

La déficience intellectuelle : du diagnostic en puces ADN à l'identification de gènes candidats

Boris Keren

► To cite this version:

Boris Keren. La déficience intellectuelle : du diagnostic en puces ADN à l'identification de gènes candidats. Médecine humaine et pathologie. Université René Descartes - Paris V, 2013. Français. NNT : 2013PA05T041 . tel-00918306

HAL Id: tel-00918306 https://theses.hal.science/tel-00918306

Submitted on 13 Dec 2013 $\,$

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.



Année 2013



Université Paris Descartes

Ecole doctorale GC2ID

Centre de Recherche de l'Institut du Cerveau et de la Moelle, UPMC/Inserm UMR-S 975/CNRS UMR 7225, GH Pitié-Salpêtrière, Paris

La déficience intellectuelle

Du diagnostic en puces ADN à l'identification de gènes candidats

Par Boris Keren

Thèse de doctorat de Génétique

Dirigée par Alexis Brice

Présentée et soutenue publiquement le 22 novembre 2013

Devant un jury composé de : Brice, Alexis, PU-PH, directeur Geneviève, David, PU-PH, rapporteur Lacombe, Didier, PU-PH, rapporteur Depienne, Christel, MCU-PH, examinatrice Héron, Delphine, PH, examinatrice Sanlaville, Damien, PU-PH, examinateur Siffroi, Jean-Pierre, PU-PH, examinateur



CC (S C Lxcept where otherwise noted, this isotrony of the second state of the second Except where otherwise noted, this work is licensed under

Keren Boris – Thèse de doctorat - Année 2013

Résumé (français) :

L'analyse chromosomique sur puce ADN (ACPA) tend à devenir le principal examen diagnostique dans la déficience intellectuelle (DI). Parmi les techniques d'ACPA, les puces SNP ont l'intérêt de pouvoir détecter les pertes d'hétérozygotie, et par conséquent d'identifier les isodisomies uniparentales (iUPD) et les zones d'identité liées à la consanguinité. Nous avons étudié une cohorte de 1 187 patients atteints de DI, dans un cadre diagnostique, sur puces SNP. Nous avons réalisé, par cette étude, 145 diagnostics (12%) dont 2 iUPD et 6 délétions n'incluant qu'un seul gène. De plus, nous avons détecté 639 CNV rares non décrits chez des sujets contrôles et incluant des séquences codantes, ce qui nous a permis d'identifier 11 gènes candidats dans la DI : CAMTA1, SP3, CNTNAP4, NUDT12, STXBP6, DOCK8, DOCK10, SMARCA2, NYAP2, ATAD3A et ATAD3B. Nous avons tenté de valider l'implication de ces gènes par séquençage, mais n'avons trouvé de seconde mutation pour aucun d'entre eux. Toutefois, des réarrangements de CAMTA1 ont été retrouvés dans 2 autres familles avec un phénotype homogène (DI et ataxie congénitale) ce qui nous a permis d'affirmer qu'il s'agit d'un gène de DI. Par ailleurs, l'homozygosity mapping, réalisé avec puces SNP, a identifié, par séquençage whole exome, une mutation non-sens homozygote du gène BUD13 dans une famille de DI syndromique. Enfin, de façon fortuite, nous avons caractérisé en ACPA une translocation familiale entraînant une disruption d'un gène d'ataxie spino-cérébelleuse, ATXN10, ce qui a permis de mieux comprendre la physiopathologie de cette maladie. Au total, notre étude démontre l'intérêt des puces SNP dans la DI, d'une part en diagnostic et d'autre part pour l'identification de nouveaux gènes responsables de DI.

Title: Intellectual disability: from microarray diagnosis to the identification of candidate genes

Abstract:

Chromosomal Microarray Analysis (CMA) has become the main diagnostic test in the field of intellectual disability (ID). Among CMA techniques, SNP arrays have the advantage of identifying losses of heterozygosity in addition to Copy Number Variants (CNVs). Therefore they can detect uniparental isodisomies (iUPD) and regions of identity by descent. We screened a cohort of 1,187 patients with ID, in diagnostic setting, by CMA using SNP arrays. Causal abnormalities, including 2 iUPDs and 6 deletions comprising only one gene, were detected in 145 patients (12%). Moreover we found 639 rare CNVs, absent from control individuals, which included coding sequences. Our results allowed us to identify 11 genes possibly involved in or contributing to ID: CAMTA1, SP3, CNTNAP4, NUDT12, STXBP6, DOCK8, DOCK10, SMARCA2, NYAP2, ATAD3A and ATAD3B. We then screened additional patients with similar phenotypes in order to find a second mutation, but no second mutation was identified in any of them. Besides, CAMTA1 rearrangements were also found among two other families with homogeneous phenotypes (ID and congenital ataxia), which confirms that this gene is involved in ID. Furthermore, thanks to homozygosity mapping made possible by SNP arrays combined with exome sequencing, we identified a homozygous nonsense mutation in the BUD13 gene, in a family with syndromic ID. Finally we incidentally characterized a familial translocation resulting in the disruption of ATXN10, a gene responsible for spinocerebellar ataxia. This translocation has helped to better understand the pathophysiology of the disease. Overall, our study shows the importance of SNP arrays for the molecular diagnosis of ID and as a tool to identify genes responsible for ID.

Mots clés (français) : déficience intellectuelle, puces SNP, séquençage de l'ADN.

Keywords: intellectual disability, SNP arrays, DNA sequencing.

Keren Boris – Thèse de doctorat - Année 2013

A ma famille et mes amis.

Keren Boris – Thèse de doctorat - Année 2013

Remerciements

J'aimerais tout d'abord remercier Alexis Brice et Christel Depienne d'avoir co-dirigé cette thèse. Leur aide et leurs conseils m'ont été précieux, et c'est grâce à leurs encouragements et leur persévérance que j'ai pu finir ce travail. Je remercie spécialement Alexis qui m'a fait confiance et m'a permis d'intégrer l'équipe de cytogénétique ; je sais que je le lui dois. Quant à Christel, c'est un immense plaisir de travailler à ses côtés : ce sont ses idées et son dynamisme qui me permettent d'avancer, et j'espère que cette collaboration continuera longtemps.

A David Geneviève et Didier Lacombe, j'exprime toute ma gratitude pour avoir accepté d'être rapporteurs de cette thèse. C'est un très grand honneur pour moi que Didier Lacombe ait consenti à cette tâche, et je souhaite que nous ayons à l'avenir l'occasion de travailler ensemble. David Geneviève m'a apporté des connaissances qui me sont utiles quotidiennement ; il m'a beaucoup appris durant ma formation en génétique clinique, et je n'oublie pas l'altruisme dont il faisait preuve, avec Damien Sanlaville, pour dispenser aux internes, dont je faisais partie, des cours d'enseignement théorique tard le soir : je leur en suis toujours reconnaissant.

Je tiens à remercier Delphine Héron pour avoir accepté d'être examinatrice de cette thèse. Je la remercie également pour beaucoup de choses. Tout d'abord de m'avoir permis de travailler à la Pitié-Salpêtrière ; c'est en effet à Delphine que je le dois, tout comme à Alexis. Merci pour notre collaboration quotidienne, à la fois sur le plan diagnostique et scientifique, pour ses avis et ses conseils, toujours d'une grande sagesse. C'est extrêmement enrichissant de travailler avec une personne de cette valeur, et je souhaite que notre coopération soit encore longue. Je lui suis également reconnaissant d'avoir bien voulu me confier ses patients pour ma thèse.

Quant à Damien Sanlaville, il est probablement la personne qui a le plus compté dans ma formation. Il m'a appris et fait aimer la cytogénétique, et c'est grâce à lui que j'ai eu envie d'en faire mon métier. Je suis très heureux que nous continuions à travailler régulièrement ensemble malgré la distance physique entre nos laboratoires, et j'espère que nous poursuivrons longtemps de la sorte. Je voudrais également le remercier d'avoir accepté de faire partie du jury de cette thèse.

De même, j'adresse mes remerciements à Jean-Pierre Siffroi, qui a lui aussi bien voulu participer à ce jury. Et je lui sais gré particulièrement de notre collaboration quotidienne : je me réjouis de tout ce que nos 2 équipes ont réalisé ensemble depuis ces 5 années. Merci pour ses conseils et son soutien ; il est toujours bon de savoir que je peux compter sur son expérience quand mes limites se font sentir.

Je tiens maintenant à témoigner ma gratitude à toute l'équipe de l'UF de Cytogénétique de la Pitié-Salpêtrière, pour la conscience professionnelle et l'enthousiasme que j'y trouve : je sais la chance que j'ai de travailler en son sein ce qui, je l'espère, durera longtemps encore. Merci en particulier à Corinne Mach, Elodie Lejeune et Patricia Leite, dont l'aide inestimable a rendu possible ce travail.Tous mes remerciements aussi à Claude Estrade dont l'expérience nous est indispensable pour faire tourner quotidiennement le laboratoire ; sans oublier

Catherine Aumont car c'est vraiment un plaisir de collaborer avec une personne aussi sérieuse dans son travail et qui a si bon caractère. Quant à Eric Fonteneau, je me félicite tous les jours de sa présence à mes côtés : ses compétences ont été essentielles pour la réalisation de cette thèse et le demeurent pour les tâches qui nous attendent.

Je souhaite exprimer ma reconnaissance à Caroline Nava, pour ce travail que nous avons en partie fait ensemble grâce à son aide efficace, ses avis et ses suggestions bien venus. Je suis heureux que nous puissions continuer de collaborer.

Je remercie tous les cliniciens de m'avoir confié leurs patients, Aurélia Jacquette, Sandra Whalen, Cyril Mignot, Fanny Mochel, Isabelle Marey, Alexandra Afenjar et Odile Boute. C'est une chance que de pouvoir travailler quotidiennement avec des collègues aussi dynamiques.

Ma reconnaissance va aussi à Sandra Chantot-Bastaraud pour notre collaboration quotidienne et pour ce que nous réalisons ensemble : c'est un très grand plaisir de travailler ainsi avec une personne de confiance.

Tous mes remerciements à Wassila Carpentier pour m'avoir formé efficacement à la technique Illumina, aide combien précieuse ; à Yannick Marie et Emeline Mundwiller, pour la formation sur le Roche 454 et pour nous avoir encadrés durant les techniques ; à l'équipe de génétique du CHU Dijon pour notre collaboration, en particulier à Julien Thévenon, Laurence Faivre, Christel Thauvin et Jean-Baptiste Rivière, avec qui j'espère avoir d'autres occasions de travailler ; ainsi qu'à Anne Lombes, pour sa collaboration et son enthousiasme : je suis content que cette thèse nous ait donné l'occasion de travailler ensemble.

Je remercie aussi Integragen et le Centre national de génotypage pour avoir effectué les techniques d'exome, et la Fondation maladies rares pour leur financement.

Enfin, toute ma reconnaissance à Florent Soubrier, qui m'a soutenu, et grâce à qui j'ai le bonheur de pouvoir rester à la Pitié-Salpêtrière et poursuivre mon travail.

Table des matières

| Remerciements | | |
|---|----|--|
| Table des matières | 9 | |
| Avertissement | 13 | |
| Liste des abréviations | 15 | |
| 1. La déficience intellectuelle | 17 | |
| 1.1. La déficience intellectuelle : généralités | 17 | |
| 1.1.1. Définitions de la déficience intellectuelle | 17 | |
| 1.1.1.1. Définition de l'APA | 17 | |
| 1.1.1.2. Définition de l'AAIDD | 18 | |
| 1.1.1.3. Définition de l'OMS | 19 | |
| 1.1.2. La déficience intellectuelle vue par la société au cours de l'histoire | 19 | |
| 1.1.3. Evaluation de la déficience intellectuelle | 21 | |
| 1.1.3.1. Evaluation du fonctionnement Intellectuel | 21 | |
| 1.1.3.2. Evaluation du fonctionnement adaptatif | 23 | |
| 1.1.4. Classifications de la déficience intellectuelle | 24 | |
| 1.1.4.1. Classification de l'APA (DSM-5) | 24 | |
| 1.1.4.2. Classification de l'AAIDD (Schalock et al. 2010) | 25 | |
| 1.1.4.3. Classification de l'OMS (CIM-10) | 25 | |
| 1.1.5. Prévalence de la déficience intellectuelle | 25 | |
| 1.1.6. Co-morbidité et mortalité de la déficience intellectuelle | 26 | |
| 1.1.7. Etiologies de la déficience intellectuelle | 29 | |
| 1.1.7.1. Etiologies acquises | 30 | |
| 1.1.7.2. Etiologies génétiques | 31 | |
| 1.2. Génétique de la déficience intellectuelle | 32 | |
| 1.2.1. Variations normales et pathologiques du génome humain | 32 | |
| 1.2.1.1 SNP | 33 | |
| 1.2.1.2. Variations de structure | 34 | |
| 1.2.1.3. Répétitions | 35 | |
| 1.2.2. Causes chromosomiques de déficience intellectuelle | 36 | |
| 1.2.2.1. Anomalies de nombre | 37 | |
| 1.2.2.2. Anomalies de structure | 38 | |

| 1.2.2.2 | 2.1. Mécanisme de survenue des CNV | . 39 |
|--|---|--|
| 1.2.2.2 | 2.2. Principaux syndromes délétionels | . 43 |
| 1.2.2.2 | 2.3. Principaux syndromes microduplicationnels | . 45 |
| 1.2.2.2 | 2.4. CNV facteurs de susceptibilité de la DI | . 46 |
| 1.2.3. | Causes géniques de déficience intellectuelle | . 47 |
| 1.2.3.1 | L. Evaluation du nombre de gènes impliqués dans la DI | . 47 |
| 1.2.3.2 | 2. Classifications des gènes de DI | . 48 |
| 1.2.3.2 | 2.1. Gènes impliqués dans les voies métaboliques | . 49 |
| 1.2.3.2 | 2.2. Gènes impliqués dans la neurogenèse | . 50 |
| 1.2.3.2 | 2.3. Gènes impliqués dans la migration neuronale | . 51 |
| 1.2.3.2 | 2.4. Gènes impliqués dans la fonction synaptique | . 52 |
| 1.2.3.2 | 2.5. Gènes impliqués dans les cascades de signalisation intracellulaire | . 56 |
| 1.2.3.2 | 2.6. Gènes impliqués dans la régulation épigénétique de la transcription | . 57 |
| 1.2.4. | Déficience intellectuelle liée à l'empreinte | . 58 |
| 1.2.5. | Connexion étiologique et physiopathologique entre déficience intellectuelle et TSA | . 60 |
| 2. OI | bjectifs du travail | . 63 |
| 2.1. | Stratégie diagnostique utilisée | . 63 |
| 2.2. | Identification et validation des gènes candidats | . 68 |
| | | |
| 2.2.1. | Délétions <i>de novo</i> | . 69 |
| 2.2.1. 2.2.2. | Délétions <i>de novo</i> Délétions transmises par un parent | . 69 . 70 |
| 2.2.1. 2.2.2. 2.2.3. | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote | . 69 . 70 . 70 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> | . 69 . 70 . 70 . 71 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> Points de cassure de remaniements équilibrés | . 69 . 70 . 70 . 71 . 72 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> Points de cassure de remaniements équilibrés Hatériels et méthodes . | . 69 . 70 . 70 . 71 . 72 . 73 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M 3.1. | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> Points de cassure de remaniements équilibrés Patériels et méthodes Série de patients | . 69 . 70 . 71 . 71 . 72 . 73 . 73 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M 3.1. 3.2. | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> Points de cassure de remaniements équilibrés Points de cassure de remaniements équilibrés Patériels et méthodes Série de patients Les puces ADN | . 69 . 70 . 70 . 71 . 72 . 73 . 73 . 73 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M 3.1. 3.2. 3.2.1. | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> Points de cassure de remaniements équilibrés Points de cassure de remaniements équilibrés Points de cassure de remaniements équilibrés Série de patients Les puces ADN Modèles de lames utilisées | . 69 . 70 . 70 . 71 . 72 . 73 . 73 . 73 . 74 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M 3.1. 3.2. 3.2.1. 3.2.2. | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> Points de cassure de remaniements équilibrés Points de cassure de remaniements équilibrés Patériels et méthodes Série de patients Les puces ADN Modèles de lames utilisées La technologie des puces Illumina | . 69 . 70 . 70 . 71 . 72 . 73 . 73 . 73 . 74 . 75 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M 3.1. 3.2. 3.2.1. 3.2.2. 3.2.3. | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> Points de cassure de remaniements équilibrés Points de cassure de remaniements équilibrés Points de cassure de remaniements équilibrés Points de patients Points de patients Points de patients Points de patients Points de patients Points de lames utilisées Points de lames utilisées Principe de la technique d'ACPA Illumina | . 69 . 70 . 70 . 71 . 72 . 73 . 73 . 73 . 74 . 75 . 78 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M 3.1. 3.2. 3.2.1. 3.2.2. 3.2.3. 3.2.4. | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> Points de cassure de remaniements équilibrés Points de cassure de remaniements équilibrés Série de patients Série de patients Les puces ADN Modèles de lames utilisées La technologie des puces Illumina Principe de la technique d'ACPA Illumina Analyse des puces Illumina | . 69 . 70 . 71 . 72 . 73 . 73 . 73 . 73 . 73 . 73 . 75 . 78 . 80 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M 3.1. 3.2. 3.2.1. 3.2.2. 3.2.3. 3.2.4. 3.2.4.1 | Délétions <i>de novo</i> Délétions transmises par un parent Délétion homozygote <i>Homozygosity mapping</i> Points de cassure de remaniements équilibrés Points de cassure de remaniements équilibrés Série de patients Série de patients Les puces ADN Modèles de lames utilisées La technologie des puces Illumina Principe de la technique d'ACPA Illumina Analyse des puces Illumina | . 69 . 70 . 71 . 72 . 73 . 73 . 73 . 73 . 73 . 73 . 73 . 73 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M 3.1. 3.2. 3.2.1. 3.2.1. 3.2.2. 3.2.3. 3.2.4. 3.2.4.1 3.2.4.2 | Délétions <i>de novo</i> | . 69 . 70 . 70 . 71 . 72 . 73 . 73 . 73 . 73 . 73 . 73 . 73 . 73 |
| 2.2.1. 2.2.2. 2.2.3. 2.2.4. 2.2.5. 3. M 3.1. 3.2.1. 3.2.1. 3.2.2. 3.2.3. 3.2.4. 3.2.4.2 3.2.4.2 3.2.4.2 | Délétions de novo Délétions transmises par un parent Délétion homozygote Homozygosity mapping Points de cassure de remaniements équilibrés Points de cassure de remaniements équilibrés Atériels et méthodes Série de patients Les puces ADN Modèles de lames utilisées La technologie des puces Illumina Principe de la technique d'ACPA Illumina Analyse des puces Illumina 2. Analyse des données 2.1. Lecture des profils | . 69 . 70 . 71 . 72 . 73 . 73 . 73 . 73 . 73 . 73 . 73 . 73 |

| 5.2.4.2 | 3. Les pertes d'hétérozygotie | 85 |
|--|---|-------------------|
| 3.2.4.2 | 4. Les anomalies en mosaïque | 86 |
| 3.2.4.2 | 5. Normalisation et référence | 89 |
| 3.2.4.2 | .6. Segmentation | |
| 3.2.4.3 | Critères qualité | |
| 3.2.5. | Intérêt des puces SNP comparées aux autres techniques d'ACPA | 94 |
| 3.5. 5 | équençage des gènes candidats | |
| 3.5.1. | Séquençage par terminaison de chaîne (méthode de Sanger) | |
| 3.5.2. | Séquençage par la méthode Roche 454 | 100 |
| 3.5.2.1 | Préparation des "libraries" | 100 |
| 3.5.2.2 | PCR en émulsion | 102 |
| 3.5.2.3 | Réaction de séquençage | 103 |
| 3.5.2.4 | Analyse des séquences | 104 |
| 3.5.3. | Séquençage de l'exome | 105 |
| 3.5.3.1 | . Technique d'enrichissement Agilent SureSelect | 105 |
| 3.5.3.2 | Séquençage par la technologie Illumina | 107 |
| 3.5.3.3 | Analyse des séquences | 108 |
| 4. Ré | sultats | 109 |
| | | |
| 4.1. F | Résultats diagnostiques | 109 |
| 4.1. F 4.1.1. | Résultats diagnostiques | 109 109 |
| 4.1. F4.1.1.4.1.2. | Résultats diagnostiques Les CNV Les disomies uniparentales | 109 |
| 4.1. F 4.1.1. 4.1.2. 4.2. C | Résultats diagnostiques Les CNV Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) | 109 |
| 4.1.1. 4.1.2. 4.2.1. | Résultats diagnostiques Les CNV Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d' <i>UBE3A</i> (article 1) | |
| 4.1.1. 4.1.2. 4.2.1. 4.2.2. | Résultats diagnostiques Les CNV Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d' <i>UBE3A</i> (article 1) Patiente PSL_9819 avec délétion de <i>FLNA</i> (article 2) | |
| 4.1.1. 4.1.2. 4.2.1. 4.2.1. 4.2.2. 4.2.3. | Résultats diagnostiques Les CNV Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d' <i>UBE3A</i> (article 1) Patiente PSL_9819 avec délétion de <i>FLNA</i> (article 2) Patiente PSL_10406 avec délétion de <i>FOXP1</i> | |
| 4.1.1. 4.1.2. 4.2.1. 4.2.2. 4.2.3. 4.2.4. | Résultats diagnostiques Les CNV. Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d'UBE3A (article 1) Patiente PSL_9819 avec délétion de FLNA (article 2) Patiente PSL_10406 avec délétion de FOXP1 Patiente PSL_12419 avec délétion de ANKRD11 | |
| 4.1.1. 4.1.2. 4.2.1. 4.2.2. 4.2.3. 4.2.4. 4.2.5. | Résultats diagnostiques Les CNV. Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d'UBE3A (article 1) Patiente PSL_9819 avec délétion de FLNA (article 2) Patiente PSL_10406 avec délétion de FOXP1 Patiente PSL_12419 avec délétion de ANKRD11 Patient PSL_10017 avec délétion de DMD | |
| 4.1. 4.1.1. 4.1.2. 4.2. 4.2.3. 4.2.4. 4.2.5. 4.3. | Résultats diagnostiques Les CNV. Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d'UBE3A (article 1) Patiente PSL_9819 avec délétion de FLNA (article 2) Patiente PSL_10406 avec délétion de FOXP1 Patiente PSL_10406 avec délétion de ANKRD11 Patiente PSL_12419 avec délétion de DMD Patient PSL_10017 avec délétion de DMD | |
| 4.1. 4.1.1. 4.1.2. 4.2. 4.2.3. 4.2.4. 4.2.5. 4.3. 4.4. | Résultats diagnostiques Les CNV Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d'UBE3A (article 1) Patiente PSL_9819 avec délétion de FLNA (article 2) Patiente PSL_10406 avec délétion de FOXP1 Patiente PSL_12419 avec délétion de ANKRD11 Patient PSL_10017 avec délétion de DMD CNV facteurs de susceptibilité. Gènes candidats (article 4) | |
| 4.1. 4.1.1. 4.1.2. 4.2. 4.2.3. 4.2.4. 4.2.5. 4.3. 4.4.1. | Résultats diagnostiques Les CNV. Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d'UBE3A (article 1) Patiente PSL_9819 avec délétion de FLNA (article 2) Patiente PSL_10406 avec délétion de FOXP1 Patiente PSL_10406 avec délétion de ANKRD11 Patiente PSL_12419 avec délétion de DMD Patient PSL_10017 avec délétion de DMD CNV facteurs de susceptibilité Gènes candidats (article 4) Gènes candidats identifiés par des CNV de novo. | |
| 4.1. 4.1.1. 4.1.2. 4.2. 4.2.1. 4.2.3. 4.2.4. 4.2.5. 4.3. 4.4.1. 4.4.1.1 | Résultats diagnostiques Les CNV Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d'UBE3A (article 1) Patiente PSL_9819 avec délétion de FLNA (article 2) Patiente PSL_10406 avec délétion de FOXP1 Patiente PSL_12419 avec délétion de ANKRD11 Patient PSL_10017 avec délétion de DMD CNV facteurs de susceptibilité Gènes candidats (article 4) Gènes candidats identifiés par des CNV de novo. Séquençage du gène CAMTA1 (article 3) | |
| 4.1. 4.1.1. 4.1.2. 4.2. 4.2.1. 4.2.3. 4.2.4. 4.2.5. 4.3. 4.4.1. 4.4.1.1 4.4.1.2 | Résultats diagnostiques Les CNV. Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d'UBE3A (article 1) Patiente PSL_9819 avec délétion de FLNA (article 2) Patiente PSL_10406 avec délétion de FOXP1 Patiente PSL_12419 avec délétion de ANKRD11 Patiente PSL_10017 avec délétion de DMD CNV facteurs de susceptibilité. Gènes candidats identifiés par des CNV de novo. Séquençage du gène CAMTA1 (article 3) Séquençage du gène SP3 | |
| 4.1. 4.1.1. 4.1.2. 4.2. 4.2.1. 4.2.3. 4.2.4. 4.2.5. 4.3. 4.4.1. 4.4.1.2 4.4.1.3 | Résultats diagnostiques Les CNV Les disomies uniparentales CNV pathogènes de petite taille (articles 1, 2 et 3) Patiente PSL_7458 avec délétion d'UBE3A (article 1) Patiente PSL_9819 avec délétion de FLNA (article 2) Patiente PSL_10406 avec délétion de FOXP1 Patiente PSL_12419 avec délétion de ANKRD11 Patiente PSL_10017 avec délétion de DMD CNV facteurs de susceptibilité Gènes candidats (article 4) Gènes candidats identifiés par des CNV de novo. Séquençage du gène CAMTA1 (article 3) Séquençage du gène SP3 Autres gènes candidats retrouvés dans des délétions de novo | |

