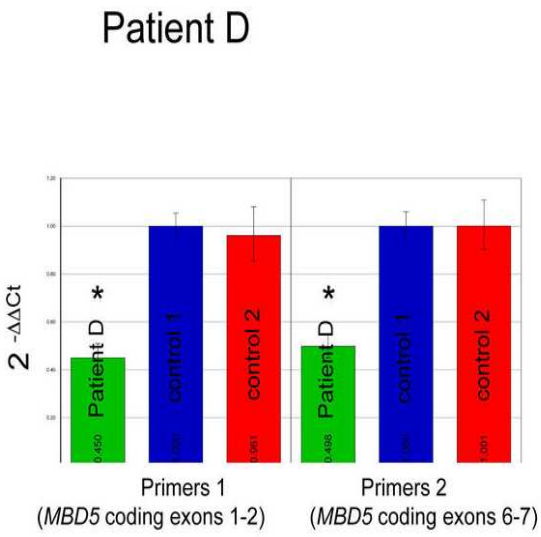


Figure 23 : Résultat de l'analyse par RT-qPCR sur cDNA

Pour la patiente D (en vert) avec l'ADN témoins (en bleu et en rouge)

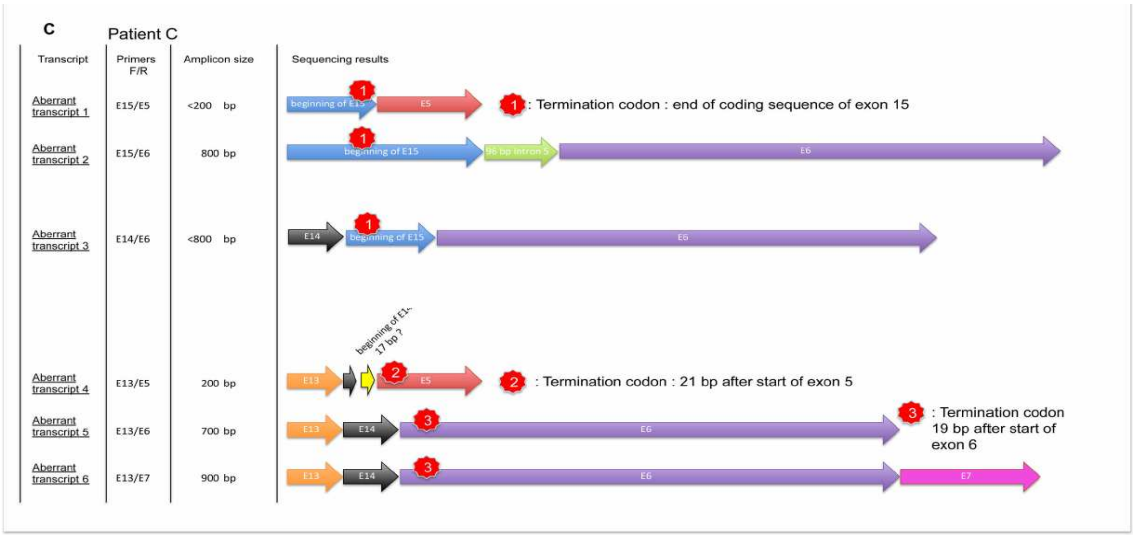


Figure 24 : Transcrits aberrants du gène *MBD5*, caractérisés par RT PCR et séquençage pour le patient C

Première colonne : numéro du transcrit ; deuxième colonne : les amorces utilisées pour l'amplification du transcrit aberrant ; troisième colonne : taille des fragments amplifiés ; quatrième colonne : structure exonique des transcrits aberrants, les étoiles rouges représentant la position des codons de terminaison.

Séquençage Sanger

Le gène *MBD5* a été séquencé dans une cohorte de 78 patients ayant un phénotype Angelman-like. Ces patients sont recrutés par le centre de référence de la maladie rare de Nancy. Nous avons détecté une mutation non sens [c.440C>G (p.Ser147*); NM_018328.3] chez le patient E. Après le séquençage des ADN parentaux de celui du frère, cette mutation a été confirmée de survenue *de novo*. (Figure 26) Le séquençage des produits de RT-PCR du gène *MBD5* pour ce patient a démontré l'expression des 2 transcrits, normal et anormal (Figure 25).

Dans cette cohorte nous avons également détecté 9 autres variations de *MBD5 :* 3 variations introniques, 3 variations faux sens et 3 variations synonymes. Toutes ces variations ne sont pas décrites dans la base de données dbSNP (build 137) [tableaux supplémentaires I et II dans annexe I]. Lorsque les prélèvements parentaux étaient accessibles, ces variants sont démontrés comme étant hérités.

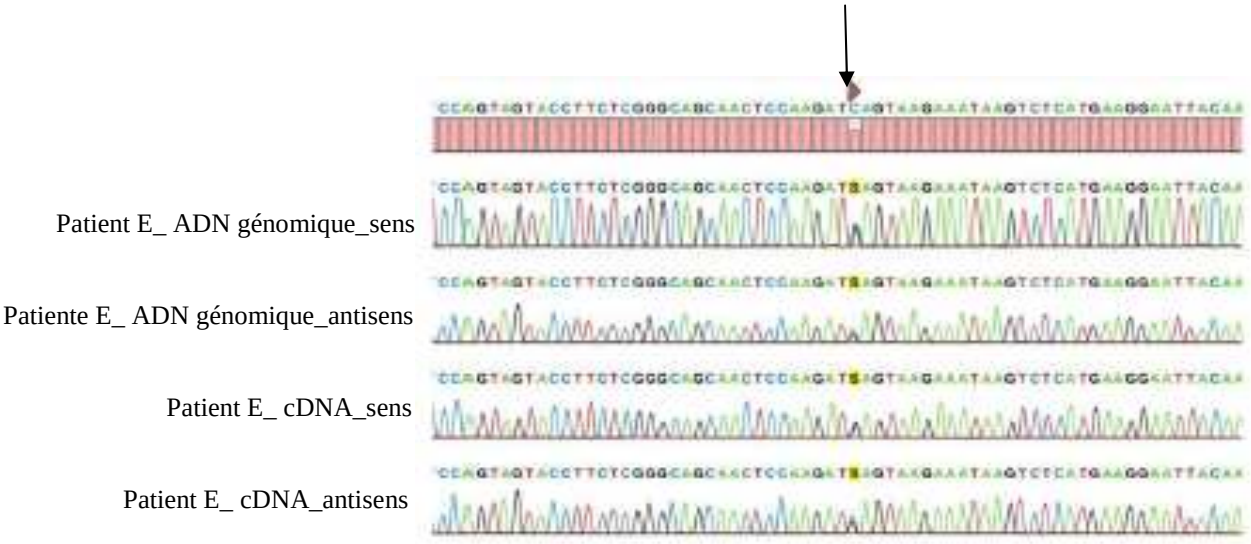


Figure 25 : Profil du séquençage Sanger sur l'ADN génomique et cDNA du patient E

La mutation non sens *MBD5* est marquée d'une flèche

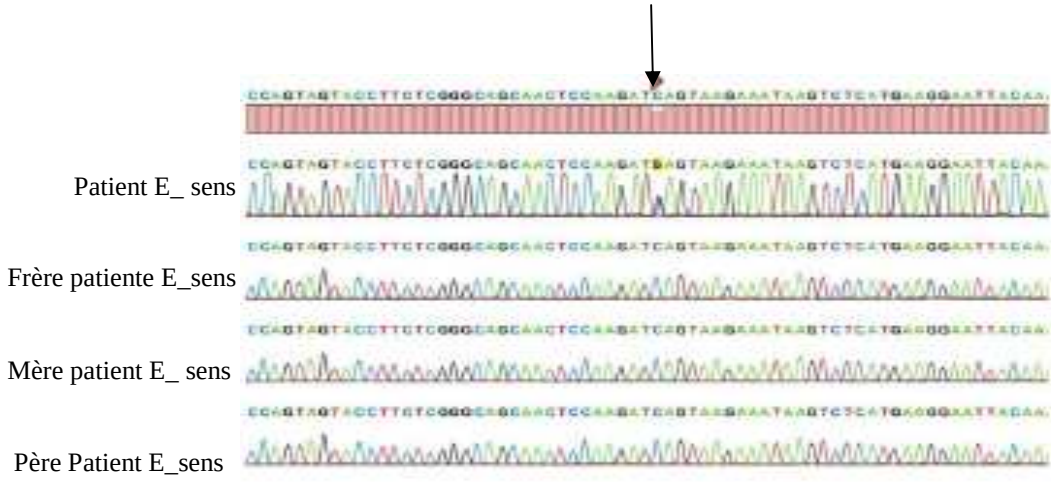


Figure 26 : Profil du séquençage d'ADN génomique du patient E, du frère, de la mère et du père

3.3 Les résultats sur le syndrome d'Aicardi

Nous avons inclus dans notre étude 22 patientes pour lesquelles le diagnostic de syndrome d'Aicardi a été porté. A la suite de nos travaux antérieurs et ceux de la littérature, notre stratégie de recherche de gènes candidats s'est orientée vers, d'une part une analyse CGH sur microréseau d'ADN haute résolution du chromosome X à la recherche d'un microremaniement de petite taille non accessible avec les microréseaux antérieurement utilisés, et d'autre part vers le séquençage nouvelle génération de l'exome du chromosome X puis de l'exome complet.

3.3.1 CGH-array 1M à façon du chromosome X

Notre hypothèse de recherche étant celle d'une mutation *de novo* d'un gène du chromosome X, nous avons utilisé notre microréseau haute résolution 1M.

3.3.1.1 L'analyse

Nous avons considéré, comme approche habituelle dans nos analyses de CGH array, le $\log_2$ ratio moyen des variations (amplification, égale ou supérieure à +0,5, et délétion égale ou inférieure à -0,5). Au total, nous avons détecté 3 681 variations pour l'ensemble de nos 22 patientes. Le filtre utilisé pour interpréter nos résultats a ciblé les variations d'au moins 3 sondes consécutives en utilisant pour l'interprétation nos résultats antérieurs de microréseaux rassemblés dans notre base de données de polymorphismes, couplées à celles issues de la littérature (Database Genomic Variants). Ces étapes nous ont permis d'éliminer 90% des variations. Nous avons retenu 354 variations dont 42 sur les autosomes et 312 sur le chromosome X. Nous avons secondairement évalué la récurrence de ces variations au sein de mêmes gènes, candidats potentiels, chez toutes les filles afin de définir une priorité pour la vérification. Nous avons mis à profit, dans l'interprétation des variations, les différentes bases de données DECIPHER, OMIM (Online Mendelian Inheritance in Man, http://www.ncbi.nlm.nih.gov/omim/), GeneCards® (http://www.genecards.org/) : lorsqu'une variation concernait une région génique annotée, nous examinions sa pertinence quant à l'existence d'un phénotype décrit en cas de mutation du gène concerné, la fonction de la protéine codée par le gène, les tissus exprimant la protéine (cerveau et rétine étant privilégiés), les éventuels modèles animaux de mutagenèse. Nous avions 312 variations dans

50 gènes du chromosome X. Après cette analyse nous avons retenu 18 variations pour vérification par la technique de qPCR. Toutes les variations exoniques ont été testées en priorité. Les résultats sont détaillés dans le tableau 3.

Tableau 3 : Résultats détectés par la puce 1M à façon du chromosome X, et leur contrôle par la technique du qPCR

patiente	gène	Localisation des sondes déviantes en pb		Nombre .de sondes déviantes	log_2 ratio	Taille minimale de l'anomalie en pb	Type d'anomalie	résultats qPCR
		début	fin					
20	*DYNLT3*	37.706.085	37.707.290	6	-2.24904	1.205	del	Faux positif
1	*DYNLT3*	37.706.085	37.707.290	6	-1.86655	1.205	del	Faux positif
15	*DYNLT3*	37.706.130	37.707.156	6	-1.67383	1.026	del	Faux positif
1	*NYX*	41.331.686	41.334.523	13	-0.642891	2.837	del	Faux positif
20	*NYX*	41.333.188	41.334.414	8	-0.845469	1.226	del	Faux positif
11	*NYX*	41.334.120	41.334.385	4	-1.13761	265	del	Faux positif
15	*NYX*	41.334.169	41.334.385	3	-1.04312	216	del	Faux positif
1	*WDR13*	48.455.508	48.456.534	7	-1.08183	1.026	del	Faux positif
20	*WDR13*	48.455.508	48.456.638	8	-0.806926	1.130	del	Faux positif
15	*WDR13*	48.455.708	48.456.491	7	-0.918836	783	del	Faux positif
1	*PLP2*	49.028.357	49.028.849	4	-1.61388	492	del	Faux positif
20	*PLP2*	49.028.357	49.028.849	4	-1.71848	492	del	Faux positif
21	*PLP2*	49.028.410	49.028.740	4	-1.25324	330	del	Faux positif
1	*GJB1*	70.442.898	70.443.348	3	-1.17396	450	del	Faux positif
1	*ZMYM3*	70.473.728	70.474.211	3	-1.25498	483	del	Faux positif
20	*PLXNA3*	153.693.325	153.695.206	11	-0.537597	1.881	del	Faux positif
1	*PLXNA3*	153.701.104	153.701.475	3	-1.32106	371	del	Faux positif
1	*VBP1*	154.444.513	154.445.002	4	-1.0056	489	del	Faux positif

Malheureusement toutes les variations retenues après l'analyse sont des faux positifs, elles ne sont pas confirmées par PCR quantitative.

3.3.2 Le séquençage haut débit de l'exome du chromosome X

En partenariat avec un prestataire extérieur (Génoscope, Evry, France) nous avons utilisé une approche de séquençage systématique de l'ensemble des gènes connus (annotés VEGA) du chromosome X et des microARN pour tester l'ADN de 3 filles atteintes d'un syndrome d'Aicardi typique (A, B et C) et les ADN du père (D) et de la mère (E), bien portants, d'une des filles (C).

Le séquençage haut débit fut réalisé par l'utilisation de la méthode de « capture de séquences » selon la technologie NimbleGen qui a permis de cibler les régions génomiques pour les séquencer ultérieurement selon la technologie Roche 454. Les régions d'intérêt ont été délimitées correspondant à 778 gènes (régions codantes et jonctions exon-intron) et les microARN pour une taille totale de 6,3 Mb (6.267.009 pb). Les puces d'oligonucléotides (60-80 mer) ont conduit à capturer les séquences et des régions d'intérêt des 5 ADN différents pour le séquençage. L'analyse a été réalisée sur équipement 454 GS-FLX de façon à obtenir une profondeur de 10x minimum pour chaque échantillon. Les séquences obtenues ont été alignées sur le génome de référence. La liste des séquences différentes par rapport au génome de référence a été accompagnée d'un indice de confiance. Les variations isolées tiennent compte des informations issues des bases de données des SNP et de HapMap. Le nombre de lectures sur chaque brin d'ADN (sens et antisens) est précisé pour chaque variation identifiée.

Statistique des résultats issus du séquençage du Génoscope

1 826 592 séquences chevauchent les régions ciblées (soit 68% de l'ensemble des lectures)
Valeurs moyennes sur les 5 individus :
Régions ciblées couvertes : 92%
Régions entièrement couvertes : 82%
Régions couvertes au moins 10x : 79%
Couverture moyenne des régions : 9x

Nous avons réalisé l'interprétation des données brutes avec l'aide du logiciel de filtrage mis à disposition par le Génoscope (Figure 27 ci-dessous).

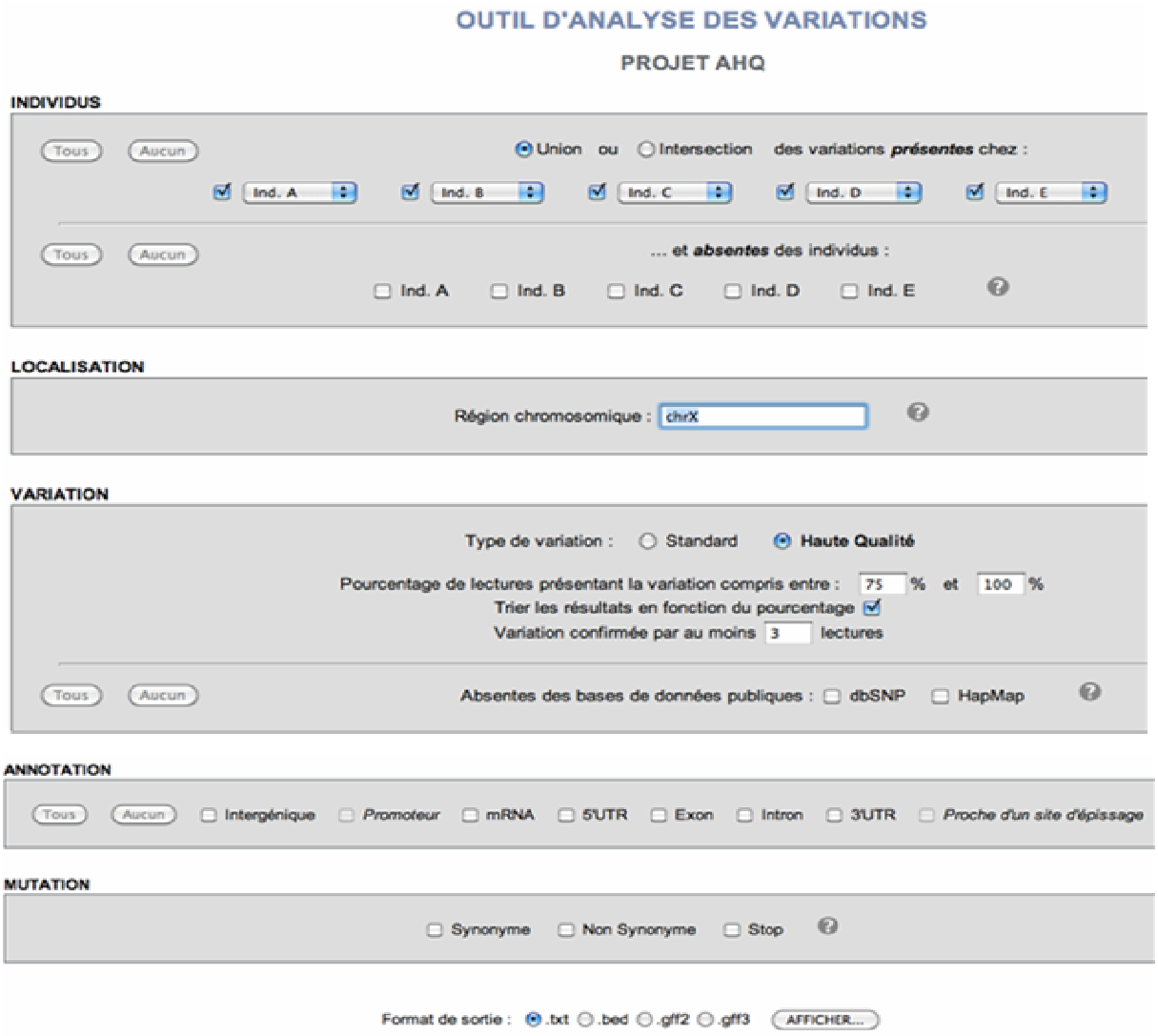


Figure 27 : Interface du logiciel d'analyse des données de séquençage haut débit de l'exome du chromosome X

Il est ainsi possible de soumettre différentes requêtes portant sur la combinaison des variations observées chez les individus A et/ou B, et/ou C, et/ou D et/ou E en union ou en intersection, variations souhaitées simultanément absentes de certains individus. La localisation des variations ici correspond au chromosome X. Les variations avec la pertinence de gène intervenant dans les voies importantes du développement neuronal et la rétine ont été en priorité choisies. Compte tenu de l'hypothèse initiale de maladie génétique dominante liée au

chromosome X évoquée *a priori* pour le syndrome d'Aicardi, une mutation présente à l'état hétérozygote chez les filles ayant 2 chromosomes X a donc été recherchée. Ainsi, la présence d'un allèle est considérée pour un pourcentage de lecture supérieur de 25%.

Pour faciliter l'interprétation des résultats, les variations préalablement décrites dans les bases de données définissant des polymorphismes dbSNP et HapMap (http://hapmap.ncbi.nlm.nih.gov/) ont été exclues. Nous avons retenu, *a priori*, toutes les localisations possibles des mutations (en sachant que les régions intergéniques n'ont pas été séquencées) et finalement, dans un premier temps seules les mutations non synonymes ont été recherchées.

Pour chaque scénario soumis, il est attendu le recueil de variations génomiques. ***L'analyse de la signification de ces variations doit tenir compte de*** :

- La présence d'une variation dans le même gène au mieux chez les 3 filles A, B et C et absente des bases de polymorphismes (dbSNP et HapMap) et aussi absente des parents de C, c'est-à-dire D et E

- Toutefois compte tenu du taux de couverture des régions d'intérêt, on ne peut éliminer un gène candidat si la variation n'est présente que chez 2 filles, d'autant plus qu'avec la technique de séquençage, seules les mutations exoniques sont détectées, les réarrangements de plus grande taille, délétion ou duplication d'un exon (chez une fille XX) ne seront pas forcément reconnues.

- Parmi les gènes porteurs d'une variation, un classement ou « ranking » peut être envisagé en se basant sur :

    - La nature de la mutation, et aussi plus particulièrement nous avons accès au positionnement exact de la mutation au sein de la séquence, à l'une des extrémités, au centre de la région d'intérêt. Notre expérience nous ayant révélé que les faux positifs de ce type de séquençage étaient particulièrement fréquents au niveau des extrémités

    - Le nombre de fois où la variation (sens et antisens) a été retrouvée par rapport au nombre de lecture de cette même région

    - L'annotation fonctionnelle du gène, en utilisant le croisement de bases de données compte tenu du phénotype clinique associé au syndrome d'Aicardi

(encéphalopathie). Une telle approche a été mise en place par l'équipe de LORIA, Nancy Université avec laquelle nous conduisons un travail collaboratif.

- Les variations sont systématiquement contrôlées par séquençage Sanger ciblé en utilisant un couple d'amorce encadrant la région modifiée, chez la ou les filles pour éliminer un éventuel faux positif.
- Les variations confirmées sont recherchées chez les parents de la fille porteuse de la variation pour connaître la survenue *de novo* ou héritée. *A priori*, seules les mutations *de novo* sont retenues comme candidates.

3.3.2.1 Scénario : <u>Intersection</u> A, B et C en l'absence de variations identiques retrouvées chez D et E

Aucune variation n'a été identifiée au seuil de 25%.
Seule variation identifiée, partagée par A, B et C mais au ***seuil de 6%***, dans le gène *TSPYL2*. Ce gène s'exprime dans le testicule et le cerveau et son produit interagit avec la protéine TGFB1. Néanmoins, cette variation est identique chez les 3 filles (récurrente), variation faux-sens et représentée à l'état de mosaïque très faible, et non retrouvée par le séquençage ciblé, rendant ainsi peu probable son implication dans la pathologie.

3.3.2.2 Scénario : <u>Union</u> A, B et C en l'absence de variations identiques retrouvées chez D et E

204 variations ont été identifiées au seuil de 25%
Le système d'analyse des variations permet de reconnaître aisément les variations identiques dans un même gène, entre les 3 filles, mais ne permet pas, de façon systématique, de discerner facilement si un même gène peut être muté de façon différente entre les 3 filles. Après une relecture séparée des variations pour chaque fille, il n'est pas apparu, de croisement entre elles pour un même gène. Ce résultat, certes quelque peu décevant, a conduit toutefois en se fondant sur l'hypothèse précitée plus haut à s'intéresser à chaque variation. Un ranking a été effectué selon le pourcentage de lecture avec la variation et les critères d'annotation fonctionnelle, suivi par un séquençage systématique de la région cible chez la fille et ses 2 parents. Ces résultats contribuent toutefois à identifier de nouveaux variants génomiques non

décrits jusqu'alors dans les bases de données telles dbSNP et HapMap. Ces variations serviront à abonder les bases de données, favorisant en retour l'interprétation prochaine des études de séquençage haut débit d'autres projets de recherche.

En fin de compte, les résultats n'ont pas montré de mutation délétère chez les 3 patientes par séquençage à haut débit : **59 mutations** dans les **51 gènes** candidats ont été retenues et contrôlées par la technique Sanger (Tableau 4):

- **13 mutations** sont **héritées de la mère**
- **8 mutations** sont **héritées du père**
- **36 mutations** sont des **faux positifs**
- **2 mutations** sont des **SNP**

Tableau 4 : Représentation des résultats de séquençage haut débit de l'exome du chromosome X contrôlés par le séquençage Sanger

	Patiente	Gène	Variation	Type de variation	Position génomique	Nombre de lecture total	Pourcentage de lecture avec la variation	localisation de la variation	Changement de l'acide aminé	Séquençage Sanger
1	A	*BMX*	T/A	MS	15 446 015	15	33	**CDS**	V/D	faux positif
								CDS	V/D	
								CDS	V/D	
2	A	*AKAP4*	C/T	MS	49 845 217	29	41	**CDS**	G/D	paternel
								CDS	G/D	
								CDS	G/D	
3	A	*TRO*	C/T	MS	54 972 122	23	61	**CDS**	S/F	paternel
								CDS	S/F	
4	A	*EDA*	G/T	MS	68 753 083	7	43	**CDS**	R/L	maternel
								CDS	R/L	
								CDS	R/L	
								CDS	R/L	
								CDS	R/L	
5	A	*TAF7L*	T/C	MS	100 425 224	12	58	**CDS**	D/G	paternel
								CDS	D/G	
								CDS	D/G	
								CDS	D/G	
6	A	*IRS4*	-/A	PTC	107 864 998	16	69	**CDS**	G/W*	maternel
7	A	*ACSL4*	T/C	MS	108 795 370	18	56	**CDS**	Y/C	paternel
								CDS	Y/C	
								CDS	Y/C	
								CDS	Y/C	
8	A	*C1GALT1C1*	C/T	MS	119 644 748	16	38	**CDS**	S/N	maternel
								CDS	S/N	

	Patiente	Gène	Variation	Type de variation	Position génomique	Nombre de lecture total	Pourcentage de lecture avec la variation	localisation de la variation	Changement de l'acide aminé	Séquençage Sanger
9	A	*IDS*	G/C	MS	148 393 595	16	69	**CDS**	A/G	paternel
								CDS	A/G	
								CDS	A/G	
								CDS	A/G	
10	A	*PASD1*	-/GCT		150 567 822	29	38	**CDS**	I/A	maternel
11	B	*GYG2*	C/A	MS	2 758 203	16	63	**5'UTR**	NA	maternel
								5'UTR	NA	
								CDS	R/S	
								5'UTR	NA	
12	B	*ARSH*	C/T	MS	2 955 389	31	58	**CDS**	R/C	paternel
13	B	*CDKL5*	C/A	MS	18 536 910	11	36	**CDS**	S/Y	faux positif
								CDS	S/Y	
14	B	*KLHL34*	A/G	MS	21 585 355	12	83	**CDS**	L/S	maternel
15	B	*DDX53*	G/A	MS	22 929 042	31	45	**CDS**	C/Y	paternel
16	B	*CXorf36*	C/A	MS	44 936 010	32	53	**CDS**	R/L	maternel
								CDS	R/L	
								CDS	R/L	
17	B	*ITIH5L*	T/A	MS	54 794 229	28	75	**CDS**	Q/L	maternel
18	B	*HDX*	A/C	MS	83 611 101	18	72	**CDS**	S/A	maternel
								CDS	S/A	
19	B	*NXT2*	G/A	MS	108 667 141	14	71	**CDS**	G/S	maternel
								5'UTR	NA	
20	B	*SMARCA1*	G/A	MS	128 458 461	24	54	**CDS**	R/C	maternel
								CDS	R/C	
								CDS	R/C	
21	B	*MST4*	T/A	MS	130 985 234	7	43	**5'UTR**	NA	faux positif
								5'UTR	NA	
								5'UTR	NA	
								5'UTR	NA	
								CDS	W/R	
								CDS	W/R	
22	B	*GPR50*	C/T	MS	150 099 809	22	50	**CDS**	R/W	paternel
23	B	*PLXNB3*	C/G	MS	152 695 954	11	45	**CDS**	S/R	maternel
24	C	*MCF2*	T/A	MS	138 542 166	13	46	**CDS**	K/N	maternel
								CDS	K/N	
								CDS	K/N	
								CDS	K/N	

	Patiente	Gène	Variation	Type de variation	Position génomique	Nombre de lecture total	Pourcentage de lecture avec la variation	localisation de la variation	Changement de l'acide aminé	Séquençage Sanger
25	A	*HDAC*	C/-	PTC	48 566 071	20	35	**CDS**	P/H*	faux positif
								CDS	P/H*	
	B	*HDAC*	C/-	PTC	48 566 071	39	26	**CDS**	P/H*	faux positif
								CDS	P/H*	
26	C	*ARHGEF6*	T/-	DEL	135 591 791	30	37	**CDS**	I/Y	faux positif
27	B	*CHIC1*	-/A	INS	72 700 026	14	36	**CDS**	E/R	faux positif
	C		-/A	INS	72 700 008	37	32	**CDS**	E/R	faux positif
28	A	*DOCK11*	T/-	DEL	117 659 280	28	25	**CDS**	F/S	faux positif
	C		A/GAGGAT	INS	117 623 270	13	31	**CDS**	L/L	faux positif
29	A	*HUWE1*	T/-	DEL	53 627 956	19	42	**3'UTR**	NA	faux positif
	A		T/-	DEL	53 627 956	19	42	**CDS**	N/T	faux positif
	C		A/-	DEL	53 636 307	24	25	**3'UTR**	NA	faux positif
	C		A/-	DEL	53 636 307	24	25	**CDS**	F/L	faux positif
30	A	*UPF3B*	A/-	DEL	118 869 769	11	45	**CDS**	S/S	faux positif
	C		A/-	DEL	118 869 769	14	29	**CDS**	S/S	faux positif
31	C	*KDM5C*	T/-	DEL	53 263 786	48	25	**CDS**	N/I	faux positif
32	A	*VCX2*	CT/GC	SUB	8 098 171	5	100	**CDS**	S/A	faux positif
	B		CT/GC	SUB	8 098 171	6	100	**CDS**	S/A	faux positif
	C		/G	INS	8 098 170	10	40	**CDS**	E/R	faux positif
33	A	*ZAMT1*	TTTCCAA/ATCTCC	SUB	101026301-307	10	40	**3'UTR**	NA	faux positif
	A		TTTCCAA/ATCTCC	SUB	101026301-307	10	40	**CDS**	L/G	faux positif
	C		T/-	DEL	101 026 118	18	28	**3'UTR**	NA	faux positif
	C		T/-	DEL	101 026 118	18	28	**CDS**	K/K	faux positif
34	B	*FLNA*	C/-	DEL	153 235 713	12	42	**5'UTR**	NA	faux positif
	B		C/-	DEL	153 235 713	12	42	**CDS**	G/E	faux positif
	C		C/-	DEL	153 231 241	37	32	**5'UTR**	NA	faux positif
	C		C/-	DEL	153 231 241	37	32	**CDS**	G/A	faux positif
35	A	*DMD*	/GA	INS	32 542 455	11	36	**CDS**	L/S	faux positif
	B		T/-	DEL	32 145 053	14	29	**CDS**	K/R	faux positif
	C		T/-	DEL	32 215 600	19	26	**CDS**	K/K	faux positif
36	C	*ACRC*	A/-	DEL	70 741 131	19	26	**CDS**	R/G	faux positif
37	C	*AR*	C/-	DEL	66 681 930	19	37	**5'UTR**	NA	SNP
	C		C/-	DEL	66 681 930	19	37	**CDS**	Q/S	SNP
38	C	*BMP15*	T/-	DEL	50 675 789	40	25	**CDS**	F/L	faux positif
39	C	*BRWD3*	/T	INS	79 819 413	16	25	**CDS**	E/R	faux positif
40	C	*FAM47A*	C/-	DEL	34 059 208	21	29	**CDS**	R/L	faux positif
41	C	*L1CAM*	G/-	DEL	152 784 104	15	27	**CDS**	H/I	faux positif
42	C	*MAP3K15*	C/-	DEL	19 290 839	15	27	**CDS**	R/R	faux positif
43	C	*PHEX*	C/-	DEL	22 025 075	22	27	**CDS**	Q/S	faux positif
44	C	*PHKA2*	T/-	DEL	18 833 851	40	28	**CDS**	N/I	faux positif
45	C	*PLXNA3*	T/-	DEL	153 342 652	22	27	**CDS**	V/G	faux positif
46	C	*PNCK*	/G	INS	152 589 833	30	27	**CDS**	G/R	faux positif
47	C	*PRICKLE3*	C/-	DEL	48 920 245	42	31	**CDS**	A/H	faux positif
48	C	*RPL36A*	A/-	DEL	100 533 465	31	32	**CDS**	K/R	faux positif
	C		A/-	DEL	100 533 465	31	32	**CDS**	K/S	faux positif
49	C	*TRO*	T/-	DEL	54 972 579	37	38	**CDS**	F/L	faux positif
50	C	*TSPYL2*	C/-	DEL	53 128 874	36	25	**CDS**	Q/R	faux positif
51	C	*ZRSR2*	A/-	DEL	15 718 986	38	26	**CDS**	K/S	faux positif

Rose : variation hérité de la mère, bleu : variation hérité du père, gris : faux positif, MS : mutation faux sens ; PTC : codon de terminaison prématuré ; INS : insertion ; DEL : délétion ; SUB : substitution ; CDS : séquence codante

3.3.3 Le Séquençage haut débit de l'exome complet

Nous avons poursuivi notre analyse avec un séquençage haut débit de l'exome complet en faisant appel à un prestataire extérieur IntegraGen. L'ADN génomique est capturé en utilisant la méthode d'Agilent d'enrichissement en solution (SureSelect Human All Exon Kits Version 2, Agilent) avec leur banque de sondes oligonucléotidiques biotinylées (Human All Exon v2 – 50 Mb, Agilent), suivi d'un séquençage haut débit des 2 extrémités sur 75 b sur Illumina HiSEQ 2000 [171].

Les ADN de cinq trios **-** patientes avec le phénotype typique du syndrome d'Aicardi et leurs parents ont été ainsi analysés. Les régions cibles sur le génome humain représente une couverture de 20.965 gènes au total, avec 334.378 exons (bases de données CCDS version 2011 ; RefSeq version 2011 ; miRBase v17 ; GENCODE v6 et UCSC Mars 2011).

Les séquences obtenues sont alignées sur le génome de référence ; un filtre de qualité des séquences est ensuite appliqué ; la profondeur moyenne de lecture obtenue, sur les régions cibles, est de 60x. Tous les résultats sont générés avec leur score de qualité, le statut homozygote/hétérozygote des bases, et comptage des bases lues. La position sur le génome, sur le gène, sur le transcrit, dans la protéine, sur les miRNA, distance à l'intron, type de mutation, changement d'acide aminé, ou de phase de lecture sont aussi précisés. Ensuite l'existence ou non de ces variants (SNP et Indels) est vérifiée dans les bases de données dbSNP (build 137), HapMap, 1000 Genomes, et de même l'expérience accumulée sur les exomes préalablement séquencés par IntegraGen. Le tableau 5 décrit la légende des annotations des SNPs selon IntegraGen

Tableau 5 : Légende des annotations des SNPs selon IntegraGen

Fichier de description de la détection et annotations des SNPs	
position	**Position du variant sur le build 137**
chrom	**chromosome**
sample.ID	**Nom de l'échantillon séquencé**
Freq.HTZ.Exomes	**Fréquence d'observation du variant avec le statut hétérozygote dans les x exomes déjà réalisés chez integraGen, hormis le projet en question.**
Freq.Hom.Exomes	**Fréquence d'observation du variant avec le statut homozygote (non référence) dans les x exomes déjà réalisés chez integraGen, hormis le projet en question.**
ref	**La base de référence**
A	**Nombre de bases A séquencées**
C	**Nombre de bases C séquencées**
G	**Nombre de bases G séquencées**
T	**Nombre de bases T séquencées**
used	**Nombre de bases utilisées pour la détection des variants**
filt	**Nombre de bases filtrées pour la détection des variants**
Q.snp.	**Score de qualité (Qphred) du SNP (probabilité que le SNP soit une erreur)**
max_gt	**Le génotype observé.**
Q.max_gt.	**Score de qualité (Qphred) du statut génotypique du SNP (Probabilité que le statut donné par max_gt soit faux)**
statut	**SNP_diff: SNP homozygote (non référent), SNP_het=SNP hétérozygote**
Gene.name	**Nom du gène incluant la position du variant**
Gene.start	**Position start du gène**
Gene.end	**Position end du gène**
strand	**Orientation du gène**
nbre.exon	**Nombre d'exon inclus dans le gène**
refseq	**Nom du refseq du gène**
typeannot	**Annotation du variant: 5-UTR, 3-UTR,intron, missense: le variant est inclu dans le cds du gène et la mutation entraîne le changement de l'acide-aminé, nonsense: le variant est inclu dans le cds et la mutation entraîne la formation d'un codon stop, synonymous: le variant est inclu dans le cds du gène et la mutation n'entaîne pas de changement de la séquence protéique.**
type.pos	**Numèro de l'exon ou de l'intron incluant le variant**
index.cdna	**Position start de l'intron incluant le variant**
index.prot	**Position end de l'intron incluant le variant**
Taille.cdna	**Taille en pb du cDNA correspondant au refseq**
Start	**Position start de l'intron ou de l'exon incluant le variant**
End	**Position end de l'intron ou de l'exon incluant le variant**
codon.wild	**Codon de référence**

aa.wild	**Acide aminé de référence**
codon.mut	**Codon observé**
aa.mut	**Acide aminé obéservé**
cds.wild	**Séquence nucléotidique cDNA de référence**
cds.mut	**Séquence nucléotidique cDNA observé**
prot.wild	**Séquence protéique de référence, présente seulement si l'acide aminé muté est différent de celui de référence**
prot.mut	**Séquence protéique observée, présente seulement si l'acide aminé muté est différent de celui de référence**
mirna	**Annotation des miRNAs (décrit dans le design de capture d'Agilent)**
region.splice.intron	**Position dans les sites donneurs (GU 5')/accepteur (AG 3') d'épissage à l'intron**
region.splice.exon	**Position dans les sites donneurs (GU 5')/accepteur (AG 3') d'épissage à l'exon**

3.3.3.1 Scénario 1 : Mutations non-sens de novo

Recherche d'une mutation non-sens *de novo* (uniquement présente chez les filles et absente chez les parents) quelle que soit la profondeur de lecture et non décrites dans les bases de données de polymorphismes connus (Tableau 6).

Tableau 6 : Mutations non-sens *de novo* détectée par Séquençage haut débit de l'exome

Gène	Chr	Position	trio	Type de variation	Variant	Prof.	A	C	G	T	Codon sauvage	Acide aminé sauvage	Codon muté	Acide aminé muté
DMAP1	chr1	44679474	3	nonsense	SNP	8	0	0	6	2	GAA	E	TAA	*
C11orf35	chr11	556543	1	nonsense	SNP	5	0	0	3	2	TCG	S	TAG	*
ARHGEF25	chr12	58005752	1	nonsense	SNP	5	2	3	0	0	TGC	C	TGA	*
ARHGEF25	chr12	58005752	1	nonsense	SNP	5	2	3	0	0				
PLEKHG3	chr14	65205475	5	nonsense	SNP	7	0	0	5	2	GAG	E	TAG	*
SH2B1	chr16	28878190	1	nonsense	SNP	7	0	0	5	2	GAG	E	TAG	*
ZBTB4	chr17	7369364	5	nonsense	SNP	5	2	3	0	0	GAG	E	TAG	*
MAFG	chr17	79880485	1	nonsense	SNP	8	0	0	6	2	TCG	S	TAG	*
C2CD4C	chr19	407992	3	nonsense	SNP	5	2	3	0	0	GAG	E	TAG	*
HMHA1	chr19	1073965	1	nonsense	SNP	5	0	0	3	2	GGA	G	TGA	*
MEIS3	chr19	47920144	4	nonsense	SNP	6	2	4	0	0	GGA	G	TGA	*
IGFBP2	chr2	217498509	1	nonsense	SNP	7	2	5	0	0	TCG	S	TAG	*
MICAL3	chr22	18314737	1	nonsense	SNP	6	2	4	0	0	GAG	E	TAG	*
COL7A1	chr3	48612798	1	nonsense	SNP	5	2	3	0	0	GGA	G	TGA	*
RNF175	chr4	154631526	5	nonsense	SNP	82	47	35	0	0	GAA	E	TAA	*
TSPAN17	chr5	176074664	4	nonsense	SNP	7	2	5	0	0	TGC	C	TGA	*
GPANK1	chr6	31630303	3	nonsense	SNP	9	2	7	0	0	GGA	G	TGA	*
USP17L2	chr8	11995066	5	nonsense	SNP	68	26	0	42	0	CGA	R	TGA	*
WASH1	chr9	15915	5	nonsense	SNP	4	2	2	0	0	GAG	E	TAG	*

De gauche à droite : le nom du gène, le chromosome, les coordonnées génomiques du variant sur le chromosome, le numéro du trio, le type de variation (non sens) et la nature SNP (à l'opposé des indel), la profondeur de lecture, les 4 nucléotides(A,C,G,T) et les proportions de variant, le codon sauvage et son acide aminé, le codon muté générant un stop.

Au total, 19 mutations sont décelées ; seules les 2 mutations avec profondeur de lecture de plus de 10x (82x et 68x) sont retenues. Pour les 17 autres variants, la profondeur de lecture est <10.

Gène *RNF175* (ring finger protein 175) (NM_173662, exon 9 sur 9)

Le gène *RNF175* est localisé sur le chromosome 4 et fortement exprimé dans le cerveau et la moelle épinière (Figure 28).

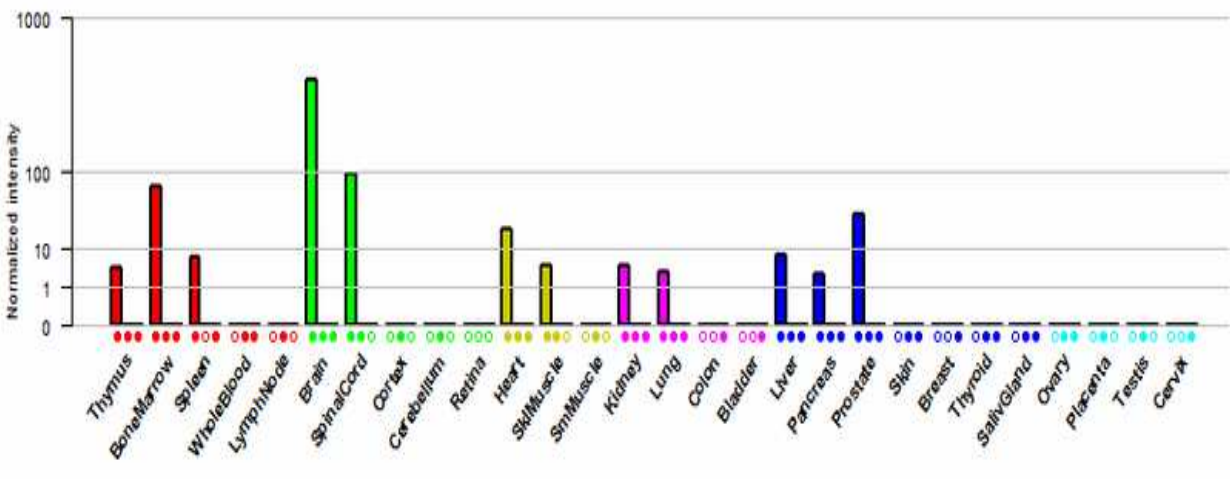


Figure 28 : Expression du gène *RNF175* selon la base de données Gene Cards®

La forte expression (en vert) dans le cerveau et la moelle épinière du *RNF175.*

Tableau 7 : Résultats du séquençage de l'exome pour le gène *RNF175* pour la patiente du trio 5 et ses parents (Extrait du tableau 6)

Gène	Chr	Position	trio	Type de variation	Variation	Prof	A	C	G	T	Codon sauvage	Acide aminé sauvage	Codon muté	Acide aminé muté
RNF175	chr4	154631526	5	Nonsense	SNP	82	47	35	0	0	GAA	E	TAA	*

- **Trio 5**

Une mutation non-sens a été détectée chez la patiente (trio 5) et non retrouvée chez les parents : elle est absente chez la mère, en revanche l'ADN paternel était non couvert (NC) lors de la lecture pour cette mutation (Tableau 7). Nous avons poursuivi par le séquençage ciblé Sanger chez le père révélant bien l'origine paternel du variant non sens détecté (Figure 29), résultat non en faveur du caractère délétère du variant.

Tableau 8 : La mutation dans le gène *RNF175* chez le trio 5

Patiente (trio5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
RNF175	47	35	0	0	82	5	360	AC	260	SNP Hétérozygote

Mère (trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp	max_gt.	Q.max_gt.	statut.
RNF175	0	88	0	0	88	1	0	CC	295	SNP Référence

Père (trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt	Q.snp	max_gt.	Q.max_gt.	statut.
RNF175	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC

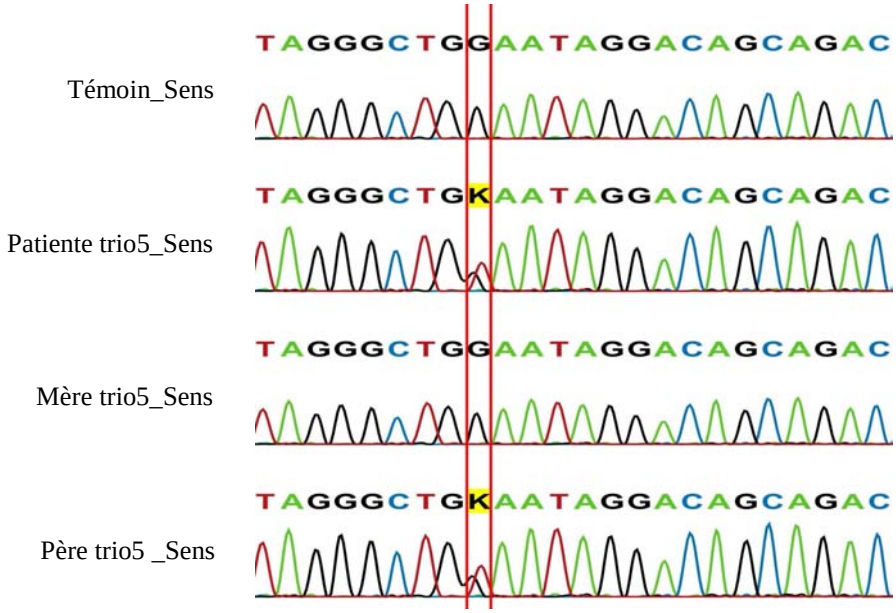


Figure 29 : Profil du séquençage ciblé Sanger du gène *RNF175*. Il s'agit du variant non sens hérité du père (un témoin, et le trio ont été séquencés)

Gènes *USP17L2* (ubiquitin specific peptidase 17-like 2) (NM_201402, exon 1 sur 1)
et ***FAM66D*** (family with sequence similarity 66, member D) (NR_027425, 5'UTR, 6 exons)

Localisé sur le chromosome 8, il possède un seul exon qui code pour la protéine de 530 acides aminés. Son transcrit de 1,7kb a été détecté dans le cerveau, le poumon et la peau (Figure 31).Le gène *USP17L2* est localisé dans l'intron 2 du gène *FAM66D* (transcrit non codant). La variation de séquence a été détectée sur la même position dans les deux gènes (Figure 30).

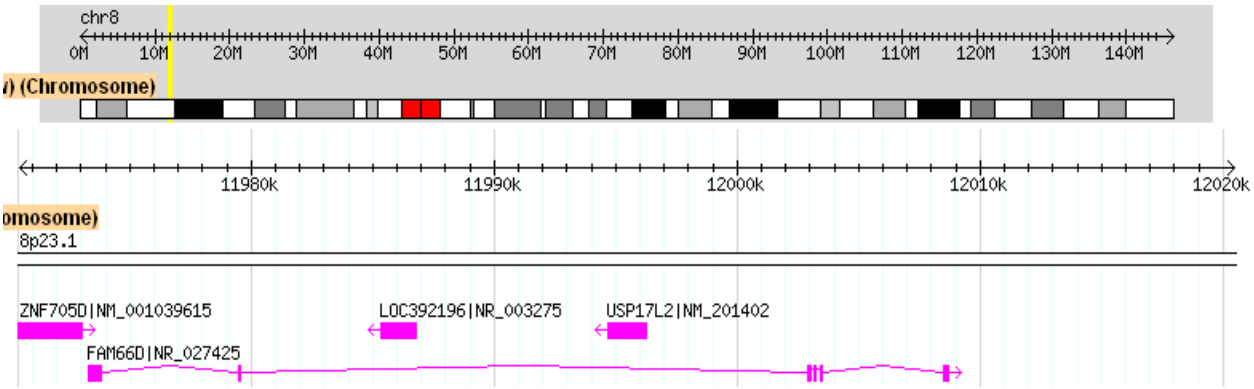


Figure 30 : Localisation des gènes *USP17L2* et *FAM66D* sur la locus 8p23.1 selon la base de données DGV

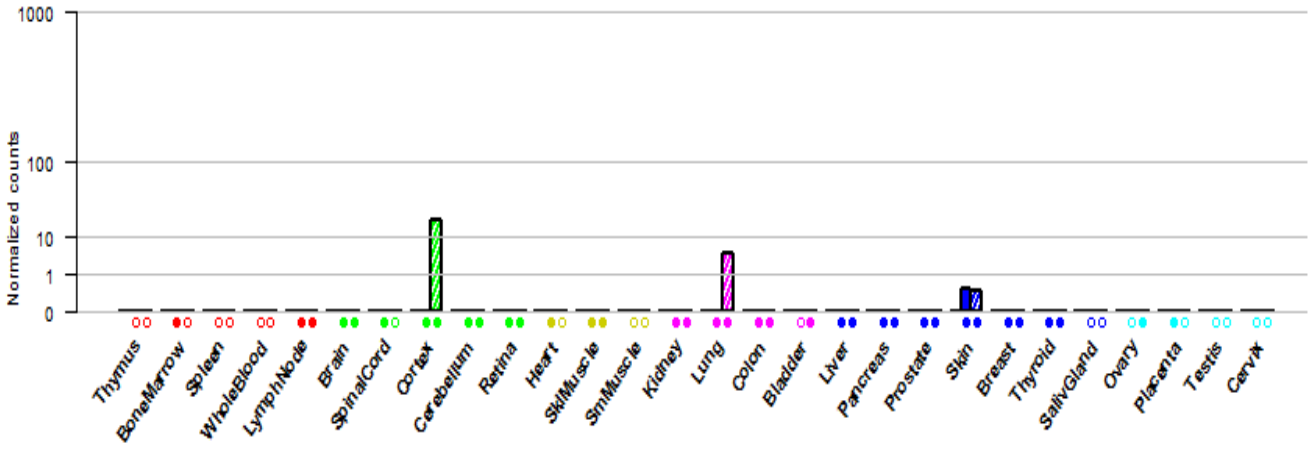


Figure 31 : Expression du gène *USP17L2* selon la base de données Gene Card®

Le gène *USP17L2* s'exprime faiblement dans la peau et le poumon. Une expression plus importante est détectée dans le cortex

Tableau 9 : Résultats du séquençage de l'exome pour le gène *RNF175* pour la patiente (trio 5) et ses parents (Extrait du tableau 6)

Gène	**Chrom**	**position**	**trio**	**Type de variation**	**Variation**	**Prof**	**A**	**C**	**G**	**T**	**Codon sauvage**	**Acide aminé sauvage**	**Codon muté**	**Acide aminé muté**
USP17L2	chr8	11995066	5	nonsense	SNP	68	26	0	42	0	CGA	R	TGA	*

Patiente (trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt	Q.snp	max_gt	Q.max_gt.	statut.
USP17L2	26	0	42	0	68	8	151	AG	151	SNP Hétérozygote
FAM66D	26	0	42	0	68	8	151	AG	151	SNP Hétérozygote

Mère (trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
USP17L2	23	0	90	0	113	9	0	GG	32	SNP Référence
FAM66D	23	0	90	0	113	9	0	GG	32	SNP Référence

Père (trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp	max_gt.	Q.max_gt.	statut.
USP17L2	0	0	11	0	11	1	0	GG	63	SNP Référence
FAM66D	0	0	11	0	11	1	0	GG	63	SNP Référence

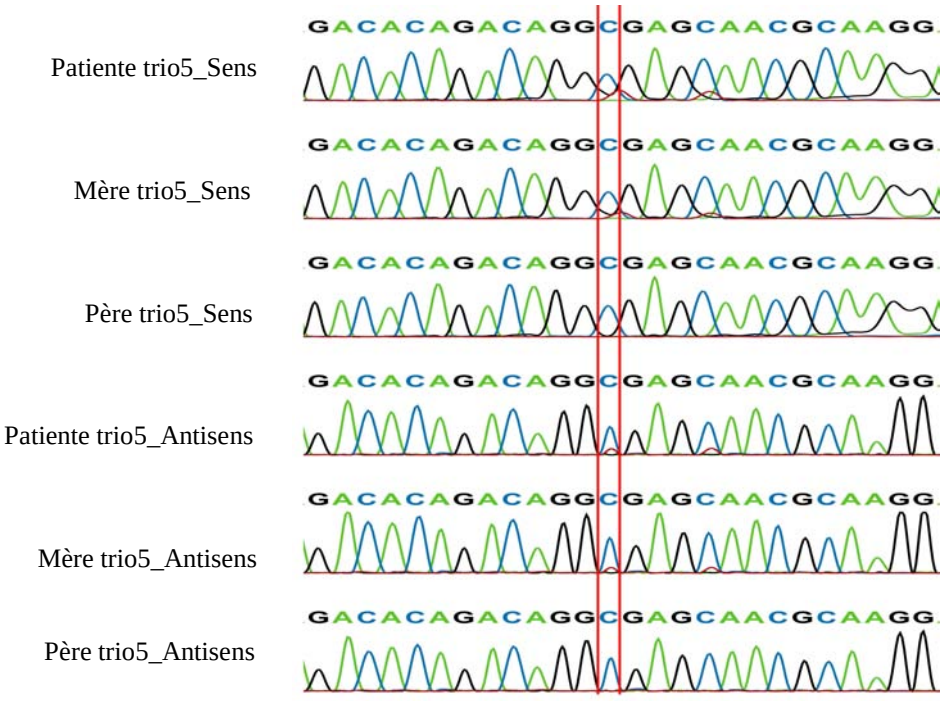


Figure 32 : Résultats du séquençage ciblé Sanger de mutation dans le gène *USP17L2*

Le séquençage ciblé Sanger révèle à la position étudiée le nucléotide G (C sur le sens de lecture ici) prédominant pour le trio et une possible hétérozygotie de très faible niveau au seuil de résolution, hérité de la mère, évocateur donc d'un polymorphisme.

3.3.3.2 Scénario 2 : Mutations de novo avec une profondeur supérieure à 20x

Les mutations *de novo* identifiée avec une profondeur de lecture supérieure à 20x et non décrites dans les bases de données de polymorphismes connus ont été analysées. Au total 56 mutations distinctes (62 lignes) sont détectées. Elles sont essentiellement localisées sur les autosomes, et seules 2 sont sur le chromosome X; elles correspondent à 56 mutations distinctes, 2 sont des mutations non sens, 54 sont des mutations faux sens.

Tableau 10 : Mutations *de novo*, avec une profondeur de lecture supérieure à 20x et non décrites dans les bases de données de polymorphismes connus

Gène	Chrom	Position	trio	Type de variation	Variation	Prof	A	C	G	T	Codon sauvage	Acide aminé sauvage	Codon muté	Acide aminé muté
FHAD1	chr1	15635220	1	missense	SNP	20	10	0	10	0	GCT	A	ACT	T
FAM176B	chr1	36787970	5	missense	SNP	22	0	9	13	0	GAC	D	CAC	H
NBPF9	chr1	144813815	5	missense	SNP	27	18	0	9	0	AGA	R	GGA	G
PDE4DIP	chr1	144930802	3	missense	SNP	245	0	0	65	180	ATC	I	CTC	L
NBPF10	chr1	145328389	5	missense	SNP	32	0	14	18	0	CAA	Q	GAA	E
NBPF10	chr1	145368460	2	missense	SNP	277	68	0	209	0	GAA	E	AAA	K
PRR9	chr1	153190780	3	missense	SNP	65	0	36	0	29	CTT	L	TTT	F
NFASC	chr1	204966474	3	missense	SNP	23	9	1	13	0	GCC	A	ACC	T
DIP2C	chr10	532440	4	missense	SNP	76	0	33	0	43	AAG	K	AGG	R
DUX4L2	chr10	135491044	4	missense	SNP	39	7	6	25	1	GCG	A	ACG	T
DUX4L3	chr10	135491044	4	missense	SNP	39	7	6	25	1	GCG	A	ACG	T
DUX4L5	chr10	135491044	4	missense	SNP	39	7	6	25	1	GCG	A	ACG	T
DUX4L6	chr10	135491044	4	missense	SNP	39	7	6	25	1	GCG	A	ACG	T
DUX4L2	chr10	135491056	4	missense	SNP	69	21	0	48	0	GGC	G	AGC	S
DUX4L3	chr10	135491056	4	missense	SNP	69	21	0	48	0	GGC	G	AGC	S
DUX4L5	chr10	135491056	4	missense	SNP	69	21	0	48	0	GGC	G	AGC	S
DUX4L6	chr10	135491056	4	missense	SNP	69	21	0	48	0	GGC	G	AGC	S
MUC5B	chr11	1265891	5	missense	SNP	35	0	22	0	13	GCC	A	GTC	V
MUC5B	chr11	1265951	5	missense	SNP	52	0	36	0	16	ACG	T	ATG	M
MRGPRG	chr11	3239878	5	missense	SNP	25	0	12	13	0	GCC	A	CCC	P
PGPEP1L	chr15	99511765	3	missense	SNP	104	0	56	0	48	AAG	K	AGG	R
KRTAP9-1	chr17	39346214	2	missense	SNP	73	35	38	0	0	CCC	P	ACC	T
DKFZP781G0119	chr18	72021088	3	missense	SNP	136	66	0	70	0	CGT	R	CAT	H
LOC390940	chr19	44085756	2	missense	SNP	135	69	0	66	0	GCC	A	ACC	T
KIR2DL1	chr19	55284821	3	missense	SNP	159	55	0	104	0	GGT	G	GAT	D
ZNF814	chr19	58385748	5	missense	SNP	61	24	0	37	0	GCT	A	GTT	V
LOC388946	chr2	46707595	5	missense	SNP	37	0	15	22	0	GAA	E	CAA	Q

Gène	Chrom	Position	trio	Type de variation	Variation	Prof	A	C	G	T	Codon sauvage	Acide aminé sauvage	Codon muté	Acide aminé muté
EML6	chr2	55096447	5	missense	SNP	101	48	0	0	53	GAT	D	GAA	E
CREG2	chr2	101971731	3	missense	SNP	25	0	0	16	9	CAA	Q	AAA	K
POTEE	chr2	131976262	4	missense	SNP	45	24	0	21	0	AAC	N	AGC	S
POTEE	chr2	131976262	5	missense	SNP	52	35	0	17	0	AAC	N	AGC	S
PYGB	chr20	25261661	4	missense	SNP	99	46	0	53	0	CGG	R	CAG	Q
TPTE	chr21	10921939	2	missense	SNP	96	0	74	0	22	GCA	A	ACA	T
KRTAP10-6	chr21	46012349	1	missense	SNP	69	35	0	34	0	ATG	M	ACG	T
WDR52	chr3	113023920	5	missense	SNP	90	0	47	0	43	AAT	N	AGT	S
WDR52	chr3	113063479	4	missense	SNP	66	36	29	0	1	GTG	V	GGG	G
MUC4	chr3	195506473	3	missense	SNP	36	17	0	19	0	GTA	V	GCA	A
MUC4	chr3	195506569	3	missense	SNP	52	27	0	25	0	GTA	V	GCA	A
MUC4	chr3	195508476	5	missense	SNP	51	1	16	34	0	CAC	H	CAG	Q
MUC4	chr3	195509077	4	missense	SNP	57	0	18	0	39	GAC	D	GGC	G
MUC4	chr3	195509212	3	missense	SNP	34	24	0	10	0	TCA	S	TTA	L
MUC4	chr3	195509381	3	missense	SNP	27	9	0	18	0	CCT	P	TCT	S
MUC4	chr3	195509560	5	missense	SNP	91	0	0	65	26	ACC	T	AAC	N
MUC4	chr3	195509566	3	missense	SNP	31	0	0	26	5	CCT	P	CAT	H
MUC4	chr3	195509861	5	missense	SNP	36	29	0	7	0	TCT	S	CCT	P
MUC4	chr3	195509918	3	missense	SNP	20	0	7	13	0	CAC	H	GAC	D
MUC4	chr3	195510107	4	missense	SNP	31	0	22	0	9	GCC	A	ACC	T
MUC4	chr3	195510767	3	missense	SNP	36	13	0	23	0	CCT	P	TCT	S
RNF175	chr4	154631526	5	nonsense	SNP	82	47	35	0	0	GAA	E	TAA	*
SORBS2	chr4	186544220	3	missense	SNP	80	0	0	35	45	CAC	H	CCC	P
ANKRD31	chr5	74491635	2	missense	SNP	200	0	110	0	90	GCA	A	ACA	T
PCDH12	chr5	141324954	4	missense	SNP	22	15	1	0	6	TGC	C	AGC	S
C6orf103	chr6	146965942	1	missense	SNP	30	0	0	10	20	TTA	L	GTA	V
FAM90A10	chr8	7629023	3	missense	SNP	22	11	11	0	0	AAC	N	ACC	T
USP17L2	chr8	11995066	5	nonsense	SNP	68	26	0	42	0	CGA	R	TGA	*
ZFHX4	chr8	77761278	5	missense	SNP	99	58	41	0	0	ACC	T	CCC	P
ANKRD18A	chr9	38593830	2	missense	SNP	58	0	26	0	32	GGT	G	GAT	D
TBC1D2	chr9	100962566	2	missense	SNP	23	12	0	11	0	CGC	R	TGC	C
C9orf169	chr9	140120188	5	missense	SNP	29	14	0	0	15	TCC	S	ACC	T
RBMXL3	chrX	114425895	1	missense	SNP	45	0	22	23	0	GAG	E	CAG	Q
GABRE	chrX	151123525	3	missense	SNP	29	1	0	14	14	ACC	T	AAC	N

De gauche à droite : le nom du gène, le chromosome, les coordonnées génomiques du variant sur le chromosome, le numéro du trio, le type de variation (faux sens ou non sens) et la nature SNP (à l'opposé des indel), la profondeur de lecture, les 4 nucléotides (A,C,G,T) et les proportions de variant, le codon sauvage et son acide aminé, le codon muté générant un substitution d'acide aminé.

3.3.3.2.1 Les mutations non-sens

Les 2 mutations non-sens dans les gènes, *RNF175* et *USP17L2* ont déjà été abordées dans le scénario 1.

3.3.3.2.2 Les mutations faux-sens

Deux de ces mutations sont localisées **sur le chromosome X** ; elles sont contrôlées en priorité.

Tableau 11 : Les deux gènes mutés localisés sur le chromosome X chez les deux patientes (trio 1 et 3) (Extrait du tableau 10)

Gène	Chr	Position	Trio	Type de variation	Variation	prof	A	C	G	T	Codon sauvage	Acide aminé sauvage	Codon muté	Acide aminé muté
RBMXL3	chrX	114425895	1	missense	SNP	45	0	22	23	0	GAG	E	CAG	Q
GABRE	chrX	151123525	3	missense	SNP	29	1	0	14	14	ACC	T	AAC	N

Une mutation a été détectée dans le **gène *RBMXL3*** (trio1). La deuxième mutation a été détectée dans le **gène *GABRE*** (trio3).

RBMXL3 (RNA binding motif protein, X-linked-like 3) (NM_001145346, exon 1 sur 1)

Ce gène est exprimé dans le cerveau (cortex et le cervelet), la moelle épinière et également dans la rétine (Figure 33).

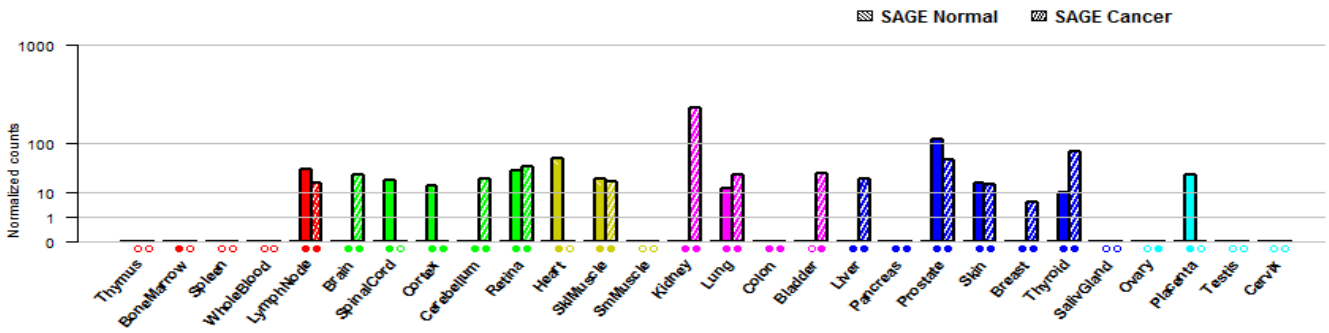


Figure 33 : Expression de gène *RBMXL3* dans les tissues neuronale, la moelle épinière et la rétine (en vert) selon la base de donnée Gene Card®

- **Trio 1**

Dans ce gène, la mutation faux sens conduit à la substitution d'un acide aminé (acide glutamique à glutamine). Pour cette variation, la couverture est incomplète pour l'ADN maternel (NC). Le contrôle par séquençage ciblé Sanger montre que la variation est héritée de la mère.