| 4.4.2.1 | . Séquençage du gène STXBP6 | . 132 |
|---------|---|-------|
| 4.4.2.2 | . Séquençage du gène CNTNAP4 | . 134 |
| 4.4.2.3 | . Séquençage du gène NUDT12 | . 135 |
| 4.4.3. | Gènes candidats identifiés par une délétion homozygote (ATAD3A et ATAD3B) | . 135 |
| 4.4.4. | Gènes candidats identifiés par homozygosity mapping et séquençage de l'exome | . 137 |
| 4.4.4.1 | . Famille de DI-S présentant un phénotype proche du Cockayne | . 137 |
| 4.4.4.1 | .1. Séquençage du gène YWHAE | . 139 |
| 4.4.4.1 | .2. Séquençage de l'exome | . 140 |
| 4.4.4.1 | .3. Séquençage du gène BUD13 | . 142 |
| 4.4.4.1 | .4. Séquençage du gène <i>RBMX2</i> | . 142 |
| 4.4.4.2 | . Séquençage de l'exome dans une famille de DI-S présentant une agénésie du corps ca | lleux |
| (PSL_10 | 0908 et PSL_10909) | . 142 |
| 4.5. 0 | Caractérisation de points de cassure de translocation (article 5) | .144 |
| 4.6. F | Résultats des patients présentant une DI et des TSA (article 6) | .147 |
| 5. Dis | scussion | .149 |
| 5.1. l | Jtilisation des puces SNP en diagnostic | .149 |
| 5.1.1. | Analyse globale des résultats | . 149 |
| 5.1.2. | Variation du nombre de CNV retrouvés en fonction du type de puce utilisé | . 150 |
| 5.1.3. | Interprétation des CNV en fonction du type d'anomalie et de la taille | . 151 |
| 5.1.4. | Apport des puces SNP en diagnostic du fait de la détection des disomies uniparentales | . 154 |
| 5.1.5. | Interprétation des CNV considérés comme des facteurs de susceptibilité | . 158 |
| 5.2. I | dentification de gènes candidats de DI | .159 |
| 5.2.1. | Identification de gènes candidats par les petites délétions de novo | . 161 |
| 5.2.2. | Identification de gènes candidats par les petites délétions transmises | . 163 |
| 5.2.3. | Identification de gènes candidats par les petites délétions homozygotes | . 164 |
| 5.2.4. | Identification de gènes candidats par homozygosity mapping | . 166 |
| 5.2.5. | Identification d'un gène candidat par caractérisation de translocation équilibrée | . 170 |
| Conclu | sion | .173 |
| Article | 1 | .191 |
| Article | 2 | .195 |
| Article | 3 | .201 |
| Article | 4 | .211 |
| Article | 5 | .219 |
| Article | 6 | .221 |

Avertissement

- L'expression "déficience intellectuelle" sera, dans ce mémoire, préférée par défaut, à la formule "retard mental", qui sera cependant employée, dans un sens totalement identique, lorsque référence sera faite à un article qui utilise ce terme.

- Toutes les coordonnées génomiques sont données selon la *build 37/hg19* sauf si précisé différemment.

- Toutes les formules chromosomiques sont données selon la nomenclature ISCN2013 (éditions Karger, Bâle, Suisse).

Keren Boris – Thèse de doctorat - Année 2013

Liste des abréviations

AAIDD : American Association on Intellectual and Developmental Disabilities ACPA : Analyse Chromosomique sur Puce ADN AD : Autosomique Dominant **APA** : American Psychiatric Association AR : Autosomique Récessif BAF : B Allele Frequency CNLOH : perte d'hétérozygotie sans perte de copie (Copy Neutral Loss of Heterozygosity) CNV : variation de nombre de copies (Copy Number Variation) del : délétion DI : Déficience Intellectuelle (substantif) ou atteint de déficience intellectuelle (adjectif) DS: Déviation Standard DSM : Diagnostic and Statistical Manual of Mental Disorders dup: duplication FISH : hybridation fluorescente in situ (Fluorescent In Situ Hybridization) GC : Gene Call Score kb : kilobases (1 000 paires de bases) LCR : Low Copy Repeats LRR : Log R Ratio LX : Lié à l'X Mb : Megabases (1 000 000 de paires de bases) MMRDR : Microhomology-Mediated Replication Dependent Recombination NAHR : recombinaison homologue non allélique (Non Allelic Homologous Recombination) NGS : séquençage massivement parallèle (*New Generation Sequencing*) NHEJ : Non Homologous End Joining MC : Malformation Congénitale OMS : Organisation Mondiale de la Santé QD : Quotient de Développement QI : Quotient Intellectuel **RM** : Retard Mental TED : Trouble Envahissant du Développement TSA : Troubles du Spectre Autistique **SNP** : Single Nucleotide Polymorphism SNV : Single Nucleotide Variant **SD** : Segmental Duplication SV : variations de structure (*Structural Variation*) UPD : disomie uniparentale (UniParental Disomy) iUPD : isodisomie uniparentale (UniParental IsoDisomy) VOUS : variant de signification inconnue (Variant of Unknown Significance)

Keren Boris – Thèse de doctorat - Année 2013

1. La déficience intellectuelle

Ce travail de thèse a porté sur des patients atteints de déficience intellectuelle (DI) : à partir de l'enquête diagnostique effectuée en analyse chromosomique, nous avons eu comme objectif l'identification de gènes candidats.

Dans cette première partie, nous ferons le point sur les données actuelles concernant la DI : les différentes définitions, un rappel historique, les classifications, l'épidémiologie et les différentes étiologies. Nous nous intéresserons ensuite plus spécifiquement aux étiologies génétiques connues : les causes chromosomiques, géniques et liées à l'empreinte. Enfin, nous examinerons les chevauchements génétiques entre la DI et les autres troubles neuropsychiatriques, en particulier les troubles du spectre autistique (TSA).

1.1. La déficience intellectuelle : généralités

1.1.1. Définitions de la déficience intellectuelle

Il existe plusieurs définitions de la déficience intellectuelle (DI), anciennement dénommée retard mental (RM).

1.1.1.1. Définition de l'APA

La définition de l'American Psychiatric Association (APA), donnée dans le manuel DSM-5, est celle qu'on utilise le plus souvent aux Etats-Unis et dans la littérature scientifique. La DI y est définie comme une faiblesse généralisée des capacités mentales avec un impact sur le fonctionnement adaptatif dans 3 domaines différents :

 le domaine conceptuel qui inclut le langage, la lecture, l'écriture, les mathématiques, le raisonnement, les connaissances, la mémoire, le temps, l'espace, l'argent ;

- le domaine social qui se réfère à l'empathie, la responsabilité sociale, le jugement social, les compétences de communication interpersonnelle, la méfiance/naïveté appropriée, la capacité à créer et maintenir des relations amicales, la capacité à résoudre les problèmes sociaux ;
- le domaine pratique qui se focalise sur l'autonomie : hygiène, utilisation de moyens de transport, responsabilités professionnelles, gestion du budget, gestion de l'emploi du temps, organisation du travail, etc.

Il n'y a pas de notion d'âge limite dans la définition du DSM-5, mais les symptômes doivent survenir durant la période du développement.

1.1.1.2. Définition de l'AAIDD

Pour l'American Association on Intellectual and Developmental Disabilities (AAIDD), la DI se définit comme une déficience caractérisée par une limitation significative du fonctionnement intellectuel et du comportement adaptatif, touchant les compétences pratiques et sociales de la vie de tous les jours. La déficience doit débuter avant l'âge de 18 ans.

Le fonctionnement intellectuel se réfère aux capacités mentales générales comme l'apprentissage, le raisonnement, la résolution de problèmes, etc. Le fonctionnement intellectuel est habituellement mesuré par le QI.

Le fonctionnement adaptatif comprend les 3 mêmes domaines de compétences que ceux de la définition du DSM-5 (cf. 1.1.1.1.) : compétences conceptuelles, sociales et pratiques.

Le diagnostic de DI doit également prendre en compte l'environnement communautaire et culturel ainsi que la diversité linguistique.

1.1.1.3. Définition de l'OMS

Dans la Classification Internationale des Maladies (CIM-10) de l'Organisation Mondiale de la Santé (OMS), la DI, encore appelée « retard mental » (RM), est définie comme un développement arrêté ou incomplet du fonctionnement mental, caractérisé essentiellement par une altération, durant la période du développement, des facultés qui déterminent le niveau global d'intelligence, c'est-à-dire des fonctions cognitives, du langage, de la motricité et des capacités sociales.

Cette classification a été la plus utilisée en Europe, mais elle est relativement ancienne (1990). La prochaine définition de la DI dans la CIM-11, prévue pour 2015, devrait se rapprocher de celles de l'APA et de l'AAIDD.

Il est déjà prévu que l'expression RM soit abandonnée au profit d'une autre terminologie (Salvador-Carulla et al. 2011).

1.1.2. La déficience intellectuelle vue par la société au cours de l'histoire

La vision qu'on a de la DI a varié au cours de l'histoire ; elle a tantôt été considérée comme une maladie, tantôt comme un handicap. De nombreux termes ont été utilisés pour désigner les individus DI, le plus souvent neutres à l'origine, puis devenus péjoratifs par leur usage : "idiots", "faibles d'esprits", "retardés mentaux", "crétins" ; et ceux-ci ont parfois été vus de façon positive comme réceptacles de la parole divine, ou avec indulgence comme des enfants ou comme objets de pitié. Mais le plus souvent on les a considérés de façon négative comme "malades mentaux", voire comme un poids et une menace envers la société.

La plus ancienne référence connue à la DI a été retrouvée à Thèbes sur un parchemin de l'Egypte ancienne datant de -1552 (Scheerenberger, 1983, Harris, 2006, Beirne-Smith et al. 2006, Harbour, 2010). Durant l'Antiquité grecque et romaine, la DI était vue comme une punition divine et en conséquence, ceux qu'elle atteignait étaient victimes de maltraitance, voire mis à mort. Au Moyen-âge, les individus ayant une DI légère et intégrés socialement n'étaient pas différenciés du reste de la population. Ceux qui avaient une DI plus sévère étaient soit objets de pitié et pris en charge par leur famille ou dans des monastères, soit confinés dans des asiles dont le seul but était leur retrait définitif de la société dite "normale".

Avec la fin du Moyen-âge, les approches religieuses et mystiques pour expliquer la DI furent progressivement remplacées par une approche médicale plus "moderne".

En 1690, le philosophe John Locke publia son « Essai sur l'entendement humain ». Il y réfutait le caractère inné des idées et y décrivait l'esprit à la naissance comme un contenant vide qui se remplit grâce à l'éducation et l'expérience. De ce fait, les troubles de l'esprit pouvaient se corriger grâce à une prise en charge appropriée. Il y faisait également, pour la première fois, la différence entre les "idiots" et les "fous".

Le premier programme documenté de prise en charge de patients atteints de DI fut développé en France, en 1799, par le Dr Jean Itard, rendu célèbre par sa tentative d'éducation et d'insertion de "l'enfant sauvage", Victor de l'Aveyron. Itard était influencé par l'idée du philosophe John Locke selon laquelle c'était l'éducation et l'expérience qui faisaient l'esprit.

Le Dr Edouard Séguin reprit les méthodes d'Itard, et établit le premier programme systématique d'éducation des « idiots », à l'Hôpital de la Salpêtrière à Paris. Pour lui, en effet, ils n'étaient pas atteints d'une maladie, mais avaient simplement été frappés d'un "arrêt du développement mental", avant ou après la naissance. Séguin émigra plus tard aux Etats-Unis où il établit plusieurs autres écoles pour éduquer les individus DI et fonda en 1876, l'*Association of Medical Officers of American Institutions for Idiotic and Feebleminded Persons*, aujourd'hui AAIDD.

Avec le développement de ces approches scientifiques, le besoin d'outils pour identifier et évaluer de façon objective la DI se fit ressentir. Alfred Binet, chercheur à la Sorbonne et son étudiant, le Dr Théodore Simon, furent les premiers à proposer, en 1905, un test objectif afin de mesurer l'intelligence : l'échelle Binet-Simon qui est encore utilisée aujourd'hui dans une version révisée : l'échelle Stanford-Binet (cf. 1.1.3).

Lors de la première partie du XIXème siècle, l'optimisme était prédominant quant à la perception qu'on avait des individus DI. On pensait que les chances de réhabilitation et d'intégration dans la société étaient élevées. Mais avec le temps, ces idéaux furent déçus. Si les patients bénéficiant des programmes de prise en charge systématique réalisaient de nets

progrès, cela ne leur permettait pas, néanmoins, de vivre une vie "normale", ce qui favorisa le développement des thèses eugénistes qui furent appliquées à la DI.

L'eugénisme est une discipline inventée par Sir Francis Galton et visant à l'amélioration de la composition génétique de la population. Galton était disciple de Charles Darwin et il voulut adapter la théorie de l'évolution à l'espèce humaine. Dans "Le Génie héréditaire", il développa l'idée que l'intelligence étant héréditaire et que la société civilisée prenant soin des plus faibles, on empêchait ainsi la sélection naturelle de s'effectuer, ce qui entraînait une "régression vers la médiocrité". L'eugénisme visait à inverser ce phénomène afin de favoriser la sélection positive des individus les plus doués et la sélection négative des individus les moins doués.

Les thèses eugénistes gagnèrent en popularité à partir de la fin du XIXème siècle et furent appliquées dans de nombreux pays : isolation, stérilisations et avortements forcés des personnes DI (mais également de celles souffrant d'autres handicaps ou appartenant à des populations considérées comme inférieures). L'application la plus extrême de l'eugénisme eut lieu, on s'en souvient, en Allemagne nazie, entraînant le meurtre d'individus DI.

Depuis 1945, l'eugénisme a été abandonné et est aujourd'hui interdit par la Charte Européenne des Droits Fondamentaux. La majorité des pays ont aujourd'hui des lois qui protègent les individus DI et leur donnent des droits.

1.1.3. Evaluation de la déficience intellectuelle

1.1.3.1. Evaluation du fonctionnement Intellectuel

Les tests évaluant le fonctionnement intellectuel sont les tests de QI. Ces tests portent sur :

- le raisonnement ;
- la résolution de problèmes ;
- la pensée abstraite ;
- la faculté de jugement ;
- l'apprentissage théorique (académique) ;
- l'apprentissage par l'expérience.

Les scores de QI sont empiriques et suivent une loi de distribution normale dans une population donnée, dont la moyenne est un score de 100 et la déviation standard (DS) est de 15. Cette loi peut être représentée par une courbe de Gauss (figure 1) :





Source : Wikimedia Commons (avril 2007)

Les tests de QI sont standardisés : l'objectif est de comparer des individus à d'autres individus d'une même population. Par conséquent, ils doivent être utilisés en fonction de l'âge et de la culture de la population dont fait partie l'individu testé.

Dans une population donnée, 68,2% des individus ont un QI entre -1 DS et +1 DS, c'est-à-dire entre 85 et 115. Par ailleurs, 95.4% des individus ont un QI entre -2 DS et +2DS, c'est-à-dire entre 70 et 130.

Classiquement, les 2,2% d'individus ayant un score de QI inférieur à 70 sont considérés comme ayant une DI, bien qu'officiellement, cette notion de seuil n'existe plus que dans la CIM-10, tandis qu'elle a disparu du DSM-5.

Les tests de QI comprennent 2 volets : l'un est verbal et l'autre porte sur les performances. Les résultats de l'une et l'autre partie peuvent diverger chez un même individu.

L'utilisation de tests de QI est critiquée et débattue, mais on ne dispose pas actuellement d'autres tests validés mesurant les performances intellectuelles.

De nombreux tests de QI différents existent. Les plus utilisés sont :

- "Wechsler Adult Intelligence Scale" (WAIS) : à partir de 16 ans ;
- "Wechsler Intelligence Scale for Children" (WISC) : entre 6 et 16 ans ;
- "Wechsler Preschool and Primary Scale of Intelligence" (WPPSI) : entre 3 et 7 ans ;

- Echelle de Stanford-Binet : entre 2 et 23 ans. C'est le test le plus ancien mais il n'est plus celui dont on se sert le plus communément.

Avant l'âge de 3 à 5 ans, on parle plutôt de quotient de développement (QD) que de QI. Et on dit "retard de développement" plutôt que DI quand le QD est diminué, car le retard de développement peut se normaliser avec l'âge et il n'évolue pas systématiquement vers une DI. Le QD peut se mesurer par l'échelle de Brunet-Lézine révisée ou on peut aussi se servir de l'échelle de Bayley (BSID) : il est égal au rapport de l'âge mesuré par le test sur l'âge réel. Ces tests s'utilisent également chez des individus plus âgés en cas de DI très profonde.

1.1.3.2. Evaluation du fonctionnement adaptatif

On teste les facultés adaptatives en mesurant la maturité sociale et émotionnelle de façon standardisée. Plusieurs domaines sont étudiés dont la communication, les compétences de la vie quotidienne, la socialisation et la motricité.

Parmi les tests les plus utilisés, on trouve d'abord l'échelle de Vineland (VABS) puis l'échelle de l'AAIDD.

1.1.4. Classifications de la déficience intellectuelle

De même que le faisait l'APA jusqu'au DSM-IVR, l'OMS utilise le QI afin de déterminer la sévérité de la DI. Depuis le DSM-5, l'APA a comme critère l'importance du déficit des facultés adaptatives. L'AAIDD, de son côté, classe la DI en fonction de l'intensité de la prise en charge nécessaire.

Ces 3 approches aboutissent à 3 classifications, comprenant chacune 4 catégories différentes de DI. En pratique, ces catégories regroupent les mêmes individus. En effet, pour donner un exemple, la DI sévère de l'APA correspond à la DI nécessitant un support étendu de l'AAIDD et au RM grave de l'OMS.

Il faut aussi noter que la réalité de la DI est un continuum. Certains individus peuvent donc se trouver à la limite entre 2 classes de DI selon ces classifications.

1.1.4.1. Classification de l'APA (DSM-5)

- DI légère : elle inclut 85% des individus DI. Ils sont autonomes, peuvent avoir un parcours scolaire avec soutien et nécessiter une aide pour gérer leurs finances, les transports et leur santé ;

- DI modérée : elle concerne 10% des individus DI. Ils ont des facultés communicatives adéquates, mais peuvent avoir besoin d'une aide dans la vie quotidienne pour des activités telles que les soins, les interactions sociales, etc. L'occupation d'un emploi aidé est possible ;

- DI sévère : elle se rencontre chez 3 à 4% des individus DI. Les facultés communicatives sont basiques, la prise en charge nécessite une aide quotidienne ;

- DI profonde : elle représente 1 à 2% des cas. Ils sont totalement dépendants d'une aide extérieure et nécessitent une prise en charge permanente.

1.1.4.2. Classification de l'AAIDD (Schalock et al. 2010)

- DI nécessitant une aide ponctuelle (pour certains moments difficiles de la vie) ;
- DI nécessitant une aide limitée (pour certaines activités) ;
- DI nécessitant un support étendu (pour la plupart des activités de la vie quotidienne) ;
- DI nécessitant un support invasif.

1.1.4.3. Classification de l'OMS (CIM-10)

- RM léger (F70) : avec un QI entre 69 et 50 qui entraîne des difficultés scolaires. Beaucoup d'adultes sont capables de travailler et de maintenir de bonnes relations sociales, et de s'intégrer dans la société ;

- RM moyen (F71) : avec un QI entre 49 et 35 qui se manifeste par un important retard de développement dans l'enfance. Néanmoins beaucoup de DI de cette catégorie peuvent acquérir des aptitudes scolaires, un certain degré d'indépendance et des capacités suffisantes pour communiquer. Les adultes ont besoin d'un soutien de niveau varié, pour travailler et vivre dans la communauté ;

- RM grave (F72) : QI entre 34 et 20 qui nécessite un besoin prolongé de soutien

- RM profond (F73) : QI inférieur à 20. On assiste à une limitation très marquée des capacités de prendre soin de soi-même, de contrôler ses besoins naturels, de communiquer et de se déplacer.

1.1.5. Prévalence de la déficience intellectuelle

La prévalence de la DI est difficile à mesurer car elle dépend beaucoup du contexte de l'étude et de la définition utilisée. Si on s'en tient à la définition théorique de l'OMS (QI<-2DS), la DI

devrait toucher 2,2% des naissances selon la distribution normale. Et en effet, dans la pratique, on constate une prévalence comparable à ce chiffre théorique.

La revue de la littérature effectuée par Roeleveld en 1997 est encore aujourd'hui considérée comme une référence (Roeleveld, 1997, Regan, 2010). Elle reprend toutes les études de prévalence ayant été conduites entre 1960 et 1995. Des variations très importantes ont été relevées allant de 2/1 000 à 85/1 000, pour les tous les RM confondus, de 4/1 000 à 79/1 000 pour les RM légers et de 3/1 000 à 7/1 000 pour les RM sévères. Ces variations étaient en grande partie liées au fait qu'aucune des 31 études de prévalence sélectionnées (principalement américaines et scandinaves) n'utilisait exactement la même méthodologie. En ne sélectionnant que les études les plus fiables, Roeleveld concluait que la prévalence était probablement aux alentours de 3% pour l'ensemble des RM et de 3/1 000 pour les RM sévères, avec un excès d'individus masculins d'environ 20%.

Une revue plus récente (Leonard et al. 2002) retrouve une fréquence proche de celle de Roeleveld : entre 1 à 3% des individus à la naissance avec un excès de garçons, d'environ 30 à 50%. Cet excès de garçons résulte en partie des causes génétiques liées au chromosome X, que nous développerons dans la partie 1.3.

1.1.6. Co-morbidité et mortalité de la déficience intellectuelle

Le phénotype des individus atteints de DI est très hétérogène. Outre d'importantes différences dans la sévérité de la DI, les patients peuvent souvent présenter d'autres symptômes associés. On distingue alors les DI syndromiques (ou associées, DI-S) des DI non syndromiques (ou isolées, idiopathiques, pures, DI-NS). La distinction entre ces 2 types de DI est floue, car certains signes associés peuvent être discrets. Il n'est pas rare que certaines DI-NS soient reclassées en DI-S une fois l'étiologie identifiée, et la caractérisation clinique affinée (Frints et al. 2002). De plus, la DI associée à des troubles uniquement neuropsychiatriques (TSA, dépression) est habituellement classée dans les DI-NS.

Les malformations congénitales (MC) sont présentes dans environ 27% des cas de DI légères à modérées, et dans 54% des DI sévères à profondes (Petterson et al. 2007) ; or elles ont un retentissement important sur le pronostic. Par ailleurs, le risque relatif qu'a un nouveau né

présentant une ou plusieurs MC d'évoluer vers une DI est de 8 à 26 (Jelliffe-Pawlowski et al. 2003, Découflé et al. 2001). Les MC les plus fréquentes sont d'abord celles de la trisomie 21 qui représente 8% des patients DI + MC (cardiaques, dont canal atrio-ventriculaire, et aussi des MC digestives, etc.). Et parmi les DI + MC non chromosomiques, les MC qu'on voit le plus souvent sont les malformations cérébrales et génito-urinaires (Petterson et al. 2007)

Les co-morbidités de la DI les plus souvent constatées sont les TSA et l'épilepsie (tableau 1). Selon certaines études, 28% des individus DI auraient des TSA associés (Bryson et al. 2008) et 67% des patients TSA sont atteints de DI (Fombonne et al. 2003). L'autisme et la DI partagent donc des mécanismes physiopathologiques communs. Les causes étiologiques et les facteurs génétiques identifiés dans les TSA jusqu'à présent sont également impliqués dans la DI, parfois sans TSA (cf. 1.5).