Tableau 12 : La mutation dans le gène *RBMXL3* pour le trio 1

Patiente (trio 1)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt	Q.snp.	max_gt	Q.max_gt.	statut
RBMXL3	0	22	23	0	45	3	201	CG	201	SNP Hétérozygote

Mère (trio 1)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt	Q.snp	max_gt	Q.max_gt	statut.
RBMXL3	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC

Père (trio 1)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt	Q.max_gt	statut.
RBMXL3	0	0	15	0	15	2	0	GG	75	SNP Référence

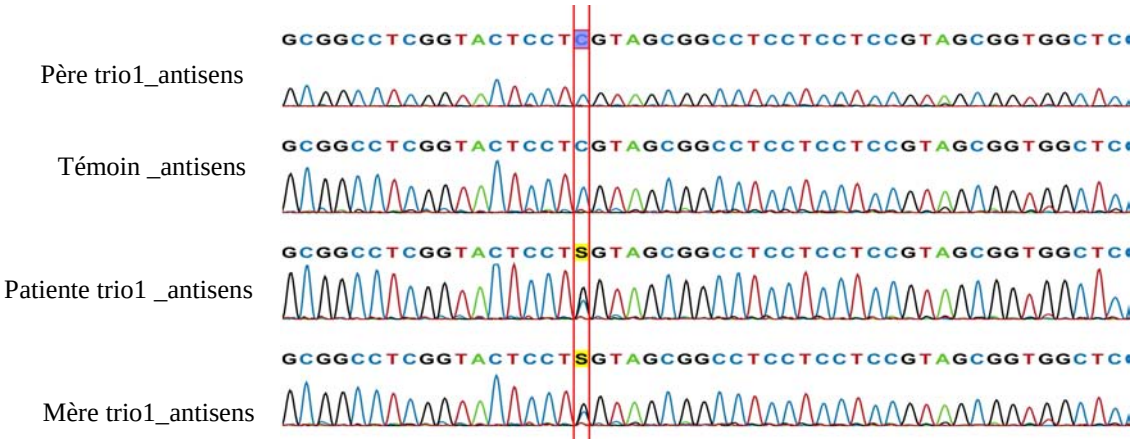


Figure 34 : Profil du séquençage ciblé Sanger a montré la varaition hérité de la mère détecté dans le gène *RBMXL3*

GABRE (gamma-aminobutyric acid (GABA) A receptor, epsilon)
(NM_004961, exon 9 sur 9)

Le gène *GABRE* fait partie de la famille du récepteur du gène GABA-A, principal inhibiteur de neurotransmetteur dans le cerveau chez les mammifères (Figure 35).

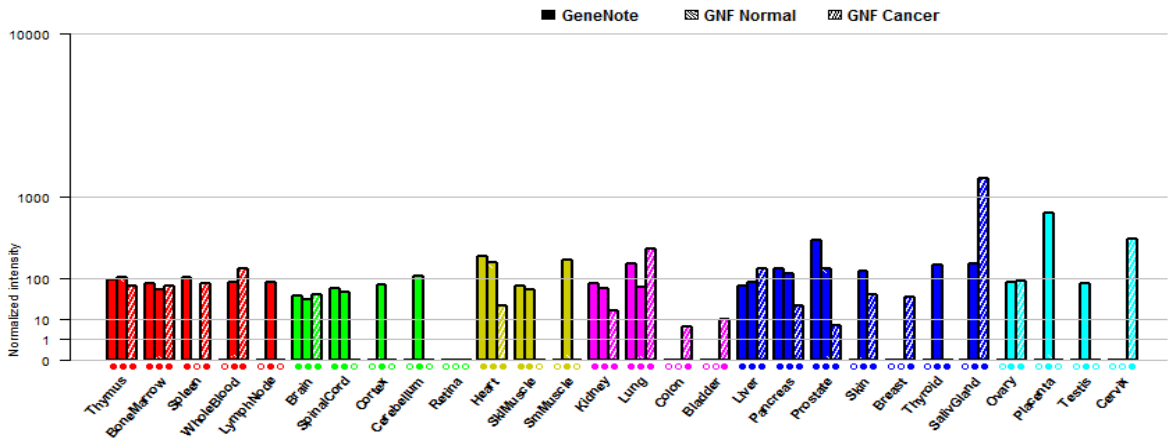


Figure 35 : Expression du gène *GABRE* selon la base de donnée Gene Card®
Expression dans le cerveau (en vert)

- **Trio 3**

La patiente du trio3 possède une variation de séquence sur un allèle GG >GT, l'acide aminé Thréonine est substitué en Asparagine ; sa mère montre un SNP hétérozygote douteux «SNP_het.douteux»; le contrôle par séquençage Sanger détecte bien le SNP chez la mère

Tableau 13 : La variation dans le gène *GABRE* chez le trio 3

Patiente (trio 3)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt	Q.max_gt.	statut.
GABRE	1	0	14	14	29	20	134	GT	134	SNP Hétérozygote

Mère (trio 3)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt	Q.snp	max_gt.	Q.max_gt	statut.
GABRE	0	0	24	6	30	10	15	GT	15	SNP Hétérozygote douteux

Père (trio 3)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
GABRE	0	0	4	0	4	3	0	GG	42	SNP Référence

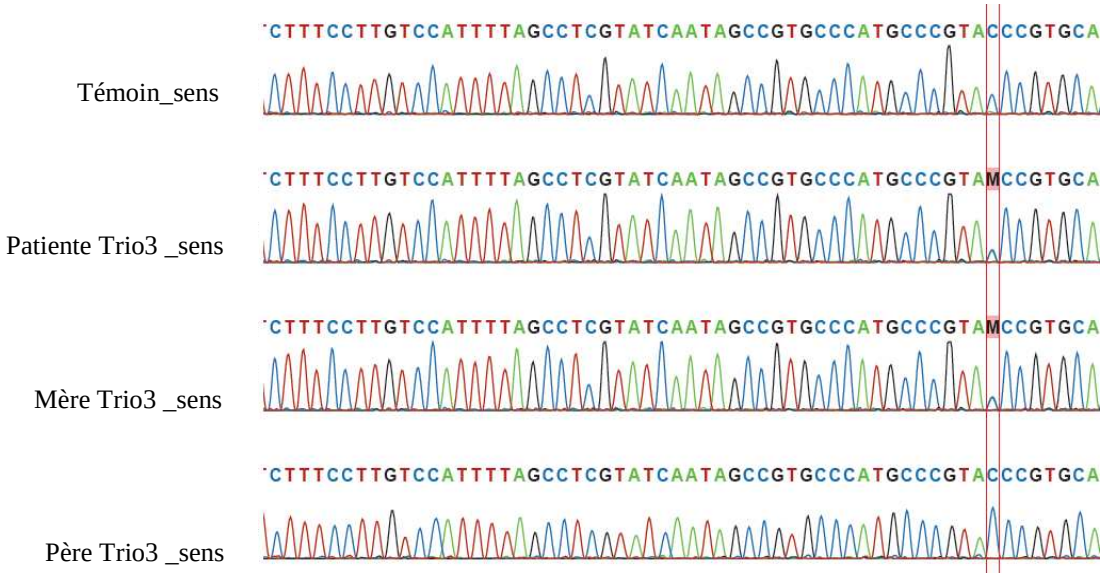


Figure 36 : Le profil du séquençage ciblé Sanger pour la variation dans le gène *GABRE*

Le séquençage Sanger a montré la variation hérité de la mère.

3.3.6.2.3 Gènes mutés chez au moins 2 patientes

Trois gènes sont mutés chez au moins deux patientes : *NBPF10*, *POTEE* et *WDR52*

NBPF10 (neuroblastoma breakpoint family, member 10) (NM_001039703, exon 33 et 84 sur 84)

Le transcrit de ce gène a été détecté dans les plusieurs tissus humains notamment dans le cerveau, la moelle épinière, le cortex et le cervelet (en vert dans la Figure 37).

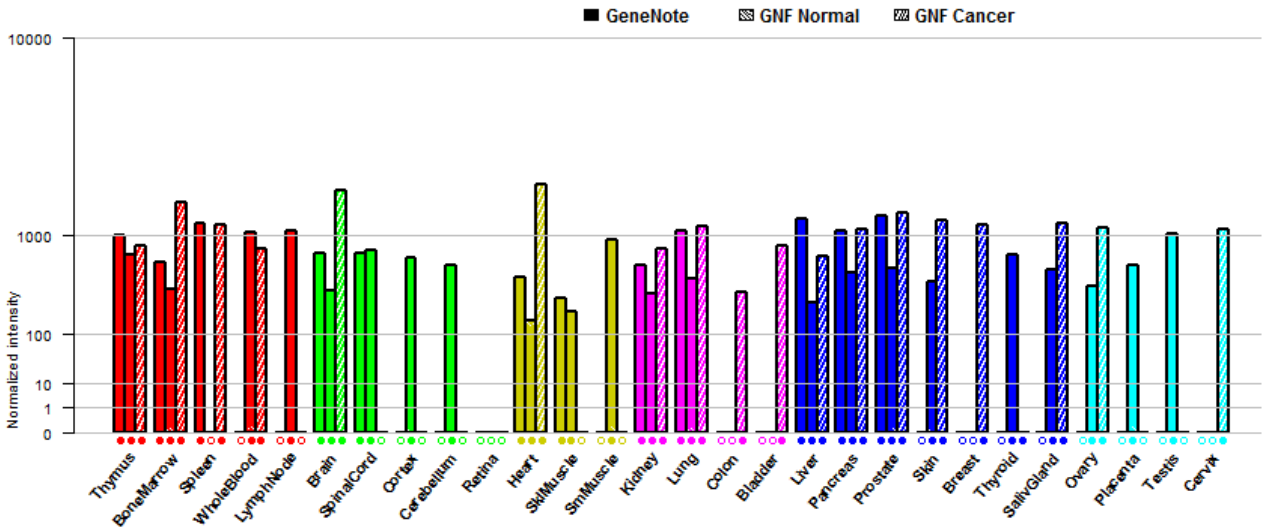


Figure 37 : Expression du gène *NBPF10* selon la base de donnée Gene Card®

NBPF10 fait partie de la famille des gènes *NBPF*, qui se compose de 22 gènes et des pseudogènes. *Les* gènes *NBPF* contiennent de nombreux éléments répétitifs et présentent une forte identité de séquence intergénique et intragénique dans des régions codante et non codante [172].

Cette mutation a été détectée dans deux familles. Ces deux mutations ne sont pas contrôlées par séquençage Sanger car une séquence répétée empêche le choix des amorces.

Tableau 14 : La mutation *de novo* du gène *NBPF10* mutés chez 2 patientes (trio 5) et (trio 2)

Gène	Chr	Position	Trio	Type de variation	Variation	Prof	A	C	G	T	Codon sauvage	Acide aminé sauvage	Codon muté	Acide aminé muté
NBPF10	chr1	145328389	5	missense	SNP	32	0	14	18	0	CAA	Q	GAA	E
NBPF10	chr1	145368460	2	missense	SNP	277	68	0	209	0	GAA	E	AAA	K

- **Trio 5**

Chez la patiente du trio 5, la séquence est hétérozygote CG alors que la mère est homozygote CC par contre la séquence du père est non couverte pour cette mutation. La glutamine est substituée par l'acide glutamique.

Tableau 15 : La mutation dans le gène *NBPF10* chez le trio 5

Patiente (Trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt	statut
NBPF10	0	14	18	0	32	3	123	CG	123	SNP Hétérozygote

Mère (Trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt	Q.max_gt.	statut.
NBPF10	0	58	14	0	72	10	0	CC	49	SNP Référence

Père (Trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
NBPF10	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC

- **Trio 2**

Chez la patiente du trio 2, dans le gène *NBPF10* à la position 145368460, la séquence est AG, avec substitution acide glutamique à lysine et l'annotation par le logiciel d'analyse est homozygote GG chez ses parents.

Tableau 16 : La mutation dans le gène *NBPF10* chez le trio 2

Patiente (Trio 2)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt	Q.snp.	max_gt	Q.max_gt.6	statut
NBPF10	68	0	209	0	277	9	24	AG	24	SNP Hétérozygote

Mère (Trio 2)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt	Q.snp.	max_gt.	Q.max_gt.	statut.
NBPF10	34	0	123	0	157	10	0	GG	26	SNP Référence

Père (Trio 2)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp	max_gt.	Q.max_gt.	statut.
NBPF10	0	0	112	0	112	11	0	GG	367	SNP Référence

POTEE (Prostate, ovary, testis, and placenta expressed gene ankyrin domain family, member E) (NM_001083538, exon 1 sur 15)

Présence d'une variation à la même position pour les deux trios. A nouveau, cette mutation n'a pu être contrôlée par séquençage Sanger du fait de séquence répétée**.** En plus les variations sont douteuses chez les parents des deux trios.

Tableau 17 : La mutation *de novo* du gène *POTEE* chez les patientes des trio 4 et trio 5

Gène	Chr	Position	Trio	Type de variation	Variation	Prof	A	C	G	T	Codon sauvage	Acide aminé sauvage	Codon muté	Acide aminé muté
POTEE	chr2	131976262	4	missense	SNP	45	24	0	21	0	AAC	N	AGC	S
POTEE	chr2	131976262	5	missense	SNP	52	35	0	17	0	AAC	N	AGC	S

- **Trio 4**

Chez la patiente du trio 4, la séquence est hétérozygote AG, substitution asparagine en sérine, et ses parents sont notés homozygotes AA, mais le statut de la mère est «SNP_référence.douteux » donc la probabilité de trouver le génotype homozygote « AA » est faible : le nucléotide « A » est détecté 24 fois et le nucléotide « G » est détecté 8 fois, le logiciel annote homozygote « AA ».

Nous concluons que cette variation est un polymorphisme.

Tableau 18 : La mutation dans le gène *POTEE* pour le trio4

Patiente (Trio 4)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
POTEE	24	0	21	0	45	10	90	AG	90	SNP Hétérozygote

Mère (Trio 4)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut
POTEE	24	0	8	0	32	8	1	AA	8	SNP Référence douteux

Père (Trio 4)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt	statut.
POTEE	35	0	12	0	47	14	0	AA	10	SNP Référence

- **Trio 5**

Même résultats chez la patiente du trio 5 : la séquence est hétérozygote AG, substitution d'asparagine à sérine, et ses parents sont homozygotes AA

Tableau 19 : La mutation dans le gène *POTEE* pour le trio5

Patiente (Trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp	max_gt.	Q.max_gt..	statut.
POTEE	35	0	17	0	52	13	20	AG	20	SNP Hétérozygote

Mère (Trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt	Q.max_gt.	statut.
POTEE	42	0	8	0	50	16	0	AA	64	SNP Référence

Père (Trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt	Q.max_gt..	statut
POTEE	42	0	14	0	56	4	0	AA	11	SNP Référence

WDR52 (WD repeat domain 52) (NM_001164496, exon 31 et 23 sur 35)

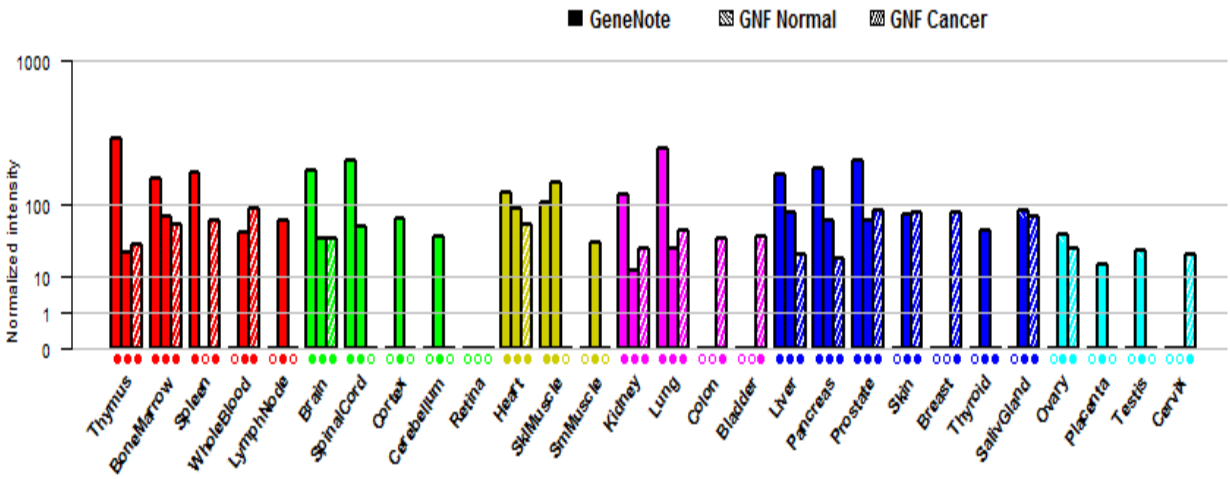


Figure 38 : Expression du gène *WDR52* selon la base de données Gene Card®

L'expression de ce gène a bien été détecté dans plusieurs tissues humaine.

Tableau 20 : La mutation *de novo* du gène *WDR52* chez les patientes des trio 5 et trio 4

Gène	Chr	Position	Trio	Type de variation	Variation	Prof	A	C	G	T	Codon sauvage	Acide aminé sauvage	Codon muté	Acide aminé muté
WDR52	chr3	113023920	5	missense	SNP	90	0	47	0	43	AAT	N	AGT	S
WDR52	chr3	113063479	4	missense	SNP	66	36	29	0	1	GTG	V	GGG	G

- **Trio 5**

Chez la patiente du trio 5, la séquence est hétérozygote CT, substitution Asparagine en Sérine, la mère est homozygote TT, par contre la séquence du père est non couverte pour cette mutation. Le séquençage Sanger détecte cette mutation comme héritée du père.

Tableau 21 : La mutation dans le gène *WDR52* pour le trio 5

Patiente (Trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
WDR52	0	47	0	43	90	3	334	CT	331	SNP Hétérozygote

Mère (Trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt	Q.max_gt.	statut.
WDR52	0	0	0	83	83	7	0	TT	280	SNP Référence

Père (Trio 5)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt	statut.
WDR52	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC

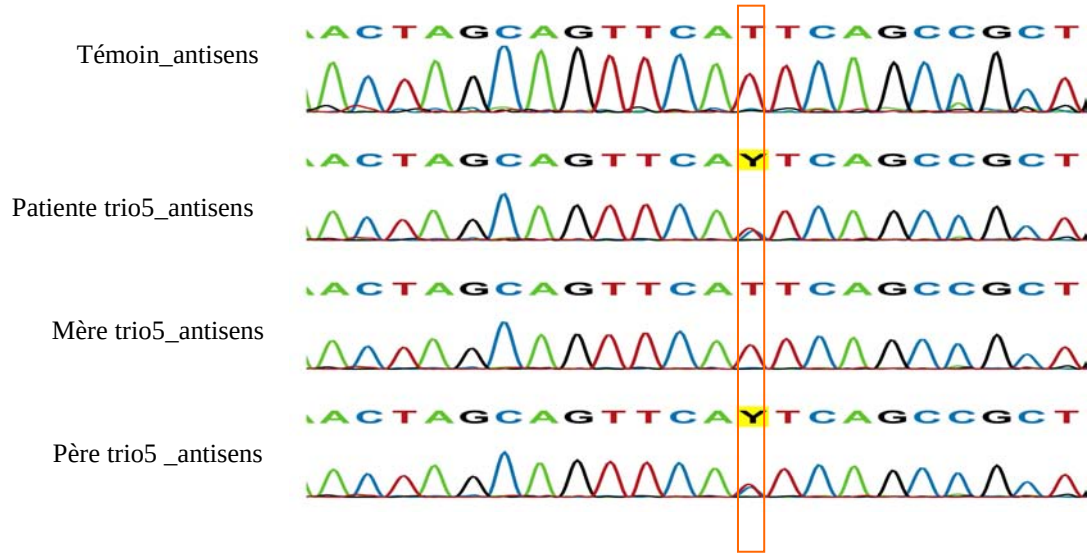


Figure 39 : Profil du séquençage Sanger pour la mutation du gène *WDR52* chez la patiente trio 5 et ses parents

Cette variation faux sens dans le gène *WDR52* a été transmis par le père à sa fille.

- **Trio 4**

Chez la patiente du trio 4, la séquence est hétérozygote AC, substitution de valine en glycine, la mère est indiquée homozygote AA par le logiciel, et le père hétérozygote AC classé douteux « SNP hétérozygote douteux».

Tableau 22 : La mutation dans le gène *WDR52* pour le trio 4

Patiente (Trio 4)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
WDR52	36	29	0	1	66	58	69	AC	69	SNP Hétérozygote

Mère (Trio 4)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp	max_gt.	Q.max_gt.	statut.
WDR52	61	30	0	0	91	59	0	AA	16	SNP Référence

Père (Trio 4)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
WDR52	51	29	0	0	80	69	10	AC	10	SNP Hétérozygote douteux

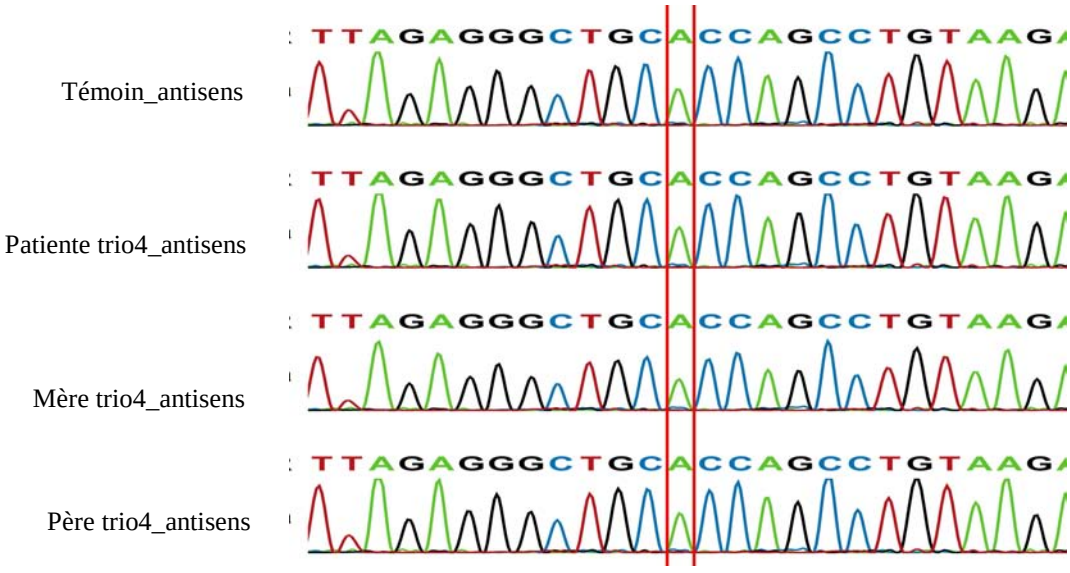


Figure 40 : Un profil « sauvage » et non muté par séquençage Sanger pour la variation du gène *WDR52* chez le trio 4

Le résultat du séquençage Sanger détecte un **faux positif** pour cette mutation, chez le père comme chez la fille

3.3.3.3 Scénario 3 : Mutations de novo quelle que soit la profondeur

Nous avons ensuite relevé toutes les mutations *de novo* avec une profondeur de lecture de plus de 1x et non décrites dans dbSNP, HapMap, 1000 Genomes et dans la base Integragen. Un même gène avec une variation chez au moins 2 patientes a été observé pour 13 gènes différents

Tableau 23 : Mutations *de novo* du même gène, quelque soit la profondeur, chez au moins 2 patientes

Gène	Chromosome	Nombre de Patientes
MUC4	chr3	3
NBPF10	chr1	3
CBX7	chr22	2
FRG2B	chr10	2
MAFG	chr17	2
MICAL3	chr22	2
MRGPRG	chr11	2
MUC12	chr7	2
POTEE	chr2	2
PPARGC1B	chr5	2
ROGDI	chr16	2
SLC15A3	chr11	2
WDR52	chr3	2

Tableau 24 : Détail de la profondeur de lecture et du type de mutation chez les patientes, pour les 13 gènes mutés chez au moins 2 patientes

Gène	Chr	Position	Nombre de Patientes	Trio	Prof	Ref Seq	A	C	G	T	Type de variation
MUC4	chr3	195 506 473	3	3	36	A	17	0	19	0	missense
MUC4	chr3	195 506 569	3	3	52	A	27	0	25	0	missense
MUC4	chr3	195 507 226	3	3	116	A	51	0	65	0	missense
MUC4	chr3	195 509 212	3	3	34	G	24	0	10	0	missense
MUC4	chr3	195 509 381	3	3	27	G	9	0	18	0	missense
MUC4	chr3	195 509 566	3	3	31	G	0	0	26	5	missense
MUC4	chr3	195 509 918	3	3	20	G	0	7	13	0	missense
MUC4	chr3	195 510 059	3	3	10	T	0	5	0	5	missense
MUC4	chr3	195 510 156	3	3	17	G	0	4	13	0	missense
MUC4	chr3	195 510 767	3	3	36	G	13	0	23	0	missense
MUC4	chr3	195 514 489	3	3	10	G	8	0	2	0	missense
MUC4	chr3	195 506 914	3	4	63	G	18	0	45	0	missense
MUC4	chr3	195 506 917	3	4	61	A	47	14	0	0	missense
MUC4	chr3	195 509 077	3	4	57	T	0	18	0	39	missense
MUC4	chr3	195 510 107	3	4	31	C	0	22	0	9	missense
MUC4	chr3	195 508 476	3	5	51	G	1	16	34	0	missense

Gène	Chr	Position	Nombre de Patientes	Trio	Prof	Ref Seq	A	C	G	T	Type de variation
MUC4	chr3	195 509 560	3	5	91	G	0	0	65	26	missense
MUC4	chr3	195 509 861	3	5	36	A	29	0	7	0	missense
MUC4	chr3	195 510 341	3	5	12	A	2	0	10	0	missense
NBPF10	chr1	145 368 460	3	2	277	G	68	0	209	0	missense
NBPF10	chr1	145 328 389	3	4	14	C	0	8	6	0	missense
NBPF10	chr1	145 326 106	3	5	6	A	1	0	0	5	missense
NBPF10	chr1	145 328 389	3	5	32	C	0	14	18	0	missense
CBX7	chr22	39 530 676	2	2	5	C	2	3	0	0	missense
CBX7	chr22	39 530 663	2	3	4	G	0	0	2	2	missense
FRG2B	chr10	135 440 122	2	3	3	T	0	2	0	1	missense
FRG2B	chr10	135 440 122	2	5	4	T	0	3	0	1	missense
MAFG	chr17	79 880 485	2	1	8	G	0	0	6	2	nonsense
MAFG	chr17	79 880 564	2	4	6	C	2	4	0	0	missense
MICAL3	chr22	18 314 737	2	1	6	C	2	4	0	0	nonsense
MICAL3	chr22	18 291 689	2	4	8	C	2	6	0	0	missense
MRGPRG	chr11	3 239 310	2	3	6	G	0	0	4	2	missense
MRGPRG	chr11	3 239 878	2	5	25	C	0	12	13	0	missense
MUC12	chr7	100 639 197	2	3	4	C	3	1	0	0	missense
MUC12	chr7	100 642 455	2	5	12	G	5	0	7	0	missense
POTEE	chr2	131 976 262	2	4	45	A	24	0	21	0	missense
POTEE	chr2	131 976 262	2	5	52	A	35	0	17	0	missense
PPARGC1B	chr5	149 109 975	2	1	8	G	3	0	5	0	missense
PPARGC1B	chr5	149 213 075	2	3	14	G	7	0	7	0	missense
ROGDI	chr16	4 847 976	2	2	3	C	2	1	0	0	missense
ROGDI	chr16	4 847 480	2	4	9	C	2	7	0	0	missense
SLC15A3	chr11	60 718 792	2	3	5	C	0	3	0	2	missense
SLC15A3	chr11	60 718 792	2	5	6	C	0	1	0	5	missense
WDR52	chr3	113 063 479	2	4	66	A	36	29	0	1	missense
WDR52	chr3	113 023 920	2	5	90	T	0	47	0	43	missense

De gauche à droite : le nom du gène, le chromosome, les coordonnées génomiques du variant sur le chromosome,le nombre de patientes, le numéro du trio, la profondeur de lecture,la séquence de la référence, les 4 nucléotides et les proportions de variant et le type de variation (faux sens)

NBPF10 : (NM_001039703, exon 33 et 84 sur 84), ***POTEE :*** (NM_001083538, exon 1 sur 15) ont été déjà abordés dans le scénario 2

MRGPRG (MAS-related GPR, member G): (NM_003816037, exon 1 sur 1)

Dong et al., en 2001 [173] a identifié la famille de protéine G couplée à des récepteurs (GPCRs). Les récepteurs sont conservés chez la souris et l'homme et corrélés avec le gène *MRGPRG*. L'expression des gènes *MRGs* ont détecté dans les neurones sensoriels.

Tableau 25 : La mutation *de novo* du gène *MRGPRG* chez les patientes des trio 5 et trio 3 (Extrait du tableau 24)

Gène	Chr	Position	Nombre de Patientes	Trio	Prof	Ref Seq	A	C	G	T	Type de variation
MRGPRG	chr11	3 239 878	2	5	25	C	0	12	13	0	Missense
MRGPRG	chr11	3 239 310	2	3	6	G	0	0	4	2	missense

Une variation est détectée chez les patientes des trio3 et trio5

- **Trio 5**

Chez le trio 5 la patiente possède une variation sur un allèle C>G

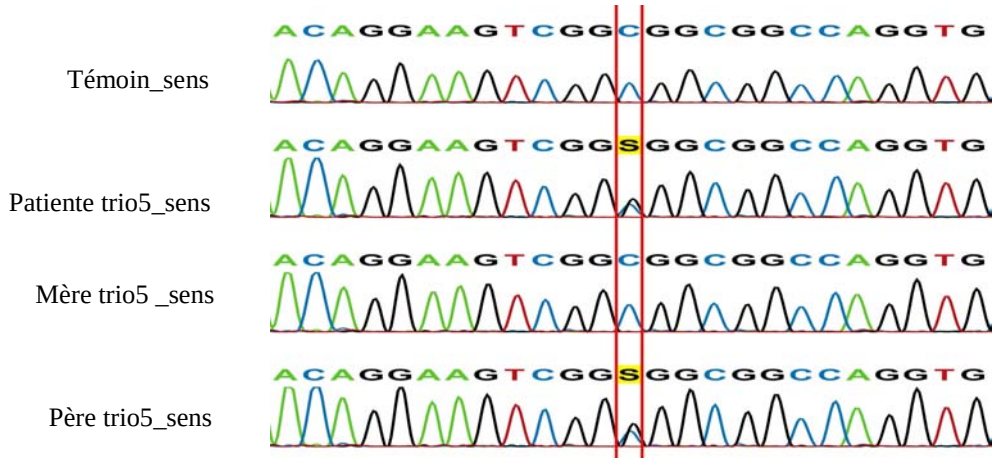


Figure 41 : Profil du séquençage Sanger pour la variation dans le gène *MRGPRG* pour le trio5

Après le séquençage Sanger la variation a été détecté d'un polymorphisme transmis par le père à sa fille.

- **Trio 3**

Chez le trio 3 le séquençage de l'exome a noté la patiente hétérozygote

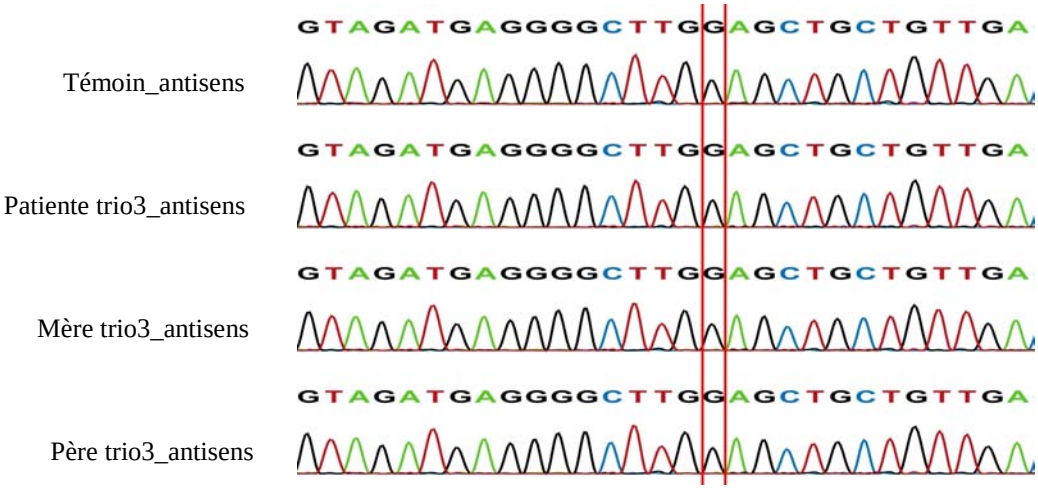


Figure 42 : Profil du séquençage Sanger pour la variation dans le gène *MRGPRG* pour le trio 3

Le séquençage Sanger a détecté un faux positif pour la variation du trio 3 dans le gène *MRGPRG*.

PPARGC1B (peroxisome proliferator-activated receptor gamma, coactivator 1 beta) (NM_ 003829037 exon1 et 5 sur 12)

Ce gène s'exprime dans les tissus neuronaux, le cœur et le muscle (Figure 43). Deux variations différentes ont été relevées chez deux patientes différentes.

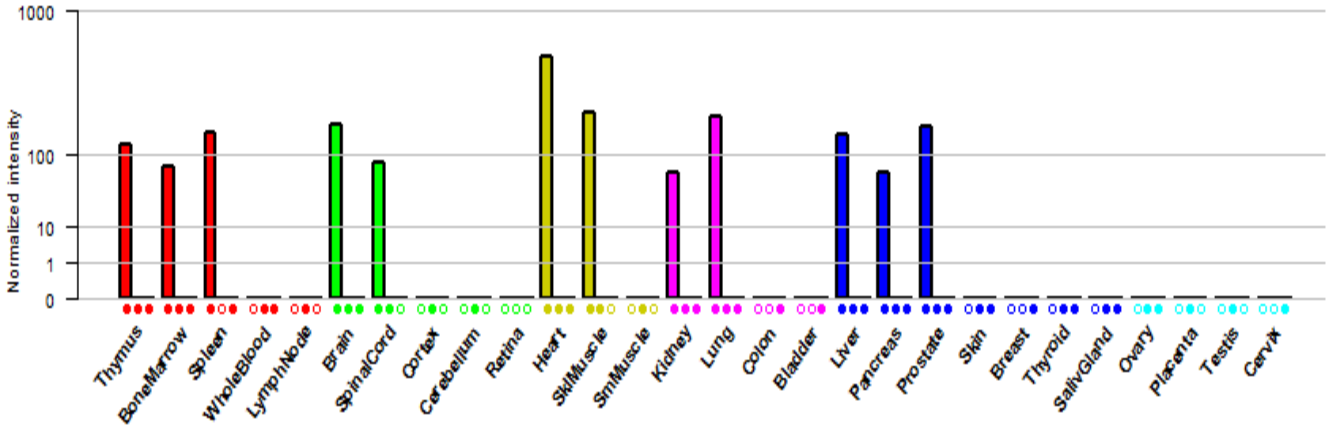


Figure 43 : Expression du gène *PPARGC1B* selon la base de données Gene Card®

Tableau 26 : La mutation du gène *PPARGC1B* chez les patientes des trio 1 et trio 3

Gène	Chr	Position	Nombre de Patientes	Trio	Profondeur	Ref Seq	A	C	G	T	Type de variation
PPARGC1B	chr5	149 109 975	2	1	8	G	3	0	5	0	missense
PPARGC1B	chr5	149 213 075	2	3	14	G	7	0	7	0	missense

- **Trio 1**

Chez la patiente du trio1, la mutation faux sens s'est avérée être, après vérification par séquençage Sanger, **un faux positif**.

Tableau 27 : La mutation dans le gène *PPARGC1B* pour le trio 1

Patiente (Trio 1)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt	Q.snp.	max_gt.	Q.max_gt	statut.
PPARGC1B	3	0	5	0	8	0	45	AG	45	SNP Hétérozygote

Père (Trio 1)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
PPARGC1B	0	0	4	0	4	0	0	GG	42	SNP Référence

Mère (Trio 1)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
PPARGC1B	2	0	1	0	3	0	5	AG	5	SNP Hétérozygote douteux

- **Trio 3**

Tableau 28 : Le détail de la mutation dans le gène *PPARGC1B* pour le trio 3

Patiente (Trio 3)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt.	statut.
PPARGC1B	7	0	7	0	14	0	86	AG	86	SNP Hétérozygote

Mère (Trio 3)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp	max_gt.	Q.max_gt	statut.
PPARGC1B	2	0	8	0	10	0	5	AG	5	SNP Hétérozygote douteux

Père (Trio 3)

Gène	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Profondeur	filt.	Q.snp.	max_gt.	Q.max_gt	statut.
PPARGC1B	0	0	6	0	6	0	0	GG	48	SNP référence

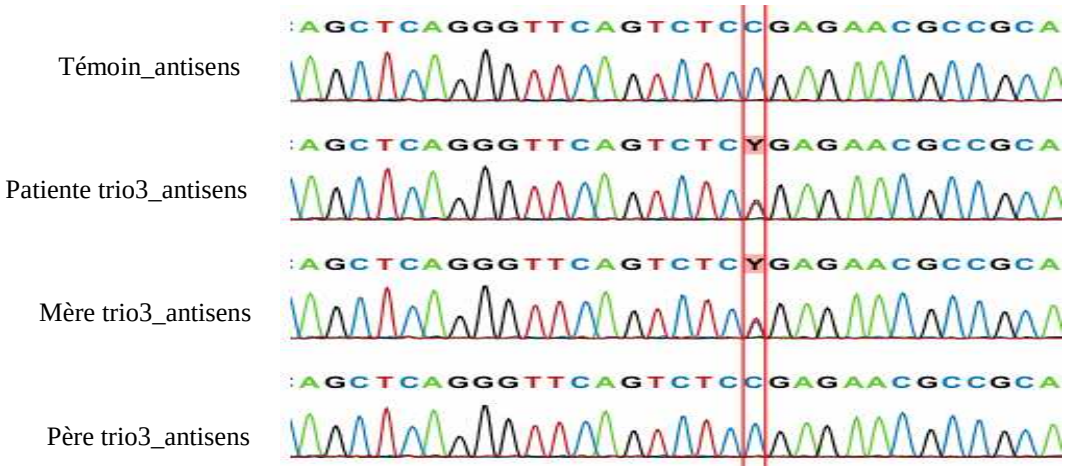


Figure 44 : Profil du séquençage Sanger pour la variation du gène *PPARGC1B* chez le trio 3

Il révèle qu'il s'agit d**'un polymorphisme hérité de la mère**

WDR52 : (NM_001164496, exon 31 et 23 sur 35)

Ce gène est déjà contrôlé dans le scénario 2.

3.3.3.4 Scénario 4 : tous types de mutation (SNP et indel) dans des gènes chez plusieurs patientes

Nous avons pris en compte tous les gènes mutés chez plusieurs patientes avec différents types de mutation, en posant l'hypothèse que le gène responsable de la maladie peut subir différents types de mutation chez différentes patientes. Par exemple, la patiente « A » peut avoir une mutation ponctuelle et la patiente « B » une insertion délétion dans le même gène.

3.3.6.4.1 Gènes mutés chez 5 patientes

Nous avons trouvé avec cette analyse 16 gènes mutés chez cinq filles.

Tableau 29 : Tous types de mutation chez 5 patientes

Gène	Chromosome	Nombre de Patientes
ABCA7	chr19	5
BAIAP3	chr16	5
BCLAF1	chr6	5
CACNA1B	chr9	5
CDC27	chr17	5
CTBP2	chr10	5
DOCK11	chrX	5
FAM104B	chrX	5
FGD6	chr12	5
HACL1	chr3	5
HYDIN	chr16	5
PTPRD	chr9	5
RFX4	chr12	5
RYR2	chr1	5
SETX	chr9	5
SNCB	chr5	5

Les deux gènes testés en priorité sont ceux localisés sur le chromosome X.

DOCK11 (Dedicator of cytokinesis 11) (NM_003812637 exon33 et intron 14 sur 53)

Deux mutations synonymes trouvées chez la patiente du trio1 et la patiente du trio 2 sur la même position. La troisième mutation était hérite du père chez la patiente du trio 4, résultats non en faveur de l'implication de ce gène dans le SAi.

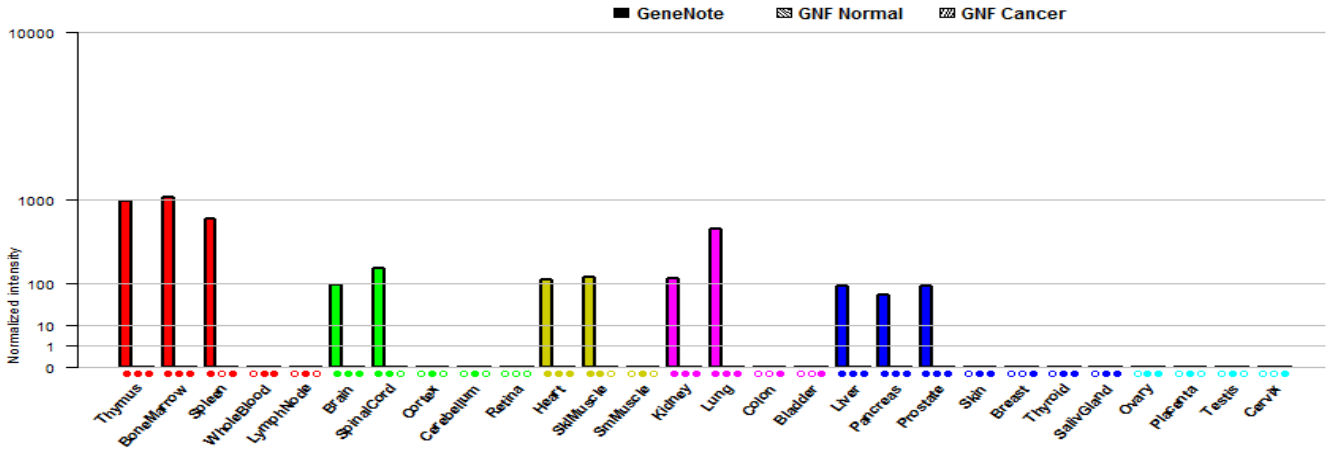


Figure 45 : Expression du gene *DOCK11* selon la base de données GeneCard®

Le transcrit de ce gène a été détecté dans le cerveau et la moelle épinière avec une expression modérée (en vert).

Tableau 30 : Les variations du gène *DOCK11* chez 3 patientes

Gène	chr	position	trio	statut	Type annot	position	Transmission	Ref seq	Prof	A	C	G	T	Codon sauvage	Acide aminé sauvage	Codon muté	Acide aminé muté	Freq hetz exomes	Freq homozygote exomes
DOCK11	chrX	117761502	1	SNP Hetz	synonymous	Exon 33	-	A	163	102	61	0	0	TCA	S	TCC	S	3.88	0.00
DOCK11	chrX	117761502	2	SNP Hetz	synonymous	Exon 33	-	A	140	83	57	0	0	TCA	S	TCC	S	3.88	0.00
DOCK11	chrX	117714175	4	SNP Hetz	intron	Intron14	père	C	26	0	15	0	11					0.00	0.00

De gauche à droite : le nom du gène, le chromosome, les coordonnées génomiques du variant sur le chromosome, le numéro du trio, la statut de la variation, l'annotation du variant, position dans le gene, l'heritage, la séquence de la référence, la profondeur de lecture, les 4 nucléotides et les proportions de variant, le codon sauvage et son acide aminé, le codon muté et son acide aminé, la fréquence du variant avec le statut heterozygote et homozygote (non référence) dans les x exomes déjà réalisés chez integraGen

Dans les deux autres trios (trios 3 et 5) toutes les variations détectées sont déjà décrites dans la base de données db SNP (build 137).

FAM104B (family with sequence similarity 104, member B) (NM_001166699)

Ce gène, situé en Xp11.22, contient 4 exons et possède 8 transcrits possibles qui codent pour une protéine de 115 à 116 acides aminés. Une variation de gène est observée chez les 5 patientes. Le profil d'expression a démontré le transcrit s'exprime bien dans le cerveau et la moelle épinière (Figure 46)

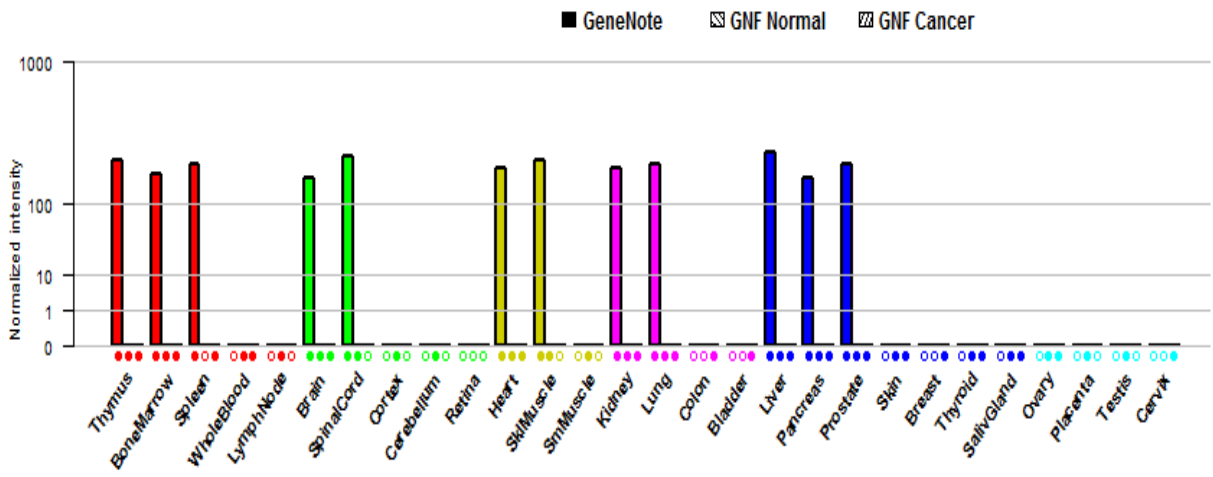


Figure 46 : Expression du gène *FAM104B* selon GeneCard®

Le gène *FAM104B* s'exprime dans le cerveau, la moelle épinière, le cœur, les reins et le foie.

Le type de variation principale est une insertion en fin d'exon 2. La position où cette insertion a lieu est la même pour toutes les patientes. Toutefois, pour certains trios, cette insertion est héritée d'un parent et fait particulier, l'insertion transmise a une situation un peu différente entre le parent et la patiente. Nous avons effectué le contrôle par séquençage Sanger chez l'ensemble des 5 patientes et leurs parents et n'avons trouvé aucune insertion de séquences dans l'exon 2 et en amont des exons 2 et 3 chez toutes les patientes ni chez les parents. Nous concluons que cette insertion est un faux positif du séquençage d'exome en rapport avec une erreur d'alignement par le logiciel.

3.3.6.4.2 Gènes mutés chez 4 patientes

Parmi les 31 gènes, un seul gène est localisé sur le chromosome X : ***ELF4***

Tableau 31 : Tout types de mutations des gènes mutés chez 4 patientes

Gène	Chromosome	Nombre de Patientes
AHNAK	chr11	4
AHNAK2	chr14	4
AKD1	chr6	4
ANKRD31	chr5	4
BTBD7	chr14	4
C11orf54	chr11	4
C12orf48	chr12	4
CCT4	chr2	4
CDK11A	chr1	4
CDK11B	chr1	4
CDON	chr11	4
DNAH10	chr12	4
ELF4	chrX	4 (indel)
EPPK1	chr8	4
FAT1	chr4	4
GPR98	chr5	4
LOC642236	chr9	4
LRP2	chr2	4
MGST3	chr1	4
MIR548N	chr2	4
MRS2	chr6	4
MUC13	chr3	4
MUL1	chr1	4
NBPF10	chr1	4
PDE4D	chr5	4
RIMS2	chr8	4
RNF20	chr9	4
ROGDI	chr16	4
SRPK2	chr7	4
TBX6	chr16	4
TXNDC2	chr18	4

ELF4 : (E74-like factor 4) (NM_003818938)

ELF4 fait partie de la famille des gènes *ETS* (facteur de transcription) [174]. C'est un activateur transcriptionnel qui se lie et active les promoteurs des gènes *CSF2, IL3, l'IL8*, et *PRF1*. La protéine codée est impliquée dans le développement des cellules immunitaires et l'induction de l'arrêt du cycle cellulaire dans les cellules CD8 + natives [175].

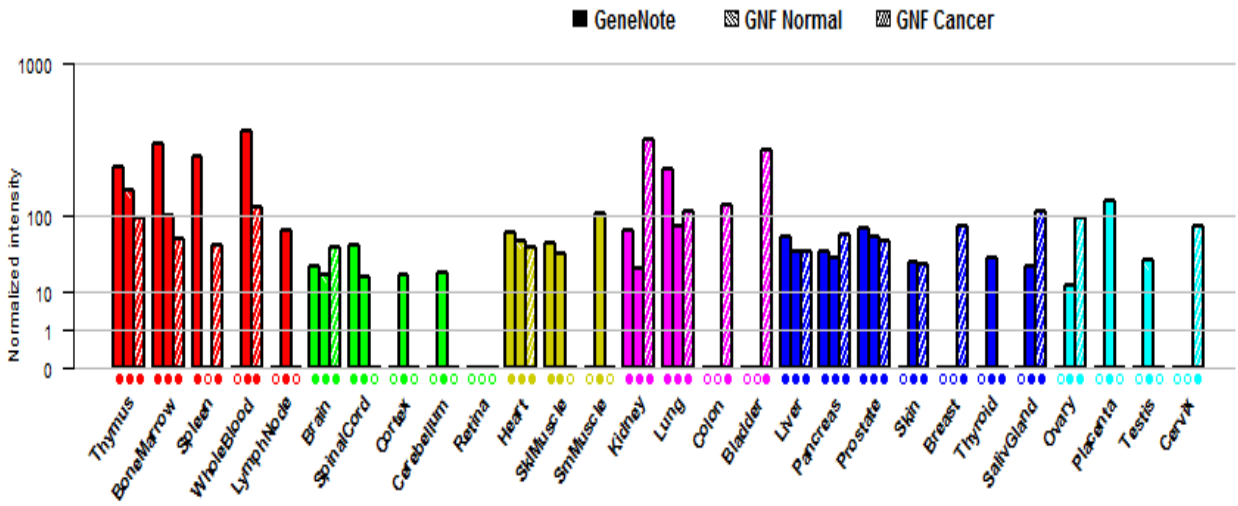


Figure 47 : Expression du gène *ELF4* selon la base de données GeneCard®

Le gène *ELF* est faiblement exprimé (en vert) dans les tissus neuronaux. Son expression est absente dans la rétine.

Le logiciel a détecté un indel dans ce gène. Comme pour le gène *FAM104B*, nous avons conclu à un disfonctionnement du logiciel d'analyse et nous n'avons pas contrôlé cette variation. De plus, le gène *ELF4* est impliqué dans le développement des cellules du système immunitaire, nous n'avons donc pas retenu ce gène.

3.3.6.4.3 Gènes mutés chez 3 patientes

120 gènes sont mutés chez 3 patientes.

Tableau 32 : Tous types de mutations des gènes chez 3 patientes

Gène	Chr	Nb_Patientes
DSG1	chr18	3
DYNC2H1	chr11	3
E2F7	chr12	3
EIF2C4	chr1	3
ERC1	chr12	3
ERLEC1	chr2	3
ESRP1	chr8	3
FABP4	chr8	3
FAM18B2	chr17	3
FANCD2	chr3	3
FIG4	chr6	3
FRG1B	chr20	3
FRMD3	chr9	3
FSTL1	chr3	3
GALNTL5	chr7	3
GPR133	chr12	3
HIVEP1	chr6	3
HNRPLL	chr2	3
HOXD9	chr2	3
HSD17B13	chr4	3
HTR4	chr5	3
INTS6	chr13	3
KAZ	chr1	3
KIAA0355	chr19	3
KIAA1033	chr12	3
KRTAP10-6	chr21	3
LAMA1	chr18	3
LGALS4	chr19	3
LMO7	chr13	3
LOC100130581	chr17	3
LRP1B	chr2	3
MAN1B1	chr9	3
MAP2K3	chr17	3
MEP1A	chr6	3
MTBP	chr8	3
MUC12	chr7	3
MUC4	chr3	3
MYO1H	chr12	3
NAALAD2	chr11	3
NBPF9	chr1	3
NCAPG2	chr7	3
NF1	chr17	3
NPIP	chr16	3
OVGP1	chr1	3

Gène	Chr	Nb_Patientes
41153	chr16	3
ABAT	chr16	3
ACACA	chr17	3
ANKRD30BP2	chr21	3
ANKRD36	chr2	3
AP3S1	chr5	3
BAGE	chr21	3
BAGE2	chr21	3
BAGE3	chr21	3
BAGE4	chr21	3
BAGE5	chr21	3
BAT5	chr6	3
BCAR1	chr16	3
BCO2	chr11	3
BLNK	chr10	3
BRCA1	chr17	3
C14orf23	chr14	3
C20orf191	chr20	3
C4orf41	chr4	3
C6orf10	chr6	3
CABP1	chr12	3
CADPS	chr3	3
CAPN8	chr1	3
CARS2	chr13	3
CDC16	chr13	3
CDYL	chr6	3
CEP170	chr1	3
CEP63	chr3	3
COG2	chr1	3
CSN3	chr4	3
CSTF3	chr11	3
CYP4A11	chr1	3
DCHS2	chr4	3
DNAH5	chr5	3

Gène	Chr	Nb_Patientes
PAPPA2	chr1	3
PDS5A	chr4	3
PHC1	chr12	3
PITRM1	chr10	3
POLA1	chrX	3
PPFIBP1	chr12	3
PRKRA	chr2	3
PROM1	chr4	3
PRPF3	chr1	3
PTRH2	chr17	3
QKI	chr6	3
QRICH2	chr17	3
RABGGTB	chr1	3
RALGAPA1	chr14	3
RAP1A	chr1	3
RBM14-RBM4	chr11	3
RBM4	chr11	3
SEC23B	chr20	3
SERPINB12	chr18	3
SET	chr9	3
SKA3	chr13	3
SLC24A4	chr14	3
SLC38A9	chr5	3
SSPO	chr7	3
SYNE1	chr6	3
TDRD9	chr14	3
THSD7B	chr2	3
TIPIN	chr15	3
TMC1	chr9	3
TPTE2	chr13	3
TSPYL1	chr6	3
TTN	chr2	3
UNC5B	chr10	3
VPS8	chr3	3
ZC3H7A	chr16	3
ZDHHC11	chr5	3
ZNF148	chr3	3
ZNF680	chr7	3
ZNF814	chr19	3
ZNF862	chr7	3
ZNF880	chr19	3
ZZEF1	chr17	3

POLA1 (Polymerase DNA Alpha1) (NM_001206065 intron 22 sur 37)

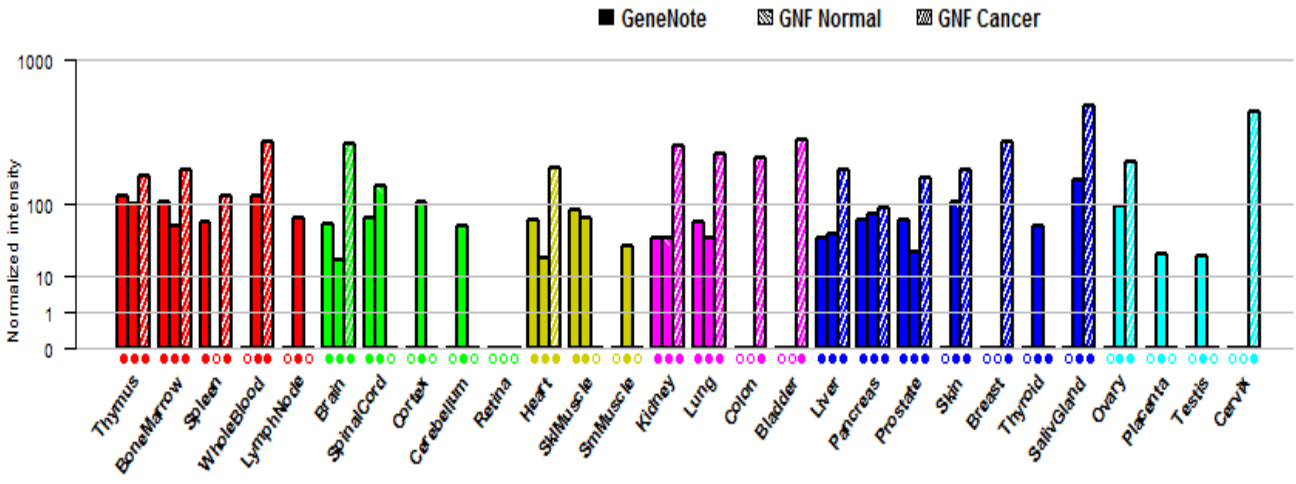


Figure 48 : Expression du gène *POLA1* selon la base de données Gene Card®

Le gène *POLA1* s'exprime dans tous les tissues humaine

Ce gène est localisé sur le chromosome X. Une mutation est détectée dans trois familles, dans l'intron 22 du gène *POLA1* et sur la même position, en faveur d'un polymorphisme

- Chez la patiente du trio3
- Chez la patiente trio4 et sa mère
- Chez la patiente du trio5 et sa mère

Tableau 33 : Détail de la mutation détectée dans le gène *POLA1* dans les trois familles

Gène	chr	position	trio	type	statut	Type annot	Position dans le gène	Séquence de référence	profondeur	A	C	G	T	Fréquence hétérozygote de l'exome
POLA1	chrX	24761320	3	patiente	SNP hétérozygote	intron	Intron22	T	34	0	0	17	17	30.77
POLA1	chrX	24761320	4	patiente	SNP hétérozygote	intron	Intron22	T	26	0	0	18	8	30.77
POLA1	chrX	24761320	4	mère	SNP hétérozygote	intron	Intron22	T	23	0	0	13	10	30.77
POLA1	chrX	24761320	5	patiente	SNP hétérozygote	intron	Intron22	T	24	0	0	14	10	30.77
POLA1	chrX	24761320	5	mère	SNP hétérozygote	intron	Intron22	T	35	0	0	18	17	30.77

3.3.6.4.4 Gènes mutés chez 2 patientes, sur le chromosome X

Nous avons détecté 13 gènes avec une mutation localisée sur le chromosome X, chez au moins de 2 patientes.

Tableau 34 : Tableau des gènes localisés sur le chromosome X, porteurs de variations chez au moins 2 patientes

Gène	Chromosome	Nombre de Patientes
AFF2	chrX	2
CASK	chrX	2
CT47B1	chrX	2
CUL4B	chrX	2
FAM47A	chrX	2
GAGE2A	chrX	2
KLF8	chrX	2
PJA1	chrX	2
PLXNA3	chrX	2
PSMD10	chrX	2
RBMXL3	chrX	2
TAF7L	chrX	2
XG	chrX	2

Les gènes sont non retenus par la présence des variations décrites dans la base de donnée db SNP (build 137). Des autre gènes sont filtrés par la faible profondeur de lecture pour la variation de plus les variation indels sont aussi non retenues.

Les gènes *FAM47A* et *RBMXL3* étant aussi détectés sur la même position par l'analyse de CGH-array 1 M du chromosome X, ils ont donc été testés en priorité par le séquençage Sanger.

FAM47A (family with sequence similarity 47, member A) (NM_ 003817712 exon1 sur 1)

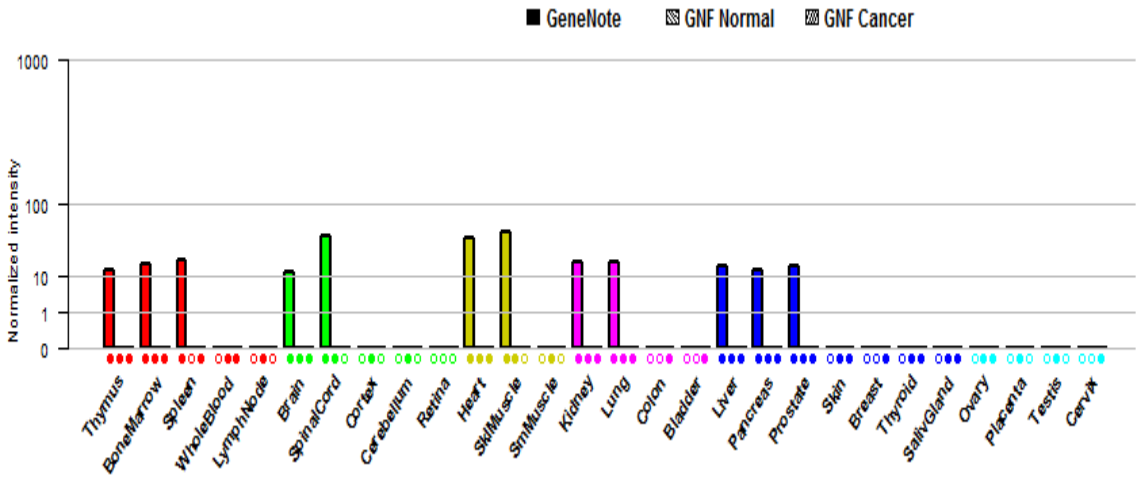


Figure 49 : Expression de gène *FAM47A* selon la base de donnée GeneCard®

L'expression la plus importante du gène *FAM47A* a été détectée dans la moelle épinière.

Tableau 35 : Détail des variations dans le gène *FAM47A* chez les patientes des trio 3 et trio 4

Gène	Chromosome	position	trio	type	statut	Type de mutation	Position dans le gène	Transmission	Séquence de référence	Profondeur	Nucléotide A	Nucléotide C	Nucléotide G	Nucléotide T	Fréquence hétérozygote exome
FAM47A	chrX	34149764	3	patiente	SNP hétérozygote	missense	Exon1	père	A	110	57	0	53	0	0.00
FAM47A	chrX	34148801	4	patiente	SNP hétérozygote	missense	Exon1	père	G	36	21	0	15	0	0.00

- **Trio 3**

Chez la patiente du trio3, le séquençage confirme bien que la variation est héritée du père, en faveur d'un polymorphisme.

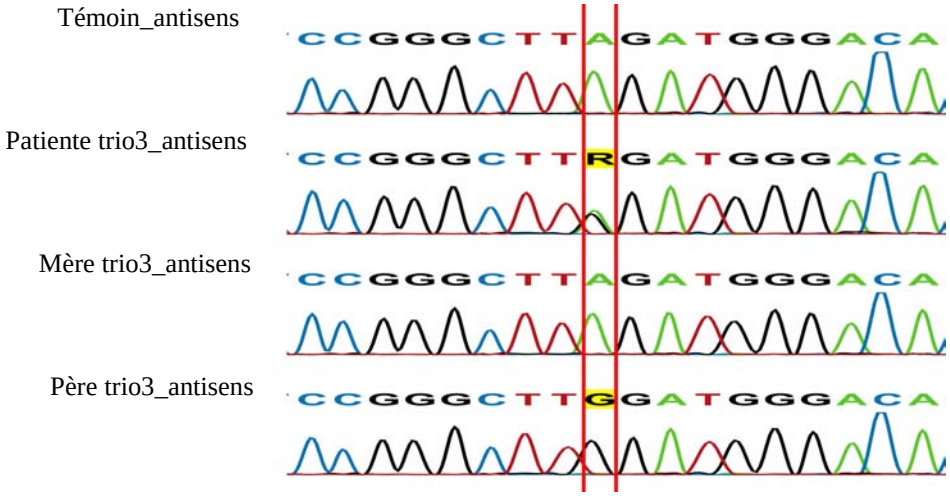


Figure 50 : Profil du séquençage Sanger pour la variation du gène *FAM47A* chez la patiente du trio 3 et ses parents

Cette variation est héritée du père.

RBMXL3 (RNA binding motif protein, X-linked-like 3) (NM_001145346, exon 1 sur 1)

Il y a trois mutations différentes détectées chez trois patientes. Toutes les mutations sont dans l'exon1 du gène *RBMXL3*. Ce gène a déjà été abordé dans le scénario 2 sans lien avec le SAi.

Tableau 36 : Détail de la mutation du gène *RBMXL3* chez les trois filles

Gène	Chr	Position	Trio	Type	Statut	Type de variation	Position dans le gène	Transmission	Ref seq	Prof	A	C	G	T	Fréquence hétérozygote exome
RBMXL3	chrX	114425895	1	patiente	SNP hétérozygote	missense	Exon1	-	G	45	0	22	23	0	0.00
RBMXL3	chrX	114426193	3	patiente	SNP hétérozygote	missense	Exon1	mère	G	79	35	0	44	0	0.00
RBMXL3	chrX	114424814	5	patiente	SNP hétérozygote	synonymous	Exon1	mère	G	28	15	0	13	0	0.00

4. DISCUSSION

Les anomalies chromosomiques sont la première cause identifiée de DI d'origine génétique.

Les avancées technologiques, analyses sur microréseaux d'ADN et séquençage de nouvelle génération, ont transformé l'approche de l'identification des gènes candidats et des réarrangements génomiques associés à la DI [148-149] .

S'agissant des techniques sur microréseaux d'ADN, nos résultats ont été interprétés et discutés en relation avec les critères antérieurement décrits [176] pour mieux appréhender l'implication des CNV dans la survenue de la DI d'autant plus que nous avons utilisé des microréseaux de haute résolution, notamment pour le chromosome X.

Tableau 37 : Critères pour mieux appréhender la distinction entre CNV bénin ou pathogène

(D'après Lee 2007 [176])

Critères	Bénin	Pathogène
CNV hérité par un parent indemne de la maladie	oui	
CNV hérité par un parent atteint de la maladie		oui
CNV *de novo*		oui
CNV décrit dans les bases de données de la population témoin	oui	
CNV décrit dans les bases de données des patients atteints de la maladie		oui
CNV dans la région d'un syndrome de microdélétion/ microduplication		oui
CNV contenant un gène associé avec une maladie décrite dans OMIM A apprécier en fonction du mode d'hérédité		oui
CNV dans des régions géniques ou intra géniques		oui
CNV dans les régions non annotées	oui	
CNV est une délétion		oui
CNV est une amplification (sans gène sensible à effet de dosage)	oui	
CNV dans la région d'éléments régulateurs de transcription		oui

4.1 Microremaniements du gène IL1RAPL1 et déficience intellectuelle

Les remaniements de ce gène ont déjà été décrits chez des garçons avec une DILX légère à modérée, non syndromique. *IL1RAPL1* est considéré comme un médiateur trans-synaptique de signalisation qui régule la formation de synapses excitatrices et d'épines dendritiques. Le Tableau 38 reprend les données de la littérature sur ces remaniements. Les résultats que nous rapportons avec les 3 patients, 2 duplications intragéniques et une délétion intragénique respectivement, confirment l'association avec une variabilité d'expression phénotypique. Il ne semble pas se dégager de dysmorphie cranio-faciale spécifique qui pourrait orienter le clinicien vers l'analyse ciblée de ce gène dans le bilan diagnostique initial.

Les remaniements rapportés (tableau 38) concernent des délétions de grande taille incluant *IL1RAPL1*, les interruptions de la structure du gène à la suite d'une inversion chromosomique, les mutations ponctuelles (non sens, frameshift). La DI est variable en intensité, de modérée à légère, avec au sein d'une même fratrie [177] une variabilité avec des garçons atteints de DI légère, modérée voire sévère. Les conséquences attendues des délétions intragéniques sont, a priori, chez les garçons hémizygotes, une perte de fonction. En colligeant les données de la littérature, il ne semble pas apparaitre de corrélation entre le nombre ou le type d'exons délétés et l'expression du phénotype. La variabilité porte aussi sur les troubles associés, du comportement, avec hyperactivité, déficit de l'attention et des symptômes de la série autistique, là encore sans corrélation avec le type d'altération du gène.