L'épilepsie est aussi fréquemment associée à la DI, présente dans 22% des DI. Cette association est d'autant plus importante que la DI est sévère : en effet, plus de la moitié des patients ayant une DI profonde sont atteints d'épilepsie (Oeseburg et al. 2011). Les crises tonico-cloniques sont le type de crises qu'on observe le plus couramment (60%) ; viennent ensuite les crises partielles complexes (20%).

Les troubles moteurs d'origine centrale (ataxie, spasticité, dystonie, etc.) sont présents chez environ 20% des patients DI, et sont également plus fréquents dans les cas sévères.

A côté des TSA, on constate des troubles neuropsychiatriques plus fréquents en cas de DI que dans la population générale (hyperactivité, déficit de l'attention, agressivité, dépression) ainsi que des déficits sensoriels (auditifs, visuels).

Tableau 1 : prévalence de la co-morbidité dans la DI, en dehors de la trisomie 21 et d'autres syndromes chromosomiques connus

| Pathologie | | Prévalence (%) |
|-----------------------|------------------------------|----------------|
| | épilepsie | 22 |
| Atteinte neurologique | troubles moteurs | 20 |
| | surdité | 5 |
| | accident vasculaire cérébral | 2 |
| | autisme | 28 |
| | autres TSA | 7 |

| Atteinte neuropsychiatrique | déficit de l'attention | 10 |
|-----------------------------|------------------------|----|
| | opposition/défiance | 12 |
| | anxiété | 17 |
| | tics | 1 |
| | cérébrales | 5 |
| | musculo-squelettiques | 3 |
| Malformations | cardiovasculaires | 2 |
| | urogénitales | 1 |
| | digestives | <1 |
| | spina bifida | <1 |

D'après Petterson et al. 2007, Bryson et al. 2008, et Oesburg et al. 2011

L'association récurrente de la DI avec certains symptômes permet de définir des syndromes et de grouper les patients dans des groupes phénotypiquement homogènes. C'est ainsi que les individus porteurs de trisomie 21 ont été identifiés comme atteints du même syndrome (Down, 1866) près de 93 ans avant la découverte de l'étiologie sous-jacente (Lejeune et al. 1959)

Le taux de mortalité est normal chez les personnes présentant une DI légère, mais il est augmenté de 20% en moyenne chez les patients présentant une DI modérée à sévère (Patja et al. 2000, Tyrer et al. 2009). L'espérance de vie est d'autant moindre qu'il existe des troubles associés, en particulier épilepsie, malformations cérébrales, troubles neuropsychiatriques, malformations cardio-vasculaires, et que le QI est bas ; l'espérance de vie peut être jusqu'à 35 ans inférieure à celle de la population générale, dans les cas de DI les plus profondes. Néanmoins, en raison d'une meilleure prise en charge des patients, l'espérance de vie de ceux-ci a beaucoup augmenté depuis les années 1960 (Patja et al. 2000).

Les principales causes de surmortalité sont, d'une part, les infections respiratoires qui résultent de troubles de la déglutition d'origine neurologique, et qu'on rencontre dans 20 à 35% des cas, et, d'autre part les malformations congénitales associées à la DI (Patja et al. 2001, Tyrer et al. 2009).

1.1.7. Etiologies de la déficience intellectuelle

L'identification de l'étiologie de la DI présente de nombreux bénéfices. Pour la famille, sur le plan psychologique, cela permet de mettre fin à la quête diagnostique et à l'incompréhension devant les troubles du sujet atteint. Cela permet aussi d'évaluer le risque qu'ont les membres de la famille de transmettre la pathologie et, lorsque cela est possible, de proposer un diagnostic prénatal lors des grossesses à risque. Enfin, même s'il est rarement possible de corriger l'anomalie qui est à l'origine de la DI, il existe des traitements symptomatiques pour certaines causes de DI.

Les étiologies de la DI sont extrêmement nombreuses et hétérogènes. De ce fait, il devient aujourd'hui difficile d'en faire une liste exhaustive, étant donné la fréquence avec laquelle on en découvre de nouvelles.

En dehors de la trisomie 21 qui représente 7 à 11% des DI (Rauch et al. 2006, Oeseburg et al. 2011), il existe un grand nombre d'étiologies de DI mais chacune prise séparément est rare ou très rare : la seconde cause la plus fréquente, l'X fragile, ne représente en effet que 1% des DI.

On estime que, chez 60% des patients, on ne retrouve pas l'étiologie de leur DI, et ceci pour plusieurs raisons. D'abord, beaucoup d'étiologies connues de DI sont trop rares pour qu'on les recherche avec les techniques actuellement utilisées en routine. Ensuite, certaines étiologies (en particulier toxiques ou infectieuses) ne peuvent pas être démontrées à distance de leur survenue : le diagnostic repose alors sur un faisceau d'arguments cliniques et paracliniques, et non sur un test diagnostique. Enfin, un très grand nombre d'étiologies de DI n'a pas encore été identifié et ne peut donc être recherché actuellement.

Schématiquement, parmi les causes connues de DI, on distingue 2 groupes : les étiologies génétiques et les étiologies acquises, qu'on appelle également exogéniques ou environnementales.

1.1.7.1. Etiologies acquises

Les étiologies acquises représentent 1 à 20% des DI selon les sources (Curry et al. 1997, van Karnebeek et al. 2005, Rauch et al. 2006). Deux explications sous-tendent cette importante variabilité : la première est que le classement dans le groupe des étiologies acquises se base sur des critères qui diffèrent selon les études, puisqu'il n'existe pas réellement de test diagnostique ; la seconde est que certaines étiologies, telles que le syndrome d'alcoolisme fœtal, sont de prévalence très variable selon l'origine géographique.

Dans le cas des étiologies acquises, la DI est le plus souvent due à une atteinte cérébrale survenue à un stade critique du développement, prénatal ou postnatal, entraînant des dommages irréversibles.

Quant à ce qui provoque les étiologiques acquises, on distingue 3 grands groupes : les causes toxiques, infectieuses et traumatiques.

Parmi les causes de DI toxiques, l'alcoolisme fœtal est la plus fréquente. Viennent ensuite l'intoxication à d'autres drogues (amphétamines, cocaïne, etc.), puis l'intoxication au plomb ou au mercure.

La phénylcétonurie chez la mère est également souvent classée dans les causes toxiques de DI, quand elle n'est pas traitée durant la grossesse.

Parmi les causes infectieuses, on compte les infections anténatales au CMV, la rubéole, la toxoplasmose, la listéria, etc. et les encéphalites anté ou postnatales.

Les hémorragies intracérébrales, l'ischémie consécutive à un accident vasculaire cérébral ou après noyade font partie des causes traumatiques de DI.

Enfin, il est à noter que quelque soit la cause, des facteurs socioculturels peuvent favoriser la DI, de sorte qu'elle est plus fréquente dans les milieux défavorisés (Emerson, 2007).

1.1.7.2. Etiologies génétiques

Il est classiquement reconnu qu'on identifie les étiologies génétiques dans environ 40% des DI : elles représentent donc un groupe beaucoup plus important que les étiologies acquises (van Karnebeek et al. 2005, Rauch et al. 2006).

Ce chiffre de 40% étant relativement ancien, il est probable qu'il soit quelque peu sous-estimé. Depuis, de nouvelles causes de DI monogéniques sont recherchées, grâce à de nouvelles techniques utilisées en routine, en particulier l'analyse chromosomique sur puce ADN (ACPA). L'ACPA permet de réaliser 5 à 20% de diagnostics étiologiques chez les patients dont la DI n'avait pas d'étiologie (Stankiewicz et al. 2007, Keren (a), 2010, Miller et al. 2010, Cooper et al. 2011, Hochstenbach et al. 2011, Schaaf et al. 2011, Shoukier et al. 2013).

On peut schématiquement classer les étiologies génétiques en 2 catégories : chromosomiques ou géniques.

Actuellement, les étiologies chromosomiques sont les plus fréquentes (environ 20%) et sont diagnostiquées par les techniques de cytogénétique : le caryotype, l'hybridation *in situ* (FISH) et l'ACPA. Elles peuvent être classifiées soit selon le type d'anomalie concerné : de nombre ou de structure, équilibrées ou avec gain ou perte de matériel (délétion, duplication, triplication, etc.) ; soit selon leur mécanisme (NAHR ou autre).

Les étiologies géniques sont dues à des mutations ponctuelles de l'ADN et leur diagnostic utilise des techniques de génétique moléculaire, en particulier le séquençage qui est la plus utilisée. Elles peuvent être classifiées en fonction du mode de transmission : autosomique récessif (AR), autosomique dominant (AD), ou lié à l'X. Elles peuvent être également classées selon la fonction du gène responsable : gène d'une voie métabolique, gènes de la synapse, facteur de transcription, molécule d'adhésion, etc.

Remarquons que certaines anomalies géniques ou chromosomiques sont également identifiées dans la population générale, mais qu'elles sont retrouvées avec une fréquence plus importante chez les individus DI. Il s'agit donc de facteurs qui favorisent la DI, mais qui ne suffisent pas à eux seuls à la provoquer. On parle alors de facteurs de susceptibilité (Gregor et al. 2011).

Figure 2 : étiologies de la DI



D'après Rauch et al. 2006

1.2. Génétique de la déficience intellectuelle

1.2.1. Variations normales et pathologiques du génome humain

Une des bases de la variabilité phénotypique normale et pathologique dans une espèce est la variabilité de séquence du génome, parmi les individus de cette espèce. Le génome humain est soumis à une dynamique mutationnelle dont il résulte un spectre complexe de variants de

séquence et de structure de l'ADN, qui vont des altérations nucléotidiques simples (mutations ponctuelles : substitutions nucléotidiques, petites insertions ou délétions) à des remaniements de grande taille concernant des segments de gènes voire de chromosomes et modifiant le nombre de copies des séquences soumises à remaniements, en passant par des expansions de répétitions en tandem. Les variations du génome humain sont, dans leur grande majorité, non pathogènes. L'objectif du diagnostic génétique est de distinguer ce qui est pathologique parmi ces variants et de le relier au phénotype des patients.

1.2.1.1 SNP

Les SNP (*Single Nucleotide Polymorphisms*) correspondent à des substitutions nucléotidiques d'une seule base. Jusqu'à présent, on a répertorié environ 53 millions de SNP, ce qui représente environ 90% des variations du génome humain (cf. base de données dbSNP 137).

Sur le plan terminologique, on distinguait classiquement les SNP des SNV (*Single Nucleotide Variants*) ; les SNV correspondaient à n'importe quelle substitution nucléotidique, alors que les SNP se rapportaient uniquement aux variants non pathogènes. Actuellement, par abus de langage, et parce qu'il existe une incertitude quant à la signification clinique de beaucoup de ces variants, les deux expressions tendent à être employées indifféremment.

La méthode classique pour identifier les SNP est le séquençage de l'ADN. En moyenne, chaque individu a environ 3,2 millions de SNP dans son génome, c'est à dire de nucléotides différents de ceux de la séquence humaine de référence hg19 (Ball et al. 2012). Et parmi ces SNP, 8 000 sont des mutations non synonymes affectant des séquences codantes et donc modifiant la séquence protéique. Le taux de survenue de SNP à chaque génération est d'environ 12.10⁻⁹ (Lynch, 2010). En conséquence, chaque nouveau-né a dans son génome, en moyenne, 70 SNP survenus *de novo*. La très grande majorité de ces SNP n'est pas pathogène. Et la probabilité qu'un SNP ait une signification clinique est d'autant plus importante qu'il entraîne une modification protéique. Les critères utilisés pour évaluer la potentielle pathogénicité d'un SNP sont expliqués à la partie 3.5.3.3.

1.2.1.2. Variations de structure

Les variations de structure (SV pour *Structural Variations*) peuvent être divisées en 2 catégories : les remaniements équilibrés, sans perte ni gain de matériel chromosomique, et les remaniements déséquilibrés, avec perte ou gain de matériel chromosomique. Parmi les SV déséquilibrés, on distingue les CNV (pour *Copy Number Variations*) qui sont des SV de grande taille > 1 kb (Redon et al. 2006) ; et les indel, de petite taille < 1 kb (Mullaney et al. 2010).

<u>CNV</u>

Les CNV retrouvés dans la population générale représentent environ 12% du génome humain (Redon et al. 2006, Kidd et al. 2008). La technique classique pour détecter les CNV est le caryotype, mais sa résolution est limitée aux anomalies de taille supérieure à 5Mb, au mieux. Actuellement, c'est l'ACPA qui tend à remplacer le caryotype.

Le taux d'apparition des CNV est difficile à évaluer. Avec une résolution de 30 kb, un CNV d*e novo* est retrouvé avec un taux de 1,2 / 100 génome à chaque génération. Même si la majorité des CNV n'est pas pathogène, ils représentent une cause majeure de maladie génétique, et en particulier de DI-S et DI-NS (cf. 1.2.2.2.). Les critères utilisés pour évaluer la pathogénicité des CNV sont expliqués à la partie 2.1.

indel

Chaque individu possède, en moyenne, 300 000 indel (dont 70% de délétions et 30% d'insertions), par comparaison au genome de référence hg19 (Ball et al. 2012), ce qui en fait la deuxième origine de variation du génome humain derrière les SNP. Le taux de survenue des indel est estimé à 0,6.10⁻⁹ pour les délétions et 0,2.10⁻⁹ pour les insertions.

Les petites indel, < 50 pb, sont généralement détectées par les méthodes de séquençage. L'identification des grandes indel est plus difficile et nécessite des techniques ciblées (southern Blot, MLPA, etc.). Les critères utilisés pour évaluer la potentielle pathogénicité des indels sont comparables à ceux utilisés pour les SNP (cf. 3.5.3.3.).

1.2.1.3. Répétitions

On distingue 3 types de répétitions : les répétitions en tandem, les LCR (*Low Copy Reapeats*) et les éléments mobiles. L'étude des répétitions est généralement difficile, nécessitant la combinaison de plusieurs techniques (PCR long range, Southern Blot, TP-PCR, etc.).

Répétitions en tandem

On en distingue 3 différentes :

les satellites qui sont des répétitions d'éléments > 60 pb, situés dans les centromères et
 l'hétérochromatine constitutive ;

- les minisatellites qui sont des répétitions d'éléments de 10 à 60 pb ;

- les microsatellites qui sont des répétitions d'éléments de taille < 10 pb.

De nombreuses pathologies sont liées à un nombre anormal de répétitions de microsatellites ; par exemple l'X fragile, première cause monogénique de DI, est dû à un nombre de répétitions d'un triplet CGG > 200 dans la région 5'UTR du gène *FMR1* ce qui entraîne une répression de son expression.

Les pathologies, liées aux autres types de répétitions en tandem que les microsatellites, sont beaucoup plus rares. On peut citer la dystrophie facio-scapulo-humérale, liée à un nombre de répétitions < 10 du satellite D4Z4 dans la région 4q35, entraînant une modification de conformation de la chromatine et une dérégulation de l'expression des gènes de cette région.
LCR

Les LCR, également appelés duplicons ou SD (*Segmental Duplications*) (Stankiewicz et al. 2002), sont des blocs d'ADN, de taille > 1 kb, situés en plusieurs localisations génomiques et présentant > 90% d'identité. Ces régions fortement homologues représentent environ 6% du génome humain et augmentent la fréquence des mésappariements à la fois en méiose et en mitose (cf. 1.2.2.2.1.).

Eléments mobiles

Les éléments mobiles, également appelés transposons, sont des séquences d'ADN d'origine virale, capables de se multiplier et de se déplacer de façon autonome. Les principaux sont les suivants :

- les séquences LINE, de 7 kb, représentent 17% du génome ;
- les séquences Alu, de 280 pb, représentent 10% du génome.

La grande majorité de ces éléments mobiles n'a pas de signification clinique connue. Ils peuvent toutefois être responsables de réarrangements chromosomiques. Et on peut aussi citer le cas du gène *FKTN*, responsable de dystrophie musculaire congénitale avec anomalies cérébrales et oculaires, dont le variant délétère le plus fréquent correspond à l'insertion d'un élément mobile dans la région 3'UTR, entraînant une répression de l'expression du gène.

1.2.2. Causes chromosomiques de déficience intellectuelle

Depuis la mise en évidence, en 1959, de la première anomalie chromosomique chez l'homme, la trisomie 21 (Lejeune et al. 1959), de nombreuses anomalies de nombre, mais aussi de structure des chromosomes ont été détectées chez des patients présentant une DI (Grouchy et al. 1984). Les anomalies chromosomiques représentent actuellement la cause majeure des retards mentaux. En effet, on évalue classiquement à 10-15% la proportion des retards mentaux dus à une anomalie chromosomique détectable avec le caryotype en bande (Rauch et al. 2006). Depuis la généralisation de l'ACPA, ce taux a augmenté, car une anomalie pathogénique est retrouvée dans 5 à 20% des DI, pour lesquelles on ne distinguait aucune cause après un bilan initial (Stankiewicz et al. 2007, Keren (a), 2010, Miller et al. 2010, Cooper et al. 2011, Hochstenbach et al. 2011, Schaaf et al. 2011, Shoukier et al. 2013).

On peut subdiviser les causes chromosomiques de DI en anomalies de nombre et anomalies de structure. Les anomalies de nombre représentent environ la moitié des causes chromosomiques de DI. Les principales anomalies de structure responsables d'une DI sont d'une part les délétions, et d'autre part les dérivés de translocation, générant le plus souvent une délétion partielle d'un chromosome, associée à une trisomie partielle d'un autre chromosome.

1.2.2.1. Anomalies de nombre

Les anomalies de nombre de chromosomes, également appelées aneuploïdies, sont très fréquentes à la conception. Elles résultent le plus souvent de non disjonctions méiotiques. Il a été estimé qu'environ 25% des blastocystes avaient un nombre de chromosomes anormal (Eiben et al. 1990, Stephenson et al. 2002), ce qui dans la très grande majorité des cas est à l'origine de fausses couches spontanées précoces.

Parmi les anomalies de nombre, celles qui sont viables et responsables de DI sont les trisomies 13, 18 et 21, qu'on appelle aussi, respectivement, syndrome de Patau, d'Edwards et de Down (tableau 2). La trisomie 21 est de très loin la plus fréquente des anomalies chromosomiques (incidence spontanée de 1 / 700 naissances, incidence en France : 1 / 4 000 en raison du diagnostic prénatal) et la première cause de DI (7 à 11% des DI).

Les anomalies de nombre des chromosomes sexuels sont également viables, mais la DI n'étant pas au premier plan du phénotype, elles ne sont habituellement pas considérées commes des étiologies classiques de DI.

| syndrome | trisomie | fréquence | OMIM | orphanet | signes classiques |
|----------|----------|------------|--------|----------|---|
| Down | 21 | 1/4 000 | 190685 | 870 | dysmorphie caractéristique, canal atrio-ventriculaire, MC digestives |
| Edwards | 18 | 1/6-8 000 | - | 3380 | hypertonie, MC cardiaques et cérébrales, décès avant 1 an dans 95% des cas. |
| Patau | 13 | 1/8-15 000 | - | 3378 | polydactylie, MC cardiaques, holoprosencéphalie, décès avant 1 an dans 95% des cas. |

Tableau 2 : résumé des principales anomalies de nombre responsables de DI.

D'après Regan et al. 2010 et Orphanet.

A côté de ces 3 syndromes, il existe d'autres trisomies responsables de DI, soit en mosaïque (par exemple trisomies 8 et 9 en mosaïque), soit incomplètes (marqueur chromosomique surnuméraire).

1.2.2.2. Anomalies de structure

Les SV sont des causes classiques de DI. On distingue les remaniements classiques des microremaniements, ces derniers étant de taille inférieure à 5 Mb, et donc non visibles au caryotype.

Pour les syndromes chromosomiques, la notion de "région critique" est très importante : il s'agit de la région minimale qui, lorsqu'elle est délétée ou dupliquée, entraîne la présence des principaux signes cliniques permettant de caractériser le syndrome. Par exemple, pour le syndrome de Wolf-Hirschhorn (ou syndrome 4p-), les délétions sont le plus souvent visibles au caryotype et font jusqu'à 25 Mb. Mais il existe une région critique de 600 kb, appelée locus WHSC-2, non visible au caryotype, et qui lorsqu'elle est délétée, suffit à provoquer un phénotype reconnaissable de 4p-.

Parfois la région critique ne correspond qu'à un seul gène, qu'on appelle en ce cas le "gène critique" du syndrome. Il est alors fréquent qu'une fois ce gène critique caractérisé, des mutations ponctuelles soient à leur tour identifiées comme étant à l'origine du syndrome.

Ainsi certains syndromes peuvent à la fois être causés par des délétions d'une région chromosomique et par des mutations ponctuelles dans le gène critique de cette région, par exemple la délétion subtélomérique 9q et les mutations d'*EHMT1*.

1.2.2.2.1. Mécanisme de survenue des CNV

L'ADN est soumis à une dynamique mutationnelle entraînant des modifications nombreuses et complexes du génome humain, allant de la simple variation mononucléotidique à de grandes variations de structure génomique. C'est dans ce dernier cas que l'on parle de remaniements chromosomiques.

Si les mutations ponctuelles sont généralement induites par des erreurs de réplication, les variations de structure sont principalement expliquées par des modèles reposant sur des mésappariements entre séquences homologues (NAHR) ou des cassures double-brin (NHEJ).

Recombinaison homologue

Un certain nombre de réarrangements génomiques sont de taille récurrente, survenant dans des régions d'instabilité liées à l'architecture génomique. C'est le cas par exemple des syndromes de Smith-Magenis ou DiGeorge (cf. 1.2.2.2.), mais également de quelques maladies mendéliennes telles que le syndrome de Charcot-Marie-Tooth de type Ia. Ces anomalies s'expliquent par le modèle NAHR pour *Non Allelic Homologous Recombination* (recombinaison homologue non allélique) qui correspond à une recombinaison et un *crossing-over* entre des LCR. Le risque de NAHR est d'autant plus fréquent que la taille des LCR est importante : en pratique, les anomalies chromosomiques récurrentes sont causées par des LCR de taille > 10 kb.

Selon la localisation des LCR en *cis* (intrachromatidiens) ou en *trans* (interchromatidiens, interchromosomiques entre chromosomes homologues ou interchromosomiques entre chromosomes non homologues) et selon leur orientation (directe ou inversée), un *crossing-over* au sein de ces mésappariements peut causer des délétions, des duplications, des inversions, des insertions ou des translocations (figure 3). Cela signifie également que pour

toute délétion récurrente causée par un NAHR, il existe une duplication réciproque avec les mêmes points de cassure, en pratique, beaucoup moins fréquente. Les LCR correspondent donc à des "*hot spots*" de microremaniements génomiques, qui ont pour un locus donné, des points de cassure récurrents. En effectuant une cartographie des LCR, de leur taille et leur homologie, il est même possible de prévoir, en théorie et avec une certaine fiabilité, les remaniements récurrents possibles et leur fréquence (Sharp et al. 2005, Dittwald et al. 2013)

D'autres séquences répétées de façon plus fréquente que les LCR (LINE, SINE, satellites, microsatellites...) peuvent également être responsables de NAHR, mais dans ce cas, les points de cassure ne sont pas récurrents.

Figure 3 : exemples de NAHR et conséquences



interchromatinien ou interchromosomique interchromosomique



D'après Chen et al. 2010

Il existe un deuxième modèle qui explique le mécanisme des réarrangements par la présence des LCR : le BIR (*Break Induced Replication*).

Le BIR suppose une cassure double-brin dans un LCR. Lors de la phase S, la fourche continue la réplication sur un LCR homologue situé sur un autre locus, provoquant ainsi un réarrangement (Smith et al. 2007). Il est très difficile de distinguer le BIR d'un NAHR, car dans les 2 cas, les points de cassure correspondent à des LCR et sont récurrents.

<u>NHEJ</u>

Le NHEJ (*Non-Homologous End Joining*) est un des principaux mécanismes de réparation de l'ADN (Lieber et al. 2003). Il a pour fonction de réparer à la fois les cassures double-brin physiologiques, telles que les réarrangements VDJ au niveau des gènes d'immunoglobuline (entre les gènes Variables, de Diversité et de Jonction des chaînes d'immunoglobuline et des récepteurs T dans les lymphocytes), et aussi les cassures double-brin pathologiques, par exemple celles qui sont induites par les radiations ionisantes. Ce mécanisme ne nécessite pas la présence de LCR, ni même de régions d'homologie. Le principe résumé du NHEJ est une ligation des brins d'ADN qui ont été séparés par la cassure double-brin. Parfois de petites délétions ou insertions (dites indel) surviennent au niveau du point de réparation : c'est ce

qu'on appelle la cicatrice d'information (*information scar*). Mais le NHEJ peut aussi entraîner la ligation de 2 brins d'ADN physiquement éloignés lorsque 2 cassures double-brin sont situées en des localisations génomiques différentes, ce qui entraîne des remaniements. Selon la localisation des 2 cassures double-brin, on peut obtenir des délétions, des translocations ou des inversions. En cas de 3 ou plus cassures double-brin, on peut obtenir des réarrangements complexes (figure 4).