Note étude permet de rapporter notamment des duplications intragéniques jusqu'alors peu décrites. La duplication identifiée chez le patient 1 comprend les exons 7 à 11, impliqués dans le domaine TIR (Toll/Interleukine-1 Receptor) qui a une relation directe dans l'initiation de la voie de signalisation. Le tableau clinique est proche de celui observé chez les patients ayant une délétion intragénique avec une DI modérée à légère. En revanche, la duplication intragénique chez le patient 2, comprenant les exons 2 et 3, s'associe à une présentation clinique, à ce jour, pour laquelle l'évolution neurocognitive et comportementale est plus favorable par rapport aux autres patients. Il faut toutefois souligné que nous n'avons pas pu obtenir de prélèvement sanguin complémentaire suffisant nous permettant d'évaluer les conséquences fonctionnelles sur le transcrit du gène et sa stabilité. Parmi les données de la littérature, à notre connaissance, seules deux publications font état d'une duplication partielle du gène [178-179]. Pour Honda et al., [178], 2 patients non apparentés, sont porteurs d'un

même microremaniement génomique, hérité d'une mère cliniquement indemne, qui outre la duplication partielle de 100 kb comprenant l'exon 2 du gène *IL1RAPL1* en Xp21.3 est associé à une duplication de 737 kb en Xq22.2 comprenant en partie les gènes *REPS2* et *NHS. REPS2* est associé à une petite protéine G et est fortement exprimé dans le cerveau, et *NHS* est le gène candidat du syndrome de Nance-Horan (OMIM 302350, associant cataracte congénitale, anomalies dentaires, dysmorphie faciale et une DI inconstante). Le tableau clinique de ces 2 garçons ayant le même double microremaniement est par ailleurs différent avec pour l'un une présentation sévère avec un syndrome de West, une DI sévère, une absence de langage et pour l'autre, une DI modérée, un retard du langage et des troubles autistiques. Il est difficile de faire la part entre l'action de la duplication partielle des gènes *IL1RAPL1*, *NHS* et *REPS2* respectivement dans le phénotype. L'autre observation de duplication intragénique du gène *IL1RAPL1* est rapportée par Melhem et al., [179] qui dans un criblage ciblée de la population océanique de Palau (République des Palaos, Micronésie) atteinte de troubles psychiatriques (schizophrénie, psychose bipolaire) ont identifié des duplications incluant les exons 3 et 4 chez des hommes, les femmes apparentées conductrices étant indemnes de signes psychiatriques pour la plupart d'entre elles. Il semble donc exister une variabilité dans l'expression clinique associée aux microremaniements intragéniques du gène *IL1RAPL1* entre la déficience intellectuelle, le plus souvent modérée à légère, un tableau d'autisme plus ou moins complet et une présentation purement psychiatrique. Ces données, se rapprochant de celles observées par Guilmatre et al., 2009 [180] pourraient suggérer que le produit du gène *IL1RAPL1* impliqué dans la formation et la maintenance des synapses est une voie de signalisation partagée par ces affections neuro développementales (déficience intellectuelle, schizophrénie, autisme).

Les microremaniements intragéniques *ILRAPL1* identifiés chez les trois garçons rapportés ici sont hérités de leur mère, indemnes de troubles cognitifs. Le profil d'inactivation de l'X des mères des patients 1 et 2 est aléatoire. Dans de très rares cas [183], des femmes porteuses d'un microremaniement *IL1RAPL1* ont été décrites avec des troubles neurocognitifs (DI légère, troubles autistiques) sans corrélation particulière avec un profil d'inactivation biaisé. Tobalocci et al., [183] décrit une femme porteuse d'un remaniement du gène *IL1RAPL1*, sans trouble cognitif, avec 98% d'expression de l'allèle muté dans le sang. Ces données évoquent des profils d'inactivation du chromosome X différents entre le tissu hématopoïétique et le tissu cérébral [127, 181, 183].

Tableau 38 : Données de la littérature sur les remaniements du gène *IL1RAPL1* et les conséquences cliniques associées chez les garçons et les mères vectrices

Référence	Réarrangement	Déficience Intellectuelle	Dysmorphie faciale	Comportement	Anomalie chez la mère
Bhat [127]	invX (p22.1q13)	modérée	+	Autisme	invX (p22.1q13) AS XCI aléatoire
Behnecke [181]	délétion exon 2	légère	+	Hyperactivité	*De novo*
	délétion exons 3 à 5	modérée	+	Déficit de l'attention	Mère vectrice AS XCI : aléatoire
Carrie [126]	Codon stop Exon 10 p.Y459X	modérée	-	-	ND
	délétion exons 3 à 5	modérée	-	-	Mère vectrice AS
Franek [128]	délétion exons 2 à 5	modérée	+	Impulsivité	ND
	délétion exons 1 à 5	modérée	-	Impulsivité	ND
Honda [178]	duplication exon 2	Modérée	ND	-	Vectrices AS XCI : aléatoire
	duplication exon 2	Sévère		trouble autistique	
Leprêtre [182]	inv X (p21.3q27.1)	modérée	+	+	Mère vectrice DI légère XCI : aléatoire
Nawara [177]	délétion exons 2 à 5	1/4 légère 2/4 modérée 1/4 sévère	+	2/4 hyperactivité agressivité	Vectrice AS
Piton [129]	Mutation frameshift Exon 9 (fille)	Fille : pas DI	-	Fille : autisme	ND
	délétion exons 3 à 7 (3 frères)	DI sévère à légère	-	autisme inconstant Hyperactivité	
Tabolacci [183]	Mutation Non sens Exon 10	modérée	-	-	Vectrices AS ou DI légère avec XCI variable
Mikhail [184]	délétion exons 2 à 6	légère à modérée	+	-	Vectrice AS XCI biaisé
Youngs [185]	délétion exons 3 à 11	légère	+	+	Vectrice AS XCI aléatoire

inv: inversion; dup: duplication; del: délétion; ex: exon; NA:non disponible; ND: non déterminé; ASD: autistic spectrum disorder; XCI: inactivation du chromosome X ; AS : asymptomatique

4.2 Microremaniements du gène MBD5 et déficience intellectuelle

La protéine MBD5 (OMIM *611472) fait partie de la famille des protéines à domaine liant les dinucléotides CpG méthylés (MDB Methyl-CpG-binding domain) dont MECP2 (OMIM *300005) qui est impliquée dans le syndrome de Rett, affection neurodéveloppementale prototypique. *MBD5* contient 5 exons non codants à son extrémité 5', suivi de 10 exons codants. Deux isoformes ont été décrites [186], la plus longue comporte 1494 acides aminés, codés par les exons 6 à 15, la seconde comportant 851 acides aminés codés par les exons 6 à 9. Les études fonctionnelles ont suggéré que *MBD5* contribuerait à la formation ou à la fonction de l'hétérochromatine [186]. L'isoforme 1 est fortement exprimée dans le cerveau et le testicule et l'isoforme 2 dans l'ovocyte, suggérant un possible rôle dans les fonctions cérébrales et dans la reprogrammation épigénétique après la fécondation. Récemment, des délétions incluant le gène *MBD5*, aussi bien que des délétions intragéniques ont été identifiées chez des individus avec un phénotype de DI, convulsions, déficience sévère du langage et troubles comportementaux [187-193]. Talkowski et al., en 2011 [194], ont suggéré un modèle mixte de mutations du gène *MBD5* : mutations délétères, de pénétrance complète, responsables d'une affection neurodéveloppementale associée aux caractéristiques du syndrome microdélétionnel 2q23.1, et des variants faux sens, de pénétrance incomplète, qui augmenteraient de façon significative le risque de troubles du spectre autistique. Dans notre étude, nous avons identifié 3 patients (A, B et D) avec une délétion *de novo* incluant les derniers exons du gène *ORC4* et seulement les deux premiers exons non traduits de l'isoforme 1 du gène *MBD5* exprimé dans le cerveau. Une délétion similaire a été rapportée [190, 194]. Le niveau d'expression de l'ARNm de *MBD5* des patients A et B est significativement réduit en comparaison des parents, de la sœur et du frère non porteurs de la délétion. Ce résultat prouve que la délétion hétérozygote des 2 premiers exons non codants de l'isoforme 1 de *MBD5* conduit à l'extinction spécifique de son expression sur l'allèle délété. Plus particulièrement, nous avons identifié au travers de cette étude, deux nouvelles altérations du gène *MBD5*, une duplication intragénique et une mutation non sens. La duplication intragénique trouvée chez le patient C s'étend de l'exon 5 non codant à l'exon 10 codant du gène *MBD5*. L'étude de la transcription a montré la présence de nombreux transcrits aberrants conduisant à un codon prématuré de terminaison. Cette duplication partielle est prédite

comme altérant la fonction de la protéine *MBD5*. La mutation non sens, *de novo*, du patient E conduit à un codon stop prématuré dans le gène *MBD5* et est prédite entrainer la formation d'une protéine tronquée qui a perdu le domaine riche en proline en plus du signal putatif de localisation nucléaire. Cette mutation n'a jamais été, à ce jour, rapportée ni dans la base de données 1000 Génomes (http://browser.1000genomes.org/) ni dans la base de SNP (bdSNP-build 137). L'analyse en RT-PCR a montré un niveau normal d'expression des transcrits MBD5 suggérant l'absence de dégradation de l'ARN. Plus spécifiquement, le séquençage des transcrits MBD5 a révélé l'expression in vivo des deux types de transcrits, muté et normal. La traduction de ce transcrit muté pourrait conduire à une protéine tronquée avec un effet dominant négatif ou cette protéine aberrante pourrait être éliminée par le système de contrôle des protéines de la cellule à l'origine d'une haploinsuffisance. Si l'on considère les mutations ponctuelles du gène *MBD5*, des variants faux sens ont été antérieurement rapportés [187, 194], principalement hérités d'un parent cliniquement indemne. Ces variants sont considérés comme des facteurs de risque de troubles du spectre autistique [194]. Nous avons aussi identifié dans notre série de patients [tableaux supplémentaires I et II, annexe I], les deux variants faux sens préalablement rapportés par Talkowski et al., 2011 [194], p.79Gly>Glu (rs34995577) pour le patient 51 et p.1048Thr>Ile pour le patient 33. Tout récemment, une mutation frameshift du gène *MBD5* [c.150del (p.Thr52Hisfs*31); NM_018328.4] entrainant un codon stop prématuré a été rapporté chez un patient avec un tableau clinique proche du spectre phénotypique du syndrome de Kleefstra [195]. Si le phénotype du patient C avec une duplication intragénique chevauche le phénotype observé chez les patients A, B et D avec une délétion intragénique, le phénotype du patient E avec la mutation non sens *MBD5* est beaucoup plus sévère sans acquisition de la marche et du langage verbal à l'âge de 10 ans. Cette mutation à forte pénétrance représente 1,2% (1/78) de notre cohorte de patients sélectionnés. En conclusion, ces résultats confirment l'implication des altérations de gène *MBD5* dans les affections neurodéveloppementales et contribue à étendre le spectre mutationnel du gène *MBD5*. Des observations complémentaires sur d'autres grandes séries seront toutefois nécessaires pour établir de façon définitive et utile des corrélations génotype-phénotype.

4.3 Recherche du gène candidat du syndrome d'Aicardi

Le syndrome d'Aicardi est une affection sporadique de la fille. En l'absence de possibilité d'analyse de liaison ou de clonage positionnel à l'aide d'un remaniement chromosomique équilibré (translocation, inversion), nous avons orienté notre recherche vers des analyses pangénomiques en utilisant les nouvelles technologies de criblage du génome humain à haute résolution progressivement développées dans notre laboratoire (analyse sur microréseau d'ADN) ou mises à disposition grâce à une collaboration (séquençage haut débit).

L'utilisation d'un microréseau d'oligonucléotides 1M du chromosome X a été basée sur l'hypothèse d'une mutation dominante, *de novo*, d'un gène du chromosome X à l'origine de la maladie. Nous n'avons pas identifié de régions ou de gènes candidats. Nous avons en revanche repéré des variations témoignant de faux positifs, reflet de notre stratégie utilisée s'agissant du « cut off » retenu pour sélectionner les réarrangements. Afin de mettre toutes les chances de notre côté, dans l'identification potentielle d'un microremaniement intragénique de petite taille, nous avons retenu toutes les variations d'au moins 3 oligonucléotides consécutifs dans les régions géniques. Pour certaines affections génétiques, le profil mutationnel comprend rarement des remaniements de la structure du gène mais plutôt essentiellement des mutations ponctuelles. Le syndrome d'Aicardi étant une affection rare, notre série « limitée » de patientes pourrait ne pas comporter une patiente avec un tel remaniement détectable par la technique d'hybridation génomique comparative. Nous continuerons cette analyse au seuil de 10 sondes consécutives déviantes pour toute nouvelle patiente ayant un syndrome d'Aicardi, dans l'hypothèse d'une telle situation exceptionnelle.

Le séquençage d'exome a été tout logiquement la suite de notre orientation de recherche en se basant sur les résultats de plus en plus nombreux des données de la littérature conduisant à l'identification de gènes candidats grâce à cette approche haut débit. L'exome du chromosome X a été initialement privilégié, en tenant compte aussi des possibilités de l'époque. Secondairement, nous avons opté pour le séquençage de l'exome complet avec la plus grande disponibilité des équipes prestataires offrant cette technologie d'analyse.

Par ces deux types d'analyse d'exome, nous avons été confronté à de nombreux faux-positifs et des variations héritées d'un parent. Les techniques de séquençage haut débit ont leurs limites et le nombre de variants détectés qui sont en fait des faux positifs augmente avec la taille de la cible séquencée (du gène à l'exome complet). Ceci est bien illustré dans une étude comparative récemment publiée dans laquelle un génome a été séquencé avec une couverture élevée (~ 150 fois) par deux plates-formes de séquençage différentes [196]. La concordance entre ces deux plates-formes n'est pas totale pour les variations détectées. Seulement 88% des SNVs (single nucleotide variants) sont retrouvées sur les deux plateformes. Les artefacts du séquençage sont particulièrement problématiques pour la détection des mutations *de novo*. Plusieurs variants sont annotés chez l'enfant comme variations *de novo* alors qu'en fait ce sont des faux positifs chez l'enfant non retrouvée chez les parents ou des réelles mutations héritées détectées chez l'enfant mais non détectées chez les parents (faux négatifs) [197]. Nous avons utilisé l'approche classique d'analyse de trios compte tenu de l'hypothèse d'une maladie dominante *de novo*. Les stratégies d'analyse employées ne nous ont pas conduit à identifier de réels gènes candidats comme dans d'autres exemples des données de la littérature, qui obtiennent plusieurs gènes répondant aux critères de présélection et qui utilisent ensuite des données d'annotation fonctionnelle qui leur permettent de prioriser les gènes candidats les uns par rapport aux autres. Nous reconnaissons les problèmes rencontrés de couverture incomplète et de profondeur parfois très faible des régions exoniques analysées. Ces limites, bien illustrées dans nos résultats, ont été à l'origine d'un grand nombre de contrôle car bien souvent la profondeur de lecture était incomplète chez l'un des deux parents. Des faux positifs du séquençage ont été rencontrés et plus particulièrement le logiciel utilisé n'a pas bien attribué l'état homozygote par rapport à l'état hétérozygote (exemple du gène *RNF175* pour le trio5). Le filtrage des données s'est basé sur dbSNP, 1000 génomes et les bases de la société Integragen. Nous reconnaissons que nous avons pu éliminer un gène candidat en raison d'une mésinformation introduite par erreur dans les bases dont dbSNP. De nombreuses études génétiques, telles que le projet 1000 Génomes, sont réalisées sur de l'ADN provenant de lignées cellulaires lymphoblastoïdes. La création de ces lignées et la culture cellulaire ultérieure sont connues pour introduire des changements génomiques qui apparaissent comme des mutations *de novo* lorsque des séquences dérivées de ces lignées cellulaires sont comparées entre parents et patient. Dans le cadre du projet 1000 Génomes, les génomes de deux trios parents-patient ont été séquencés en utilisant de l'ADN dérivé de lignées cellulaires lymphoblastoïdes [198]. Dans l'un des trios, les auteurs ont identifié 643 mutations *de novo* dans l'ADN issu de lignée cellulaire qui n'ont pas été observées dans

l'ADN dérivé du sang de ces personnes. En revanche, seulement 35 des mutations *de novo* ont été observées à la fois dans le sang et l'ADN dérivé des lignées cellulaires lymphoblastoïdes, ce qui démontre que la majorité de ces mutations *de novo* potentielles étaient en fait causées par transformation de la lignée cellulaire et la culture, une constatation qui a été récemment confirmée [199]. Nous avons aussi été surpris par les résultats concernant les indels. Nous avions des indels tout à fait inhabituelles résultant manifestement d'un problème d'alignement des séquences générant de tels artefacts. Pour les régions de séquence répétée, l'analyse des indels est particulièrement délicate. Dans la même étude publiée par Roach et al., en 2010 [197], les indels sont détectées de manière beaucoup moins fiable que les SNVs. En fait la concordance entre les deux plateformes n'est que de 28% pour les indels. Le gène *FAM104B* sur les données du séquençage haut débit révélait des indels tant chez les 5 patientes et leur parents mais d'un aspect différent pour chaque individu. Le séquençage Sanger concluait en un faux positif. Malgré tous nos efforts, à partir ce ces expériences à ce jour, nous n'avons pas identifié de gène candidat du syndrome d'Aicardi.

CONCLUSION ET PERSPECTIVES

Manifestement, l'expérience de ces deux exomes nous conduit à réévaluer l'analyse des indels, à partir des données brutes de séquençage pour un nouvel alignement auprès d'une autre équipe de référence.

Nous avons posé aussi l'hypothèse d'une mutation exonique à l'origine de la maladie, en raison de la fréquence d'un tel type de mutation dans les maladies monogéniques. Toutefois, si la mutation siège dans une région non explorée par le séquençage, elle ne sera pas identifiée. Ainsi, l'expansion de triplet propre au gène *FMR1*, première cause de DI héréditaire, n'aurait pas été relevée par la technique de l'exome. Une autre voie de recherche pourrait concerner le rôle d'une mutation épigénétique altérant l'expression du gène candidat. Ce type de mutation pourrait être abordée par l'analyse du méthylome. S'agissant de l'analyse de l'expression, le séquençage haut débit du transcriptome pourrait être aussi une voie de recherche et notamment en s'intéressant à ces cellules fibroblastiques dont le profil d'expression des gènes est plus proche de celui du tissu cérébral par rapport au tissu hématopoïétique.

L'hypothèse d'une mutation en mosaïque, post zygotique est aussi à envisager, comme cela a été récemment identifié pour le syndrome de Protée avec le gène *AKT1* [200]. Sa détection sera techniquement plus complexe nécessitant une profondeur de lecture de qualité, et sans doute plusieurs cas index séquencés.

Des modèles d'analyse des résultats de variants rares hérités parentaux dont la transmission correspondrait à un modèle de bigénisme devraient aussi tenir compte de l'expression féminine essentielle du syndrome d'Aicardi, nécessitant, a priori une mutation sur le chromosome X, révélée par une autre mutation dans un autre gène autosomique [201].



« Chercher l'aiguille dans une botte de foin »

REFERENCES INTERNET

- 1000Genomes

http://www.1000genomes.org/

- A.A.L Asociaton du syndrome d'Aicardi

www.aicardi.info

- Agilent

http://www.home.agilent.com

- CCDS Bases de données

http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi

- dbSNP

http://www.ncbi.nlm.nih.gov/projects/SNP/

- DECIPHER (DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources)

https://decipher.sanger.ac.uk/application/

- DGV (Database of Genomic Variants)

http://projects.tcag.ca/variation

- ECARUCA

http://www.ecaruca.net

- Ensembl

http://www.ensembl.org

- GENCODE

http://www.gencodegenes.org/

- GeneCards ®

http://www.genecards.org/

- Génoscope, Evry, France

http://www.genoscope.cns.fr/

- Greenwood Genetics center

http://www.ggc.org/xlmr.htm

- Illumina

http://www.illumina.com

- Integragen

http://www.integragen.fr/home.php

- International HapMap Project

http://hapmap.ncbi.nlm.nih.gov/

- miRBase

http://www.mirbase.org/

- NCBI

http://www.ncbi.nlm.nih.gov/

- Online Mendelian Inheritance in Man (OMIM)

http://www.ncbi.nlm.nih.gov/omim/

- Primer 3 (v. 0.4.0)

http://frodo.wi.mit.edu/

- RefSeq

http://www.ncbi.nlm.nih.gov/RefSeq/

- Roche life science

http://www.454.com

- Sanger Institute

http://www.sanger.ac.uk

- SNP Check3 NGRL (National Genetics Reference Laboratory)

https://ngrl.manchester.ac.uk/SNPCheckV3/snpcheck.htm

- UCSC Genome Browser

http://genome.ucsc.edu/

- VEGA

http://vega.sanger.ac.uk/index.html

BIBLIOGRAPHIE

1. *Manuel de Diagnostic et Statistiques des Troubles Mentaux- Quatrième Edition (DSM-IV).* Vol. IV. 1994, Washington DC: American Psychiatric Association, APA.
2. *World Health Organization: International Statistical Classification of Diseases and Related Health Problems*. 10th revision ed. Vol. vol 1. 1992, Geneva: World Health Organization.
3. Tredgold, A.F., *A Textbook of Mental Deficiency (Amentia)*. 6th ed. ed. 1937, London: Bailliêne, Tindall & Cox.
4. Leonard, H. and X. Wen, *The epidemiology of mental retardation: challenges and opportunities in the new millennium.* Ment Retard Dev Disabil Res Rev, 2002. **8**(3): p. 117-34.
5. Kwok, H. and P.W. Cheung, *Co-morbidity of psychiatric disorder and medical illness in people with intellectual disabilities.* Curr Opin Psychiatry, 2007. **20**(5): p. 443-9.
6. Kelly, M.P., et al., *Wechsler Adult Intelligence Scale and Wechsler Adult Intelligence Scale-Revised in a neurologically impaired population.* J Clin Psychol, 1984. **40**(3): p. 788-91.
7. Ryan, J.J. and S.J. Rosenberg, *Validity of the Verbal IQ as a short form of the Wechsler Adult Intelligence Scale-Revised.* J Clin Psychol, 1984. **40**(1): p. 306-8.
8. Westerinen, H., et al., *Prevalence of intellectual disability: a comprehensive study based on national registers.* J Intellect Disabil Res, 2007. **51**(Pt 9): p. 715-25.
9. Roeleveld, N., G.A. Zielhuis, and F. Gabreels, *The prevalence of mental retardation: a critical review of recent literature.* Dev Med Child Neurol, 1997. **39**(2): p. 125-32.
10. Simonoff, E., et al., *The Croydon Assessment of Learning Study: prevalence and educational identification of mild mental retardation.* J Child Psychol Psychiatry, 2006. **47**(8): p. 828-39.
11. Stromme, P. and T.H. Diseth, *Prevalence of psychiatric diagnoses in children with mental retardation: data from a population-based study.* Dev Med Child Neurol, 2000. **42**(4): p. 266-70.
12. Matilainen, R., et al., *A population-based study on the causes of mild and severe mental retardation.* Acta Paediatr, 1995. **84**(3): p. 261-6.
13. Cans, C., et al., *Aetiological findings and associated factors in children with severe mental retardation.* Dev Med Child Neurol, 1999. **41**(4): p. 233-9.
14. Murphy C, B.C., Schendel D, Decouflé P, Yeargin-Allsopp M. , *Epidemiology of mental retardation in children.* Ment Retrd Dev Disabil Res Rev, 1998. **4**(1): p. 6-13.
15. McLaren, J. and S.E. Bryson, *Review of recent epidemiological studies of mental retardation: prevalence, associated disorders, and etiology.* Am J Ment Retard, 1987. **92**(3): p. 243-54.
16. Mandel, J.L. and J. Chelly, *Monogenic X-linked mental retardation: is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations.* Eur J Hum Genet, 2004. **12**(9): p. 689-93.
17. Chelly, J., et al., *Genetics and pathophysiology of mental retardation.* Eur J Hum Genet, 2006. **14**(6): p. 701-13.
18. Curry, C.J., et al., *Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics.* Am J Med Genet, 1997. **72**(4): p. 468-77.

19. Frints, S.G., et al., *X-linked mental retardation: vanishing boundaries between non-specific (MRX) and syndromic (MRXS) forms.* Clin Genet, 2002. **62**(6): p. 423-32.
20. Ropers, H.H. and B.C. Hamel, *X-linked mental retardation.* Nat Rev Genet, 2005. **6**(1): p. 46-57.
21. Stevenson, R.E., et al., *Genetic syndromes among individuals with mental retardation.* Am J Med Genet A, 2003. **123A**(1): p. 29-32.
22. *RKI: Infektionsepidemiologisches Jahrbuch fur 2008*. 2009, Robert Koch Institut: RKI, Berlin.
23. Abel, E.L. and J.H. Hannigan, *Maternal risk factors in fetal alcohol syndrome: provocative and permissive influences.* Neurotoxicol Teratol, 1995. **17**(4): p. 445-62.
24. Spohr, H.L., J. Willms, and H.C. Steinhausen, *Fetal alcohol spectrum disorders in young adulthood.* J Pediatr, 2007. **150**(2): p. 175-9, 179 e1.
25. Anderson, J.L., et al., *Maternal obesity, gestational diabetes, and central nervous system birth defects.* Epidemiology, 2005. **16**(1): p. 87-92.
26. Ropers, H.H., *New perspectives for the elucidation of genetic disorders.* Am J Hum Genet, 2007. **81**(2): p. 199-207.
27. Winnepenninckx B, R., L., Kooy, R.F., *Mental retardation : a review of the genetic causes.* Brit J Dev Disabil, 2003. **49**: p. 29-44.
28. Lambert, L., et al., *MEF2C mutations are a rare cause of Rett or severe Rett-like encephalopathies.* Clin Genet, 2012.
29. Ropers, H.H., *Genetics of intellectual disability.* Curr Opin Genet Dev, 2008. **18**(3): p. 241-50.
30. Flint, J. and S. Knight, *The use of telomere probes to investigate submicroscopic rearrangements associated with mental retardation.* Curr Opin Genet Dev, 2003. **13**(3): p. 310-6.
31. Stromme, P., P.G. Bjornstad, and K. Ramstad, *Prevalence estimation of Williams syndrome.* J Child Neurol, 2002. **17**(4): p. 269-71.
32. Vissers, L.E., B.B. de Vries, and J.A. Veltman, *Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis.* J Med Genet, 2010. **47**(5): p. 289-97.
33. Sebat, J., D.L. Levy, and S.E. McCarthy, *Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders.* Trends Genet, 2009. **25**(12): p. 528-35.
34. Cook, E.H., Jr. and S.W. Scherer, *Copy-number variations associated with neuropsychiatric conditions.* Nature, 2008. **455**(7215): p. 919-23.
35. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-12.
36. Koolen, D.A., et al., *A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism.* Nat Genet, 2006. **38**(9): p. 999-1001.
37. Shaw-Smith, C., et al., *Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability.* Nat Genet, 2006. **38**(9): p. 1032-7.
38. Zollino, M., et al., *Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype.* Nat Genet, 2012. **44**(6): p. 636-8.
39. Saccone, S., et al., *The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes.* Proc Natl Acad Sci U S A, 1992. **89**(11): p. 4913-7.
40. Biesecker, L.G., *The end of the beginning of chromosome ends.* Am J Med Genet, 2002. **107**(4): p. 263-6.

41. Zahir, F. and J.M. Friedman, *The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility.* Clin Genet, 2007. **72**(4): p. 271-87.
42. Mefford, H.C., et al., *Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes.* N Engl J Med, 2008. **359**(16): p. 1685-99.
43. Brunetti-Pierri, N., et al., *Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities.* Nat Genet, 2008. **40**(12): p. 1466-71.
44. Miller, D.T., et al., *Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders.* J Med Genet, 2009. **46**(4): p. 242-8.
45. Ramalingam, A., et al., *16p13.11 duplication is a risk factor for a wide spectrum of neuropsychiatric disorders.* J Hum Genet, 2011. **56**(7): p. 541-4.
46. Albers, C.A., et al., *Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome.* Nat Genet, 2012. **44**(4): p. 435-9, S1-2.
47. Girirajan, S., et al., *Phenotypic Heterogeneity of Genomic Disorders and Rare Copy-Number Variants.* N Engl J Med, 2012.
48. Weiss, L.A., et al., *Association between microdeletion and microduplication at 16p11.2 and autism.* N Engl J Med, 2008. **358**(7): p. 667-75.
49. Lupski, J.R., *Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits.* Trends Genet, 1998. **14**(10): p. 417-22.
50. Liu, P., et al., *Mechanisms for recurrent and complex human genomic rearrangements.* Curr Opin Genet Dev, 2012. **22**(3): p. 211-20.
51. Gu, W., F. Zhang, and J.R. Lupski, *Mechanisms for human genomic rearrangements.* Pathogenetics, 2008. **1**(1): p. 4.
52. Chen, K.S., et al., *Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome.* Nat Genet, 1997. **17**(2): p. 154-63.
53. Woodward, K.J., et al., *Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination.* Am J Hum Genet, 2005. **77**(6): p. 966-87.
54. Hastings, P.J., G. Ira, and J.R. Lupski, *A microhomology-mediated break-induced replication model for the origin of human copy number variation.* PLoS Genet, 2009. **5**(1): p. e1000327.
55. Stevenson, R.E., *Splitting and lumping in the nosology of XLMR.* Am J Med Genet, 2000. **97**(3): p. 174-82.
56. Turner, G., *Finding genes on the X chromosome by which homo may have become sapiens.* Am J Hum Genet, 1996. **58**(6): p. 1109-10.
57. Ropers, H.H., *Genetics of early onset cognitive impairment.* Annu Rev Genomics Hum Genet, 2010. **11**: p. 161-87.
58. Gecz, J., C. Shoubridge, and M. Corbett, *The genetic landscape of intellectual disability arising from chromosome X.* Trends Genet, 2009. **25**(7): p. 308-16.
59. Aicardi, J., Lefebvre, J. & Lerique-Koechlin, A. , *A new syndrome: spasms in flexions, callosal agenesis, ocular abnormalities.* Electroenceph Clin Neurophysiol 1965. **19**: p. 606-612.
60. Aicardi, J., J.J. Chevrie, and F. Rousselie, *[Spasma-in-flexion syndrome, callosal agenesis, chorioretinal abnormalities].* Arch Fr Pediatr, 1969. **26**(10): p. 1103-20.
61. Donnenfeld, A.E., et al., *Clinical, cytogenetic, and pedigree findings in 18 cases of Aicardi syndrome.* Am J Med Genet, 1989. **32**(4): p. 461-7.

62. Sutton, V.R., et al., *Facial and physical features of Aicardi syndrome: infants to teenagers.* Am J Med Genet A, 2005. **138A**(3): p. 254-8.
63. Rosser, T.L., M.T. Acosta, and R.J. Packer, *Aicardi syndrome: spectrum of disease and long-term prognosis in 77 females.* Pediatr Neurol, 2002. **27**(5): p. 343-6.
64. Menezes, A.V., R.W. Enzenauer, and J.R. Buncic, *Aicardi syndrome--the elusive mild case.* Br J Ophthalmol, 1994. **78**(6): p. 494-6.
65. Aicardi, J., *Aicardi syndrome: Old and new findings.* Int Pediatr, 1999. **14**: p. 5-8.
66. Aicardi, J., *Aicardi syndrome.* Brain Dev, 2005. **27**(3): p. 164-71.
67. Carney, S.H., et al., *Aicardi syndrome: more than meets the eye.* Surv Ophthalmol, 1993. **37**(6): p. 419-24.
68. Menezes, A.V., D.L. MacGregor, and J.R. Buncic, *Aicardi syndrome: natural history and possible predictors of severity.* Pediatr Neurol, 1994. **11**(4): p. 313-8.
69. Grosso, S., et al., *Aicardi syndrome with favorable outcome: case report and review.* Brain Dev, 2007. **29**(7): p. 443-6.
70. Yacoub, M., et al., *[Aicardi syndrome with favorable outcome].* Arch Pediatr, 2003. **10**(6): p. 530-2.
71. King, A.M., et al., *Aicardi syndrome.* Br J Ophthalmol, 1998. **82**(4): p. 457.
72. Matlary, A., et al., *Aicardi syndrome in a girl with mild developmental delay, absence of epilepsy and normal EEG.* Clin Dysmorphol, 2004. **13**(4): p. 257-60.
73. Lee, S.W., et al., *An atypical case of Aicardi syndrome with favorable outcome.* Korean J Ophthalmol, 2004. **18**(1): p. 79-83.
74. Prats Vinas, J.M., et al., *Callosal agenesis, chorioretinal lacunae, absence of infantile spasms, and normal development: Aicardi syndrome without epilepsy?* Dev Med Child Neurol, 2005. **47**(6): p. 419-20; discussion 364.
75. Neidich, J.A., et al., *Heterogeneity of clinical severity and molecular lesions in Aicardi syndrome.* J Pediatr, 1990. **116**(6): p. 911-7.
76. Hopkins, B., et al., *Neuroimaging aspects of Aicardi syndrome.* Am J Med Genet A, 2008. **146A**(22): p. 2871-8.
77. Hopkins, I.J., et al., *The Aicardi syndrome in a 47, XXY male.* Aust Paediatr J, 1979. **15**(4): p. 278-80.
78. Molina, J.A., et al., *Aicardi syndrome in two sisters.* J Pediatr, 1989. **115**(2): p. 282-3.
79. Hoag, H.M., et al., *Evidence that skewed X inactivation is not needed for the phenotypic expression of Aicardi syndrome.* Hum Genet, 1997. **100**(3-4): p. 459-64.
80. Lyon, M.F., *Sex chromatin and gene action in the mammalian X-chromosome.* Am J Hum Genet, 1962. **14**: p. 135-48.
81. Wieacker, P., J. Zimmer, and H.H. Ropers, *X inactivation patterns in two syndromes with probable X-linked dominant, male lethal inheritance.* Clin Genet, 1985. **28**(3): p. 238-42.
82. Wettke-Schafer, R. and G. Kantner, *X-linked dominant inherited diseases with lethality in hemizygous males.* Hum Genet, 1983. **64**(1): p. 1-23.
83. Migeon, B.R., et al., *Selection against lethal alleles in females heterozygous for incontinentia pigmenti.* Am J Hum Genet, 1989. **44**(1): p. 100-6.
84. Eble, T.N., et al., *Non-random X chromosome inactivation in Aicardi syndrome.* Hum Genet, 2009. **125**(2): p. 211-6.
85. Zubairi, M.S., R.F. Carter, and G.M. Ronen, *A male phenotype with Aicardi syndrome.* J Child Neurol, 2009. **24**(2): p. 204-7.
86. Chen, T.H., et al., *Aicardi syndrome in a 47, XXY male neonate with lissencephaly and holoprosencephaly.* J Neurol Sci, 2009. **278**(1-2): p. 138-40.
87. Curatolo, P., G. Libutti, and B. Dallapiccola, *Aicardi syndrome in a male infant.* J Pediatr, 1980. **96**(2): p. 286-7.