A l'inverse du NAHR, les remaniements causés par le NHEJ n'ont pas de points de cassure récurrents.



Figure 4 : exemples de NHEJ et conséquences

D'après Chen et al. 2010

MMRDR ou FoSTeS

Le MMRDR (*Microhomology-Mediated Replication Dependent Recombination*) ou FoSTeS (*Fork Stalling and Template Switching*) est un modèle proposé pour expliquer des réarrangements intrachromatidiens très complexes, où plusieurs régions en perte de copies sont séparées par des régions en gains de copies ou en nombre de copies normal. Ce type d'anomalie est par exemple fréquent au niveau du gène *PLP1* de la maladie de Pelizaeus-Merzbacher (Lee et al. 2007a).

D'après ce modèle, lors de la réplication de l'ADN, la fourche peut se décrocher de sa matrice et reprendre la polymérisation sur une nouvelle matrice, en raison de régions de micro homologies. Si la nouvelle matrice est en amont, cela provoquera un gain de copies ; si elle est en aval, cela provoquera une perte de copie. Ce mécanisme peut survenir plusieurs fois lors de la réplication, provoquant ainsi un réarrangement complexe fait d'une succession de régions en gain, perte ou nombre de copies normal.

1.2.2.2.2. Principaux syndromes délétionels

La majorité des délétions et microdélétions pathologiques sont dites privées, c'est-à-dire spécifiques de l'individu qui les porte, avec des points de cassure pouvant survenir potentiellement n'importe où sur le génome. Il est donc très difficile de faire le compte des réarrangements possibles. Toutefois, un certain nombre d'anomalies sont plus fréquentes que d'autres et peuvent être corrélées à un phénotype caractéristique chez les patients atteints. Ces anomalies qualifiées de classiques peuvent donc être évoquées cliniquement devant une présentation syndromique spécifique (tableau 3). La très grande majorité des microdélétions récurrentes entraîne une DI, à l'exception de la délétion 17p12, comportant le gène *PMP22* et responsable de la neuropathie par hypersensibilité à la pression (Weise et al. 2012, Keren (a) et al. 2013).

Tableau 3 : principaux syndromes délétionnels et microdélétionnels avec DI ; en gras noir : les gènes pour lesquels des mutations ponctuelles ont été identifiées comme responsables d'une partie du phénotype ; en gras rouge : les gènes pour lesquels des mutations ponctuelles ont été identifiées comme responsables du phénotype complet

| locus | syndrome | taille (Mb) | mécanisme | gènes critiques | OMIM | orphanet | principaux signes (en plus de la DI) |
|--------|----------------------------------|----------------|-----------|--------------------|--------|----------|---|
| 1p36 | - | variable | variable | - | 607872 | 1606 | dysmorphie typique, spasmes |
| 2q37 | osteodystrophie Albright-like | variable | variable | HDAC4 | 600430 | 1001 | brachymétacarpie, brachymétatarsie, obésité |
| 4p16 | Wolf-Hirschhorn | variable | variable | WHSC1, LETM1 | 194190 | 280 | dysmorphie en casque grec, petite taille, épilepsie |
| 5p15 | cri-du-chat | variable | variable | SEMAF, CTNND2 | 123450 | 281 | cri caractéristique, visage lunaire |
| 7q11.2 | Williams-Beuren | 1,5-1,8 | NAHR | ELN1, | 194050 | 904 | dysmorphie |

| | | | | LIMK1 | | | typique, sténose |
|----------|-----------------|-----------|----------|---------------------|--------|--------|----------------------|
| | | | | | | | aortique, |
| | | | | | | | hyperlaxité, |
| | | | | | | | atteinte visuo- |
| | | | | | | | spatiale |
| | | | | | | | cardiopathie, |
| 8p23.1 | - | 5 | NAHR | GATA4 | - | 251071 | hernie |
| | | | | | | | diaphragmatique |
| | | | | 70064 | | | syndrome tricho- |
| 8q23q24 | Langer-Giedon | variable | variable | TRPS1, | 150230 | 502 | rhino-phalangien, |
| | | | | EXII | | | exostoses |
| | | | | | | | dysmorphie |
| 9q34.3 | Kleefstra | variable | variable | EHMT1 | 610523 | 96147 | caractéristique, |
| | | | | | | | synophrys |
| | | | | DAYC | | | aniridie, |
| 11p13 | WAGR | variable | variable | PAX0, | 194072 | 89 | nephropathie, |
| | | | | VV 1 1 | | | nephroblastome |
| | | | | EVT2 | | | foramens |
| 11p11.2 | Potocki-Shaffer | variable | variable | EX12, | 601224 | 52022 | bipariétaux, |
| | | | | ALA4 | | | exostoses |
| 15011.2 | Prader_Willi | 5 2-6 | МАНР | SNIPDN | 176270 | 720 | hypotonie, |
| 15411.2 | Flauer-Willi | 5,5-0 | NAIM | SINITEN | 1/02/0 | 735 | hyperphagie |
| 15a11 2 | Angelman | 5 3-6 | NAHR | LIBE3A | 105830 | 72 | microcéphalie, |
| 15911.2 | Angeiman | 5,5 0 | NAIIN | ODESA | 105050 | 72 | épilepsie |
| 16n13 3 | ATR-16 | variable | variable | HBA1, | 141750 | 9879 | α-thalassémie |
| 10013.5 | //// 10 | variable | Variable | HBA2 | 141750 | 5075 | a maassenne |
| 17p13.3 | Miller-Dieker | variable | variable | ple PAFAH1B1 | 247200 | 531 | lissencéphalie, |
| 1, 11010 | | | | YWHAE | | | épilepsie |
| | | | | | | | dysmorphie |
| 17p11.2 | Smith-Magenis | 3,7 | NAHR | RAI1 | 182290 | 819 | typique, troubles |
| | | | | | | | du sommeil |
| 17a12 | RCAD | 1.5 | NAHR | HNF1B | 189907 | 261265 | diabète, kyste |
| | | _/- | | | | | rénaux |
| 17a21.3 | Koolen-DeVries | 0.7 | NAHR | KANSL1 | 610443 | 96169 | hypotonie, visage |
| | | , | | | | | allongé |
| | | | | | | | atrésie des |
| 18q23 | - | variable | variable | - | 601808 | 1600 | conduits auditifs |
| | | | | | | | externes |
| | | | | | | | cardiopathie cono- |
| 22q11.2 | DiGeorge | 1,5-3 | NAHR | TBX1 | 188400 | 567 | troncale, |
| | -1 -1 | | | | | | insuffisance vélaire |
| 22q13.3 | Phelan- | variable | variable | SHANK3 | 606232 | 48652 | TSA, avance |
| | McDermid | · · · · · | | | | | staturale |

D'après Keren(a) et al. 2013

1.2.2.2.3. Principaux syndromes microduplicationnels

Comme expliqué précédemment, le NAHR est responsable de microremaniements chromosomiques de tous types et pas uniquement de délétions. Par conséquent, les syndromes microdélétionnels à points de cassure récurrents ont chacun une microduplication correspondante (Weise et al. 2012, Keren (a) et al. 2013).

Ces microduplications ont plusieurs caractéristiques :

- leur fréquence est jusqu'à 10 fois inférieure à leur microdélétion correspondante. Cela est lié, en partie, au fait qu'elles ne peuvent résulter que de NAHR en *trans* alors que les microdélétions peuvent résulter de NAHR en *cis* et en *trans* ;

- il existe souvent une très grande hétérogénéité phénotypique parmi les patients porteurs de la même anomalie, avec absence de tableau clinique réellement caractéristique, ce qui rend leur diagnostic clinique très difficile ;

leur phénotype est le plus souvent moins sévère que la microdélétion correspondante.
La DI est inconstante dans certaines microduplications ;

- la plupart des microduplications, mais pas toutes, sont plus souvent transmises et plus rarement *de novo* que leur microdélétion correspondante.

| locus | syndrome | taille (Mb) | OMIM | Orphanet | principaux signes cliniques (en plus |
|---------|-----------------------|-------------|--------|----------|--------------------------------------|
| | | | | | de la DI) |
| 7q11.2 | Somerville-Van der Aa | 1,5-1,8 | 609757 | 96121 | retard de langage |
| 8p23.1 | - | 5 | - | 251076 | - |
| 15q11.2 | | 5.3-8 | 608636 | 238446 | - |
| 17p11.2 | Potocki-Lupski | 3,7 | 610883 | 1713 | DI constante |
| 22q11.2 | - | 1,5-3 | 608863 | 1727 | MC cardiaque, insuffisance vélaire |

Tableau 4 : principaux syndromes duplicationnels récurrents par mécanisme NAHR

D'après Keren(a) et al. 2013

1.2.2.2.4. CNV facteurs de susceptibilité de la DI

Avec le développement de l'ACPA, on a décrit des microremaniements récurrents, relativement fréquents, et avec des points de cassure récurrents qui se situent au niveau de LCR et qui entraînent des phénotypes très hétérogènes d'un patient à l'autre. Pour un même locus, on constate des CNV en gain et perte. Ces CNV sont le plus souvent transmis par un parent sain, et on les retrouve aussi chez des individus contrôles. Leur fréquence chez les patients atteints est toutefois significativement supérieure à celle qu'on a constatée chez les populations contrôles, raison pour laquelle on parle d'enrichissement, de facteurs de susceptibilité, de facteurs de risque, de facteurs associés à la DI ou de microremaniements à pénétrance incomplète (tableau 5). Il existe probablement d'autres facteurs, génétiques ou environnementaux, nécessaires pour exprimer un phénotype pathologique spécifique chez les patients porteurs du réarrangement.

Certains gènes, tels que *NRXN1, CNTNAP2*, et d'autres, sont des "*hot spots*" de remaniements chromosomiques présentant les mêmes cractéristiques que les réarrangements évoqués précédemment (généralement transmis par un parent sain, présents à faible fréquence dans des populations contrôles et associés à des phénotypes variables). Ces remaniements ne sont pas causés par un NAHR et ils présentent donc des points de cassure non récurrents (Gregor et al. 2011, Chen et al. 2013). Il est possible que ces réarrangements constituent des facteurs de susceptibilité, contribuant à la DI en association avec d'autres causes génétiques ou environnementales, bien que ceci n'ait pas encore été démontré.

Le nombre potentiel de CNV agissant en tant que facteurs de susceptibilité est probablement beaucoup plus important que ce que nous connaissons aujourd'hui (Cooper et al. 2011, Kaminsky et al. 2011). Mais en raison, soit de leur rareté (en particulier s'ils ne résultent pas de NAHR), soit de leur risque relatif peu élevé, beaucoup n'ont pas été encore été clairement caractérisés.

Il reste également à déterminer la valeur prédictive de ces CNV sur la DI. Des facteurs supplémentaires (génétiques ou environnementaux), non encore identifiés, sont probablement nécessaires pour causer la DI. On n'est donc actuellement pas capable d'établir un pronostic sur le développement intellectuel, lorsqu'un de ces CNV est retrouvé en bas âge ou en prénatal.

| locus | CNV | taille | OMIM | Orphanet | principaux signes associés (en plus de la DI) |
|----------|-------|--------|--------|----------|---|
| | | (Mb) | | | |
| 1q21.1 | perte | 1,35 | 612474 | 250989 | microcéphalie, cataracte, MC cardiaque, TSA, déficit de |
| | | | | | l'attention, schizophrénie |
| 1q21.1 | gain | 1,35 | 612475 | 250994 | macrocéphalie, TSA, schizophrénie |
| 15q13.3 | perte | 2 | 612001 | 199318 | épilepsie, TSA, déficit de l'attention |
| 16p11.2 | perte | 0,58 | 611913 | 261197 | TSA, épilepsie, obésité |
| 16p11.2 | gain | 0,58 | 614671 | 261204 | TSA, déficit de l'attention, absence de prise de poids |
| 16p13.11 | perte | 1,6 | - | 261236 | microcéphalie, épilepsie |
| 16p13.11 | gain | 1,6 | | 261243 | TSA, déficit de l'attention |

Tableau 5 : principaux locus de CNV facteurs de susceptibilité à la DI

D'après Keren(a) et al. 2013

1.2.3. Causes géniques de déficience intellectuelle

En raison de leur incidence plus élevée dans la DI et de leur détection par une méthode de diagnostic ancienne (le caryotype), les causes chromosomiques ont été les premières étiologies génétiques identifiées de DI, dans les années 1950 et 1960. Pourtant, dès les années 1940, certaines DI familiales étaient décrites avec une transmission mendélienne, en particulier récessive liée au chromosome X (Martin et al. 1943). Ce qui conduisit, dans les années 1970, à attribuer l'excès de DI observé chez les garçons, à des anomalies monogéniques sur le chromosome X (Lehrke et al. 1972). Finalement, l'observation au caryotype d'un site fragile récurrent en Xq27.3 chez des enfants atteints de DI permit l'identification de l'expansion nucléotidique responsable de l'X fragile (Oberlé et al. 1991), établissant ainsi que la DI pouvait être provoquée par des anomalies monogéniques.

1.2.3.1. Evaluation du nombre de gènes impliqués dans la DI

Aujourd'hui, une anomalie génique est retrouvée dans 1 cas de DI sur 10 à peu près (Rauch et al. 2006). Or, des recherches récentes ont identifié de très nombreux gènes responsables de DI (Chelly et al. 2006), et dont le nombre augmente de plus en plus rapidement avec le développement des techniques de séquençage massivement parallèles (ou NGS pour *New Generation Sequencing*) (Tarpey et al. 2009, Kaufman et al. 2010, van Bokhoven et al. 2011,

De Ligt et al. 2012, Bessa et al. 2012, Lubs et al. 2012). Un nombre croissant de ces gènes ne concerne souvent que très peu de familles, voire une seule, si bien qu'il devient difficile d'en faire une revue exhaustive (Vissers et al. 2010, Rauch et al. 2012).

Beaucoup de ces gènes connus (environ 70) sont situés sur le chromosome X. Cela peut s'expliquer par le mode de transmission récessif lié à l'X qui favorise les grands pédigrées familiaux avec membres atteints sur plusieurs générations, ce qui offre de plus grandes possibilités pour réaliser des études de liaison. La DI étant associée à une diminution des capacités de procréation, les patients dont la DI est AD sont le plus souvent des formes sporadiques. Quant aux formes AR, étant donné le mode de transmission, ceux qui en sont atteints se concentrent le plus souvent sur une seule génération, en dehors des familles avec plusieurs boucles de consanguinité.

En faisant une projection grossière à partir des 70 gènes de DI sur l'X, on peut évaluer le nombre total de gènes impliqués dans la DI à environ 1 400, nombre qui est proche de celui des pathologies répertoriées dans OMIM avec DI-S ou DI-NS, à savoir 1 557 (Raymond et al. 2010). Le nombre de gènes dont des mutations ont été impliquées au moins une fois dans la DI était évalué à 450 dans une review datant de début 2013 (Ellison et al. 2013). Ce chiffre a certainement grandement augmenté depuis.

1.2.3.2. Classifications des gènes de DI

Les gènes de DI peuvent être classés de différentes façons. Premièrement sur une base clinique, les gènes responsables de DI-NS d'une part, et les gènes responsables des différentes DI-S d'autre part, peuvent être distingués. Cette classification est très utile en pratique clinique, afin d'orienter les examens diagnostiques vers un ou plusieurs gènes spécifiques. Cependant, un nombre croissant de gènes expliquent à la fois des formes de DI-NS et de DI-S, démontrant une importante variabilité phénotypique. C'est le cas, par exemple, des gènes *RPS6KA3* (syndrome de Coffin-Lowry), *ATRX* (α -thalassémie et retard mental lié à l'X), *ARX* (lissencéphalie et syndrome de Partington), *ARID1B* (syndrome de Coffin-Siris) : ces gènes, tout comme d'autres, sont aussi responsables de DI-NS (Guerini et al. 2000, Field et al. 2006, Poirier et al. 2006, Hoyer 2012). De plus, certains gènes, qui étaient initialement classés

comme étant la cause de DI-NS, ont été reclassés en causes de DI-S après une réévaluation clinique des patients mutés. Il en est ainsi de la DI liée aux mutations d'*OPHN1* : une dysgénésie vermienne ainsi qu'une atrophie cortico-sous corticale ont été retrouvées à l'IRM, à postériori, chez les patients mutés (des Portes et al. 2004). Il devient donc difficile de classer les différents gènes de DI en fonction seulement des caractéristiques cliniques des patients.

On peut également classer les gènes de DI en fonction du mode de transmission (AD, AR et LX) de leurs mutations. Cette classification est utile en pratique clinique dans les formes familiales, car elle permet d'orienter les examens diagnostiques vers certains gènes dont les mutations ont un mode de transmission correspondant. Elle est également capitale pour l'identification de nouveaux gènes de DI : en effet, la plupart des méthodes nécessitent, à priori, de poser l'hypothèse du mode de transmission des mutations responsables (analyse de liaison, *homozygosity mapping*, séquençage de trio pour recherche de mutations *de novo...*). Néanmoins, avec les avancées récentes des techniques moléculaires, certains gènes ont été identifiés comme impliqués chez des patients DI, avec une transmission complexe. On parle alors, comme pour certains CNV récurrents (cf. 1.2.2.2.4.), de facteurs de susceptibilité ou d'étiologie multifactorielle / polygénique, l'hypothèse étant que la DI résulte de l'accumulation de ces facteurs.

Enfin, il est également possible de classer les gènes selon le mécanisme physiopathologique dans lequels ils sont impliqués. Etant donné les très nombreux gènes identifiés dans la DI, beaucoup de processus cellulaires et tissulaires différents ont été mis en évidence. Il est toutefois possible de dégager quelques principaux mécanismes biologiques dans lesquels les gènes de DI semblent préférentiellement jouer un rôle. Cette dernière classification est de plus en plus utilisée dans les revues listant les gènes de DI car elle permet d'ordonner les gènes d'une manière fonctionnellement logique (Chelly et al. 2006, Kaufman et al. 2010, van Bockoven et al. 2011, Bessa et al. 2012).

1.2.3.2.1. Gènes impliqués dans les voies métaboliques

Bien que les DI liées à des gènes des voies métaboliques, qui représentent 1 à 5% des DI, soient dans leur très grande majorité d'étiologie monogénique et de transmission AR, elles

sont le plus souvent classées dans un groupe à part, dit des maladies métaboliques (van Karnebeek et al. 2005, Rauch et al. 2006, tableau 6), en raison de leur clinique et de leur physiopathologie particulières, ainsi que de leur méthode diagnostique reposant principalement sur des tests biochimiques. Dans le cas des maladies métaboliques, la DI est causée soit par une accumulation d'un substrat toxique, soit par un déficit d'un substrat nécessaire au développement cérébral, soit par un déficit énergétique, lors d'une étape critique du développement cérébral.

La maladie métabolique classiquement la plus fréquente est la phénylcétonurie, liée à des mutations bialléliques dans le gène *PAH*. Mais dans les pays où un dépistage est mis en place à la naissance, elle est traitée et ne provoque plus de DI.

| gène | locus | Voie métabolique | Principaux signes associés |
|---------|----------|-----------------------------------|--|
| ALDH5A1 | 6p22.2 | métabolisme des acides organiques | TSA, épilepsie, ataxie |
| L2HGDH | 14q22.1 | métabolisme des acides organiques | épilepsie, ataxie, lissencéphalie |
| NAGLU | 17q21.31 | métabolisme des polysaccharides | régression, agressivité, traits épais |
| SGSH | 17q25.3 | métabolisme des polysaccharides | régression, agressivité, traits épais |
| ADSL | 22q13.1 | métabolisme des purines | TSA, épilepsie, ataxie |
| PMM1 | 22q13.2 | glycosylation des protéines | ataxie, atrophie cérébelleuse |
| SLC16A2 | Xq13.2 | transporteur des monocarboxylates | épilepsie, ataxie |
| ATP7A | Xq21.1 | métabolisme du cuivre | anomalies musculo-squelettiques, épilepsie |
| SLC6A8 | Xq28 | transporteur de la créatine | TSA, épilepsie |

D'après Garcia-Cazorla et al. 2009

1.2.3.2.2. Gènes impliqués dans la neurogenèse

Durant le développement, la neurogenèse et la migration neuronale surviennent selon un schéma spatio-temporel strictement contrôlé. Ces mécanismes dépendent à la fois de mécanismes intracellulaires et d'interactions intercellulaires. Les principales protéines jouant un rôle dans la neurogenèse contrôlent la division cellulaire. Toute disruption de ce mécanisme peut entraîner une DI et une microcéphalie (tableau 7).

| gène | locus | fonction | signes associés ou syndrome |
|----------|----------|--|-----------------------------|
| ASPM | 1q31.3 | régulation du fuseau mitotique des neuroblastes | microcéphalie |
| MCPH1 | 8p23.1 | microcéphaline, réponses aux dommages de l'adn, blocage de la mitose des neuroblastes | microcéphalie |
| CDK5RAP2 | 9q33.2 | régulation du fuseau mitotique des neuroblastes | microcéphalie |
| CENPJ | 13q12.12 | régulation du fuseau mitotique des neuroblastes | microcéphalie |

Tableau 7 : exemples de gènes de DI impliqués dans la neurogenèse

D'après Bessa et al. 2012

1.2.3.2.3. Gènes impliqués dans la migration neuronale

Durant le développement, les cellules nerveuses migrent des zones ventriculaires et sousventriculaires jusqu'à leur localisation définitive. Comme la neurogenèse, la migration neuronale doit être strictement contrôlée. Un grand nombre de protéines jouent un rôle dans la migration neuronale : des protéines du cytosquelette, des molécules de la matrice extracellulaire, des protéines de glycosylation qui sont impliquées dans l'arrêt de la migration neuronale. Une disruption de ce mécanisme entraîne généralement une DI associée à des anomalies de gyration cérébrale (tableau 8).

| gène | locus | fonction | signes associés ou syndrome |
|---------|--------|-------------------------------------|--|
| POMGNT1 | 1p34.1 | glycosylation des protéines | lissencéphalie pavimenteuse, anomalies |
| | | | congénitales oculaires et musculaires |
| TUBB2B | 6p25.2 | β-tubuline, sous-unité des | polymicrogyrie |
| | | microtubules | |
| RELN | 7q22.1 | reeline, protéine de la matrice | lissencéphalie, hypoplasie du cervelet |
| | | extracellulaire | |
| VLDLR | 9p24.2 | récepteur des vldl, dépendant de la | lissencéphalie, hypoplasie du cervelet |
| | | reeline | |
| FKTN | 9q31.2 | glycosylation des protéines | lissencéphalie pavimenteuse, anomalies |
| | | | congénitales oculaires et musculaires |

| POMT1 | 9q34.13 | glycosylation des protéines | lissencéphalie pavimenteuse, anomalies |
|----------|----------|-------------------------------------|--|
| | | | congénitales oculaires et musculaires |
| TUBA1A | 12q13.12 | α-tubuline, sous-unité des | lissencéphalie, hétérotopies sous-corticales |
| | | microtubules | |
| POMT2 | 14q24.3 | glycosylation des protéines | lissencéphalie pavimenteuse, anomalies |
| | | | congénitales oculaires et musculaires |
| PAFAH1B1 | 17p13.3 | protéine associée aux microtubules | lissencéphalie, hétérotopies sous-corticales |
| FKRP | 19q13.32 | glycosylation des protéines | lissencéphalie pavimenteuse, anomalies |
| | | | congénitales oculaires et musculaires |
| LARGE | 22q12.3 | glycosylation des protéines | lissencéphalie pavimenteuse, anomalies |
| | | | congénitales oculaires et musculaires |
| ARX | Xp21.3 | facteur de transcription, rôle dans | lissencéphalie, encéphalopathie épileptique, |
| | | la migration neuronale | syndrome de Partington |
| DCX | Xq23 | régulation des microtubules | lissencéphalie, hétérotopies sous-corticales |

D'après Verrotti et al. 2010

1.2.3.2.4. Gènes impliqués dans la fonction synaptique

La plupart des synapses du système nerveux sont chimiques et répondent à un stimulus en libérant un neurotransmetteur. De très nombreuses vésicules contenant les neurotransmetteurs sont présentes dans la terminaison présynaptique. Après stimulation, ces vésicules sont transloquées au niveau de la zone active et arrimées à la membrane présynaptique. Cette circulation des vésicules est dépendante des protéines Rab qui sont des GTPases, actives quand elles lient un GTP et inactives quand elles lient un GDP. Puis, en réponse à un influx intracellulaire de calcium, ces vésicules subissent une exocytose qui entraîne la libération du neurotransmetteur dans la fente synaptique.