88. Chappelow, A.V., et al., *Aicardi syndrome in a genotypic male.* Ophthalmic Genet, 2008. **29**(4): p. 181-3.
89. Collins, F.S., *Positional cloning: let's not call it reverse anymore.* Nat Genet, 1992. **1**(1): p. 3-6.
90. Weissenbach, J., et al., *A second-generation linkage map of the human genome.* Nature, 1992. **359**(6398): p. 794-801.
91. Ropers, H.H., et al., *Agenesis of corpus callosum, ocular, and skeletal anomalies (X-linked dominant Aicardi's syndrome) in a girl with balanced X/3 translocation.* Hum Genet, 1982. **61**(4): p. 364-8.
92. Van den Veyver, I.B., et al., *Presence of filamin in the astrocytic inclusions of Aicardi syndrome.* Pediatr Neurol, 2004. **30**(1): p. 7-15.
93. Yilmaz, S., et al., *Screening of subtle copy number changes in Aicardi syndrome patients with a high resolution X chromosome array-CGH.* Eur J Med Genet, 2007. **50**(5): p. 386-91.
94. Wang, X., et al., *A genome-wide screen for copy number alterations in Aicardi syndrome.* Am J Med Genet A, 2009. **149A**(10): p. 2113-21.
95. Bursztejn, A.C., et al., *Molecular characterization of a monosomy 1p36 presenting as an Aicardi syndrome phenocopy.* Am J Med Genet A, 2009. **149A**(11): p. 2493-500.
96. Fishburn, J., et al., *The diagnosis and frequency of X-linked conditions in a cohort of moderately retarded males with affected brothers.* Am J Med Genet, 1983. **14**(4): p. 713-24.
97. Tarpey, P.S., et al., *A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation.* Nat Genet, 2009. **41**(5): p. 535-43.
98. de Brouwer, A.P., et al., *Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium.* Hum Mutat, 2007. **28**(2): p. 207-8.
99. Froyen, G., et al., *Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation.* Am J Hum Genet, 2008. **82**(2): p. 432-43.
100. Solomon, N.M., et al., *Array comparative genomic hybridisation analysis of boys with X linked hypopituitarism identifies a 3.9 Mb duplicated critical region at Xq27 containing SOX3.* J Med Genet, 2004. **41**(9): p. 669-78.
101. Mimault, C., et al., *Proteolipoprotein gene analysis in 82 patients with sporadic Pelizaeus-Merzbacher Disease: duplications, the major cause of the disease, originate more frequently in male germ cells, but point mutations do not. The Clinical European Network on Brain Dysmyelinating Disease.* Am J Hum Genet, 1999. **65**(2): p. 360-9.
102. Van Esch, H., et al., *Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males.* Am J Hum Genet, 2005. **77**(3): p. 442-53.
103. Bauters, M., et al., *Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair.* Genome Res, 2008. **18**(6): p. 847-58.
104. Sanlaville, D., C. Schluth-Bolard, and C. Turleau, *Distal Xq duplication and functional Xq disomy.* Orphanet J Rare Dis, 2009. **4**: p. 4.
105. Hamdan, F.F., et al., *Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation.* N Engl J Med, 2009. **360**(6): p. 599-605.
106. Hamdan, F.F., et al., *De novo STXBP1 mutations in mental retardation and nonsyndromic epilepsy.* Ann Neurol, 2009. **65**(6): p. 748-53.
107. Shoichet, S.A., et al., *Haploinsufficiency of novel FOXG1B variants in a patient with severe mental retardation, brain malformations and microcephaly.* Hum Genet, 2005. **117**(6): p. 536-44.

108. Barbosa, A.C., et al., *MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function.* Proc Natl Acad Sci U S A, 2008. **105**(27): p. 9391-6.
109. Bittles, A.H., *A community genetics perspective on consanguineous marriage.* Community Genet, 2008. **11**(6): p. 324-30.
110. Molinari, F., et al., *Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation.* Science, 2002. **298**(5599): p. 1779-81.
111. Higgins, J.J., et al., *A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation.* Neurology, 2004. **63**(10): p. 1927-31.
112. Basel-Vanagaite, L., et al., *The CC2D1A, a member of a new gene family with C2 domains, is involved in autosomal recessive non-syndromic mental retardation.* J Med Genet, 2006. **43**(3): p. 203-10.
113. Motazacker, M.M., et al., *A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation.* Am J Hum Genet, 2007. **81**(4): p. 792-8.
114. Garshasbi, M., et al., *A defect in the TUSC3 gene is associated with autosomal recessive mental retardation.* Am J Hum Genet, 2008. **82**(5): p. 1158-64.
115. Mochida, G.H., et al., *A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly.* Am J Hum Genet, 2009. **85**(6): p. 897-902.
116. Raymond, F.L., *X linked mental retardation: a clinical guide.* J Med Genet, 2006. **43**(3): p. 193-200.
117. Ropers, H.H., *X-linked mental retardation: many genes for a complex disorder.* Curr Opin Genet Dev, 2006. **16**(3): p. 260-9.
118. Laumonnier, F., P.C. Cuthbert, and S.G. Grant, *The role of neuronal complexes in human X-linked brain diseases.* Am J Hum Genet, 2007. **80**(2): p. 205-20.
119. Hall, A., *Rho GTPases and the actin cytoskeleton.* Science, 1998. **279**(5350): p. 509-14.
120. Zipkin, I.D., R.M. Kindt, and C.J. Kenyon, *Role of a new Rho family member in cell migration and axon guidance in C. elegans.* Cell, 1997. **90**(5): p. 883-94.
121. Dickson, B.J., *Rho GTPases in growth cone guidance.* Curr Opin Neurobiol, 2001. **11**(1): p. 103-10.
122. Daniels, R.H., P.S. Hall, and G.M. Bokoch, *Membrane targeting of p21-activated kinase 1 (PAK1) induces neurite outgrowth from PC12 cells.* EMBO J, 1998. **17**(3): p. 754-64.
123. Bagrodia, S., et al., *Cdc42 and PAK-mediated signaling leads to Jun kinase and p38 mitogen-activated protein kinase activation.* J Biol Chem, 1995. **270**(47): p. 27995-8.
124. Salomons, G.S., et al., *X-linked creatine transporter defect: an overview.* J Inherit Metab Dis, 2003. **26**(2-3): p. 309-18.
125. Salomons, G.S., et al., *X-linked creatine-transporter gene (SLC6A8) defect: a new creatine-deficiency syndrome.* Am J Hum Genet, 2001. **68**(6): p. 1497-500.
126. Carrie, A., et al., *A new member of the IL-1 receptor family highly expressed in hippocampus and involved in X-linked mental retardation.* Nat Genet, 1999. **23**(1): p. 25-31.
127. Bhat, S.S., et al., *Disruption of the IL1RAPL1 gene associated with a pericentromeric inversion of the X chromosome in a patient with mental retardation and autism.* Clin Genet, 2008. **73**(1): p. 94-6.
128. Franek, K.J., et al., *Deletion of the immunoglobulin domain of IL1RAPL1 results in nonsyndromic X-linked intellectual disability associated with behavioral problems and mild dysmorphism.* Am J Med Genet A, 2011. **155A**(5): p. 1109-14.

129. Piton, A., et al., *Mutations in the calcium-related gene IL1RAPL1 are associated with autism.* Hum Mol Genet, 2008. **17**(24): p. 3965-74.
130. Gambino, F., et al., *IL1-receptor accessory protein-like 1 (IL1RAPL1), a protein involved in cognitive functions, regulates N-type Ca2+-channel and neurite elongation.* Proc Natl Acad Sci U S A, 2007. **104**(21): p. 9063-8.
131. Verpelli, C. and C. Sala, *Molecular and synaptic defects in intellectual disability syndromes.* Curr Opin Neurobiol, 2012. **22**(3): p. 530-6.
132. Pavlowsky, A., et al., *A postsynaptic signaling pathway that may account for the cognitive defect due to IL1RAPL1 mutation.* Curr Biol, 2010. **20**(2): p. 103-15.
133. Bagni, C. and W.T. Greenough, *From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome.* Nat Rev Neurosci, 2005. **6**(5): p. 376-87.
134. Bassell, G.J. and S.T. Warren, *Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function.* Neuron, 2008. **60**(2): p. 201-14.
135. De Rubeis, S. and C. Bagni, *Fragile X mental retardation protein control of neuronal mRNA metabolism: Insights into mRNA stability.* Mol Cell Neurosci, 2010. **43**(1): p. 43-50.
136. Antar, L.N., et al., *Local functions for FMRP in axon growth cone motility and activity-dependent regulation of filopodia and spine synapses.* Mol Cell Neurosci, 2006. **32**(1-2): p. 37-48.
137. Antar, L.N., et al., *Metabotropic glutamate receptor activation regulates fragile x mental retardation protein and FMR1 mRNA localization differentially in dendrites and at synapses.* J Neurosci, 2004. **24**(11): p. 2648-55.
138. Ferrari, F., et al., *The fragile X mental retardation protein-RNP granules show an mGluR-dependent localization in the post-synaptic spines.* Mol Cell Neurosci, 2007. **34**(3): p. 343-54.
139. Edbauer, D., et al., *Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132.* Neuron, 2010. **65**(3): p. 373-84.
140. Tarpey, P., et al., *Mutations in the DLG3 gene cause nonsyndromic X-linked mental retardation.* Am J Hum Genet, 2004. **75**(2): p. 318-24.
141. Zanni, G., et al., *A novel mutation in the DLG3 gene encoding the synapse-associated protein 102 (SAP102) causes non-syndromic mental retardation.* Neurogenetics, 2010. **11**(2): p. 251-5.
142. Chen, B.S., et al., *NMDA receptor-dependent regulation of dendritic spine morphology by SAP102 splice variants.* J Neurosci, 2011. **31**(1): p. 89-96.
143. Wilson, H.L., et al., *Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms.* J Med Genet, 2003. **40**(8): p. 575-84.
144. Manning, M.A., et al., *Terminal 22q deletion syndrome: a newly recognized cause of speech and language disability in the autism spectrum.* Pediatrics, 2004. **114**(2): p. 451-7.
145. Bonaglia, M.C., et al., *Disruption of the ProSAP2 gene in a t(12;22)(q24.1;q13.3) is associated with the 22q13.3 deletion syndrome.* Am J Hum Genet, 2001. **69**(2): p. 261-8.
146. Kernohan, K.D. and N.G. Berube, *Genetic and epigenetic dysregulation of imprinted genes in the brain.* Epigenomics, 2010. **2**(6): p. 743-63.
147. Pinkel, D., T. Straume, and J.W. Gray, *Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization.* Proc Natl Acad Sci U S A, 1986. **83**(9): p. 2934-8.

148. Sagoo, G.S., et al., *Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects.* Genet Med, 2009. **11**(3): p. 139-46.
149. Shaffer, L.G., et al., *Targeted genomic microarray analysis for identification of chromosome abnormalities in 1500 consecutive clinical cases.* J Pediatr, 2006. **149**(1): p. 98-102.
150. Christian, S.L., et al., *Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder.* Biol Psychiatry, 2008. **63**(12): p. 1111-7.
151. Pinto, D., et al., *Functional impact of global rare copy number variation in autism spectrum disorders.* Nature, 2010. **466**(7304): p. 368-72.
152. Mefford, H.C. and E.E. Eichler, *Duplication hotspots, rare genomic disorders, and common disease.* Curr Opin Genet Dev, 2009. **19**(3): p. 196-204.
153. Metzker, M.L., *Emerging technologies in DNA sequencing.* Genome Res, 2005. **15**(12): p. 1767-76.
154. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.
155. Dressman, D., et al., *Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8817-22.
156. Fedurco, M., et al., *BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies.* Nucleic Acids Res, 2006. **34**(3): p. e22.
157. Mardis, E.R., *Next-generation DNA sequencing methods.* Annu Rev Genomics Hum Genet, 2008. **9**: p. 387-402.
158. Rothberg, J.M., et al., *An integrated semiconductor device enabling non-optical genome sequencing.* Nature, 2011. **475**(7356): p. 348-52.
159. Batzoglou, S., *The many faces of sequence alignment.* Brief Bioinform, 2005. **6**(1): p. 6-22.
160. Topper, S., C. Ober, and S. Das, *Exome sequencing and the genetics of intellectual disability.* Clin Genet, 2011. **80**(2): p. 117-26.
161. Ng, S.B., et al., *Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.* Nat Genet, 2010. **42**(9): p. 790-3.
162. Hoischen, A., et al., *De novo mutations of SETBP1 cause Schinzel-Giedion syndrome.* Nat Genet, 2010. **42**(6): p. 483-5.
163. Roach, J.C., et al., *Analysis of genetic inheritance in a family quartet by whole-genome sequencing.* Science, 2010. **328**(5978): p. 636-9.
164. Lynch, M., *Rate, molecular spectrum, and consequences of human mutation.* Proc Natl Acad Sci U S A, 2010. **107**(3): p. 961-8.
165. Vissers, L.E., et al., *A de novo paradigm for mental retardation.* Nat Genet, 2010. **42**(12): p. 1109-12.
166. Najmabadi, H., et al., *Deep sequencing reveals 50 novel genes for recessive cognitive disorders.* Nature, 2011. **478**(7367): p. 57-63.
167. Krawitz, P.M., et al., *Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome.* Nat Genet, 2010. **42**(10): p. 827-9.
168. Kuhlenbaumer, G., J. Hullmann, and S. Appenzeller, *Novel genomic techniques open new avenues in the analysis of monogenic disorders.* Hum Mutat, 2011. **32**(2): p. 144-51.

169. Allen, R.C., et al., *Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation.* Am J Hum Genet, 1992. **51**(6): p. 1229-39.
170. Smith, L.M., et al., *Fluorescence detection in automated DNA sequence analysis.* Nature, 1986. **321**(6071): p. 674-9.
171. Gnirke, A., et al., *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing.* Nat Biotechnol, 2009. **27**(2): p. 182-9.
172. Vandepoele, K., et al., *A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution.* Mol Biol Evol, 2005. **22**(11): p. 2265-74.
173. Dong, X., et al., *A diverse family of GPCRs expressed in specific subsets of nociceptive sensory neurons.* Cell, 2001. **106**(5): p. 619-32.
174. Aryee, D.N., et al., *Cloning of a novel human ELF-1-related ETS transcription factor, ELFR, its characterization and chromosomal assignment relative to ELF-1.* Gene, 1998. **210**(1): p. 71-8.
175. Yamada, T., et al., *Transcription factor ELF4 controls the proliferation and homing of CD8+ T cells via the Kruppel-like factors KLF4 and KLF2.* Nat Immunol, 2009. **10**(6): p. 618-26.
176. Lee, C., A.J. Iafrate, and A.R. Brothman, *Copy number variations and clinical cytogenetic diagnosis of constitutional disorders.* Nat Genet, 2007. **39**(7 Suppl): p. S48-54.
177. Nawara, M., et al., *Novel mutation of IL1RAPL1 gene in a nonspecific X-linked mental retardation (MRX) family.* Am J Med Genet A, 2008. **146A**(24): p. 3167-72.
178. Honda, S., et al., *Copy-number variations on the X chromosome in Japanese patients with mental retardation detected by array-based comparative genomic hybridization analysis.* J Hum Genet, 2010. **55**(9): p. 590-9.
179. Melhem, N., et al., *Copy number variants for schizophrenia and related psychotic disorders in Oceanic Palau: risk and transmission in extended pedigrees.* Biol Psychiatry, 2011. **70**(12): p. 1115-21.
180. Guilmatre, A., et al., *Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation.* Arch Gen Psychiatry, 2009. **66**(9): p. 947-56.
181. Behnecke, A., et al., *Intragenic deletions of IL1RAPL1: Report of two cases and review of the literature.* Am J Med Genet A, 2011. **155A**(2): p. 372-9.
182. Lepretre, F., et al., *Dissection of an inverted X(p21.3q27.1) chromosome associated with mental retardation.* Cytogenet Genome Res, 2003. **101**(2): p. 124-9.
183. Tabolacci, E., et al., *A truncating mutation in the IL1RAPL1 gene is responsible for X-linked mental retardation in the MRX21 family.* Am J Med Genet A, 2006. **140**(5): p. 482-7.
184. Mikhail, F.M., et al., *Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders.* Am J Med Genet A, 2011. **155A**(10): p. 2386-96.
185. Youngs, E.L., et al., *IL1RAPL1 gene deletion as a cause of X-linked intellectual disability and dysmorphic features.* Eur J Med Genet, 2012. **55**(1): p. 32-6.
186. Laget, S., et al., *The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA.* PLoS One, 2010. **5**(8): p. e11982.
187. Wagenstaller, J., et al., *Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation.* Am J Hum Genet, 2007. **81**(4): p. 768-79.

188. Jaillard, S., et al., *2q23.1 microdeletion identified by array comparative genomic hybridisation: an emerging phenotype with Angelman-like features?* J Med Genet, 2009. **46**(12): p. 847-55.
189. Williams, S.R., et al., *Haploinsufficiency of MBD5 associated with a syndrome involving microcephaly, intellectual disabilities, severe speech impairment, and seizures.* Eur J Hum Genet, 2010. **18**(4): p. 436-41.
190. van Bon, B.W., et al., *The 2q23.1 microdeletion syndrome: clinical and behavioural phenotype.* Eur J Hum Genet, 2010. **18**(2): p. 163-70.
191. Chung, B.H., et al., *2q23 de novo microdeletion involving the MBD5 gene in a patient with developmental delay, postnatal microcephaly and distinct facial features.* Am J Med Genet A, 2011. **155A**(2): p. 424-9.
192. Chung, B.H., et al., *Severe intellectual disability and autistic features associated with microduplication 2q23.1.* Eur J Hum Genet, 2012. **20**(4): p. 398-403.
193. Noh, G.J. and J.M. Graham, Jr., *2q23.1 microdeletion of the MBD5 gene in a female with seizures, developmental delay and distinct dysmorphic features.* Eur J Med Genet, 2012. **55**(5): p. 354-7.
194. Talkowski, M.E., et al., *Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder.* Am J Hum Genet, 2011. **89**(4): p. 551-63.
195. Kleefstra, T., et al., *Disruption of an EHMT1-Associated Chromatin-Modification Module Causes Intellectual Disability.* Am J Hum Genet, 2012. **91**(1): p. 73-82.
196. Lam, H.Y., et al., *Performance comparison of whole-genome sequencing platforms.* Nat Biotechnol, 2012. **30**(1): p. 78-82.
197. Roach, J.C., et al., *Analysis of genetic inheritance in a family quartet by whole-genome sequencing.* Science, 2010. **328**(5978): p. 636-9.
198. Conrad, D.F., et al., *Variation in genome-wide mutation rates within and between human families.* Nat Genet, 2011. **43**(7): p. 712-4.
199. Londin, E.R., et al., *Whole-exome sequencing of DNA from peripheral blood mononuclear cells (PBMC) and EBV-transformed lymphocytes from the same donor.* BMC Genomics, 2011. **12**: p. 464.
200. Lindhurst, M.J., et al., *A mosaic activating mutation in AKT1 associated with the Proteus syndrome.* N Engl J Med, 2011. **365**(7): p. 611-9.
201. Buhler, E.M., N.J. Malik, and M. Alkan, *Another model for the inheritance of Rett syndrome.* Am J Med Genet, 1990. **36**(1): p. 126-31.

Liste des publications:

- Bonnet, C., **Khan, A. A**., Bresso E., Vigouroux C., Béri-Dexheimer, M., Lejczak S., Deemer B., Andrieux J., Philippe, C., Moncla A., Giurgea I., Leheup B., and Jonveaux, P. *Extended spectrum of MBD5 mutations in neurodevelopmental disorders*.
  Article Soumis dans European Journal of Human Genetics

- Bonnet, C., Masurel-Paulet, A., **Khan, A. A**., Béri-Dexheimer, M., Callier, P., Mugneret, F., Philippe, C., Thauvin-Robinet, C., Faivre, L. and Jonveaux, P. (2012), *Exploring the potential role of disease-causing mutation in a gene desert: Duplication of noncoding elements 5′ of GRIA3 is associated with GRIA3 silencing and X-linked intellectual disability*. Hum. Mutat., 33: 355–358. doi: 10.1002/humu.21649

- Huynh, M. T., Béri-Dexheimer, M., Bonnet, C., Bronner, M., **Khan, A. A.**, Allou, L., Philippe, C., Vigneron, J. and Jonveaux, P. (2012), *RUNX1T1, a chromatin repression protein, is a candidate gene for autosomal dominant intellectual disability*. Am. J. Med. Genet., 158A: 1782–1784. doi: 10.1002/ajmg.a.35386

Liste des posters:

- **Ali Khan A**, Béri-Dexheimer M, Valduga M, Vibert M, Thauvin-Robinet C, Callier P, Mugneret F, Amsallem D, Roze V, Bonnet C, Jonveaux P. *Microrearrangements génomiques de IL1RAPL1 et déficience intellectuelle liée au chromosome X.* « 6ème Assises de génétique humaine et médicale » Marseille, 2012.

- Huynh MT, Béri-Dexheimer M, Bonnet C, Bronner M, **Ali Khan A**, Allou L, Philippe C, Vigneron J, Jonveaux P. *Les mutations du gène RUNX1T1, acteur de la répression chromatinienne, sont associées à une déficience intellectuelle autosomique dominante modérée à légère.* « 6ème Assises de génétique humaine et médicale » Marseille, 2012.

- Bonnet C, **Ali Khan A**, Nemos C, Devignes MD, Jonveaux P. *Recherche du gène responsable du syndrome d'Aicardi.* « Conférence Jacques Monod, Le retard mental : des gènes aux synapses, fonctions et dysfonctions » Roscoff, 2010.

- **Ali Khan A**, Bursztejn AC, Philippe C, Jonveaux P, Nemos C. *Validation des gènes candidats impliqués par array CGH dans le syndrome d'Aicardi.* **«** 5ème Assises de génétique Humaine et Médicale » Strasbourg, 2010

ANNEXE I

Extended spectrum of *MBD5* mutations in neurodevelopmental disorders

Céline Bonnet[1,8], Asma Ali Khan[1,8], Emmanuel Bresso[2], Charlène Vigouroux[1], Mylène Béri[1], Sarah Lejczak[1], Bénédicte Deemer[3], Joris Andrieux[4], Christophe Philippe[1], Anne Moncla[5], Irina Giurgea[6], Bruno Leheup[1,7], Philippe Jonveaux[1]

(1) Laboratoire de Génétique, EA 4368, Université de Lorraine, Centre Hospitalier Universitaire de Nancy, Vandoeuvre les Nancy, France

(2) LORIA UMR7503, CNRS, INRIA, Nancy-Université, BP239, 54506 Vandoeuvre-les-Nancy cedex, France

(3) Service de Pédiatrie et Génétique, Hôpital Nord, Amiens, France

(4) Laboratoire de Génétique Médicale, Hôpital Jeanne de Flandre, CHU de Lille, Lille, France

(5) Département de Génétique Médicale, Hôpital d'Enfants de la Timone, Marseille, France

(6) Service de Biochimie-Génétique, APHP, Groupe hospitalier Henri Mondor, Créteil, France

(7) Service de Médecine Infantile III et génétique clinique, Hôpital d'enfants, Centre Hospitalier Universitaire de Nancy, Vandoeuvre les Nancy, France

(8) These authors contributed equally to this work

Corresponding author: Philippe Jonveaux, MD, PhD

Laboratoire de Génétique, EA 4368, Université de Lorraine, CHU de Nancy, rue de Morvan, 54511 Vandoeuvre les Nancy, France.

Tel : 33 3 83 15 37 71, Fax : 33 3 83 15 37 72, E-mail : p.jonveaux@chu-nancy.fr

Key words: *MBD5*, nonsense mutation, intragenic duplication, intellectual disability

ABSTRACT

Intellectual disability (ID) is a clinical sign reflecting diverse neurodevelopmental disorders that are genetically and phenotypically heterogeneous. Just recently, partial or complete deletion of *MBD5* gene has been implicated as causative in the phenotype associated with 2q23.1 microdeletion syndrome. In the course of systematic whole genome screening of individuals with unexplained ID by array-based Comparative Genomic Hybridization, we identified *de novo* intragenic deletions of *MBD5* in three patients leading, as previously documented, to haploinsufficiency of *MBD5*. In addition, we described a patient with an unreported de novo *MBD5* intragenic duplication. RT-PCR and sequencing analyses showed the presence of numerous aberrant transcripts leading to premature termination codon. To further elucidate the involvement of *MBD5* in ID, we sequenced the 10 coding, 5 non-coding exons and an evolutionary conserved region in intron2, in a selected cohort of 78 subjects with a phenotype reminiscent of 2q23.1 microdeletion syndrome. Besides variants most often inherited from an healthy parent, we identified for the first time a *de novo* nonsense mutation associated with a much more damaging phenotype. Taken together, these results extend the mutation spectrum in *MBD5* gene and contribute to refine the associated phenotype of neurodevelopmental disorder.

INTRODUCTION

MBD5 protein (OMIM *611472) is a member of the methyl-CpG-binding domain (MBD) protein family in which MECP2 (OMIM *300005) is involved in Rett Syndrome, a prototypical neurodevelopmental disorder. *MBD5* contains 5 noncoding exons at its 5-prime end, followed by 10 coding exons. Two isoforms have been described[1], the longer one contains 1494 amino-acids and is encoded by exons 6 to 15, the second one contains 851 amino-acids and is encoded by exons 6 to 9. Functional studies suggested that MBD5 is likely to contribute to the formation or function of heterochromatin[1]. Isoform 1 of *MBD5* is highly expressed in brain and testis and isoform 2 is highly expressed in oocytes, which suggests a possible role in cerebral functions and in epigenetic reprogramming after fertilization. Recently, deletions encompassing *MBD5*, as well as intragenic *MBD5* deletions have been identified in individuals with a phenotype of intellectual disability, seizures, significant speech impairment, and behavioral problems[2,3,4,5,6,7,8]. In this study, we used pangenomic array-Comparative Genomic hybridization (array-CGH), capillary sequencing of *MBD5* gene to investigate DNAs from patients with unexplained intellectual disability. We further extend the mutational spectrum of *MBD5* with damaging intragenic duplication and nonsense mutation associated with a clinical spectrum of neurodevelopmental disorder.

SUBJECTS AND METHODS

Ascertainment of the patients

The patients took part in a clinical diagnostic testing for genomic imbalance using array-CGH following initial testing for karyotype (results normal) thanks to the national aCGH network funded by the French Ministry of Health. To further elucidate the involvement of *MBD5* point mutations, we collected a clinically defined cohort of 78 individuals with moderate to severe intellectual disability without a known genetic cause (genomic copy-number variants larger

than 200 kb were previously excluded) and with significant clinical overlap with 2q23.1 deletion syndrome, reminiscent of Angelman-like phenotype or Smith-Magenis-like syndrome. Informed consents were available for all tested patients.

Array-Comparative Genomic Hybridization Analysis

Microarray-CGH analysis was carried out using 44K or 105K-oligonucleotide array (Agilent Technologies, Santa Clara, CA) as previously described[9]. The array was analyzed with the Agilent scanner and the Feature Extraction software (v9.5.3.1). A graphical overview was obtained using the CGH analytics software (v3.5.14) (Agilent Technologies, Santa Clara, CA).

Genomic quantitative PCR

Quantitative PCR (qPCR) was performed on genomic DNA, using an ABI PRISM 7500 Sequence Detection System (Applied Biosystems, Foster City, CA). We designed primer sets in *MBD5* gene (all primer sequences used in this study are available on request). qPCR was carried out as previously described[9]. The *RPPH1* gene was selected as the control amplicon. Validation experiments demonstrated that amplification efficiency of the control and all target amplicons were approximately equal. All samples were run in triplicate. The dosage of each amplicon relative to *RPPH1* and normalized to control male DNA was determined using the $2^{-\Delta\Delta Ct}$ method.

Genomic sequencing

MBD5 10 coding, 5 non-coding exons (NM_018328) and one ECR (evolutionary conserved region) in intron 2 were PCR amplified using standard procedures (available on request). PCR products were then purified and subjected to sequencing using BigDye Terminator kit (Applied Biosystems, Foster City, CA).

mRNA isolation, RT-qPCR

Total RNAs were isolated from PaxGen blood RNA tubes using RNeasy® mini kit (Qiagen). RNA was reverse transcribed through the use of random primers (Superscript, Invitrogen). Reverse transcriptase quantitative real-time PCR (RT-qPCR) was performed on an ABI PRISM 7500 Sequence Detection System (Applied Biosystems). We designed primer sets within *MBD5* (available on request). RT-qPCR was carried out in a total volume of 20 µl containing 10 µl of SYBR Green Master Mix (Applied Biosystems), 0.4 mM of each primer and 5 µl of cDNA. Thermal cycling conditions were 95°C for 20 sec, followed by 40 cycles with 95°C for 3 sec and 60°C for 30 sec. The *ESD* and *ABL* genes were selected as control amplicons. Validation experiments demonstrated that amplification efficiency of control and all target amplicons were approximately equal. All samples were run in triplicate. The dosage of each amplicon relative to *ESD* and *ABL* and normalized to control male cDNA was determined using the $2^{-\Delta\Delta Ct}$ method.

cDNA sequencing

Primers were selected in *MDB5* exons. RT-PCR products were electrophoresed on agarose gels, purified with NucleoSpin Extract II kit (Macherey-Nagel, SARL) and sequenced using BigDye Terminator kit (Applied Biosystems).

RESULTS

Clinical reports

Patients A and B are monozygotic twin sisters. The father and the 2 first siblings are healthy. The mother was treated for epilepsy but treatment was interrupted during pregnancies. They were born prematurely without foetal distress. Z scores of birth weight and length were at -1 and head circumference was in the normal range. They were noted to have global developmental delay. Patient B sat independently at 16 months and walked at 2 years 6 months, patient C sat independently at 17 months and walked at 3 years. Both spoke only

single words and presented with stereotypies and autistic features. A brain MRI was normal. At the age of 3 years 6 months, both heights were at -3 SD whereas weights and head circumferences were in the normal range. There is isolated nostril anteversion on facial examination (Figure 1A).

Patient C. This male proband was born following an uncomplicated, full term pregnancy. Parents were non-consanguineous and healthy. Family history is otherwise unremarkable. Neonatal adaptation was normal. Birth weight (3300 g), birth length (49 cm) and head circumference (37 cm) were within the normal range. He presented with hypospadias and developed multiple bronchiolitis. He was operated on inguinal hernia. He was noted to have global developmental delay. He walked at the age of 22 months, and language milestones were delayed. At the age of 4 years, height was 99 cm (median), weight 15kg (median) and head circumference 52 cm (+1DS). No particular facial features were reported. Although his intelligence had not been formally evaluated, his intellectual disability was estimated to be mild to moderate.

Patient D is the only child of healthy nonconsanguineous parents. Pregnancy was uneventful to the exception of haemorrhage related to partial placental detachment at 3 months of gestation. She was born at term with normal growth parameters. Initial developmental milestones were reported normal. She walked at 19 months. The first words were pronounced at 13 months. Between 24 and 30 months of age, regression of language skills occurred with concomitant regression of response to social overture. She gradually developed problematic behavior with periods of hyperactivity and attention deficit. She was seen at the neuropediatric department at 2 years and 4 months and at 3 years and 6 months. There was no motor deficit. Slight symptoms of cerebelar syndrome were noted with oral dyspraxia. She was also seen at the outpatient genetics clinic at 3 years and 11 months. Growth parameters

were within the normal range and clinical examination showed a round face, nostril anteversion and down-turned corners of the mouth (Figure 1A).

Patient E is a 10-year-old boy first seen at the age of 14 months because of developmental delay. He is the second of three children of healthy unconsanguineous parents. Pregnancy was reported as normal. He was born at 40 weeks by caesarean section because of placenta praevia. Neonatal adaptation was normal. Birth weight (2,560 g), birth length (48 cm) and head circumference (34.5 cm) were in the low normal range. Since the first days of life, parents reported feeding difficulties. The boy developed opisthotonos during the first months of life. Unmotivated laughter was also reported. When first seen at 14 months sitting was unstable, hand movements were poor, and language was absent. Eye contact was reported as easy. Some jerky movements were described. Length, weight, and head circumference were at the 50th percentile. Craniofacial examination was not specific with slightly broad forehead. EEG and cerebral MRI were normal. A screen for metabolic abnormalities and methylation analysis for Angelman syndrome were normal. *UBE3A* gene analysis (Dr Moncla, Marseille) was normal. At the age of 2 years 7 months there was still no verbal language. Hypotonia was severe without walking. He was last seen at the age of 10 years 3 months. Height was at 147.5 cm (+2SD), weight and head circumference were in the normal range. He stood independently for a short period of time but did not walk. There was no verbal language. His parents reported him as happy with very frequent smiles. Craniofacial examination showed unspecific hypotonic characteristics with long face, open mouth and slightly everted lower lip. Ear lobules were large (Figure 1A). He presented with generalized tonico-clonic seizures at the age of 4 years. Treatment with valproate was initiated. At the age of 8 years clonus of both legs were reported associated with tongue and mouth clonus. EEG reported focal spikes and spike-wave complexes in the frontal and temporal left area leading to the diagnosis of partial epilepsy.