Un grand nombre de gènes mutés dans ces différents mécanismes sont responsables de DI (tableau 9).

| gène | locus | fonction | signes associés ou syndrome |
|----------|--------|--|----------------------------------|
| NRXN1 | 2p16.3 | neurexine, adhésion entre membranes pré et post- | TSA, syndrome de tourette |
| | | synaptiques, interagit avec les neuroligines | |
| RAB3GAP1 | 2q21.3 | voie rab-gtpase, impliquée dans le trafic des | syndrome warburg-micro |
| | | vésicules | |
| PRSS12 | 4q26 | neurotrypsine | - |
| STXBP1 | 9q34.1 | arrimage des vésicules à la membrane pré- | encéphalopathie épileptique, |
| | 1 | synaptique, interagit avec la syntaxine. | syndrome d'ohtahara |
| CDH15 | 16q24. | adhésion entre membranes pré et post- | - |
| | 3 | synaptiques | |
| IL1RAPL1 | Xp21 | récepteur inhibiteur de l'exocytose | - |
| CASK | Xp11.4 | kinase calmoduline dépendante impliquée dans | microcéphalie et atrophie ponto- |
| | | l'exocytose (des neurotransmetteurs ou des | cérébelleuse et FG syndrome |
| | | récepteurs post-synaptiques). | |
| SYN1 | Xp11.2 | synapsine, se lie aux vésicules pré-synaptiques | épilepsie |
| | 3 | | |
| GDI1 | Xq28 | voie rab-GTPase, impliquée dans le trafic des | - |
| | | vésicules | |
| RAB39B | Xq28 | voie rab-GTPase, impliquée dans le trafic des | TSA, épilepsie |
| | | vésicules | |

Tableau 9 : exemples de gènes de DI de fonction présynaptique

La densité post-synaptique est la région intracellulaire (dense en microscopie électronique) immédiatement sous-membranaire, contenant la partie intracytoplasmique des récepteurs des neurotransmetteurs, des canaux ioniques, ainsi que le cytosquelette sous-membranaire. Cette structure est très importante pour la transduction post-synaptique du signal. La DI (ainsi que les TSA) peuvent résulter de la disruption de la transduction du signal, tant pour les neurones excitateurs (glutaminergiques) que pour les neurones inhibiteurs (GABAergiques). La majorité des gènes de DI identifiés, qui sont impliqués dans ce mécanisme physiologique,

sont liés à la liaison post-synaptique du glutamate dont il existe 2 types de récepteur : NMDA et AMPA (tableau 10).

Tableau 10 : exemples de gènes de DI impliqués dans l'organisation de la densité postsynaptique

| gène | locus | fonction | signes associés ou syndrome |
|---------|---------|---|-----------------------------|
| CNTNAP2 | 7q36.1 | adhésion entre membranes pré et post- | TSA, syndrome de tourette, |
| | | synaptiques, interagit avec les contactines | dysplasie corticale. |
| SHANK2 | 11q13.3 | association des récepteurs des neurotransmetteurs | TSA |
| | | à d'autres protéines de membrane | |
| GRIN2B | 12p13.2 | sous-unité du récepteur NMDA | epilepsie |
| GRIN2A | 16p13.2 | sous-unité du récepteur NMDA | epilepsie |
| SHANK3 | 22q13.3 | association des récepteurs des neurotransmetteurs | TSA, syndrome de phelan-mc |
| | | à d'autres protéines de membrane | dermid |
| NLGN4 | Xp22.3 | neuroligine, adhésion entre membranes pré et | TSA |
| | | post-synaptiques, interagit avec la Neurexine. | |
| NLGN3 | Xp13.1 | neuroligine, adhésion entre membranes pré et | TSA |
| | | post-synaptiques, interagit avec la Neurexine. | |
| DLG3 | Xp13.1 | guanylate kinase, favorise le regroupement des | - |
| | | récepteurs NMDA | |
| GRIA3 | Xq25 | sous-unité du récepteur AMPA | - |

D'après van Bockoven et al. 2011

La régulation de la densité protéique post-synaptique est très importante pour la plasticité synaptique. Cette régulation peut s'effectuer par 2 mécanismes : d'une part l'action des facteurs de transcription qui augmentent l'expression de certaines protéines, et, d'autre part, la protéolyse entraînant la dégradation de ces protéines, en particulier la protéolyse du système ubiquitine-protéasome, ou protéolyse UPS. Une dysfonction des gènes impliqués dans cette régulation peut avoir pour résultat une DI (tableau 11).

| gène | locus | fonction | signes associés ou syndrome |
|-------|---------|---|-------------------------------|
| UBE3A | 15q11.2 | ubiquitine ligase de la protéolyse UPS | syndrome d'Angelman |
| UBR1 | 15q15.2 | ubiquitine ligase de la protéolyse UPS | syndrome de Johanson-Blizzard |
| HUWE1 | Xp11.22 | ubiquitine ligase de la protéolyse UPS | DI de type Turner |
| UBE2A | Xq24 | ubiquitine ligase de la protéolyse UPS | TSA |
| CUL4B | Xq24 | ubiquitine de la protéolyse UPS | petite taille, hypogonadisme |
| FMR1 | Xq27.3 | trafic des ARNm entre le noyau et le cytoplasme | syndrome de l'X fragile |

Tableau 11 : exemples de gènes de DI impliqués dans la régulation des protéines postsynaptiques

D'après van Bockoven et al. 2011

L'épine dendritique peut changer de taille et de forme en quelques secondes. Cette importante plasticité, qui résulte de la polymérisation/dépolymérisation rapide des filaments d'actine et de microtubules, dépend de la voie de signalisation Rho-GTPase. Une dysfonction des protéines régulant ces mécanismes peut entraîner une DI (tableau 12).

Tableau 12 : exemples de gènes de DI impliqués dans la dynamique du cytosquelette dendritique.

| gène | locus | fonction | signes associés ou syndrome |
|---------|---------|---|--|
| MEGAP | 3p25.3 | activation de la voie Rho-GTPase | dans la région critique du syndrome |
| | | | 3р- |
| LIMK1 | 7q11.23 | kinase, inhibition de la dépolymérisation de | dans la région critique de la délétion |
| | | l'actine, activée par PAK3 et voie Rho-GTPase | du syndrome de Williams-Beuren |
| AP1S2 | Xp22.2 | adaptine, régulation de la polymérisation de | hydrocéphalie, calcifications des |
| | | l'actine et de l'endocytose des vésicules | noyaux de la base |
| IQSEC2 | Xp11.22 | activation de la voie ARF-GTPase, régulation | - |
| | | l'endocytose des vésicules | |
| FGD1 | Xp11.22 | activation de la voie Rho-GTPase | syndrome d'Aarskog-Scott |
| ARHGEF9 | Xq11.1 | activation de la voie Rho-GTPase | encéphalopathie épileptique |

| OPHN1 | Xq12 | activation de la voie Rho-GTPase | hypoplasie cérébelleuse |
|---------|--------|--|-------------------------|
| РАКЗ | Xq22.3 | kinase, liaison entre les Rho-GTPase et le | - |
| | | cytosquelette, activation de LIMK1 | |
| OCRL1 | Xq25 | activation de la voie Rho-GTPase | syndrome de Lowe |
| ARHGEF6 | Xq26.3 | activation de la voie Rho-GTPase | - |

1.2.3.2.5. Gènes impliqués dans les cascades de signalisation intracellulaire

L'activation des récepteurs des neurotransmetteurs entraîne des cascades de signalisation postsynaptique, en particulier de la voie Ras-MAPK-ERK et PI3K-AKT-mTOR. L'inactivation des gènes impliqués dans ces voies de signalisation représente un certain nombre d'étiologies connues de DI génique (tableau 13).

| Tableau 13 : exer | nples de gènes de | e DI impliqués dans le | es cascades de signalisation o | cellulaire. |
|-------------------|-------------------|------------------------|--------------------------------|-------------|
| | 1 0 | | 0 | |

| gène | locus | fonction | signes associés ou syndrome |
|--------|----------|--|------------------------------------|
| SOS1 | 2p22.1 | inactivation des GTPase de la voie Ras- | syndrome de Noonan |
| | | MAPK-ERK | |
| RAF1 | 3p25.1 | kinase de la voie Ras-MAPK-ERK | syndromes de Noonan, LEOPARD |
| BRAF | 7q34 | kinase de la voie Ras-MAPK-ERK | syndromes de Noonan, LEOPARD, |
| | | | Cardio-Facio-Cutané |
| TSC1 | 9q34.13 | hamartine, inactivation de la voie PI3K-AKT- | sclérose tubéreuse de Bourneville |
| | | mTOR, forme un complexe avec TSC2 | |
| PTEN | 10q23.31 | phosphatase de la voie PI3K-AKT-mTOR | TSA, macrocéphalie, syndromes de |
| | | | Cowden et de Bannayan-Riley- |
| | | | Ruvalcaba |
| SHOC2 | 10q25.2 | protéine de structure impliquée dans la voie | syndrome Noonan-like avec cheveux |
| | | Ras-MAPK-ERK | anagènes caducs |
| HRAS | 11p15.5 | GTPase de la voie Ras-MAPK-ERK | syndrome de Costello |
| KRAS | 12p12.1 | GTPase de la voie Ras-MAPK-ERK | syndromes de Noonan, Cardio-Facio- |
| | | | Cutané |
| PTPN11 | 12q24.3 | phosphatase de la voie Ras-MAPK-ERK | syndromes de Noonan, LEOPARD |
| SPRED1 | 15q14 | régulation de la voie Ras-MAPK-ERK | syndrome de Legius |

| MAP2K1 | 15q22.31 | kinase de la voie Ras-MAPK-ERK | syndrome Cardio-Facio-Cutané |
|------------------|---------------------|--|---|
| TSC2 | 16p13.3 | tubérine, inactivation de la voie PI3K-AKT- | sclérose tubéreuse de Bourneville |
| | | mTOR, forme un complexe avec TSC2 | |
| NF1 | 17q11.2 | inhibition de la voie Ras-MAPK-ERK | neurofibromatose, syndrome NF1- |
| | | | |
| | | | Noonan |
| MAP2K2 | 19p13.3 | kinase de la voie Ras-MAPK-ERK | Noonan syndrome Cardio-Facio-Cutané |
| MAP2K2 DYRK1A | 19p13.3 21q22.13 | kinase de la voie Ras-MAPK-ERK kinase, régulation de la voie Ras-MAPK-ERK | Noonan syndrome Cardio-Facio-Cutané microcéphalie |

1.2.3.2.6. Gènes impliqués dans la régulation épigénétique de la transcription

Le fonctionnement cérébral repose de façon critique sur la régulation de la transcription génique. Un très grand nombre de gènes intervenant dans ce mécanisme sont impliqués dans les DI. Il s'agit souvent de facteurs de transcription, de protéines intervenant dans la méthylation de l'ADN, dans la méthylation et l'acétylation des histones, ou dans le remodelage de la chromatine (tableau 14).

| Tableau 14 : exemp | les de gènes de D | I impliqués dans | l'expression génique. |
|--------------------|-------------------|------------------|-----------------------|
| | 0 | | |

| gène | locus | fonction | signes associés ou syndrome |
|--------|----------|--|------------------------------------|
| HDAC4 | 2q37.3 | histone déacétyltransférase | brachydactylie, brachymétacarpie, |
| | | | obésité |
| NSD1 | 5q35.3 | histone méthyltransférase | syndromes de Sotos et de Weaver |
| MED23 | 6q23.2 | sous-unité du complexe Mediator, rôle dans | - |
| | | la préinitiation de la transcription | |
| ARID1B | 6q25.3 | liaison à l'ADN, sous-unité du complexe | - |
| | | SWI/SNF, remodelage de la chromatine | |
| CHD7 | 8q12.2 | hélicase, régulation de la transcription | syndrome CHARGE |
| EHMT1 | 9q34.3 | histone méthyltransférase | syndrome de Kleefstra |
| MED17 | 11q21 | sous-unité du complexe Mediator, rôle dans | microcéphalie, épilepsie, atrophie |
| | | la préinitiation de la transcription | cérébrale |
| MLL2 | 12q13.12 | histone méthyltransférase | syndrome Kabuki |
| CREBBP | 16p13.3 | histone acétyltransférase | syndrome de Rubinstein-Taybi |

| RAI1 | 17p11.2 | facteur de transcription syndrome de Smith-Mage | | |
|--------|----------|--|---|--|
| SETBP1 | 18q12.3 | réplication de l'ADN | syndrome de Schinzel-Giedion | |
| TCF4 | 18q21.2 | facteur de transcription (hélice-boucle- syndrome de Pitt-Ho | | |
| | | hélice) | | |
| DNMT3B | 20q11.21 | DNA méthyltransférase | syndrome ICF | |
| EP300 | 22q13.2 | histone acétyltransférase | syndrome de Rubinstein-Taybi | |
| CDKL5 | Xp22.13 | sérine-thréonine kinase, trafic et épissage | encéphalopathie épileptique | |
| | | des ARN | | |
| BCOR | Xp11.4 | répression du facteur de transcription BCL6 | microphtalmie | |
| ZNF41 | Xp11.3 | facteur de transcription (doigts de zinc) | - | |
| ZNF674 | Xp11.3 | facteur de transcription (doigts de zinc) | - | |
| ZNF81 | Xp11.23 | facteur de transcription (doigts de zinc) | - | |
| KDM5C | Xp11.22 | histone déméthylase | paraplégie spastique, petite taille, | |
| | | | microcéphalie | |
| PHF8 | Xp11.22 | facteur de transcription (doigts de zinc) | fente palatine | |
| MED12 | Xq13.1 | sous-unité du complexe Mediator, rôle dans | Mediator, rôle dans syndromes FG et Lujan-Fryns | |
| | | la préinitiation de la transcription | | |
| ATRX | Xq21.1 | hélicase, sous-unité du complexe SWI/SNF, | WI/SNF, α-Thalassémie, retard mental lié à l'X | |
| | | remodelage de la chromatine | | |
| BRWD3 | Xq21.1 | modification de la chromatine | - | |
| ZNF711 | Xq21.1 | facteur de transcription (doigts de zinc) | - | |
| PHF6 | Xq26.2 | facteur de transcription (doigts de zinc) | syndrome de Borjeson-Forssman- | |
| | | | Lehmann | |
| MECP2 | Xq28 | liaison à l'ADN méthylé, répression de la | syndrome de Rett | |
| | | transcription | | |

1.2.4. Déficience intellectuelle liée à l'empreinte

Les gènes soumis à empreinte ont une expression monoallélique, la copie active étant soit sur l'allèle paternel, soit sur l'allèle maternel.

Environ une centaine de gènes humains soumis à empreinte sont connus (Dulac et al. 2010), et une grande partie d'entre eux sont exprimés dans le cerveau. Ces gènes sont groupés dans des régions spécifiques du génome : les régions soumises à empreinte ou différentiellement méthylées, qui sont sous le contrôle de séquences d'ADN particulières appelées "centres de l'empreinte".

Les maladies liées à l'empreinte résultent d'une inactivation de l'allèle physiologiquement exprimé, que ce soit par délétion chromosomique, mutation ponctuelle (génique ou du centre de l'empreinte), épimutation (déméthylation ou hyperméthylation d'un îlot CpG méthylé différentiellement) ou disomie uniparentale. Ainsi les délétions d'un même locus et des disomies uniparentales d'un même chromosome peuvent entraîner 2 syndromes totalement différents selon que c'est l'allèle paternel ou maternel qui est perdu. C'est le cas des syndromes d'Angelman et de Prader-Willi, tous 2 dus à la délétion de la même région, en 15q11.2 : le premier résulte de la perte de l'allèle maternel et le second de l'allèle paternel.

La majorité des syndromes connus liés à l'empreinte sont des pathologies de la croissance, mais certains sont impliqués dans la DI (tableau 15).

| syndrome | locus | allèle | principaux mécanismes | signes associés |
|--------------|---------|----------|------------------------|--------------------------|
| | | inactivé | | |
| Prader-Willi | 15q11.2 | paternel | délétion, disomie | hypotonie, obésité |
| | | | uniparentale | |
| Angelman | 15q11.2 | maternel | délétion, disomie | épilepsie, microcéphalie |
| | | | uniparentale, mutation | |
| | | | d' <i>UBE3A</i> | |

Tableau 15 : principaux syndromes liés à l'empreinte avec DI au premier plan

D'après Hosoki et al. 2009, Yamazawa et al. 2010

1.2.5. Connexion étiologique et physiopathologique entre déficience intellectuelle et TSA

Comme nous l'avons vu à la partie 1.6., la DI et les TSA sont 2 affections fréquemment associées : 28% des individus DI ont des TSA (Bryson et al. 2008), 67% des patients TSA sont atteints de DI (Fombonne et al. 2003), et quasiment tous les patients ayant une DI sévère répondent aux critères des TSA (Bryson et al. 2008). Bien que les TSA et la DI correspondent à 2 diagnostics différents, ce chevauchement phénotypique est remarquable.

Les TSA ont une incidence élevée dans de nombreuses DI-S : 25 à 47% dans l'X fragile, 5 à 10% dans la trisomie 21, et 16 à 48% des scléroses tubéreuses de Bourneville (Kaufman et al. 2010).

Durant la dernière décennie, les importants progrès technologiques, (puces ADN, NGS) ont permis d'identifier un grand nombre d'anomalies de l'ADN, qu'il s'agisse des variations structurales ou des variations ponctuelles qui sont impliquées dans la DI, dans les TSA et dans d'autres pathologies neuropsychiatriques comme la schizophrénie. Ces avancées ont permis d'identifier des facteurs étiologiques communs à ces pathologies, qu'ils aient un fort effet pathogène avec une hérédité mendélienne, ou qu'ils soient des facteurs de susceptibilité nécessitant d'autres facteurs génétiques ou environnementaux pour s'exprimer (Mefford et al. 2012).

Les CNV impliqués à la fois dans les TSA et la DI sont principalement la délétion de *SHANK3*, la délétion 3q29, la duplication 15q11.2 et la majorité des CNV classés comme facteurs de susceptibilité, *i.e.* les délétions et duplications 1q21.1, 16p11.2, 16p13.3, la délétion 15q13.3, la duplication 15q11.2 (cf. 1.2.2.2.4., Malhotrat et al. 2012), etc.

Au niveau génique, ce sont principalement les gènes codant les protéines de structure et du fonctionnement synaptique qui sont mutés à la fois dans les TSA et dans la DI, de sorte qu'on parle de « synaptopathie » (cf. 1.2.3.2.4. Zoghbi et al. 2012). *FMR1*, qui est impliqué dans la régulation des protéines post-synaptiques, est la cause monogénique la plus fréquemment retrouvée dans la DI et dans les TSA. On retrouve également des gènes de protéines de l'organisation de la synapse qui sont mutés dans la DI et comme dans les TSA : il s'agit de certaines neuroligines et contactines, les neurexines, *UBE2A*, *SHANK2* et *SHANK3*. Une autre cause classique de TSA et de DI est la sclérose tubéreuse de Bourneville, causée par des

mutations de *TSC1* et *TSC2*, responsables d'une disruption de cascades de signalisation postsynaptique.

Il faut signaler que certaines mutations géniques, associées à la fois à la DI et aux TSA, peuvent être associées aussi à la schizophrénie. Ainsi, les avancées récentes dans l'étude des étiologies de ces troubles neuropsychiatriques semblent démontrer l'existence de voies pathologiques communes qui permettraient de les considérer comme appartenant au même spectre de pathologies (*Cross-Disorder Group of the Psychiatric Genomics Consortium*, Lee et al. 2013).

Une théorie intéressante, qui explique que les mêmes anomalies génétiques occasionnent des pathologies cliniquement très différentes, est celle de l'homéostasie neuronale (Ramocki et al. 2008). Lorsqu'une mutation ou un CNV entraîne une dysfonction cellulaire, le phénotype résultant va dépendre de la fonction du ou des groupes de neurones atteints dans le cerveau en développement. L'anomalie génétique va activer des mécanismes de compensation (homéostasie) dans les différents réseaux neuronaux atteints, ce qui va entraîner une réduction de la flexibilité synaptique. Et selon le réseau neuronal touché, il en découlera des anomalies différentes du phénotype neuropsychiatrique.

Keren Boris – Thèse de doctorat - Année 2013

2. Objectifs du travail

Nous avons effectué une ACPA en puces SNP dans une série de patients atteints de DI sans étiologie.

Les objectifs ont été :

1) de détecter, dans un cadre diagnostic, les CNV responsables de la DI des patients ;

2) d'identifier, via l'analyse des CNV, des gènes candidats potentiellement impliqués dans la DI;

3) d'étudier, chez des patients sélectionnés, la séquence de certains de ces gènes afin de valider leur implication dans la DI.

2.1. Stratégie diagnostique utilisée

Le caryotype était, jusqu'à peu, la seule technique d'étude globale du génome utilisée en routine. Depuis la première mise en évidence d'une anomalie chromosomique chez l'homme, à savoir la trisomie 21 (Lejeune et al. 1959), de nombreuses anomalies de nombre, mais également de structure des chromosomes ont été détectées chez des patients présentant une DI (Grouchy et al. 1984). En effet, on évalue actuellement à 10-15% la proportion des DI dues à une anomalie chromosomique détectable avec le caryotype en bande (Rauch et al 2006). Pour les 60% de patients pour lesquels aucune autre étiologie (infectieuse, néonatale, génique, environnementale...) n'est retrouvée, la question se pose alors d'une anomalie chromosomique infra microscopique, car la sensibilité actuelle de cette technique est de l'ordre de 5 millions de paires de bases (5Mb).

Afin d'essayer de mettre en évidence des anomalies chromosomiques plus fines, de taille inférieure à 5 Mb, les techniques de cytogénétique ont été combinées à celles de biologie moléculaire, dès la fin des années 1980 : on a ainsi mis au point les techniques dites de cytogénétique moléculaire. L'ensemble de ces techniques repose sur la capacité qu'a l'ADN de se dénaturer et de s'hybrider de façon spécifique dans des conditions de température, de salinité et de pH particulières. Se sont ainsi développées certaines de ces techniques, comme la FISH, la MLPA, la QMPSF ou la MAPH qui permettent de détecter des anomalies déséquilibrées avec une résolution bien meilleure que le caryotype. En effet, elles détectent de petits déséquilibres de l'ordre de la base nucléotidique pour les plus résolutives, mais ont pour inconvénient de n'explorer que certains locus spécifiques et en conséquence de n'être pas bien adaptées au diagnostic de la majorité des anomalies chromosomiques qui sont des anomalies privées.

L'avancée technologique la plus significative a été le développement de l'ACPA qui permet de façon reproductible une exploration pangénomique comme le caryotype, mais avec un niveau de résolution beaucoup plus important. Plusieurs méthodes existent. La plus classique est la "CGH array" mais, les puces SNP sont également utilisées pour les mêmes applications. Ce sont ces puces que nous avons utilisées afin de retrouver les CNV responsables de la DI de nos patients.

Afin de distinguer les anomalies potentiellement pathogènes des anomalies normalement présentes chez des individus sains, nous avons suivi une stratégie issue des différentes recommandations (Lee (b) et al. 2007, Mannings et al. 2010, Miller et al. 2010, South et al. 2011, Kearney et al. 2011).

Plusieurs critères peuvent être utilisés pour juger de la pathogénicité d'un CNV :

- L'analyse des parents : si l'un des 2 parents est bien portant et porteur du CNV retrouvé chez le cas index DI, cela est en défaveur du caractère pathogène. Les limites de ce critère sont les CNV touchant les gènes de pathologies récessives et les facteurs de susceptibilité. Une délétion peut être portée par un parent sain, mais être pathogène chez le cas index en raison de la présence d'une mutation ponctuelle délétère sur l'allèle restant. Ce critère est également pris en défaut pour les CNV facteurs de risque à expréssivité variable et

pour les CNV incluant des gènes liés à l'empreinte : ils sont retrouvés aussi bien chez des individus DI que chez des personnes en bonne santé.

- La taille du CNV et son contenu en gènes : plus le CNV est grand et contient de nombreux gènes et plus son caractère pathogène est probable. Mais ce critère n'est pas absolu, car une délétion ponctuelle dans une séquence codante d'un gène responsable de DI peut être pathogène. A l'inverse, des CNV de plusieurs Mb peuvent être bénins (Filges et al. 2009).

- La nature du CNV : les gains de copies sont classiquement mieux supportés et donc en défaveur du caractère pathogène. Il ne s'agit toutefois pas d'un critère absolu, certains gains de copies étant pathogènes (cf. 1.2.2.2.3.).