Molecular investigations

Array-Comparative Genomic Hybridization analysis demonstrated, according to UCSC build 36/hg18 (Figure 1B-1C): (i) In patients A and B, an interstitial deletion at 2q23.1: arr 2q23.1(148,447,496-148,515,776)x1 with a minimal size of 68,280 bp. The region includes the end of *ORC4* and the 2 first non-coding exons of *MBD5*; (ii) In patient C, an interstitial duplication at 2q23.1: arr 2q23.1(148,944,718-148,979,574)x3 with a minimal size of 34,856 bp. This duplication affects only *MBD5*, the minimal duplicated region including 4 exons (5 to 8) and the maximal region including 11 exons (nc5 to 10); (iii) In patient D, an interstitial deletion at 2q23.1: arr 2q23.1(148,496,551-148,515,776)x1 with a minimal size of 19,225 bp including the end of *ORC4* and the 2 first non-coding exons of *MBD5*. This region had never been described as a copy number polymorphism (CNP) in the database of genomic variants (http://projects.tcag.ca/variation/?source=hg18). Except for polymorphic regions, no copy number alterations were observed in other chromosomes. Using qPCR analysis on genomic DNA from patients A, B, D and their respective parents, we confirmed the biological relationships and revealed that genomic imbalances arose *de novo*. For patient C, parental DNAs were not available. However, qPCR on his genomic DNA allowed determining more precisely the extent of the duplication from non-coding exon 5 to coding exon 10. Thanks to Sanger sequencing, we screened *MBD5* for point mutations in the selected cohort of 78 individuals with ID. We identified a nonsense mutation [c.440C>G (p.Ser147*); NM_018328.3] within coding exon 4 in patient E (Figure 2A). Analysis of parental DNA confirmed the biological relationships and *de novo* occurrence of the mutation. In this series of patients, we also detected 9 variants in protein-coding exons, not annotated in dbSNP (build 137), 3 intronic variants, 3 synonymous variants and 3 missense variants. In ECR and non-coding exons, 5 different variations were found. Detailed sequencing results are

displayed in supplementary tables I and II. When parental material was available, we could show segregation from a healthy parent in all cases. RT-qPCR analysis showed (Figure 2B): (i) a notable reduction of *MBD5* expression for both sisters A and B and for patient D (ii) a significantly increased level for duplicated MBD5 exons in patient C, and (iii) a normal level of expression for patient E. RT-PCR analysis in patient C, with forward primers in coding exons 8, 9 and 10 and reverse primer in exon nc5, coding exons 1 and 2 of *MBD5,* amplified different aberrant transcripts. Sequencing analysis of these fragments (Figure 2C) showed three aberrant transcripts leading to premature termination codon. For patient E, RT-PCR and sequencing analysis of exon 4 showed that both normal and mutated alleles were expressed (figure 2A).

DISCUSSION

Recently, Talkowski et al, 2011[10] suggested a mixed model of deleterious, fully penetrant *MBD5* deletions causing a neurodevelopmental disorder associated with features of 2q23.1 microdeletion syndrome, and reduced penetrance missense variants that significantly increase risk for autism spectrum disorder (ASD). In our work, we identified 3 patients (A, B and D) with a deletion including the last exons of *ORC4* and only the two first untranslated exons of the brain-expressed isoform 1 of *MBD5*. A similar deletion has been reported[5,10]. Expression level of *MBD5* mRNA in patients A and B was significantly reduced in comparison to their non-deleted parents, sister and brother. This result proves that heterozygous deletion of the two first non-coding exons of *MBD5* isoform 1 specifically leads to extinction of its expression on deleted allele. Interestingly, two novel *MBD5* genetic alteration types were identified, an intragenic duplication and a nonsense mutation. Patient C intragenic duplication affects non-coding exon 5 to coding exon 10 of *MBD5*. Transcriptional study showed the presence of numerous aberrant transcripts leading to premature termination codon. This

partial duplication is predicted to affect *MBD5* protein function. Patient E *de novo* nonsense mutation leads to premature termination codon in *MBD5* gene and is predicted to result in a truncated protein that lacks the Proline-rich domain in addition to the putative nuclear localization signal. This mutation was reported neither in the 1000 Genomes project (http://browser.1000genomes.org/) nor in dbSNP (build 137). RT-PCR analysis showed a normal level of expression of *MBD5* suggesting that RNA decay did not occur. Notably, *MBD5* transcripts sequencing showed *in vivo* expression of both normal and mutated transcripts. Translation of this mutated transcript might lead to a truncated protein with a dominant negative effect or this aberrant protein might be eliminated by cellular protein quality control systems leading to haploinsufficiency. Regarding *MBD5* point mutations, missense variants have been reported[2,10], mainly inherited from a healthy parent. They are hypothesized as a risk factor for ASD[10]. We also identified in our series of patients (supplementary table I), the two previously reported missense variants[10] [p.79Gly>Glu (rs34995577) in patient 51 and p.1048Thr>Ile in patient 33]. Just recently, a *MBD5* frameshift mutation [c.150del (p.Thr52His*fs**31); NM_018328.4] resulting in a premature stop codon has been reported in a patient with Kleefstra syndrome phenotypic spectrum[11]. If the phenotype of patient C with intragenic duplication overlaps with that observed in patients A, B, and D with intragenic deletion, the phenotype of patient E with *MDB5* nonsense mutation is much more damaging without walking and verbal speech at the age of 10 years. This fully penetrant mutation represents 1.2% (1/78) of our selected cohort.

In conclusion, these findings confirm the involvement of MBD5 mutations in neurodevelopmental disorders and extent the mutational spectrum of *MBD5*. Additional observations will be needed to establish fine genotype-phenotype correlations.

Acknowledgements

The authors thank the patients and their family for their kind cooperation. We thank the cytogenetics and molecular genetics staff at the Nancy University Hospitals for their expert technical assistance. This study was supported by grants from the French Ministry of Health (DGOS) and the "Fondation Jérôme Lejeune".

REFERENCES

1. Laget S, Joulie M, Le Masson F *et al*:. The Human Proteins MBD5 and MBD6 Associate with Heterochromatin but They Do Not Bind Methylated DNA. *PLoS One* 2010; **5**: e11982.

2. Wagenstaller J, Spranger S, Lorenz-Depiereux B *et al*:. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. *Am J Hum Genet* 2007; **81**, 768-779.

3. Jaillard S, Dubourg C, Gerard-Blanluet M *et al*:. 2q23.1 microdeletion identified by array comparative genomic hybridisation: an emerging phenotype with Angelman-like features? *J Med Genet* 2009; **46**: 847-855.

4. Williams SR, Mullegama SV, Rosenfeld JA *et al*:. Haploinsufficiency of MBD5 associated with a syndrome involving microcephaly, intellectual disabilities, severe speech impairment, and seizures. *Eur J Hum Genet* 2010; **18**: 436-441.

5. Van Bon BW, Koolen DA, Brueton L *et al*. The 2q23.1 microdeletion syndrome: clinical and behavioural phenotype. *Eur J Hum Genet* 2010; **18**: 163-170.

6. Chung BH, Stavropoulos J, Marshall CR *et al*: 2q23 de novo microdeletion involving the *MBD5* gene in a patient with developmental delay, postnatal microcephaly and distinct facial features. *Am J Med Genet A* 2011; **155**: 424-429.

7. Noh GJ, Graham Jr JM. 2q23.1 microdeletion of the *MBD5* gene in a female with seizures, developmental delay and distinct dysmorphic features. *Eur J Med Genet* 2012; **55**: 354-357

8. Chung BH, Mullegama S, Marshall CR *et al*:. Severe intellectual disability and autistic features associated with microduplication 2q23.1. *Eur J Hum Genet* 2012 ; **20**: 398-403.

9. Bonnet C, Masurel-Paulet A, Khan AA *et al*:. Exploring the potential role of disease-causing mutation in a gene desert: duplication of noncoding elements 5' of GRIA3 is associated with GRIA3 silencing and X-linked intellectual disability. *Hum Mutat* 2012; **33**: 355-358.

10. Talkowski ME, Mullegama SV, Rosenfeld JA *et al*:. Assessment of 2q23.1 microdeletion syndrome implicates *MBD5* as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. *Am J Hum Genet 2011*; **89**: 551-563.

11. Kleefstra T, Kramer JM, Neveling K *et al:* Disruption of an EHMT1-Associated Chromatin-Modification Module Causes Intellectual Disability. *Am J Hum Genet* 2012; **91**: 1-10

LEGENDS TO FIGURES

Figure 1:

(A) Frontal and lateral views of patients A, B, D and E, without specific facial features.

(B) 105 K array-based CGH results showing the extent of *MBD5* intragenic deletion in patients A and B (7 probes) and for patient D (3 probes) and the extent of intragenic duplication for patient E (3 probes using 44 K array CGH)

(C) Map of genomic alterations: deletions, duplication and nonsense mutation (snapshot of Database of Genomic Variants (http://projects.tcag.ca/variation/)

Figure 2:

(A) Left panel: Genomic sequencing results for patient E, his brother and parents showing (arrowhead) a *de novo* nonsense mutation in exon 4 (NM_018328.3:c.440C>G) Right panel: comparison of DNA and cDNA sequencing results in patient E showing that both normal and mutated alleles are expressed (arrowhead).

(B) RT-qPCR results with primer set 1(*MBD5* coding exons 1-2) and primer set 2 (*MBD5* coding exons 6-7): left panel, a normal level of expression of *MBD5* in patient E compared to that four controls, middle panel and right panel, a decreased level of expression of *MBD5* in patients A and B (compared to that healthy sister and parents), and in patient D (compared to that two controls), respectively.

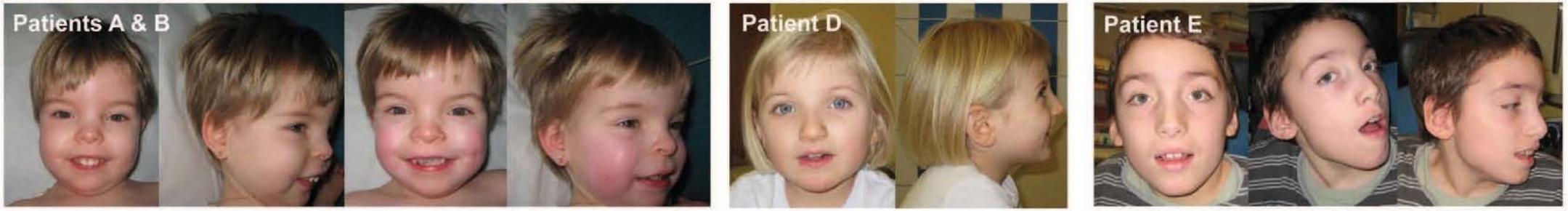
(C) Aberrant transcripts (1 to 6) characterized by RT-PCR and sequencing analysis for patient C. First column : transcript number, second column : primers used for amplification of the aberrant transcript, third column : size of the amplified fragment, 4th column : exon structure of the amplified aberrant transcript, red stars indicate position of termination codon.

Supplementary table legends

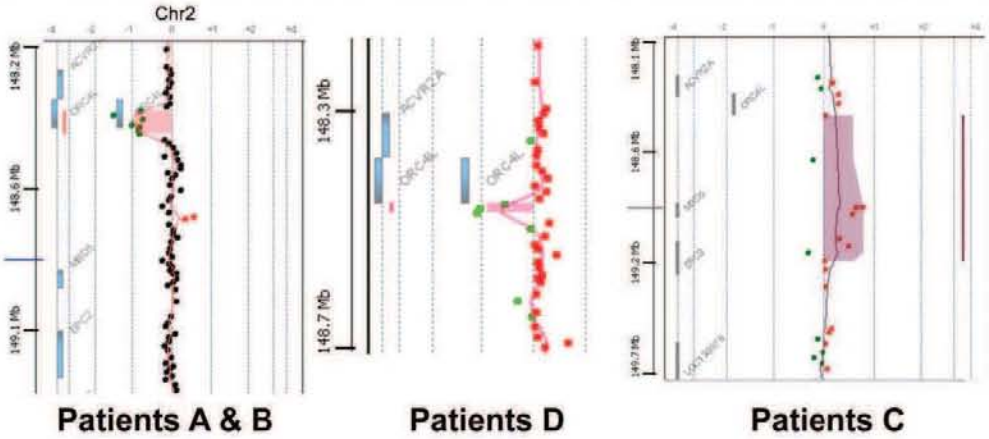
Table I: Variants detected in coding exons (1 to 10)

Table II: Variants detected in ECR and non-coding exons (nc1 to nc5)

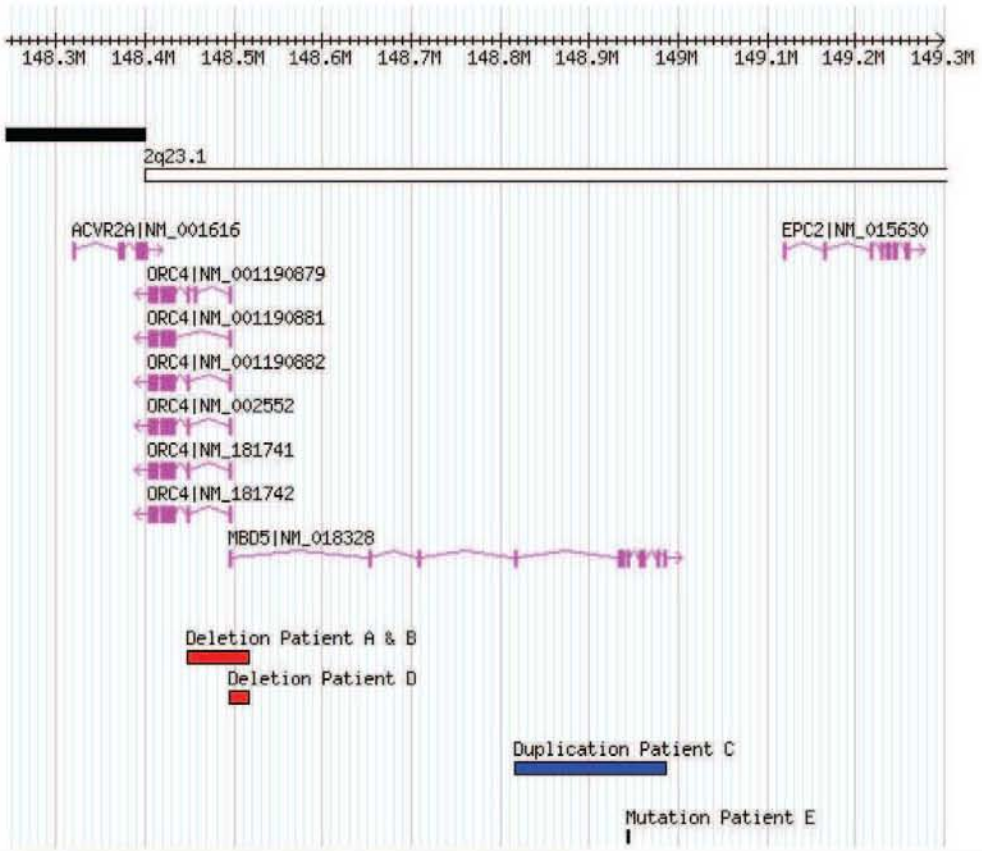
A

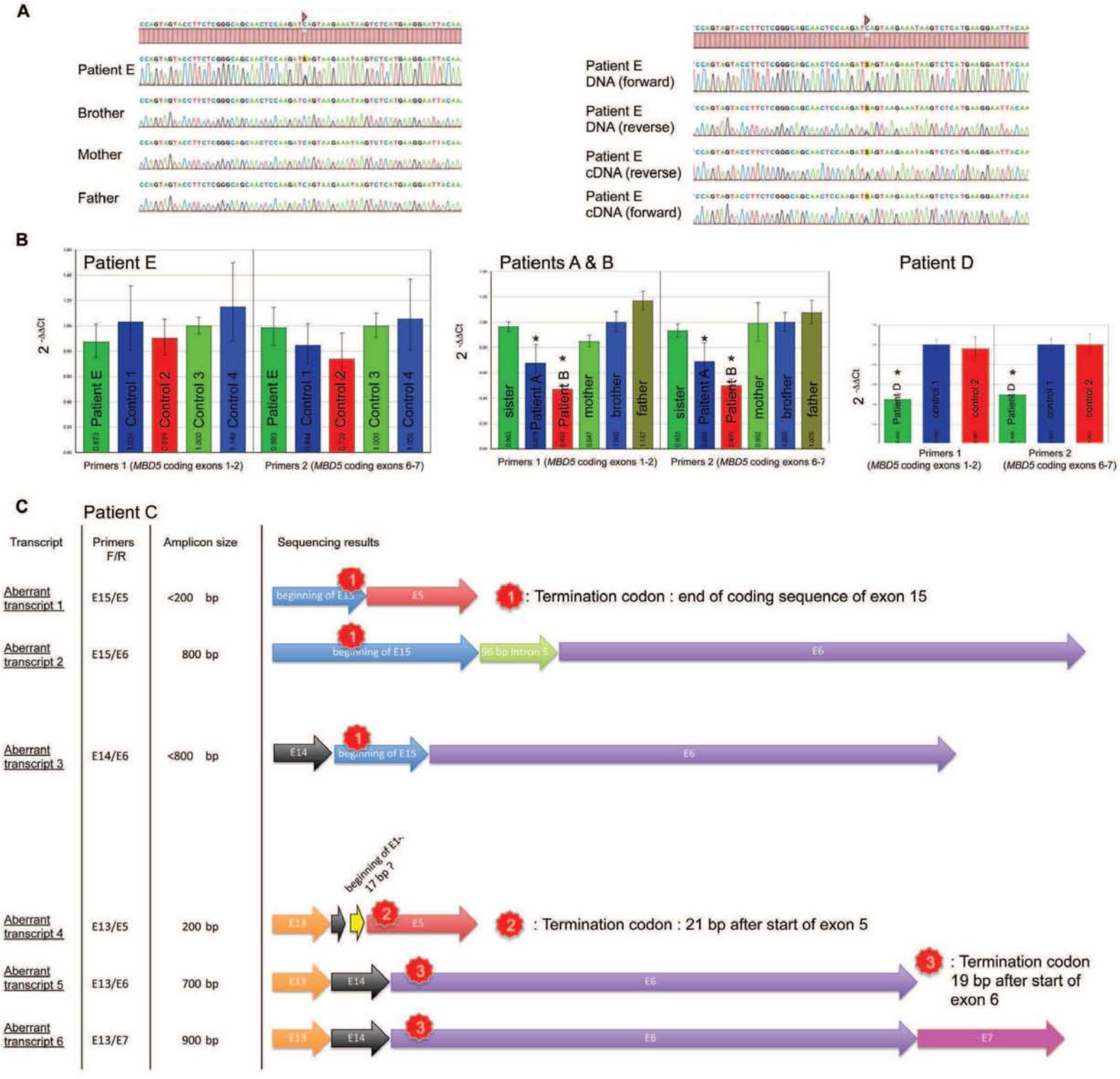

Patients A & B
Patient D
Patient E

B


Chr2
148.3 Mb
148.7 Mb
Patients A & B
Patients D
Patients C

C


148.3M
148.4M
148.5M
148.6M
148.7M
148.8M
148.9M
149M
149.1M
149.2M
149.3M
2q23.1
ACVR2A|NM_001616
EPC2|NM_015630
ORC4|NM_001190879
ORC4|NM_001190881
ORC4|NM_001190882
ORC4|NM_002552
ORC4|NM_181741
ORC4|NM_181742
MBD5|NM_018328
Deletion Patient A & B
Deletion Patient D
Duplication Patient C
Mutation Patient E


A
Patient E
Brother
Mother
Father
Patient E DNA (forward)
Patient E DNA (reverse)
Patient E cDNA (reverse)
Patient E cDNA (forward)
B
Patient E
2 -ΔΔCt
Patient E
Control 1
Control 2
Control 3
Control 4
Primers 1 (MBD5 coding exons 1-2)
Primers 2 (MBD5 coding exons 6-7)
Patients A & B
sister
Patient A
Patient B
mother
brother
father
Patient D
Patient D
control 1
control 2
C
Patient C
Transcript
Primers F/R
Amplicon size
Sequencing results
Aberrant transcript 1
E15/E5
<200 bp
beginning of E15
E5
: Termination codon : end of coding sequence of exon 15
Aberrant transcript 2
E15/E6
800 bp
96 bp Intron 5
E6
Aberrant transcript 3
E14/E6
<800 bp
E14
Aberrant transcript 4
E13/E5
200 bp
beginning of E1.. 17 bp ?
E13
: Termination codon : 21 bp after start of exon 5
Aberrant transcript 5
E13/E6
700 bp
: Termination codon 19 bp after start of exon 6
Aberrant transcript 6
E13/E7
900 bp
E7

SUPPLEMENTARY TABLE LEGENDS

Table I: Variations detected in coding exons (1 to 10)

Table II: Variations detected in ECR and non-coding exons (nc1 to nc5)

Table I:

Patient ID	Exon/Intron	Variation	Protein change	Type
Patient 1	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Intron 6	NM_018328.3:c.3055-48dup		
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 2	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 3	Intron 2	NM_018328.3:c.216+165G>A		intronic
	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 4 (A)	Exon 4	NM_018328.3:c.440C>G	p.Ser147X	Premature Stop Codon (PTC), *de novo*
	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 5	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.[*200A>G]+[*200A>G]		SNP rs16828708
Patient 6	Intron 6	NM_018328.3:c.3055-48dup		
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 7	Intron 6	NM_018328.3:c.3055-48dup		
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 8	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Intron 6	NM_018328.3:c.3055-48dup		
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 9	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 10	Intron 6	NM_018328.3:c.3055-48dup		
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 11	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Intron 6	NM_018328.3:c.3055-48dup		
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 12	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 13	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 14	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 15	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 16	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 17	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 18	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 19	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 20	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 21	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 22	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 23	Intron 6	NM_018328.3:c.[3055-48dup]+[3055-48dup]		
	Intron 7	NM_018328.3:c.[4264-56G>A]+ [4264-56G>A]		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 24	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 25	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 7	NM_018328.3:c.4251C>T	p.=	synonymous
Patient 26	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 27	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 28	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 29	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344

Patient ID	Exon/Intron	Variation	Protein change	Type
Patient 30	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 31	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 32				
Patient 33	Intron 6	NM_018328.3:c.3055-48dup		
	Exon 7	NM_018328.3:c.3143C>T	p.Thr1048>Ile	missense, maternal (0/660[1])
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 34	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 35	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 36	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 37	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 38	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Intron 5	NM_018328.3:c.2846-43G>A		intronic, paternal
Patient 39	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 40	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 41	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 42	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 43	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Intron 6	NM_018328.3:c.3055-48dup		
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 44	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Intron 6	NM_018328.3:c.3055-48dup		
	Exon 7	NM_018328.3:c.3531C>T	p.=	synonymous
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
Patient 45	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 46	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 47	Exon 4	NM_018328.3:c.2030G>A	p.Ser677>Asn	missense (3/660[1])
	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Intron 6	NM_018328.3:c.3055-48dup		
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 48	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 49	5'UTR	NM_018328.3:c.-240C>T		intronic, maternal
	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 50	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 51	Exon 3	NM_018328.3:c.236G>A	p.79Gly>Glu	SNP rs34995577
	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 52	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 53	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 54	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.[*200A>G]+ [*200A>G]		SNP rs16828708
Patient 55	Intron 6	NM_018328.3:c.3055-48dup		
	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
Patient 56	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 57	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 58	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344

[1] Wagenstaller et al., 2007

Patient ID	Exon/Intron	Variation	Protein change	Type
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 59	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 60	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	Exon 10	NM_018328.3:c.[*200A>G]+ [*200A>G]		SNP rs16828708
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 61	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 62	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 63	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 64	Exon 4	NM_018328.3:c.2254A>G	p.Ile752Val	missense, maternal
	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 65	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 66	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 67	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 68	Exon 3	NM_018328.3:c.297A>G	p.=	synonymous
	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 69	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
Patient 70				
Patient 71	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 72	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 73	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.*200A>G		SNP rs16828708
Patient 74	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344
	Exon 10	NM_018328.3:c.[*200A>G]+[*200A>G]		SNP rs16828708
Patient 75	Intron 7	NM_018328.3:c.4264-56G>A		SNP rs73003524
	3'UTR	NT_022135.15:g.37979066T>G		SNP rs16828718
Patient 76				
Patient 77	Intron 5	NM_018328.3:c.[2846-51G>A]+[2846-51G>A]		SNP rs2121344
Patient 78	Intron 5	NM_018328.3:c.2846-51G>A		SNP rs2121344

Table II:

Patient ID	Exon/Intron	Variation	Protein change	Type
Patient 1	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 2	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 3				
Patient 4 (A)	ND[2]	ND	ND	ND
Patient 5				
Patient 6	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 7				
Patient 8	ND	ND	ND	ND
Patient 9	89 bp before NC3	NT_022135.16:g.[38684843A>C]+[38684843 A>C]		SNP rs12465610
Patient 10				
Patient 11	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 12	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 13	89 bp before NC3	NT_022135.16:g.[38684843A>C]+[38684843 A>C]		SNP rs12465610
Patient 14	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 15				
Patient 16	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 17				
Patient 18	89 bp before NC3	NT_022135.16:g.[38684843A>C]+[38684843 A>C]		SNP rs12465610
Patient 19	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 20	89 bp before NC3	NT_022135.16:g.[38684843A>C]+[38684843 A>C]		SNP rs12465610
Patient 21				
Patient 22				
Patient 23				
Patient 24				
Patient 25	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 26	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 27	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 28	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 29				
Patient 30				
Patient 31				
Patient 32	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 33				
Patient 34	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 35				
Patient 36				
Patient 37				
Patient 38				
Patient 39	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 40				
Patient 41	ECR	Chr2(NCBI 36):g.148503081A>G		ND
Patient 42	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 43				
Patient 44	89 bp before NC3	NT_022135.16:g.[38684843A>C]+[38684843 A>C]		SNP rs12465610
Patient 45				
Patient 46	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 47				
Patient 48	89 bp before NC3 ECR	NT_022135.16:g.38684843A>C Chr2(NCBI 36):g.148503081A>G		SNP rs12465610 ND
Patient 49	NC4	Chr2(NCBI 36) :g.148707285A>G		paternal
Patient 50				
Patient 51	89 bp before NC3	NT_022135.16:g.[38684843A>C]+[38684843 A>C]		SNP rs12465610
Patient 52	89 bp before NC3	NT_022135.16:g.[38684843A>C]+[38684843 A>C]		SNP rs12465610
Patient 53	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 54	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 55				
Patient 56				
Patient 57	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610

[2] ND: no data

Patient ID	Exon/Intron	Variation	Protein change	Type
Patient 58				
Patient 59	ECR	Chr2(NCBI 36):g.148503081A>G		maternal
Patient 60				
Patient 61	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 62	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 63	89 bp before NC3 NC4	NT_022135.16:g.38684843A>C Chr2(NCBI 36) :g.148707285A>G		SNP rs12465610 ND
Patient 64	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 65	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 66	89 bp before NC3 102bp after NC4 69 and 104 bp after NC5	NT_022135.16:g.38684843A>C Chr2(NCBI 36):g.148707536C>T Chr2(NCBI 36):g.148816444T>G and Chr2(NCBI 36):g.148164479G>A		SNP rs12465610 ND ND
Patient 67	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 68	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 69	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 70				
Patient 71	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 72				
Patient 73	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 74	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 75	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 76	89 bp before NC3	NT_022135.16:g.[38684843A>C]+[38684843 A>C]		SNP rs12465610
Patient 77	89 bp before NC3	NT_022135.16:g.38684843A>C		SNP rs12465610
Patient 78	89 bp before NC3	NT_022135.16:g.[38684843A>C]+[38684843 A>C]		SNP rs12465610

ANNEXE II

BRIEF REPORT

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Exploring the Potential Role of Disease-Causing Mutation in a Gene Desert: Duplication of Noncoding Elements 5′ of *GRIA3* is Associated with *GRIA3* Silencing and X-Linked Intellectual Disability

Céline Bonnet,[1] Alice Masurel-Paulet,[2] Asma Ali Khan,[1] Mylène Béri-Dexheimer,[1] Patrick Callier,[3] Francine Mugneret,[3] Christophe Philippe,[1] Christel Thauvin-Robinet,[2] Laurence Faivre,[2] and Philippe Jonveaux[1]*

[1] *Laboratoire de Génétique, EA 4368, Nancy Université, Centre Hospitalier Universitaire de Nancy, France;* [2] *Centre de génétique et Centre de Référence Anomalies du développement et Syndromes malformatifs, Hôpital d'enfants, CHU Dijon, France;* [3] *Service de Cytogénétique, Plateau technique de Biologie, CHU Le Bocage, Dijon, France*

Communicated by Stylianos E. Antonarakis

Received 14 June 2011; accepted revised manuscript 1 November 2011.
Published online 3 November 2011 in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.21649

ABSTRACT: ***GRIA3*** **encodes glutamate receptor ionotropic AMPA (alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) subunit 3 and has been previously involved in X-linked intellectual disability (ID). We report on a male proband with ID and epilepsy associated with a duplication mapping within a gene desert, 874-kb upstream of the** ***GRIA3*** **gene. This 970-kb duplication is maternally inherited. The proband's mother has a skewed X chromosome-inactivation pattern in agreement with her normal cognitive function. Quantitative polymerase chain reaction analysis indicates absence of** ***GRIA3*** **mRNA in the proband lymphocytes relative to a wild-type control. Centromeric to the duplicated region, comparative genomic analysis showed a 2268-bp evolutionarily conserved region that could be a critical transcription factor binding-site for** ***GRIA3*** **expression. The repositioning of distant-acting sequences, rather a missense/nonsense mutation, is considered to be causative for** ***GRIA3*****-linked ID. This study illustrates the importance of high-resolution array-Comparative Genomic Hybridization analysis in exploring the potential role of disease-causing mutation in functional noncoding sequences.**
Hum Mutat 33:355–358, 2012. © **2011 Wiley Periodicals, Inc.**

KEY WORDS: **intellectual disability;** ***GRIA3*****; duplication; position effect**

The role of mutations not involving primary gene structural sequences has been minimally explored, largely owing to our inability to recognize relevant noncoding sequences, much less predict their function. A more precise knowledge of the higher order chromatin structure induced by copy number variation changes would likely bring important insights for understanding the complexity of the regulation of human genome. In the course of a systematic whole-genome screening of patients with unexplained intellectual disability (ID) by array-Comparative Genomic Hybridization analysis (aCGH), we have identified a mentally retarded boy with a microduplication 5′ of *GRIA3* (MIM# 305915) associated with gene silencing. Disruption of distant-acting transcriptional enhancer may contribute to this *GRIA3* loss of function-associated phenotype.

The male proband was the first child of healthy unrelated parents and his family history was unremarkable. The pregnancy was reported as normal with delivery at term. Growth parameters at birth were normal. He presented two status epilepticus at day 2 and day 20 of life. Metabolic screening and cerebral magnetic resonance imaging (MRI) were normal. Follow-up revealed severe psychomotor delay, with axial hypotonia, poor eye contact, and pharmacoresistant epilepsy with numerous tonico-clonic seizures. The child was referred to the Genetics unit at the age of 11. He had a hypotonic face with mild ptosis, deep-set eyes, a short philtrum with open mouth, and large ears. Extremities were normal and he had no malformation. Neurologic examination showed severe axial hypotonia with kyphosis, the sitting position and the walk were not acquired. A mild spasticity was identified with positive Babinsky sign. There were persistent difficulties in catching the look, but ocular motricity was normal. Language was still limited to sounds. His karyotype was normal (550-band level of resolution). New metabolic investigations (plasma amino acid and urine organic acid chromatographies; very long chain fatty acids; serum copper and ceruloplasmin concentrations; urine concentration of AICAR (5-aminoimidazdle-4-carboxamide ribotide), SAICAR (succinyl aminoimidazole carboxamide ribotide), creatine, and guanidinoacetate; as well as lactates; catecholamine; GBA; and amino acid in the cerebrospinal fluid) showed no abnormalities. Ophthalmologic investigations revealed astigmatism. Brain MRI with T1- and T2-weighted sequences revealed a bilateral frontal cortical atrophy, hypoplastic corpus callosum, hypoplastic white matter predominantly on the parietal lobes, with normal spectroscopy (Supp. Fig. S1).

Blood samples were obtained with informed consent from the proband's parents. aCGH was carried out using a 105-K oligonucleotide array (Agilent, Santa Clara, CA) as previously described [Bonnet et al., 2009]. Fluorescent in situ hybridization (FISH) analysis was performed with bacterial artificial chromosone (BAC) clones RP11-155H03 (chrX:120,486,934-120,661,408) and RP11-328F18

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Philippe Jonveaux, Laboratoire de Génétique, EA 4368, Nancy Université, Centre Hospitalier Universitaire de Nancy, Hôpitaux de Brabois, rue de Morvan, 54511 Vandoeuvre les Nancy, France. E-mail: p.jonveaux@chu-nancy.fr

Contract grant sponsor: French Ministry of Health; Fondation Jérome Lejeune.

© 2011 WILEY PERIODICALS, INC.

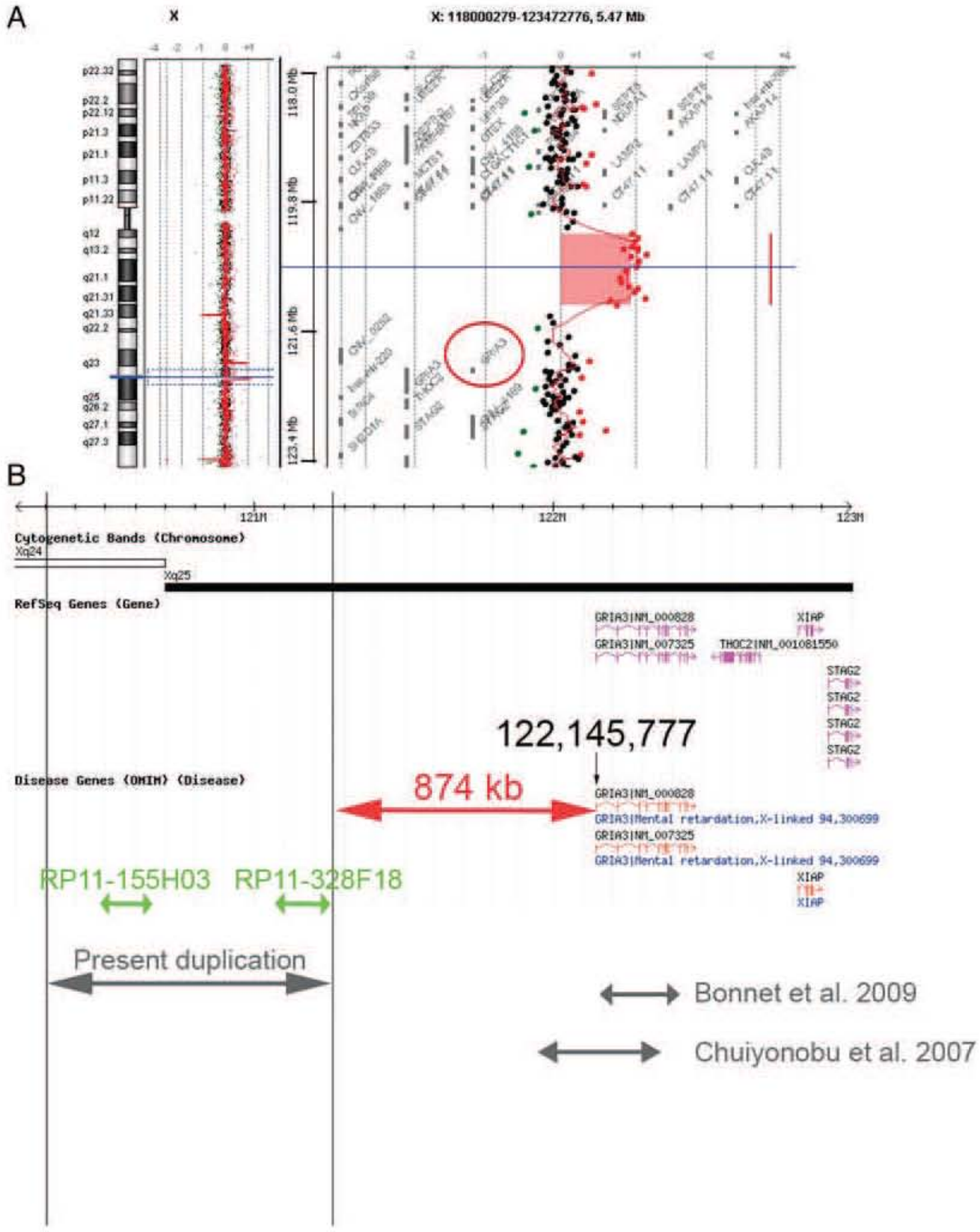


Figure 1. **A:** Array-Comparative Genomic Hybridization (aCGH) result using the CGH analytics software (version 3.5.14; Agilent) shows extend of the Xq24q25 duplication in the proband. **B:** Map of the duplicated region and previously reported duplications snapshot of Database of Genomic Variants (http://projects.tcag.ca/variation/). Present duplication ends 874 kb before start of *GRIA3* at position 122,145,777. BAC clones RP11-328 and RP11-155 are indicated.

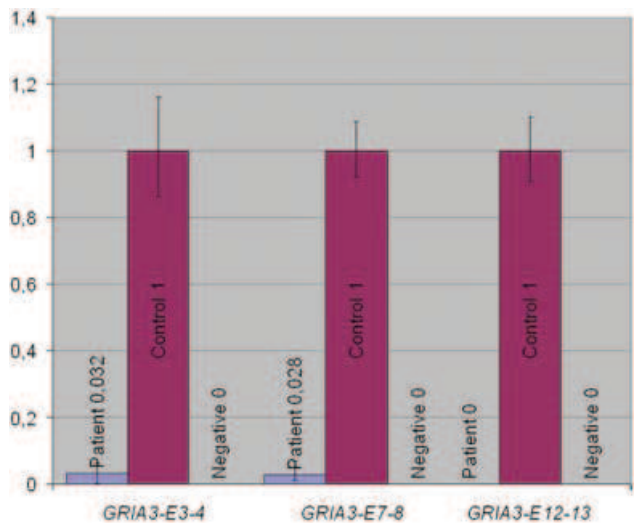


Figure 2. Reverse transcriptase quantitative real-time PCR analysis using three primer sets (exons 3–4, exons 7–8, exons 12–13) shows loss of *GRIA3* expression for the patient relative to a male control.

(chrX:121,066,300-121,250,723) obtained from CHORI (CHORI, Oakland, CA).

Quantitative polymerase chain reaction (qPCR) using SYBR-Green PCR Master Mix (ABI SYBR® Green PCR Master Mix; Applied Biosystems, Foster City, CA) was performed on genomic DNA with ABI Prism® 7500 Sequence Detection System (Applied Biosystems). Primer sequences are given in Supp. Table S1. All samples were run in triplicate. The dosage of each amplicon relative to *RPPH1* (MIM# 608513) (control amplicon) and normalized to control male DNA was determined using the 2 method. X-chromosome inactivation (XCI) status was investigated in the proband's mother using the $(CAG)_n$ repeat at the androgen receptor locus according to standard protocols. In addition, the proband was also screened for mutations by direct sequencing of *GRIA3* exons amplified using flanking intronic primers (available on request). PCR products were purified and sequenced using BigDye Terminator v3.1 (Applied Biosystems). Reverse transcriptase quantitative real-time PCR (RT-qPCR) was performed thanks to primer sets within *GRIA3, STAG2* (MIM# 300826), *THOC2* (MIM# 300395), *XIAP* (MIM# 300079), and *CULAB* (MIM# 300304) (Supp. Table S1). All samples were run in triplicate. The dosage of each amplicon relative to *ESD* (MIM# 133280) and *ABL* (MIM# 189980) (control amplicons) and normalized to control male cDNA was determined using the $2^{-\Delta\Delta Ct}$ method.

aCGH analysis identified a 970-kb duplication at Xq24q25 in the proband (Fig. 1A): arr Xq24q25(chrX:120,301,930-121,272,067)x2. The duplicated region is a gene desert and is 874-kb upstream of the *GRIA3* gene (Fig. 1B). This region had never been described as a copy number polymorphism in the database of genomic variants (http://projects.tcag.ca/variation/?source = hg18). Except for polymorphic regions, no copy number alterations were observed in autosomal chromosomes. On patient's metaphase spreads and interphase nuclei, FISH analysis showed inverted intrachromosomal duplication at Xq24q25 and no signal on any other chromosome (Supp. Fig. S2). The duplicated region was confirmed using qPCR study and was showed to be inherited from the proband's mother who was described as having normal cognitive function. The XCI assay revealed in the proband's mother a skewed pattern (10–90%), in agreement with her phenotype. Skewed XCI is relatively common in the peripheral blood cells of X-linked ID carriers.

RT-qPCR analysis of mRNA from the proband revealed a complete loss of *GRIA3* expression (Fig. 2) relative to male control. In addition to *GRIA3*, we performed RT-qPCR analysis for four other genes located in the vicinity, *THOC2*, *XIAP*, *STAG2,* and *CUL4B*. We did not observe any long-distance dysregulation of these genes (Supp. Table S2). Finally, sequence analysis excluded a point mutation within all 17 coding exons of the *GRIA3* gene.

Glutamate receptors mediate most of the excitatory neurotransmission in the mammalian brain and also participate in processes of synaptic plasticity and efficacy in learning and memory. Several male patients with ID were found to be hemizygous for genomic rearrangements or deleterious point mutations within the *GRIA3* gene [Chiyonobu et al., 2007; Bonnet et al., 2009; Guilmatre et al., 2009; Wu et al., 2007]. These rearrangements are associated with moderate to severe ID, hypotonia and inconstantly, seizures, behavior disorders, and minor facial anomalies whereas the phenotype of heterozygous female carriers depends on XCI. Important functional properties are embedded in noncoding regions of the human genome, but identifying and defining these features remains a major challenge. Variation in distant-acting sequences and the resultant changes in their activities can contribute to human disorders [Kleinjan et al., 2005]. Present duplication (chrX:120,301,930-121,272,067) is 874-kb upstream of the *GRIA3* gene. Analysis of Evolutionary Conserved Regions (ECRs) in genomes of 13 sequenced species [Loots et al., 2007] revealed the existence of a 2268-bp region with 86.2% identity between human and mouse at position chrX:120,164,374-120,166,641 (Supp. Fig. S3A). This region was also conserved in zebrafish (75%), fugu (72%), tetraodon (74%), frog (75%), chicken (80%), rhesus macaque (95.8%), and chimpanzee (99%). Although we cannot confirm that the DNA sequence conservation in that example is related to coevolution of regulatory sequences, it is well known that cross-species sequence comparison allows identifying functional noncoding sequences as critical transcriptional regulatory signals. Transcription factor binding-site predictions with the Transfac Matrix Database (version 7.0) created by Biobase (Wolfenbuettel, Germany) identified five conserved transcription factor binding sites in the 2268-bp ECR in the human/mouse/rat alignment: HMX1, COMP1, CDP, XBP1, and MRF2. In the present case, the duplication separates the *GRIA3* gene from this 2268-bp ECR that could be a critical transcription factor binding site for *GRIA3* expression.

Regulatory potential (RP) scores computed from alignments of seven species, human, chimpanzee (panTro2), macaque (rheMac2), mouse (mm8), rat (rn4), dog (canFam2), and cow (bosTau2), compare frequencies of short alignment patterns between known regulatory elements and neutral DNA [King et al., 2005; Kolbe et al., 2004]. For the same chrX:120,164,374-120,166,641 region, RP scores were between 0.1 and 0.5, which indicate very marked resemblance to alignment patterns typical of regulatory elements (Supp. Fig. S3B). Collectively, these data argue for the presence of a putative regulator of *GRIA3* expression within the 2268-bp ECR. The position effect seems evident as the duplication event occurs in the vicinity of a gene, for which the associated phenotype is well known. We confirmed that the expression of other genes in the vicinity, *THOC2, XIAP,* and *STAG2,* which have never been involved in ID or epilepsy, is not affected by the duplication. More specifically, the gene *CUL4B* for which mutations are responsible for syndromic mental retardation is not dysregulated by the duplication.

In conclusion, duplication sparing *GRIA3* sequence associated with *GRIA3* silencing is described and allowed the identification of a potential distant regulator of *GRIA3.* This duplication may put

GRIA3 under different regulatory constraints resulting in a phenotype similar to that caused by *GRIA3* null mutations. Further studies in additional mentally retarded patients with *GRIA3* silencing will help to characterize this long-range regulator.

Acknowledgments

The authors thank the patient and his family for their kind cooperation. We thank the cytogenetics and molecular genetics staff at the Nancy and Dijon University Hospitals for their expert technical assistance.

References

Bonnet C, Leheup B, Béri M, Philippe C, Grégoire MJ, Jonveaux P. 2009. Aberrant GRIA3 transcripts with multi-exon duplications in a family with X-linked mental retardation. Am J Med Genet A 149A:1280–1289.

Chiyonobu T, Hayashi S, Kobayashi K, Morimoto M, Miyanomae Y, Nishimura A, Nishimoto A, Ito C, Imoto I, Sugimoto T, Jia Z, Inazawa J, Toda T. 2007. Partial tandem duplication of *GRIA3* in a male with mental retardation. Am J Med Genet A 143A:1448–1455.

Guilmatre A, Dubourg C, Mosca AL, Legallic S, Goldenberg A, Drouin-Garraud V, Layet V, Rosier A, Briault S, Bonnet-Brilhault F, Laumonnier F, Odent S, Le Vacon G, Joly-Helas G, David V, Bendavid C, Pinoit JM, Henry C, Impallomeni C, Germano E, Tortorella G, Di Rosa G, Barthelemy C, Andres C, Faivre L, Frébourg T, Saugier Veber P, Campion D. 2009. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. Arch Gen Psychiatry 66:947–956.

King DC, Taylor J, Elnitski L, Chiaromonte F, Miller W, Hardison RC. 2005. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. Genome Res 15:1051–1060.

Kleinjan DA, van Heyningen V. 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76:8–32.

Kolbe D, Taylor J, Elnitski L, Eswara P, Li J, Miller W, Hardison R, Chiaromonte F. 2004. Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. Genome Res 14:700–707.

Loots G, Ovcharenko I. 2007. ECRbase: database of evolutionary conserved regions, promoters and transcription factor binding sites in vertebrate genomes. Bioinformatics 23:122–124.

Wu Y, Arai AC, Rumbaugh G, Srivastava AK, Turner G, Hayashi T, Suzuki E, Jiang Y, Zhang L, Rodriguez J, Boyle J, Tarpey P, Raymond FL, Nevelsteen J, Froyen G, Stratton M, Futreal A, Gezk J, Stevenson R, Schwartz CE, Valle D, Huganir RL, Wang T. 2007. Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. Proc Natl Acad Sci USA 104:18163–18168.

ANNEXE III

RESEARCH LETTER

AMERICAN JOURNAL OF medical genetics PART A

RUNX1T1, a chromatin repression protein, is a candidate gene for autosomal dominant intellectual disability

Minh Tuan Huynh,[1] Mylène Béri-Dexheimer,[1] Céline Bonnet,[1] Myriam Bronner,[1] Asma Ali Khan,[1] Lila Allou,[1] Christophe Philippe,[1] Jacqueline Vigneron,[2] and Philippe Jonveaux[1]*

[1]Laboratoire de Génétique, EA4368, Nancy Université, CHU Nancy, France

[2]Service de Néonatalogie-Génétique, Maternité Régionale Universitaire, Nancy, France

Manuscript Received: 29 November 2011; Manuscript Accepted: 4 March 2012

TO THE EDITOR:

Intellectual disability (ID) is a variable and heterogeneous manifestation of central nervous system dysfunction, affecting 2–3% of the Western population [Raymond, 2010]. Although several genes have been shown to be mutated in patients with ID, only a small number of mutated autosomal genes have been identified thus far. Balanced de novo chromosomal translocations in patients with ID are a valuable resource in the search for genes causally related to disease. Recently, the *RUNX1T1* gene at 8q21.3 was found to be disrupted in a patient with ID with a balanced translocation t(5;8)(q31;q21) [Zhang et al., 2009]. Analysis of *RUNX1T1* expression in human embryonic and fetal tissues has suggested a role in brain development and cognitive impairment [Zhang et al., 2009]. Here, we report on a patient with apparent ID of a mild degree and a de novo deletion within the *RUNX1T1* gene establishing defects in this transcription regulator as a likely cause of autosomal dominant mild to moderate ID.

The patient took part in a clinical diagnostic testing for genomic imbalance using array-CGH following initial testing for karyotype (result normal) thanks to the national array-CGH network funded by the French Ministry of Health. This 38-year-old female is the first child of healthy non-related parents. She has a healthy sibling with no reported learning problems. The pregnancy was reported as normal with delivery at term. Growth parameters at birth were normal and congenital malformations were not apparent. No delay in major motor milestones was reported. She had a normal speech development but her learning disability was recognized at the age of 6 years. She was diagnosed with a mild ID but full scale IQ testing was not performed. She attended special education. Non-verbal social interactions were within the normal range. At the age of 38 years, her height was 158 cm (−1 SD), her weight was 69.8 kg (+2 SD), and OFC was 52.5 cm (−1.5 SD). The cardiac physical exam was normal. Only minor facial anomalies such as broad forehead, full eyebrows with mild synophrys, a right protruding ear, and a low posterior hairline were noted (Fig. 1). She was friendly and cooperative with a good participation in daytime activities but parental guidance was mandatory for her practical life. Standard resolution G-banded chromosome analysis and *FMR1* analysis were normal. Blood samples were obtained with informed consent from the proband and her parents. Array-CGH was carried out using 180K-oligonucleotide array (Agilent, Santa Clara, CA) as described [Bonnet et al., 2010]. Quantitative PCR (qPCR) was performed on genomic DNA as described [Bonnet et al., 2010]. Primer sequences from *RUNX1T1* are available upon request. All samples were run in triplicate. The dosage of each amplicon relative to *RPPH1* (Gene ID 85495) (control amplicon) and normalized to control DNA was determined using the $2^{-\Delta\Delta Ct}$ method [Yuan et al., 2008]. Total RNA was isolated from a blood sample using TRIzol reagent (Invitrogen, Carlsbad, CA). RNA was reverse transcribed according to the manufacturer's protocol (Superscript, Invitrogen). Specific primers spanning from exons 2 to 8 of the *RUNX1T1* gene

How to Cite this Article:
Huynh MT, Béri-Dexheimer M, Bonnet C, Bronner M, Khan AA, Allou L, Philippe C, Vigneron J, Jonveaux P. 2012. *RUNX1T1*, a chromatin repression protein, is a candidate gene for autosomal dominant intellectual disability.
Am J Med Genet Part A.

Additional supporting information may be found in the online version of this article.

Grant sponsor: French Ministry of Health; Grant sponsor: The "Fondation Jérome Lejeune".

*Correspondence to:
Philippe Jonveaux, MD, PhD, Laboratoire de Génétique, EA4368, Nancy Université, Centre Hospitalier Universitaire de Nancy, Hôpitaux de Brabois, rue du Morvan, 54511 Vandoeuvre les Nancy, France.
E-mail: p.jonveaux@chu-nancy.fr

Article first published online in Wiley Online Library (wileyonlinelibrary.com): 00 Month 2012
DOI 10.1002/ajmg.a.35386

© 2012 Wiley Periodicals, Inc.



FIG. 1. Front and lateral views of the patient: Features include broad forehead, full eyebrows, low posterior hairline and right protruding ear. Patient consent was obtained for the publication of this figure.

were designed to investigate the deleted transcript (available upon request). RT-PCR products were sequenced using BigDye Termination kit (Applied Biosystems, Foster City, CA). In addition, the proband was also screened for mutations by direct sequencing of *RUNX1T1* exons amplified using flanking intronic primers (NM_001198633.1) (available upon request). PCR products were sequenced using BigDye Terminator v3.1.

Array-CGH analysis identified a 37.85 kb deletion at 8q21.3 in the proband (Fig. 2): 46,XX.arr 8q21.3(93,079,398 - 93,117,255) × 1, UCSC build 36/hg18. The deleted region included the coding sequence of the *RUNX1T1* gene. This region has not been described as a copy number polymorphism in the database of genomic variants (http://projects.tcag.ca/variation/?source=hg18). Except for polymorphic regions, no copy number alterations were observed elsewhere in the genome (data not shown). Quantitative PCR with a set of primers from exons 2 to 8 confirmed that the deleted region included exons 3 to 7, while the parents showed a normal result, suggesting de novo occurrence of the deletion (Supplementary Fig. 1). Sequencing analysis performed on cDNA showed an abnormal transcript with contiguity of exons 2 to 8 without a frame shift (Supplementary Fig. 2). Finally, sequence analysis excluded a point mutation within the coding exons of *RUNX1T1*.

Reporting individuals with ID who share mutations in the same gene supports a causal relationship of mutations in a gene to a disorder. We identified an apparently de novo intragenic deletion in *RUNX1T1* in a patient with apparent mild ID. The previously described patient [Zhang et al., 2009] also presented a mild to moderate ID in addition to minor craniofacial dysmorphism and a ventricular septal defect. A translocation breakpoint was found within intron 1b of the *RUNX1T1* gene, suggesting that the patient's phenotype could be due to *RUNX1T1* haploinsufficiency. Noticeably, some resemblance was noted, both patients had a broad forehead, full eyebrows, and a low posterior hairline. However, the present patient had a normal cardiac physical examination although echocardiogram has not been performed. According to Zhang et al. [2009], only three patients with ID and isolated 8q21.3q22 deletion have been described, but the exact size of the rearrangements was not precisely determined. Since then, four patients were reported to DECIPHER (2399, 4103, 248172, and 253842) with overlapping submicroscopic deletions of 8q21.3 associated with a more complex phenotype (https://decipher.sanger.ac.uk). However, those genomic imbalances were larger than that of the present case (3.6, 11.13, 12.8, and 5.8 Mb, respectively) and involved numerous genes, which hamper a precise genotype–phenotype correlation. *RUNX1T1* encodes a member of the myeloid translocation gene (MTG) family, which mediates multiple complex protein interactions leading to multiprotein networks involving transcriptional corepressors, histone-modifying enzymes, and DNA-binding transcription factors [Koyano-Nakagawa and Kintner, 2005]. *RUNX1T1* is highly expressed during brain development and recent data have suggested that *RUNX1T1* is required for executing specific gene repression programs in brain [Aaker et al., 2010]. The modular structure of wild-type RUNX1T1 protein is characterized by four evolutionarily conserved domains termed nervy homology regions (NHR1-4) [Davis et al., 2003]. Owing to the present deletion from exons 3 to 7, the expected loss of NHR1 and NHR2 domains of interaction with nuclear corepressors and chromatin-modifying enzymes may impair *RUNX1T1* function as a transcriptional corepressor.

When chromosome rearrangements affect autosomal genes, the associated disorder is generally a dominant rather than recessive form. In the present patient, sequencing analysis excluded a point mutation within the coding exons of *RUNX1T1* in favor of a dominant form of ID. In conclusion, the fact that two patients with ID have truncated copies of *RUNX1T1* is significant, and it is thus very likely that *RUNX1T1* is a valuable candidate to explain the apparent cognitive disorder.

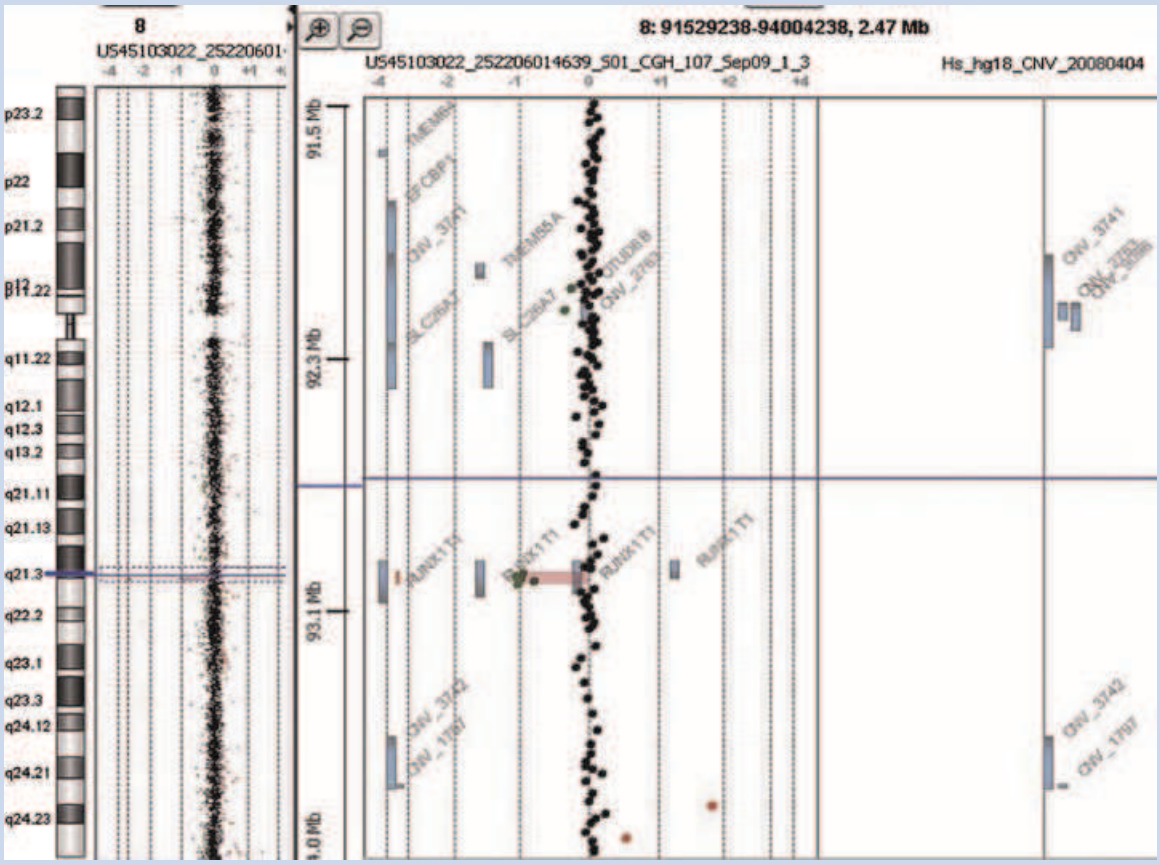


FIG. 2. Array comparative genomic hybridization results, showing the extent of *RUNX1T1* intragenic deletion (five probes).

ACKNOWLEDGMENTS

The authors thank the patient and her family for their kind cooperation. We thank the cytogenetics and molecular genetics staff at the Nancy University Hospital for their expert technical assistance. This study was supported by grants from the French Ministry of Health and the "Fondation Jérome Lejeune".

REFERENCES

Aaker JD, Patineau AL, Yang HJ, Ewart DT, Nakagawa Y, McLoon SC, Koyano-Nakagawa N. 2010. Interaction of MTG family proteins with NEUROG and ASCL1 in the developing nervous system. Neurosci Lett 474:46–51.

Bonnet C, Andrieux J, Béri-Dexheimer M, Leheup B, Boute O, Manouvrier S, Delobel B, Copin H, Receveur A, Mathieu M, Thiriez G, Le Caignec C, David A, de Blois MC, Malan V, Philippe A, Cormier-Daire V, Colleaux L, Flori E, Dollfus H, Pelletier V, Thauvin-Robinet C, Masurel-Paulet A, Faivre L, Tardieu M, Bahi-Buisson N, Callier P, Mugneret F, Edery P, Jonveaux P, Sanlaville D. 2010. Microdeletion at chromosome 4q21 defines a new emerging syndrome with marked growth restriction and absent or severely delayed speech. J Med Genet 47:377–384.

Davis JN, McGhee L, Meyers S. 2003. The ETO (MTG8) gene family. Gene 303:1–10.

Koyano-Nakagawa N, Kintner C. 2005. The expression and function of MTG/ETO family proteins during neurogenesis. Dev Biol 278:22–34.

Raymond FL. 2010. Monogenic causes of mental retardation. Genetics of Mental Retardation. In: Knight SJL, editors: Monogr Hum Genet Basel, Karger. 18:89–100.

Yuan JS, Wang D, Stewart CN Jr. 2008. Statistical methods for efficiency adjusted real-time PCR quantification, Biotechnol J 3:112–123.

Zhang L, Tümer Z, Møllgård K, Barbi G, Rossier E, Bendsen E, Møller RS, Ullmann R, He J, Papadopoulos N, Tommerup N, Larsen LA. 2009. Characterization of a t(5;8)(q31;q21) translocation in a patient with mental retardation and congenital heart disease: Implications for involvement of RUNX1T1 in human brain and heart development. Eur J Hum Genet 17:1010–1018.

Les analyses pangénomiques dans l'exploration génétique de la déficience intellectuelle : de la recherche de gènes candidats du syndrome d'Aicardi, à la caractérisation du spectre mutationnel des gènes *IL1RAPL1* et *MBD5*

L'exploration génétique de la déficience intellectuelle (DI) a été révolutionnée par l'amélioration des technologies de séquençage depuis ces dernières années avec la caractérisation du spectre mutationnel des gènes impliqués dans la DI, ainsi que l'identification de nouveaux gènes associés. Dans notre étude, nous avons utilisé la technique d'analyse sur microréseau d'ADN (Hybridation Génomique Comparative ou CGH-array) à haute résolution, puis celle du séquençage haut débit pour rechercher une altération à l'origine de la DI inexpliquée.

Le syndrome d'Aicardi est une maladie neurodéveloppementale rare et sporadique, caractérisée par la triade: spasmes infantiles, agénésie du corps calleux, et lacunes choriorétiniennes. Ce syndrome est décrit exclusivement chez les filles avec l'hypothèse la plus probable d'une mutation dominante liée au chromosome X. Nous avons d'abord analysé les ADN de 22 patientes atteintes du syndrome d'Aicardi par CGH-array, à l'aide d'un microréseau d'oligonucléotides haute résolution (1M) spécifique du chromosome X, sans identifier de remaniements ou CNV pouvant être impliqués dans la maladie. Un premier séquençage haut débit de l'exome du chromosome X, a été effectué sur l'ADN d'un trio (patiente et parents) et deux autres patientes présentant des signes typiques du syndrome d'Aicardi. Les résultats ont révélé 59 mutations dans 51 gènes. Il s'agit de 13 variants hérités de la mère, 8 hérités du père, de 36 faux positifs, et 2 SNP. Un deuxième séquençage haut débit, sur l'exome complet, a ensuite été réalisé, à partir de l'ADN de cinq trios (patientes et parents). Nous présentons et commentons les différentes stratégies d'analyses utilisées à la recherche d'un gène candidat. Les résultats obtenus pour les SNP soulignent les difficultés rencontrées en terme de profondeur du séquençage générant de nombreux contrôles et les difficultés d'alignement des séquences ne rendant pas performant l'analyse des indels.

Parallèlement, dans notre cohorte de patients du centre de référence maladies rares, la CGH-array a identifié des altérations intragéniques du gène *IL1RAPL1* dont deux duplications originales et une délétion. Nous analysons les corrélations génotype –phénotype au regard des données de la littérature avec notamment la variabilité d'expression clinique. Deux délétions intragéniques et une duplication intragénique du gène *MBD5*, survenue *de novo* ont été aussi détectées chez des patients atteints de DI. Cette duplication conduit à des transcrits aberrants avec codon stop prématuré. Le gène *MBD5* a été séquencé sur une cohorte de 78 patients phénotypiquement sélectionnés révélant une mutation non-sens *de novo* détecté chez un garçon associé avec un phénotype sévère.

Nos travaux témoignent des avantages de ces stratégies d'analyse pangénomiques, dont l'analyse sur microréseau, mais souligne aussi la complexité, les limites en terme d'interprétation des résultats, tout particulièrement pour le séquençage de nouvelle génération.

Mots clés : Déficience intellectuelle, Microremaniements génomique, Syndrome d'Aicardi, *IL1RAPL1*, *MBD5*.

Extensive pangenomic analysis for genetic exploration of intellectual disability; in search of candidate gene for Aicardi syndrome and characterization of mutational spectrum of *IL1RAPL1* and *MBD5* genes

The genetic exploration of intellectual disability (ID) has been revolutionized with the improvement in sequencing technologies during last decade with characterization of mutational spectrum of genes involved in ID as well as to identify new genes associated with it. In this study we used high resolution (comparative genomic hybridization array) CGH-array and high throughput sequencing technique to find the genetic cause in patients with unexplained ID.

Aicardi syndrome is a rare sporadic neurodevelopmental syndrome, characterized by classic triad of agenesis of corpus callosum, chorioretinal lacunes and infantile spasms. This syndrome is exclusively present in females with plausible hypothesis of X linked dominant mutation. We first tested DNA of 22 patients diagnosed with Aicardi syndrome by using a high resolution oligonucleotide CGH-array 1M specifically designed for X-chromosome without identifying any pathogenic CNV or deleterious rearrangements involved in the disease.

High throughput sequencing for exome of X chromosome was carried out in one trio (patient-parents) and two patients with typical Aicardi syndrome diagnosis. Sequencing results detected 59 mutations in 51 genes. 13 mutations were inherited from mother, 8 inherited from father, 36 false positive and 2 were SNP's. Second approach was based on High throughput sequencing for complete exome of five trios (patient-parents) DNA. We presented and commented different strategies for data analysis in search of a candidate gene. These results highlighted difficulties in terms of depth and alignment of sequencing reads which generated various false positive SNP's and indels.

In second cohort from reference centre of rare diseases CGH-array has identified two intragenic rearrangements of *IL1RAPL1* gene: two unique duplications and one deletion. We analyze genotype-phenotype correlations with cases described in literature which emphasizes the clinical variability of expression in these patients.

Two *de novo* intragenic deletions and a *de novo* intragenic duplication were detected in *MBD5* gene in patients with ID. The *de novo* duplication of *MBD5* resulted in an aberrant transcripts leading to a premature termination codon. A selected cohort of 78 patients were sequenced for *MBD5* gene which revealed a *de novo* nonsense mutation in a male patient associated with a much more damaging phenotype.

This study highlighted the advantages of pangenomic analysis by CGH-array and at the same time it identified the complexity and limitations in interpretation of results particularly for High throughput sequencing.

Key words: Intellectual disability, Genomic microrearrangement, Aicardi syndrome, *IL1RAPL1*, *MBD5*.