- Les précédentes études rapportant le CNV et les bases de données ISCA, DECIPHER, DGV, Genereviews : si le CNV a déjà été identifié chez des patients DI et / ou s'il est absent de populations contrôles de même origine géographique, cela est en faveur du caractère pathogène. A l'inverse, si le CNV a déjà été retrouvé chez des individus sains, cela est en défaveur du caractère pathogène. Il existe toutefois des limites à ce principe : des CNV potentiellement pathogènes peuvent être retrouvés dans la population générale quand ils contiennent des gènes de maladies AR (une mutation sur l'allèle restant est alors nécessaire pour développer la maladie), ou quand ils ont une expressivité variable (facteurs de susceptibilité). En raison de leur moindre fiabilité, nous n'avons pas pris en compte les CNV de DGV détectés uniquement sur des puces anciennes (de type BAC).

- La présence d'un ou de plusieurs gènes connus étant pathogéniques dans le CNV (base de données OMIM). Dans ce cas, il faut vérifier que le type de mutation répertoriée correspond au type de CNV. Par exemple, un gain de copies d'un gène, dont les mutations pathogènes sont inactivatrices, peut être bénin. De même, une perte de copie monoallélique d'un gène n'a classiquement pas de conséquence phénotypique, lorsque les mutations de ce gène sont responsables d'une maladie AR.

Afin de faciliter l'interprétation des CNV, nous avons développé un outil bioinformatique d'aide à l'analyse (figure 5). Cet outil, appelé CNVAP, permet de répertorier tous les CNV que nous retrouvons, d'y associer la clinique des patients, et les éléments provenant des bases de données publiques en termes de CNV (ISCA, DGV) et de gènes (OMIM, Refseq).

Figure 5 : capture d'écran de la fiche d'un individu sur CNVAP : (a) numéro d'échantillon, (b) pathologie du patient (nomenclature Orphanet), (c) signes cliniques du patient (nomenclature LDDB), (d) différents CNV retrouvés (nomenclature ISCN), (e) nombre de gènes dans le CNV (base de données Refseq), (f) nombre de gènes de pathologies connues (base de données OMIM), (g) CNV décrit dans la population générale (base de données DGV), (h) CNV répertorié par l'ISCA (avec le type de CNV : bénin, probablement bénin, probablement pathogène, pathogène), (i) CNV répertorié dans CNVAP, (j) interprétation, (k) exportation de CNV vers DECIPHER.



En cliquant sur les liens UCSC ou DGV, on peut obtenir des informations supplémentaires sur les gènes inclus (nom, fonction, pathologie associée, expression tissulaire) ou être redirigé vers les *genome browsers* UCSC, *NCBI map viewer* ou DGV.

A partir de cette aide, nous avons établi la stratégie diagnostique comme suit :

- lorsqu'un CNV ne contenait aucune séquence codante ou est répertorié chez des individus contrôle (DGV), nous l'avons considéré comme bénin et nous ne sommes pas allés plus loin dans son étude ;

- les CNV non présents dans DGV et contenant des séquences codantes ont été considérés comme des CNV d'intérêt. Dans ce cas, une étude des parents, a été réalisée lorsqu'elle était possible afin de déterminer si le CNV est transmis ou *de novo*.

Nous avons ensuite interprété en fonction des données de la littérature, de la taille, et des gènes inclus. Cette stratégie diagnostique est schématisée sur la figure 6.

Lorsque le CNV était transmis par un parent sain, nous l'avons habituellement considéré comme probablement bénin. Il est toutefois possible qu'un gène de pathologie autosomique récessive se trouve dans une délétion transmise (Boone et al. 2013). Dans ce cas, la délétion de ce gène peut avoir une expression clinique chez le patient si l'allèle controlatéral est muté, alors qu'elle n'en a pas chez le parent sain dont l'allèle n'est pas muté.

Figure 6 : schéma simplifié de la conduite diagnostique devant la découverte d'un CNV en ACPA. VOUS : Variant of Unknown Significance (variant de signification inconnue). Ce schéma part de la situation la plus fréquente, le CNV est sur un autosome et les parents sont sains. Néanmoins, certains cas particuliers n'obéissent pas à cet algorithme.



Tous les CNV d'intérêt ont été contrôlés chez les patients et recherchés chez les parents quand l'ADN de ceux-ci était disponible :

- pour les délétions > 500 kb et les duplications > 2 Mb, nous avons utilisé la FISH afin de les rechercher chez les parents et les contrôler chez les patients ;

pour les délétions < 500 kb et les duplications < 2 Mb (plus difficilement visibles en FISH) nous avons commencé par rechercher le CNV en ACPA chez les parents. Si le CNV n'a pas été retrouvé, nous avons vérifié la paternité puis refait une ACPA chez le patient.

2.2. Identification et validation des gènes candidats

En étudiant notre série passée en ACPA, nous avons eu comme objectif d'identifier de nouveaux gènes candidats de DI, et pour ce faire, selon le type de CNV, nous avons utilisé

des stratégies différentes. En outre, étant donné la détection des LOH par les puces SNP que nous avons utilisées (cf. 3.2.4.2.3.), nous avons également exploité ce type de données.

2.2.1. Délétions de novo

Un certain nombre de gènes de DI monogénique ont été identifiés en séquençant dans des cohortes de patients à phénotype identique, le ou les gènes retrouvés chez un patient présentant une délétion *de novo*.

C'est le cas, par exemple du gène *CHD7*, responsable du syndrome CHARGE : deux délétions *de novo* chevauchantes incluant *CHD7* ont été retrouvées, en CGH array, chez 2 patients. Et en effet, un séquençage du gène dans des séries de patients atteints a retrouvé des mutations délétères et permis de valider la responsabilité de ce gène dans le syndrome (Vissers et al. 2004).

C'est aussi de cette façon que le gène *MBD5* a été identifié comme responsable de DI, par la détection de délétions *de novo* en ACPA chez des sujets atteints (Wagenstaller et al. 2007, Talkowski et al. 2011). Notre équipe a également identifié des mutations dans le gène *PCDH19*, chez des patientes atteintes d'une encéphalopathie épileptique proche du syndrome de Dravet, conduisant à une DI de sévérité variable limitée aux filles, après avoir retrouvé une délétion *de novo* chez un patient qui s'est avéré être mosaïque (Depienne et al. 2009).

Durant notre travail, nous avons détecté plusieurs petites délétions *de novo* ne contenant qu'un gène (cf. 4.4.1.).

Nous avons utilisé les critères suivants afin de sélectionner les gènes d'intérêt :

- le gène est exprimé dans le cerveau ;

- le gène a une fonction parmi celles impliquées dans la DI (cf.1.2.3.);

- le gène est d'une taille limitée (permettant ainsi son analyse de façon relativement aisée et peu couteuse);

- le phénotype permet d'obtenir facilement une cohorte de réplication via le Centre de référence des déficiences intellectuelles de causes rares, ou via des collaborations extérieures.

Lorsque ces critères étaient réunis, nous avons séquencé (en méthode Sanger ou NGS) le gène d'intérêt dans une série de patients possédant le même phénotype que le cas index porteur de la délétion *de novo*.

2.2.2. Délétions transmises par un parent

Dans cette hypothèse, nous avons sélectionné plusieurs délétions (cf. 4.4.2.), ne contenant qu'un gène nous paraissant intéressant. Les critères que nous avons utilisés pour juger de l'intérêt du gène délété étaient les mêmes que ceux utilisés précédemment pour les délétions *de novo*.

Lorsque la plupart de ces critères étaient rassemblés nous avons séquencé (en méthode Sanger) le gène d'intérêt sur l'allèle non délété afin de détecter une mutation délétère, qui entraînerait l'expression d'une pathologie AR.

2.2.3. Délétion homozygote

Les délétions homozygotes emportant des séquences codantes sont rares, en dehors des polymorphismes communs. Elles peuvent, quand elles sont présentes, être responsables d'une maladie autosomique récessive (Boone et al. 2013). Ainsi, nous avons suspecté qu'une délétion homozygote des gènes *ATAD3B* et *ATAD3C* était pathogène, et réalisé un séquençage d'une série de patients phénotypiquement sélectionnés, à la recherche de mutations homozygotes ou hétérozygotes composites.

2.2.4. Homozygosity mapping

Comme nous l'expliquerons plus loin (cf. 3.2.4.2.3.), l'un des intérêts des puces SNP est la détection des pertes d'hétérozygotie à nombre de copies normal (*Copy Neutral Loss of Heterozygosity* ou CNLOH). Pour les cas familiaux de DI autosomique récessive avec consanguinité, il nous a ainsi été possible d'utiliser ces CNLOH afin de réaliser de "l'*homozygosity mapping*". Le principe est de sélectionner, dans une famille consanguine, les CNLOH portés par les individus atteints mais absents chez ceux qui ne le sont pas, en émettant l'hypothèse que la mutation homozygote, responsable de la pathologie, est dans une de ces régions (Lander et al. 1987, Hildebrandt et al. 2009, Kaufman et al. 2010). La taille globale des CNLOH d'intérêt est souvent d'autant plus petite qu'il y a d'individus à analyser dans la famille et que la consanguinité est éloignée.

Nous avons ensuite sélectionné les familles pour lesquelles une étude génique a été réalisée selon les critères suivants (cf. 4.4.4.) :

- l'ensemble des CNLOH sélectionnés dans la famille comporte moins de 100 gènes ;

- le phénotype permet d'obtenir facilement une cohorte de réplication via le Centre de référence des déficiences intellectuelles de causes rares, ou via des collaborations extérieures.

Une fois les CNLOH sélectionnés, si un gène inclus est fortement suspect d'être impliqué dans la DI (d'après la littérature), nous l'avons séquencé selon la méthode Sanger. Dans le cas contraire, nous avons fait réaliser un séquençage de toutes les séquences codantes (l'exome) en NGS, en nous focalisant sur les mutations retrouvées dans les régions sélectionnées par l'*homozygosity mapping*.

Lorsqu'une mutation potentiellement pathogène a été retrouvée en NGS, nous avons séquencé ce gène en méthode Sanger chez les patients afin de valider le résultat, puis dans une série de patients sélectionnés avec un phénotype comparable.
2.2.5. Points de cassure de remaniements équilibrés

La caractérisation des points de cassure des remaniements équilibrés est une méthode classique d'identification de gènes pathogènes. Cette méthode a par exemple permis d'identifier les gènes *NSD1* et *NIPBL* comme responsables, respectivement, des syndromes de Sotos (Korotaki et al. 2002) et de Cornelia de Lange (Krantz et al. 2004), ou plus récemment, les gènes *CDH15* et *KIREL3* (Bhalla et al. 2008) comme gènes de DI.

L'ACPA ne nous permet pas de caractériser les points de cassure des remaniements équilibrés car elle ne détecte que les CNV. Nous avons toutefois pu caractériser un remaniement déséquilibré cryptique porté par une patiente DI, dont le père et la soeur non atteints étaient porteurs du remaniement sous sa forme équilibrée avec un point de cassure coupant le gène *ATXN10* (cf. 4.5).

3. Matériels et méthodes

Dans cette partie, nous décrirons la série de patients DI que nous avons étudiée, ainsi que les techniques que nous avons utilisées, à savoir l'ACPA et le séquençage de l'ADN.

3.1. Série de patients

Nous avons étudié l'ADN de 1 187 patients, atteints de DI sans diagnostic étiologique, par puces ADN. Les patients ont tous été vus en consultation de génétique, dans le Centre national de référence des DI de causes rares (coordonné par le Dr Delphine Héron à l'hôpital de la Pitié-Salpêtrière). Lors de cette consultation, comme il est de coutume, les patients ont fait l'objet d'un interrogatoire détaillé sur leur histoire personnelle et familiale, ainsi que d'un examen clinique complet. Puis, en fonction de la clinique, a été réalisé un bilan étiologique paraclinique qui pouvait comprendre un caryotype, une recherche d'X fragile, une IRM cérébrale, un bilan malformatif et électrophysiologique, et un bilan métabolique (Verloes et al. 2012). Si ce premier bilan était négatif, une analyse sur puce ADN a été effectuée, après avoir reçu le consentement éclairé des patients.

Les patients de la série avaient 17 ans d'âge moyen, avec des extrêmes allant de 6 mois (retard de développement) à 75 ans (tableau 16) ; 61% étaient de sexe masculin et à 39% de sexe féminin ; 68% avaient une DI-S, et 32% une DI-NS. Et enfin, sur les 1 187 patients, 256 (21%) présentaient des TSA, et 144 (12%) étaient nés de parents apparentés.

| | DI- | NS | DI-S | | total | |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| sexe | nombre | âge moyen | nombre | âge moyen | nombre | âge moyen |
| masculin | 277 | 13 ans | 453 | 17 ans | 730 (61%) | 15 ans |
| féminin | 101 | 18 ans | 356 | 20 ans | 457 (39%) | 19 ans |
| total | 378 (32%) | 14 ans | 809 (68%) | 18 ans | 1 187 | 17 ans |

Tableau 16 : résumés des caractéristiques des patients DI passés en ACPA pour ce travail.

3.2. Les puces ADN

3.2.1. Modèles de lames utilisées

Pour notre étude, nous avons utilisé 3 différents modèles de puces (tableau 17, figure 7) :

- les Illumina HumanCNV370-duo, de décembre 2007 à février 2008 (61 patients) ;
- les Illumina HumanCNV370-quad de mars 2008 à mars 2009 (92 patients) ;
- les Illumina HumanCytoSNP-12 d'avril 2009 à mai 2013 (1034 patients).

Tableau 17 : différents types de puces Illumina utilisées dans cette étude

| | HumanCNV370-duo | HumanCNV370-quad | HumanCytoSNP-12 |
|-------------------|-------------------|-------------------------|-------------------------|
| marqueurs | 370 404 | 370 404 | 299 140 |
| marqueurs SNP | 353 203 | 353 203 | 297 623 |
| échantillons/lame | 2 | 4 | 12 |
| écart médian | 5 kb | 5 kb | 6,2 kb |
| nom du protocole | Infinium II Assay | Infinium HD Assay Super | Infinium HD Assay Ultra |
| ADN de départ | 750ng | 300ng | 300ng |

Source : Illumina

Figure 7 : de gauche à droite : puces Illumina HumanCNV-370-duo, HumanCNV370-quad et HumanCytoSNP-12



Source : Illumina

3.2.2. La technologie des puces Illumina

Pour notre travail, nous avons utilisé la technique des puces à ADN développée par Illumina, appelée "*beadchips*" ou "*bead arrays*". Les puces Illumina sont des puces SNP, c'est-à-dire que chaque sonde analyse un SNP. Contrairement à la technique d'ACPA la plus utilisée, à savoir la CGH array, chaque sonde donne non seulement une indication sur le nombre de copies (CNV), mais aussi sur le génotype du SNP correspondant (génotype homozygote AA, homozygote BB ou hétérozygote AB).

Par rapport aux autres puces oligonucléotidiques, les puces Illumina reposent sur une conception différente (Gunderson et al. 2004) : en effet, les sondes ne sont pas synthétisées *in situ*, mais indépendamment de la puce. Les oligonucléotides sont tous longs de 50 bases nucléotidiques et fixés sur des billes (*beads*) de silice de 3µm (Steinberg et al. 2004). Chaque

bille contient environ 10^6 oligonucléotides porteurs de la même séquence et servant ainsi de sonde. Les sondes s'assemblent ensuite de façon aléatoire sur une lame de silicium contenant des puits de même diamètre que les billes (Fan et al. 2006). Les billes sont présentes sur la lame avec une redondance de 15 à 20 fois en moyenne (figure 8).

Figure 8 : exemple de puce Illumina (HumanCNV370-quad). Chaque champ est formé de centaines de milliers de micropuits dans lesquels se sont assemblées des billes où sont fixés les oligonucléotides.



D'après Gunderson et al. 2004

L'intérêt de cette méthode de production est qu'elle permet un très haut degré de miniaturisation. En revanche, son inconvénient est que les sondes ne sont pas dirigées sur le support lors de la fabrication de la puce et sont, de ce fait, localisées de façon aléatoire et variable d'une puce à l'autre. Chaque puce doit donc subir une étape de décodage (ou d'adressage) avant utilisation (figure 9).

Pour cela, sont effectuées plusieurs étapes d'hybridation / déshybridation séquentielles de la puce, avec des oligonucléotides spécifiques marqués en FAM ou cy3, en faisant varier le

marquage d'un oligonucléotide donné d'une étape à l'autre. Ainsi, chaque sonde sera associée à une signature correspondant à la séquence de couleurs obtenue par les différentes hybridations.

Figure 9 : processus de décodage d'une puce Illumina. Chaque bille possède sa propre signature suite aux étapes d'hybridation successives, permettant de l'adresser sur la puce. Ici, la bille #2 est reconnue par la séquence vert/rouge/vert.



Source : Gunderson et al. 2004

Ainsi, à chaque puce correspond une série de fichiers de décodage, qu'on appelle "*Decode Data*", informant sur la localisation de chaque sonde et qui diffère d'une puce à l'autre.

Les puces Illumina peuvent avoir plusieurs applications : génotypage de SNP pour des études d'association ou de liason par exemple, recherche de CNV, étude d'expression, étude de méthylation, etc. Ces applications n'utilisent pas toutes les mêmes modes opératoires ni les mêmes techniques d'analyse, mais le principe de production et de décodage des puces est identique.

3.2.3. Principe de la technique d'ACPA Illumina

L'analyse pangénomique sur puces Illumina (et plus généralement sur puces SNP) diffère de la technique de CGH array qui est la plus utilisée actuellement en ACPA.

La différence essentielle est qu'il ne s'agit pas d'une hybridation comparative de l'ADN testé avec un ADN témoin : seul l'ADN testé est utilisé dans cette méthode. Le protocole s'appelle "*Infinium Assay*" et repose sur la détection de l'extension d'une seule base sur la sonde (*single-base extension assay*) (Steemers et al. 2006).

Les principales étapes du processus sont décrites dans la figure 10. Elles comportent une amplification de l'ADN génomique, suivie d'une fragmentation et d'une hybridation sur la puce. Les sondes présentes sur les puces, et hybridées avec les ADN testés, subissent alors l'étape d'extension d'une base permettant le génotypage SNP, suivie d'une révélation par des anticorps fluorescents. Le matériel de départ est 750 ng d'ADN pour les puces HumanCNV370-Duo et 300 ng pour les puces HumanCNV370-quad et HumanCytoSNP-12.



Figure 10 : technique Infinium Assay

D'après Steemers et al. 2006

3.2.4. Analyse des puces Illumina

3.2.4.1. Lecture des lames

Une fois la révélation réalisée, les lames ont été lues par un scanner Illumina : l'Illumina BeadArray Reader, ou l'Illumina iScan (figure 11).

Figure 11 : à gauche : le BeadArray Reader ; à droite : l'iScan.



Source : Illumina

Grâce à 2 lasers, l'un rouge et l'autre vert, ces scanners vont détecter la fluorescence au niveau de chaque sonde de la puce. Les sondes dont le SNP est homozygote apparaîtront en rouge ou en vert, et les sondes dont le SNP est hétérozygote apparaîtront en jaune (figure 12).

Figure 12 : image de puce en train d'être scannée sur le logiciel Illumina BeadScan. Chaque point correspond à une sonde, rouge, verte ou jaune, selon le génotype du SNP correspondant.



Après ce scan, les données brutes, composées du fichier image de la puce (extensions .jpeg ou .tiff) et des fichiers d'intensité (extensions .idat) ont été récupérées.

3.2.4.2. Analyse des données

Les fichiers d'analyse de puces ont été créés avec le logiciel Illumina GenomeStudio à partir des fichiers d'intensité : fichiers .idat comprenant les valeurs de fluorescence, en vert et rouge.

3.2.4.2.1. Lecture des profils

L'analyse des résultats a été faite dans le Genome Viewer du logiciel Illumina GenomeStudio. Deux graphiques ont été utilisés pour cela ; et dans les 2 cas, chaque point représente une sonde (figure 13) :

- B allele freq (BAF) : correspond à la fréquence de l'allèle B pour chaque sonde, dont on peut déduire le génotype SNP. Un SNP de génotype AA aura une BAF à 0 ; un AB sera à 0,5 et un BB à 1 ;

log R ratio (LRR) : correspond à l'intensité du signal par rapport à la référence interne suivant la formule ln(R test / R référence). Si le nombre de copies est normal la valeur sera à ln1 soit 0 ;

Figure 13 : image d'un chromosome 7 normal sur le Genome Viewer. Graphe du haut : BAF à 0 ; 0,5 et 1 selon que le génotype SNP est AA, AB ou BB. Graphe du bas : LRR à 0 (nombre de copies normal)



Le chromosome X des échantillons masculins est particulier, car, en dehors des régions pseudoautosomiques, il n'y a qu'une copie pour chaque sonde. Par conséquent, la BAF est soit à 0 (SNP de génotype A), soit à 1 (SNP de génotype B) (figure 14).

Figure 14 : image d'un chromosome X normal chez un homme. A noter l'absence de SNP avec BAF à 0,5 en dehors des régions pseudoautosomiques.



3.2.4.2.2. Visualisation des anomalies chromosomiques homogènes

En cas de délétion homogène d'un autosome, c'est-à-dire en cas de perte d'une copie, le LRR devient négatif, en théorie ln 0,5 = -0,7. Chaque SNP n'a qu'un allèle : le génotype est donc soit A (BAF = 0), soit B (BAF = 1) ; il n'y a plus de SNP avec BAF = 0,5. On appelle cela une perte d'hétérozygotie (*Loss of Heterozygosity* ou LOH, figure 15)

Figure 15 : délétion terminale 6q de 3,5Mb (patient PSL_12844). Il n'y a plus de SNP avec BAF à 0,5 (LOH), et le LRR est négatif



En cas de duplication homogène d'un autosome, le LRR devient positif ; en théorie, on a ln1,5 = 0,4. Les génotypes possibles pour les SNP sont : AAA (BAF = 0) ; AAB (BAF = 0,33) ; ABB (BAF = 0,67) ou BBB (BAF = 1) (figure 16).

Figure 16 : duplication 7q11.2 (patient PSL_11799). Le LRR est positif, et la BAF de certains SNP est à 0,33 ou 0,67.



En cas de délétion biallélique (0 copie), le LRR plonge à des valeurs très négatives (LRR théorique : $\ln 0 = -\infty$). Quant à la BAF, elle a des valeurs aléatoires, car il n'y a ni allèle A, ni allèle B ; par conséquent la valeur théorique est 0 / 0, donc une forme indéterminée. En cas de délétion monoallélique sur l'X d'un homme, on a la même image, car il ne reste également aucune copie (figure 17).

Figure 17 : délétion sur le chromosome X d'un homme. La LRR est très négative et les BAF se répartissent de façon aléatoire.



En cas de duplication biallélique ou de triplication monoallélique (4 copies), le LRR est positif à environ 0,7 (ln2), et les SNP ont, selon leur génotype, une BAF à 0 (AAAA) ; 0,25 (AAAB) ; 0,5 (AABB) ; 0,75 (ABBB) ou 1 (BBBB) (figure 18).

Figure 18 : triplication de la région 15q proximale (patient PSL_12810). Le LRR est positif et les BAF sont à 0 ; 0,25 ; 0,5 ; 0,75 et 1.



3.2.4.2.3. Les pertes d'hétérozygotie

Un des intérêts majeurs des puces SNP, comparées à la CGH array, est que, grâce à l'analyse des génotypes, il est possible de détecter les pertes d'hétérozygotie sans modification du nombre de copies (*Copy Neutral LOH* ou CNLOH).

Ces CNLOH peuvent s'expliquer de 2 façons :

- une consanguinité : le CNLOH correspond à une région héritée de l'ancêtre commun des parents apparentés ;

- une isodisomie uniparentale (iUPD) : le CNLOH correspond à une région dont les 2 allèles proviennent du même chromosome parental.

Il est souvent possible de distinguer une iUPD du CNLOH à partir de plusieurs éléments :

- la consanguinité est généralement connue grâce à l'interrogatoire, et les CNLOH sont habituellement répartis sur plusieurs chromosomes ;

- les CNLOH d'une iUPD sont de grande taille et sur un seul chromosome (figure 19).

Avant d'affirmer qu'il s'agit d'une iUPD, un génotypage des parents est toutefois nécessaire.

Figure 19 : CNLOH complet du chromosome 16, dû à une upd(16) (patient PSL_7798).



3.2.4.2.4. Les anomalies en mosaïque

Les anomalies en mosaïque sont responsables d'images particulières sur le Genomeviewer de GenomeStudio.

En cas de perte de copies en mosaïque, le LRR reste négatif, mais est plus proche du 0 que les délétions homogènes. Et certains SNP ont une BAF intermédiaire, comprise entre 0 et 0,5 (mosaïque A/AB) et entre 0,5 et 1 (mosaïque B/AB) (figure 20).

Figure 20 : délétion 7q interstitielle en mosaïque. Le LRR est négatif, mais moins que dans le cas d'une délétion interstitielle homogène. Certains SNP ont une BAF intermédiaire.



Il est possible de calculer le taux de mosaïque en utilisant les valeurs de BAF intermédiaires entre 0 et 0,5 pour les SNP en mosaïque A / AB (qu'on appelle BAF mineures, figure 21).

Figure 21 : calcul du taux de mosaïque (M) d'une délétion en utilisant la valeur de la BAF mineure.

BAF = nombres d'allèles B / nombre d'allèles totaux = (1 - M) / (M + 2 (1 - M))M = (1 - 2BAF) / (1 - BAF)

En reprenant la délétion de la figure 20, la BAF mineure est environ à 0,35 ; donc le taux de mosaïque est autour de 46%.



En cas de gain de copies en mosaïque, le LRR reste positif, mais est plus proche du 0 que les délétions homogènes. Et certains SNP ont une BAF intermédiaire, comprise entre 0,33 et 0,5 (mosaïque AAB/AB) et entre 0,5 et 0,67 (mosaïque ABB/AB) (figure 22).

Figure 22 : trisomie 12 complète en mosaïque. Le LRR est positif, mais moins que pour un gain homogène. Certains SNP ont une BAF intermédiaire.



Comme pour les délétions en mosaïque, il est possible de calculer le taux de mosaïque, en utilisant les valeurs de BAF mineure (entre 0,33 et 0,5, figure 23).

Figure 23 : calcul du taux de mosaïque (M) d'une duplication, en utilisant la valeur de la BAF mineure.

BAF = nombres d'allèles B / nombre d'allèles totaux = (M + 1 - M) / (3M + 2 (1-M))M = (1 - 2BAF) / BAF

En reprenant la trisomie de la figure 22, la BAF mineure est environ à 0,4 ; donc le taux de mosaïque est environ à 50%.

Population cellulaire dupliquée Population cellulaire normale AAB M
1-M

Enfin, en cas de disomie uniparentale en mosaïque, le LRR est toujours à 0. Et certains SNP ont une BAF intermédiaire entre 0 et 0,5 (mosaïque AA / AB) et entre 0,5 et 1 (mosaïque BB/AB) (figure 24).

Figure 24 : iUPD 11p terminale en mosaïque. Le LRR est à 0. Certains SNP ont une BAF intermédiaire.



Comme pour les délétions et duplications en mosaïque, il est possible de calculer le taux de mosaïque, en utilisant les valeurs de BAF mineure (entre 0 et 0,5, figure 25).

Figure 25 : calcul du taux de mosaïque (M) d'une iUPD en utilisant la valeur de la BAF mineure.

BAF = nombres d'allèles B/nombre d'allèles totaux = (1 - M) / (2M + 2(1-M))

M = 1 - 2BAF

En reprenant l'iUPD de la figure 24, la BAF mineure est environ à 0,25 ; donc le taux de mosaïque est environ à 50%.



3.2.4.2.5. Normalisation et référence

Comme signalé précédemment, les puces Illumina, contrairement à la CGH array, ne comparent pas l'ADN testé à un ADN contrôle. La référence est bioinformatique et peut être réalisée de 2 façons différentes :

- en utilisant un fichier de référence fourni par Illumina : le fichier *cluster* (.egt) correspondant à plusieurs milliers de patients analysés par Illumina ;

- en utilisant la méthode du "*reclustering*" : la référence est calculée par un algorithme (Illumina Genetrend) en utilisant des patients de notre choix.

La méthode du *reclustering* est selon nous la meilleure, lorsqu'elle est utilisée avec des échantillons dont la technique a été effectuée lors du même "*run*", car elle permet d'avoir une référence réalisée exactement dans les mêmes conditions expérimentales que le patient test. Avec cette procédure, la dispersion des points est moindre (figure 26), ce qui permet d'être plus sensible et de détecter des CNV non décelés avec la méthode du fichier référence (figure 28). Elle permet en outre, en reclusterisant les femmes et les hommes séparément, de pouvoir

analyser simplement le chromosome X. En effet, avec le fichier *cluster* d'Illumina comme référence, l'X des hommes apparaît délété, car le fichier a été réalisé uniquement avec des données provenant de femmes (figure 27).

Figure 26 : graphes d'un même SNP (rs10000037), dans GenomeStudio avec la méthode du fichier *cluster* (a) et après *reclustering* (b). En abscisse : la *norm theta* qui correspond à la couleur normalisée (fonction du génotype) ; et en ordonnée : l'intensité normalisée (fonction du nombre de copies). Chaque point correspond à un échantillon du même projet. On voit 3 groupes de points sur l'abscisse, qui correspondent aux échantillons dont le génotype est AA (en rouge), AB (en violet), BB (en bleu). Les ellipses rouges, violettes et bleues correspondent à la référence. On voit qu'avec la méthode du *reclustering*, les ellipses sont mieux centrées sur le groupe de points





Figure 27 : graphe d'un chromosome X normal d'un même échantillon masculin, avec la méthode du fichier *cluster* (a) et après *reclustering* (b). Avec la première méthode, le chromosome X apparaît délété, sauf dans les régions pseudoautosomiques, alors qu'il apparaît tout à fait normal après *reclustering*.



Figure 28 : zoom en 15q proximal au niveau d'une délétion de 76 kb du gène UBE3A (patiente PSL 7458), avec la méthode du fichier *cluster* (a) et après *reclustering* (b).

La délétion avait été manquée avec la première méthode, à la fois par l'algorithme de segmentation (cf. infra) et en lecture visuelle, du fait de l'importante dispersion des points. Après que la délétion eut été trouvée par une autre technique (MLPA), nous avons changé notre méthode d'analyse en faisant un *reclustering*, et cette fois la délétion a été retrouvée par l'algorithme de segmentation.



L'inconvénient de la méthode du *reclustering*, est que l'algorithme *Genetrend* nécessite un grand nombre de patients pour établir la référence de façon fiable, sinon il est source d'erreurs (figure 29). C'est pour cette raison que depuis mars 2009, nous faisons la technique par *runs* de 96 échantillons, de façon à pouvoir effectuer un *reclustering* fiable.

Figure 29 : inversion duplication délétion 8p, après un *reclustering* de 96 patients (a) et de 10 patients (b). On constate en b des erreurs de *clustering* (points avec LRR à 0 et BAF à 0,5) dans le réarrangement.



3.2.4.2.6. Segmentation

L'objectif de la segmentation est de détecter automatiquement les régions du génome qui présentent un nombre anormal de copies. Nous avons systématiquement effectué une lecture visuelle des profils, comme expliqué précédemment, mais la segmentation a constitué une aide pour détecter certains CNV de petite taille (figure 30).

Nous nous sommes servis de l'algorithme d'Illumina, cnvPartition, qui utilise la segmentation binaire linéaire (Olshsen et al. 2004, technote Illumina DNA Copy Number and Loss of Heterozygosity Algorithm). Schématiquement, cet algorithme analyse les valeurs de LRR et BAF des sondes de façon ordonnée (par position génomique) et les compare à leur valeur théorique selon 5 hypothèses : 0 copie, 1 copie, 2 copies, 3 copies et 4 copies. En fonction de

ces valeurs, il attribue à chaque segment du génome l'hypothèse du nombre de copies la plus probable et un score de confiance qu'on dénomme la *CNVconfidence*.

Plus les valeurs de LRR et de BAF sont proches des théoriques, plus il y a de sondes incluses dans le segment, et plus ce score de confiance est élevé ; l'hypothèse posée par l'algorithme est alors d'autant plus vraisemblable. Pour l'analyse, nous avons conservé la valeur seuil par défaut de *CNVconfidence*, soit 35.

Les CNV détectés par cnvPartition sont indiqués sur le Genome Viewer de GenomeStudio, sous forme de marques appelées *bookmarks*.

Figure 30 : Petite délétion 15q proximale de 76 kb au niveau du gène *UBE3A* (patiente PSL_7458), signalée par cnvPartition sous forme d'un *bookmark* orange.



CnvPartition peut également détecter les CNLOH. Quand une analyse des LOH a été réalisée, nous avons fixé un seuil à 1 Mb.

3.2.4.3. Critères qualité

La qualité du résultat pour chaque sonde est calculée par le *GeneCall Score* (GC). En reprenant la figure 26, ce score est fonction de la distance entre la position de la sonde et le centre de l'ellipse qui est considéré comme la position "idéale". Si la sonde est dans cette position idéale, le GC est à 1, et plus le point est éloigné, plus le GC s'approche de 0.

Lorsque le GC est inférieur à 0,15 alors le SNP ne peut être génotypé ; il est dit *no call*. La proportion de SNP génotypés avec succès pour un échantillon est appelée le *Call Rate* et c'est ce taux qui est classiquement le plus utilisé comme score qualité pour les puces SNP. Nous avons considéré que les résultats avec un *Call Rate* > 0,98 étaient de bonne qualité, avec un

contrôle de la bonne qualité des profils lors de la lecture visuelle (dispersion des points, présence de "vagues" ou *Genomic Waves*). Lorsque les échantillons étaient de mauvaise qualité, l'ACPA a été refaite.

3.2.5. Intérêt des puces SNP comparées aux autres techniques d'ACPA

Le fait de ne pas nécessiter l'utilisation d'un ADN témoin présente un avantage important : lors de la lecture d'un profil d'ACPA, on analyse uniquement l'ADN du patient et non celui de l'ADN contrôle, ce qui facilite l'interprétation du résultat.

Outre qu'elles n'utilisent pas d'ADN témoin, les puces SNP ont pour avantage, par rapport à la CGH array, de donner les génotypes des SNP au niveau des sondes. Ces génotypes rendent possibles des études de liaison ou d'association, mais cela n'a pas été exploité dans notre travail. Elles permettent aussi de mettre en évidence des pertes d'hétérozygotie dues à des iUPD ou à la consanguinité, qui ne sont pas visibles par la technique de CGH array.

Grâce à l'identification des UPD, on peut faire certains diagnostics de maladies liées à l'empreinte, qui ne seraient pas détectées en CGH classique (Altug-Teber et al. 2005). Si le chromosome touché n'a pas de pathologie liée à l'empreinte connue, cela peut également orienter vers une maladie AR. Le taux de diagnostics d'iUPD dans de grandes séries d'ACPA constitutionnelles (principalement DI-S et DI-NS) varie de 0,1 à 0,7% (Papenhausen et al. 2011, De Leeuw et al. 2011, Bruno et al. 2011).

L'identification des LOH dans la consanguinité permet de faire de l'*homozygosity mapping* dans des familles où une pathologie AR est suspectée, afin d'identifier les locus candidats liés à la pathologie (cf. 2.2.4.).

3.5. Séquençage des gènes candidats

Le séquençage des gènes candidats dans des séries de patients DI a été effectué soit grâce à la méthode de Sanger (par terminaison de chaîne), soit par la méthode de pyroséquençage (Roche 454).

La recherche systématique de mutations dans les régions codantes candidates, définies par *homozygosity mapping* ou présentes dans un CNV d'intérêt, a été faite par séquençage de l'exome (séquençage nouvelle génération). Celui-ci a été réalisé en service, soit par la société Integragen, soit par le Centre National de Génotypage (CNG).

3.5.1. Séquençage par terminaison de chaîne (méthode de Sanger)

Le séquençage de l'ADN par la méthode de Sanger est la technique de séquençage la plus classiquement utilisée (Sanger et al. 1977). Elle repose sur le principe d'une polymérisation avec terminaison de chaîne par ddNTP fluorescents (appelés *dye-terminators*), couplée à une séparation des fragments obtenus par électrophorèse.

Pour cette technique, nous avons d'abord réalisé une PCR de chaque fragment à séquencer, sur un thermocycleur Bio-Rad C1000, avec un kit Roche Fastart DNA Polymerase. Les produits de PCR ont été purifiés sur un gel de filtration Bio-Rad Bio-Gel P100. La réaction de séquence a été ensuite effectuée avec le kit Applied Biosystem Big Dye Terminator V3.1 puis purifiée avec le gel de filtration GE Healthcare G50. Enfin les réactions de séquence ont été analysées sur un séquenceur Applied Biosystem 3730.

La lecture des électrophorégrammes a été réalisée avec le logiciel Applied Biosystem Seqscape v2.5.

Par cette méthode, nous avons séquencé les gènes *CAMTA1* (tableau 18), *SP3* (tableau 19), *STXBP6* (tableau 20), *CNTNAP4* (tableau 21), *NUDT12* (tableau 22), *ATXN10* (tableau 23), *YWHAE* (tableau 24), *BUD13* (tableau 25), *RBMX2* (tableau 26), *ATAD3A* (tableau 27) et *ATAD3B* (tableau 28).

Tableau 18 : séquences des amorces utilisées pour le séquençage de l'exon 1 de CAMTA1chez les 50 patients de Dijon.

| | Amorces de PCR | Amorces internes de séquence | | |
|------------|----------------------|------------------------------|----------------------|--|
| CAMTA1_1PF | TCCCGGAGAGTAGTGAGACC | CAMTA1_1SF | CGGAGAGTAGTGAGACCCCT | |
| CAMTA1_1PR | GTCACCTCAGGACTCGGC | CAMTA1_1SR | TCACCTCAGGACTCGGCC | |

Tableau 19 : séquences des amorces utilisées pour le séquençage de SP3 dans notre série.

| SP3_1F | GTTTCCCGCACAGTCAATC | SP3_4.3R | GAAGAATGAACCTGCCCACT |
|----------|---------------------------|----------|---------------------------|
| SP3_1R | AGTGCAGCTTTCTGCCTCTC | SP3_4.4F | TGAGCGGGTTTCTCCTGATA |
| SP3_2F | TTGGGAGTAATTTCGCTTCG | SP3_4.4R | GGTCACTGTCTGTGCCTGAA |
| SP3_2R | AGGGGAGAGACGAGGAGG | SP3_4.5F | AGTCTCAGCAGCCAACCAGT |
| SP3_3F | GGACCGAGTCGTACTTCCTG | SP3_4.5R | TGTGAGTCTGATGCTGATAAATACC |
| SP3_3R | GGCTGGCTTCAGATCCAAAC | SP3_5F | TGCAAAGTTGTAACATTGGGAG |
| SP3_4.1F | TTTGAACTCCATTGTGTTTTAAGAT | SP3_5R | CTTCTGCCACACGCATACAC |
| SP3_4.1R | GCCAGGAATGATCTGAATTTG | SP3_6F | TGGTTTTGGTTTCTGATTGACTT |
| SP3_4.2F | TCTTCCCCTTCAGAATTTGC | SP3_6R | TCAAGGCAGTTATGTAGCCCTAT |
| SP3_4.2R | CTGAGAACTGCCCGAGAGTC | SP3_7F | ACTTTCTGTTTTGGGGGAGA |
| SP3_4.3F | TTGCCTCTGGAACACCTTCT | SP3_7R | AAATTTTTGCACATCAATAAAAAGA |

Tableau 20 : séquences des amorces utilisées pour le séquençage de *STXBP6*.

| STXBP6_2F | ATGTTGTCAATTCCAACTTTCC | STXBP6_4R | TATGCTGCTGAACTCCTAACTTC |
|-----------|---------------------------|-----------|--------------------------|
| STXBP6_2R | CTGTGAATGTAAACCTCCTGG | STXBP6_5F | TACAGGATACTCACCCTGCC |
| STXBP6_3F | GTAGAAACAGCACTGAATGTGG | STXBP6_5R | CTGAAATAAAGTTTCTTGGCACTG |
| STXBP6_3R | CCACCACTACCACTACCAAGG | STXBP6_6F | GGGAGCAGCACAGTGAAATC |
| STXBP6_4F | GCCCTTATAAATCTTTTGACAGGAG | STXBP6_6R | AAAAGAAGCAAGCGGAGGTC |

Tableau 21 : séquences des amorces utilisées pour le séquençage de CNTNAP4.

| CNTNAP4_1F | AGAGCTGGCAGAGCTACTGAG | CNTNAP4_13R | GCAAATTACGTCAACATTCTGC |
|------------|-------------------------|-------------|---------------------------|
| CNTNAP4_1R | TGCAAATAATCATCATCAAACC | CNTNAP4_14F | AACATTTGACTCAGATAAAAGGTGA |
| CNTNAP4_2F | TTTGGGGCTGATAGGATTTG | CNTNAP4_14R | GACACCACCACTTTGCAGAC |
| CNTNAP4_2R | AAGCAACCAGTGCAAGAAGC | CNTNAP4_15F | AAGGAAATGCCAGCAATCAT |
| CNTNAP4_3F | TGTTGTTGTTTACTTGTTGGTTG | CNTNAP4_15R | GAAAATAACTCAAAAAGGCATTGA |
| CNTNAP4_3R | CCCTTTCACATTCCAGAACAG | CNTNAP4_16F | TTTTGTAGGTGCAAATTTAGTTCAG |
| CNTNAP4_4F | CAGCAAGACCCCTAGTTCTCC | CNTNAP4_16R | TGATAGAAACGTGATGCTTAAGTG |

Keren Boris – Thèse de doctorat - Année 2013

| CNTNAP4_4R | TCTTGGAAAATGTTTTCAATGTAAG | CNTNAP4_17F | GGATTTCAATTCCTGTAAAATCTC |
|-------------|---------------------------|-------------|--------------------------|
| CNTNAP4_5F | GCACCATTCATAGTAAAATGCAG | CNTNAP4_17R | GGCGCAATGAAAAGAATCAG |
| CNTNAP4_5R | AAGCTTTCCACAAAAATTATTCC | CNTNAP4_18F | TTCAGCGATTGTTACGCTGT |
| CNTNAP4_6F | TGAGGTTTGGTGGGTTACAT | CNTNAP4_18R | CAGGCTGGTTATGCAGAAGA |
| CNTNAP4_6R | CAAGCCCTTGAATCTGAAAA | CNTNAP4_19F | TTTTTGCCGTTCTCTGTAAGG |
| CNTNAP4_7F | TGAAAGATGGTGGTCCACAG | CNTNAP4_19R | TTGCCAAAGCAATTCAGAAA |
| CNTNAP4_7R | GATTTACTGAAAAGGTGGGCTC | CNTNAP4_20F | TTTGATTCTATTTTGGGGTTGG |
| CNTNAP4_8F | TGAAAAATTGATCTGTGTGTGATG | CNTNAP4_20R | TGAGAAGCCAGAACACGAGA |
| CNTNAP4_8R | TGGGAACCTAAAATGGAGGA | CNTNAP4_21F | TCCCTCGAAATGAATAAATAGC |
| CNTNAP4_9F | CGCGGATAATGTGAGATTGA | CNTNAP4_21R | GACCAATGAAAGTGTCTCTGC |
| CNTNAP4_9R | GCAAATATTACATTTTTGGTTATGC | CNTNAP4_22F | ATGTTTGGGGCCATGCTAAC |
| CNTNAP4_10F | TCCAATTTGTCTCCCTTTCC | CNTNAP4_22R | CGAAGTGGGTGTGTGTAATATG |
| CNTNAP4_10R | AGTCCACCTTGCAGAAATGC | CNTNAP4_23F | TTGCATTAGCTCAGAATCCTC |
| CNTNAP4_11F | TCATTATTCCCTCCTTTAATACAGA | CNTNAP4_23R | GGCAATGACCCCCATAAATA |
| CNTNAP4_11R | GAATGGGGAAGAGGAAAGAA | CNTNAP4_24F | CACATTCGCCTCCATAATTC |
| CNTNAP4_12F | TGACATCTGTCTTTCTCATGACC | CNTNAP4_24R | GGCACTTGGAGTATGTGCAG |
| CNTNAP4_12R | CCCTTTGCCAAGTTTGTTTC | CNTNAP4_25F | GCTTGAGCCAGAAATACCTGA |
| CNTNAP4_13F | AATTTCAAGATTTGCGGTATTTC | CNTNAP4_25R | CAAGCCTGTACATGGCAAGA |

Tableau 22 : séquences des amorces utilisées pour le séquençage de NUDT12.

| NUDT12_2F | GCCTTGGAAACCATTGCTATT | NUDT12_4R | GGAAAACTCTTTAGCTCCTCACA |
|-------------|-----------------------------|-----------|---------------------------|
| NUDT12_2R | TCAACAATATAAATATGGAAACAAGCA | NUDT12_5F | GGGCAGAAAGAAGCAATCTC |
| NUDT12_3.1F | AGAAGAACTTTAATATGTATCTGGAAA | NUDT12_5R | TGAATTTTTAACAAAGCAGACCAA |
| NUDT12_3.1R | CTCAGGCTGGGCCAAATAA | NUDT12_6F | TGAACCATTCAGGACACAAAG |
| NUDT12_3.2F | CCATCCAGCCACAGTTTTTA | NUDT12_6R | CCTTGCTCTGTGACCAAACC |
| NUDT12_3.2R | AAAGAGTGCCCCTGGAAAGT | NUDT12_7F | TGGATGTTGTCACTTTCTACTGTTG |
| NUDT12_4F | AAAACCCCTGATCACACAGC | NUDT12_7R | CAACATCGTATTTTGTGTTGTG |

Tableau 23 : séquences des amorces utilisées pour séquencer les points de cassure de la translocation 2;22, cassant *ATXN10*.

| Dérivé 22 | chr2-der22F | CCCAAGGGTTTTGCTTGTAG | chr22-der22R | AGGGCTTCTGGTCTTCCTG |
|-----------|-------------|----------------------|--------------|----------------------|
| Dérivé 2 | chr2-der2F | GGACAGCAAGGGTAGACTGC | chr22-der2R | AACAGTACTACCCGTGGGGA |

Tableau 24 : séquences des amorces utilisées pour le séquençage de *YWHAE*. Les amorces en bleu n'ont été utilisées que pour la réaction de PCR et les amorces en rouge sont des amorces internes utilisées uniquement pour la réaction de séquençage.

| YWHAE_1F | CTTTCGGCAGTCGCAGTTC | YWHAE_4R | GCCAGTGGATGTGATGAAAC | YWHAE_8F | TCCTTTGTCCTGCTCTGTTG |
|----------|--------------------------|----------|---------------------------|------------|-------------------------|
| YWHAE_1R | GGTCACATTCCAGGGCATAG | YWHAE_5F | ACCTCAAAACCAGAGATAGAATATG | YWHAE_8R | AACAGGACGATCATACGCAG |
| YWHAE_2F | CTCGTAGTGAGGGGTTTCGG | YWHAE_5R | CAAGTGGTTCCAAGGAATGG | YWHAE_9F | GACACCTTTGTTTGTGGCTG |
| YWHAE_2R | GAGCGCCTGGAATAGGTTAAG | YWHAE_6F | TGAAACGAAGCCAGTTATTCC | YWHAE_9R | GGTTCTCAGTGACAATGGGG |
| YWHAE_3F | CTTCCAACAGCAAAATATTATCAC | YWHAE_6R | ACAGGCCCAAGAAACAACAC | YWHAE_S_1F | CAGAGGAGAGGCGTCTCG |
| YWHAE_3R | CAATTCCCATGTACCCGAAC | YWHAE_7F | TTGCTGTCTAGCATGTCATTTC | YWHAE_S_1R | AGGCCATTTCCTGCTTCCC |
| YWHAE_4F | TCTGGAGTCTTTGTATTGGCTC | YWHAE_7R | GCTTGATATAACGACAAGCCAAG | YWHAE_S_6F | GAGAGTGTTTAAAATTATTGAAG |

Tableau 25 : séquences des amorces utilisées pour le séquençage de *BUD13* chez des patients Cockayne.

| BUD13_1F | TACGCGTTGAGGCATTTTG | BUD13_5F | GTTCTTCTCCCAGCCATAGG | BUD13_7R | TGCATATAGCTATGTATCAGCATCC |
|----------|-------------------------|----------|--------------------------|-----------|---------------------------|
| BUD13_1R | AGTTGGGAAGGGGTGGC | BUD13_5R | TTGGCTTGATTCTCTGGGAC | BUD13_8F | TAGCAGTATGGTCCATCTCGTG |
| BUD13_2F | AAGCACCAAAGGATGACCAG | BUD13_6F | CACAAATGCAACTAACTGTGTACC | BUD13_9F | AATTCCACATTGCCATCCAG |
| BUD13_2R | AATTTGAGAAGCTTGGCCTG | BUD13_6R | AACCAGAGCTGAAATTTGATAGC | BUD13_9R | GCAACTTACGAGTGGTTAGGG |
| BUD13_3F | TTGAGACAGTGCTTCTTGGC | BUD13_7F | TCCCAAGCGGTTGATTTG | BUD13_10F | TTGAAACAAATGCCCCTTTC |
| BUD13_3R | TGCTAAAACAAACTATGCATTCC | BUD13_7R | GCAGAGAAATGACTTGTTCCAG | BUD13_10R | GATATCTCGCTGCCTATGCC |

Tableau 26 : séquences des amorces utilisées pour le séquençage de *RBMX2* chez des patients Cockayne.

| RBMX2_1-2F | GTCACCGAATCGACTGGC | RBMX2_4R | CACCAAGCTCACAAGTGCTG |
|------------|--------------------------|----------|-----------------------------|
| RBMX2_1-2R | CGAAATAGTCGGAAACCGAG | RBMX2_5F | TCAAATAGTTTGACAGTTTGGTATGAG |
| RBMX2_3F | TTGCTTACTGTTTCCACCCC | RBMX2_5R | GCAAGGAAGAGCAGACATTAAG |
| RBMX2_3R | TCAGCCTAGAGTGAAAATAAGTGG | RBMX2_6F | CACCCGGCTGTGAATGTTTT |
| RBMX2_4F | TGCAACCTTTCTATGTCATCTTG | RBMX2_6R | CCACCCAACTGAATTTGACTGC |

Tableau 27 : séquences des amorces utilisées pour le séquençage de ATAD3A.

| ATAD3A_1F | AGTCAGACTCGGGTGGGG | ATAD3A_9F | AGGTCACTGGGTAGGTGGTT |
|-----------|----------------------|------------|----------------------|
| ATAD3A_1R | TGTAACCTATGAGCGCCACC | ATAD3A_9R | GTCGCATCTCAGGACCCAAA |
| ATAD3A_2F | CTCACAAATGCATCAGGCCG | ATAD3A_10F | TTTGGGTCCTGAGATGCGAC |
| ATAD3A_2R | GAGAAGATCCAGGCCATCCG | ATAD3A_10R | CCTGGAAGCCACAGCCTG |

Keren Boris – Thèse de doctorat - Année 2013

| ATAD3A_3F | CTGTGGGTCTGGATTCCTCC | ATAD3A_11F | GCGCCTCCCATCTTCCAG |
|-----------|----------------------|------------|-------------------------|
| ATAD3A_3R | GCCATCAGCCTCACACAAAC | ATAD3A_11R | CACAAGGACAGAGCTGCAGA |
| ATAD3A_4F | CCTGTAAGGTCCCTCACTGC | ATAD3A_12F | GTGTGTGGTTGGTTTCCCAC |
| ATAD3A_4R | CCTGTACACGTGTGCTGGT | ATAD3A_12R | AAAGCTGTTTTTCACACAAAGCT |
| ATAD3A_5F | CACACTCGGGCACAGTCAT | ATAD3A_13F | CAGCCTCCCAAAATCCAGGT |
| ATAD3A_5R | CTCCCTGAACAGATCCTGGC | ATAD3A_13R | CCGCATGTTCCTCAGTGACA |
| ATAD3A_6F | CTCCCTGAACAGATCCTGGC | ATAD3A_14F | TCGTGTTTCCCGCCACTTTA |
| ATAD3A_6R | CACCAAGGAGTCAGCCTACG | ATAD3A_14R | GATGGTGGACTGTGGGTCG |
| ATAD3A_7F | CTCACCCTCAACCTGCTCTC | ATAD3A_15F | CTGGTGTGAGGAGTGGTTGG |
| ATAD3A_7R | GCCACCACACCCAGCTAATA | ATAD3A_15R | ACTGTCAGCCACCGACATG |
| ATAD3A_8F | CACCTGCCTCCTGTAACCG | ATAD3A_16F | GAAGCCTGTGTTTCACGCTC |
| ATAD3A_8R | GGGAAGGGTCTCACTGATGC | ATAD3A_16R | GTGAGCCAGACGGAAAATGC |

Tableau 28 : séquences des amorces utilisées pour le séquençage de ATAD3B.

| ATAD3B_1F | CGTGGAGGCTGCTCCCAG | ATAD3B_9F | GGAACCGGAAAATGCCCCTA |
|-----------|----------------------|------------|----------------------------|
| ATAD3B_1R | ACCGCTTTTCGCACCTGA | ATAD3B_9R | CGTGGGCAGCCTATCTATCC |
| ATAD3B_2F | CCAGGCTTTGGTATCCGTGT | ATAD3B_10F | GGTCCTGAGATGCAACTGCT |
| ATAD3B_2R | CTCCCTCAGGACGACCAGAT | ATAD3B_10R | CACACCAGCACCCTGAGC |
| ATAD3B_3F | ACCTGCTGCACACACTAGTC | ATAD3B_11F | GCGCCTCCCATCTTCCAG |
| ATAD3B_3R | TCAGCCTCACACAAACGTGG | ATAD3B_11R | TCCACACTCAGGGGACTTTC |
| ATAD3B_4F | AGTCAGCCATTAGAGCTGCC | ATAD 12.1F | GCTTCTGTCGAGTCCAGGAC |
| ATAD3B_4R | CTCTAGCACCCACCGTCAG | ATAD 12.1R | ACTAGGTGTGGTGGCAAACA |
| ATAD3B_5F | CTTGTCAGGGAAGCCACAGT | ATAD3B_13F | CTGGTCAGGATTTTGAGTTTAGATCC |
| ATAD3B_5R | CTCCCTGAACAGATCCTGGC | ATAD3B_13R | GCCCCTGCATGTTCCTCAAT |
| ATAD3B_6F | TCTGTGGCGTTGGTCTGTC | ATAD3B_14F | TCCCCATGTTTCCTGTGCTC |
| ATAD3B_6R | GGAAGGACACGAAGATCCGT | ATAD3B_14R | CGGACCCCACACCATGAG |
| ATAD3B_7F | CGTGGTACGGATCTTCGTGT | ATAD3B_15F | CCCCAATCCAGGCACCATAT |
| ATAD3B_7R | TCTCGAGTAGCAGGACCACA | ATAD3B_15R | CTCCTGAGGGCACGAGCAG |
| ATAD3B_8F | CTGGGAATTCGGGTTCCTGT | ATAD3B_16F | CCTGGCGTGCATTAAGGGT |
| ATAD3B_8R | CAGACAGGGGAAGGGTCTCA | ATAD3B_16R | GTGAGCCAGACGGAAAATGC |

3.5.2. Séquençage par la méthode Roche 454

Pour le gène *CAMTA1* (cf. 4.4.1.1.), nous avons utilisé la méthode de pyroséquençage (NGS 454 de Roche) avec une approche "amplicon", dont le principe est expliqué ci-après (Margulies et al. 2005).

3.5.2.1. Préparation des "libraries"

L'objectif est d'obtenir un ensemble avec tous les fragments à séquencer, en nombre de molécules égal, et préparés de façon à ce que les extrémités 5' et 3' soient constituées d'adaptateurs. Ces adaptateurs sont composés de 3 oligonucléotides : un oligonucléotide servant d'amorce à l'emPCR (cf. 3.5.2.2.), une "clé" de séquençage, et un MID (*Multiplex Identifier*) spécifique à chaque patient et qui permet donc de les identifier.

Pour cela, nous avons effectué deux réactions de PCR avec le kit Roche FastStart High Fidelity PCR, conformément au protocole *Roche Amplicon Library Preparation Manual* (figure 31) :

- la première PCR a consisté en une amplification de tous nos amplicons à séquencer. Les amorces ont été désignées avec l'outil GS FLX Titanium Fusion Primers tool d'Integrated DNA Technology (http://eu.idtdna.com/scitools/applications/fusionprimers/default.aspx). Chaque amorce comporte une séquence spécifique de l'amplicon en 3' et une séquence consensus en 5' ;

- la deuxième PCR a consisté à réamplifier les produits de la première PCR avec des amorces contenant l'adaptateur.

Tous les amplicons ont ensuite été ajustés à la concentration de 10^9 molécules/µL, puis rassemblés.

Figure 31 : schéma de la préparation des *libraries* pour le séquençage par méthode 454 avec l'approche amplicon.



Les primers utilisés pour le séquençage du gène CAMTA1 sont fournis ci-après (tableau 29)

Tableau 29 : séquences spécifiques des amorces utilisées pour le séquençage de CAMTA1dans notre série de 47 patients (les séquences communes sont dans le protocole RocheemPCR Amplification Method Manual).

| CAMTA1_1F | TCCCGGAGAGTAGTGAGACC | CAMTA1_9.2F | AGGCCGTGTACACCATGTC | CAMTA1_16F | GCACTTTGTGTCTCTCAGGC |
|-----------|--------------------------|-------------|----------------------|--------------|-------------------------|
| CAMTA1_1R | GTCACCTCAGGACTCGGC | CAMTA1_9.2R | GTCCTTGTTGGAGTCGATGG | CAMTA1_16R | GGAATTCCACTTTCCTACAGGC |
| CAMTA1_2F | TGGAATGTTCCTTGATTTGG | CAMTA1_9.3F | AGCTCCCTCACCCTGACC | CAMTA1_16.2F | CCAGTGAAGGAGTGAGGGAC |
| CAMTA1_2R | GCTGAAGGCTGTTTCCACTG | CAMTA1_9.3R | GTAGAGTCCCTGCACCACG | CAMTA1_16.2R | CACAACCTTCGGAAATGGAC |
| CAMTA1_3F | CCTCTGATGCTCTTGGTCTC | CAMTA1_9.4F | TTGCAGCTCTGAGCACTACC | CAMTA1_17F | GCGAGGCAGGGAGGTAGAC |
| CAMTA1_3R | AAAGTAGGACTATGTGAAGCAACC | CAMTA1_9.4R | ATCCCTTAGCTGCAGGGTG | CAMTA1_17R | TGACACGGCATTCATCAAATAC |
| CAMTA1_4F | GAAAACAGCAGCTGGGAAAC | CAMTA1_10F | AGCAGTGAAGCCTCAGGGG | CAMTA1_18F | GGCAGATACCATCTTGGCTC |
| CAMTA1_4R | AACATGGCATTTTCAACTCG | CAMTA1_10R | AAGGCCAACTCCGAGTCC | CAMTA1_18R | TCGCATGATTCTGACTTTCATC |
| CAMTA1_5F | TGTAGAGACTTTTACTGGTCGATG | CAMTA1_11F | GTAGTGGGGAGGGGACATTC | CAMTA1_19F | TGGTAAACCTGGGATCTCAAC |
| CAMTA1_5R | TCCAAGCAATTCCATTCTCC | CAMTA1_11R | ACACCTTGTGCAAGAGTCCC | CAMTA1_19R | ACAAATTAAGTCTCTGTAATGAG |
| CAMTA1_6F | CCCTTCTTGCTCCTTCTCTG | CAMTA1_12F | GCATTTGGATGCTGGTCC | CAMTA1_20F | TAAGCTCGAAGGACGCTGAG |

| CCTCAGTGCCCTCCACAG | CAMTA1_12R | GATCCATCCACTGCCTCG | CAMTA1_20R | ACCCCAGACTCCAATGTCAG |
|--------------------------|--|---|--|---|
| CTCAGTCTCTGCCTGCACC | CAMTA1_13F | TGTTTCCCCTACATCGAAGC | CAMTA1_21F | TCACTGCTGACCAGTTTTCC |
| AAGGCCACAGAGGACCTTC | CAMTA1_13R | GGGTAGATTTCTTCCACGGG | CAMTA1_21R | TGAGGGTTAAGAAGGTCAGGG |
| CCTGGGTCCTACCCCTTG | CAMTA1_14F | GTTGGGTTTCATCTTGGTGG | CAMTA1_22F | CCCACCATCACTCTCCTTTTC |
| TACTCCCACAGAGGACAGCC | CAMTA1_14R | GTGGGAACCAGGCAAATC | CAMTA1_22R | TGGATGTGAAACTTACTGGCATAC |
| GGTCCTAGCTCTGACTCTCTTTTG | CAMTA1_15F | GCAATGGCAAAAGTCAGGTC | CAMTA1_23F | CCCCTTTCCTTTACTTGGTC |
| TGGCCATCAGAGCTCACG | CAMTA1_15R | AGGGCAGCTGCATGAAAG | CAMTA1_23R | GTGCGTGTGTGTTGTTGTTG |
| | CCTCAGTGCCCTCCACAG CTCAGTCTCTGCCTGCACC AAGGCCACAGAGGACCTTC CCTGGGTCCTACCCCTTG TACTCCCACAGAGGACAGCC GGTCCTAGCTCTGACTCTCTTTTG TGGCCATCAGAGCTCACG | CCTCAGTGCCCTCCACAG CAMTA1_12R CTCAGTCTCTGCCTGCACC CAMTA1_13F AAGGCCACAGAGGAGACCTTC CAMTA1_13R CCTGGGTCCTACCCCTTG CAMTA1_14F TACTCCCACAGAGGACAGCC CAMTA1_14R GGTCCTAGCTCTGACTCTTTTG CAMTA1_15F TGGCCATCAGAGCTCACG CAMTA1_15R | CCTCAGTGCCCTCCACAG CAMTA1_12R GATCCATCCACTGCCTCG CTCAGTCTCTGCCTGCACC CAMTA1_13F TGTTTCCCCTACATCGAAGC AAGGCCACAGAGGGACCTTC CAMTA1_13R GGGTAGATTTCTTCCACGGGG CCTGGGTCCTACCCCTTG CAMTA1_14F GTTGGGTTTCATCTTGGTGG TACTCCCACAGAGGACAGCC CAMTA1_14R GTGGGAACCAGGCAAATC GGTCCTAGCTCTGACTCTCTTTG CAMTA1_15F GCAATGGCAAAAGTCAGGTC TGGCCATCAGAGGCTCACG CAMTA1_15R AGGGCAGCTGCATGAAAAGTCAGGTC | CCTCAGTGCCCTCCACAGCAMTA1_12RGATCCATCCACTGCCTCGCAMTA1_20RCTCAGTCTCTGCCTGCACCCAMTA1_13FTGTTTCCCCTACATCGAAGCCAMTA1_21FAAGGCCACAGAGGACCTTCCAMTA1_13RGGGTAGATTTCTTCCACGGGCAMTA1_21RCCTGGGTCCTACCCCTTGCAMTA1_14FGTTGGGTTCATCTTGGTGGCAMTA1_22FTACTCCCACAGAGGACAGCCCAMTA1_14RGTGGGAACCAGGCAAATCCAMTA1_22RGGTCCTAGCTCTGACTCTTTTGCAMTA1_15FGCAATGGCAAAAGTCAGGTCCAMTA1_23FTGGCCATCAGAGCTCACGCAMTA1_15RAGGGCAGCTGCATGAAAGCAMTA1_23R |

3.5.2.2. PCR en émulsion

L'objectif est d'amplifier les amplicons à séquencer sur des billes, pour obtenir un grand nombre de molécules du même amplicon par bille. Nous avons, pour cela, utilisé le kit Roche GS Junior Titanium emPCR Kit (Lib-A) et le protocole *Roche emPCR Amplification Method Manual* dont voici les étapes (figure 32) :

- capture de *library* : le principe est de mixer la *library* avec les billes, chaque bille étant recouverte d'oligonucléotides complémentaires à la partie 5' des adaptateurs ;

 PCR en émulsion (emPCR) : cette étape crée une émulsion de gouttelettes d'eau dans de l'huile (1000 gouttelettes/µL), constituant chacune un milieu de réaction (qu'on appelle "microréacteur") où un amplicon va être amplifié sur une bille ;

- enrichissement : après l'emPCR, les fragments amplifiés ont été marqués par de la biotine. Puis des billes magnétiques avec de la streptavidine ont été utilisées de façon à récupérer préférentiellement à l'aide d'un aimant, les billes où une amplification a eu lieu.



Figure 32 : schéma de la PCR en émulsion par la technologie 454.

D'après Margulies et al. 2005

3.5.2.3. Réaction de séquençage

La réaction de séquençage s'est déroulée sur une plaque composée de micropuits de $44\mu m$, appelée "*PicoTiter Plate*" dans un séquenceur Roche GS Junior. Chaque micropuit constitue un milieu de réaction avec l'amplicon sur la bille, servant de matrice au séquençage.

La méthode de séquençage n'est pas une terminaison de chaîne, mais un pyroséquençage ou "séquençage par synthèse" (Ronaghi et al. 1998).

En résumé, les 4 différents dNTP ont été ajoutés, chacun leur tour, à chaque cycle de séquence. Lorsque le dNTP est complémentaire de la base à élonger sur la matrice, il se

produit une polymérisation de l'ADN et une libération de pyrophosphate qui va émettre de la lumière lors d'une réaction couplée avec une sulfurylase et une luciférase.

Ainsi, à chaque cycle, les puits émettant de la lumière sont ceux dont la base séquencée correspond au dNTP ajouté au milieu de réaction.

Pour cette étape, nous avons utilisé les kits Roche GS Junior Titanium Sequencing Kit et GS Junior Titanium PicoTiterPlate Kit, en suivant le protocole *Roche GS Junior Sequencing Method protocol* (figure 33).

Figure 33 : réaction de séquençage par la technologie 454



D'après Margulies et al. 2005

3.5.2.4. Analyse des séquences

Les séquences obtenues ont été alignées sur les séquences de référence et les variants détectés grâce au logiciel d'analyse Roche *GS Amplicon Variant Analyzer* (AVA).

3.5.3. Séquençage de l'exome

Le séquençage de l'exome (ensemble des exons du génome) a été fait dans 2 familles consanguines. La partie technique a été réalisée sous forme de service par la société Integragen (patients PSL_10908 et PSL_10909), et le Centre National de Génotypage (CNG) (patients PSL_9730 et PSL_9731). Dans les 2 cas, nous avons récupéré les variants annotés par les prestataires sous forme de fichiers grâce à une interface web.

La méthode comporte plusieurs parties que nous allons développer ci-dessous.

3.5.3.1. Technique d'enrichissement Agilent SureSelect

La "capture d'exons" s'effectue en phase liquide, par la technique Agilent SureSelect, (kit Agilent Human All Exon V4 + UTRs – 70 Mb, figure 34) permettant d'obtenir une *library* contenant tous les exons du génome à partir de l'ADN génomique (Gnirke et al. 2009).



Figure 34 : technique d'enrichissement Agilent SureSelect

Source : Protocole Agilent SureSelect Target Enrichment System

3.5.3.2. Séquençage par la technologie Illumina

La technique de séquençage Illumina, anciennement technologie Solexa, inclut d'abord l'étape de *clustering*, qui consiste en l'amplification de la *library* dans une *flow cell* comportant 8 canaux, chaque canal servant d'enceinte de réaction pour une *library* (figure 35). Puis se produit la réaction de séquençage en phase solide proprement dite, toujours dans la *flow cell* (figure 36). Cette technique a été effectuée avec un séquenceur Illumina HiSeq 2000, qui produit des longueurs de lecture de 2x75 paires de bases, par une méthode utilisant une terminaison de chaîne réversible (Bentley et al. 2006).

Figure 35 : étape de *clustering* (séquençage Illumina). La *library* est capturée sur une *flow cell* où sont présentes des amorces, complémentaires des adaptateurs ligués lors de la capture.



Source : Illumina


Figure 36 : réaction de séquençage par la technologie Illumina.



3.5.3.3. Analyse des séquences

La suite logicielle Illumina CASAVA 1.8 a permis l'alignement des séquences obtenues (dites lectures ou *reads*) et la détection des variants. L'annotation des variants a ensuite été réalisée par des outils internes à Integragen et le CNG, et ces résultats nous ont été rendus sous forme d'une interface web *http://eris.integragen.com* et *https://www.genoscope.cns.fr/sadc/*.

Etant donné le nombre très important de variants retrouvés par cette technique, nous avons utilisé différents filtres afin de sélectionner les variants d'intérêt :

- la rareté du variant (ressources dbSNP, HapMap, 1000 genomes, EVS) : nous avons sélectionné les variants absents des différentes bases de données ou de fréquence < 1% ;

- l'effet fonctionnel du variant sur la protéine : nous avons sélectionné en priorité les variants conduisant à un stop prématuré (non-sens, indel avec décalage du cadre de lecture, mutation de séquence consensus d'épissage), puis les variants faux-sens ayant un effet délétère selon les principaux algorithmes de prédiction (SIFT, PolyPhen-2, MutPred).

Dans cette partie, nous présenterons d'abord les résultats de l'ACPA en diagnostic sur la série de patients complète. Nous nous focaliserons ensuite sur certains types de CNV particuliers : les anomalies de petite taille (ne contenant que 1 ou 2 gènes) pathogènes (articles 1, 2 et 3) et les facteurs de susceptibilité (article 4). Puis nous nous intéresserons aux différents gènes candidats identifiés grâce à la stratégie utilisée dans notre série (articles 1 et 5). Enfin, nous nous arrêterons spécifiquement sur les résultats retrouvés chez les patients de notre série DI présentant des TSA (article 6).

4.1. Résultats diagnostiques

4.1.1. Les CNV

Les CNV retrouvés en ACPA dans notre série de 1 187 patients ont été classés en 4 groupes :

- CNV bénins (B), généralement présents dans DGV et/ou sans séquence codante incluse ;

- CNV *de novo* (DN), que nous avons considérés comme pathogènes dans la grande majorité des cas, ou VOUS dans quelques cas ;

- VOUS, absents de DGV mais non vérifiés car nous n'avons pas obtenu d'échantillons des parents ;

- CNV absents de DGV, vérifiés et transmis (VT) par un parent sain, considérés comme probablement bénins, mais possiblement impliqués dans une maladie AR ou comme facteurs de susceptibilité.

Sur les 1 187 patients DI analysés, nous avons retrouvé 143 CNV *de novo* considérés comme pathogènes, soit un total de diagnostics établis chez 12% des patients (tableau 30, figure 37).

| | DN | VOUS | VT | В | Total |
|--------------------------------------|-------|--------|--------|--------|--------|
| CNV totaux | 143 | 200 | 296 | 4 355 | 4 994 |
| nombre de pertes | 91 | 79 | 96 | 2 424 | 2 690 |
| taille moyenne des pertes | 6,5Mb | 0,37Mb | 0,36Mb | 0,18Mb | 0,41Mb |
| nombre moyen de gènes délétés | 39,4 | 4,0 | 2,3 | 1,4 | 2,8 |
| nombre de gains | 52 | 121 | 200 | 1 931 | 2 304 |
| taille moyenne des gains | 11Mb | 0,53Mb | 0,32Mb | 0,27Mb | 0,53Mb |
| nombre moyen de gènes dupliqués | 81,7 | 4,9 | 3,7 | 2,5 | 4,5 |
| CNV retrouvés avec les puces CNV370 | 15 | 6 | 39 | 1 600 | 1 660 |
| CNV retrouvés avec les puces CytoSNP | 128 | 194 | 257 | 2 755 | 3 334 |

Tableau 30 : résultats des CNV retrouvés dans notre série DI

Concernant nos résultats en fonction de la clinique, un CNV *de novo* a été retrouvé chez 14,6% de nos 809 patients DI-S (n = 118) et chez 6,6% de nos 378 patients DI-NS (n = 25). Et parmi nos 256 patients présentant des TSA, 5% (n = 13) ont un CNV *de novo*.



Figure 37 : localisation des différents CNV retrouvés, chromosome par chromosome, dans UCSC Genome Browser

Keren Boris - Thèse de doctorat - Année 2013





Keren Boris - Thèse de doctorat - Année 2013



114





4.1.2. Les disomies uniparentales

Sur les 1 187 patients DI analysés, nous avons retrouvé 5 iUPD.

- Une disomie uniparentale complète et homogène du chromosome 9, avec un mélange d'isodisomie et d'hétérodisomie, d'origine maternelle (figure 38).

Figure 38 : profil de la patiente PSL_9378 avec disomie uniparentale du 9.

| | M | GT Samples 9 | 78 [65] (1pxI=95KB) | and the second | |
|-------------|------------------|-------------------------|---|--|--------------------------|
| Viele Freq | 75 50 | 2. 2. | and the second state of the second states and the | | |
| 8 | | à. | Although the second state | | |
| 112 | | GT Samples 9 | 378 [65] (1pxl=95KB) | | |
| Log R Ratio | | | | | |
| 9 | p21.1 p15.3 | p12 p11/2 p11.1 q11 q12 | q21.13 | q31.1 q31.2 q31.3 q32 q3 | 3.1 933.2 933.3 9 934.3) |

Etant donné qu'il n'y a pas de pathologie connue provoquée par l'UPD du 9, et que les UPD peuvent résulter d'une correction de trisomie, nous avons effectué une recherche de trisomie 9 en mosaïque dans les fibroblastes de la patiente, par une FISH avec la sonde centromérique. Cette recherche a retrouvé une trisomie 9 dans 4% des 200 cellules analysées.

Nous avons donc considéré que cette trisomie du 9 était l'étiologie de la DI-S de cette patiente.

- Une isodisomie, segmentaire en mosaïque, de l'extrémité terminale du bras court du chromosome 11, d'origine paternelle (figure 39). Le taux de mosaïque est évalué à 70%, et la sonde la plus proximale informative dans l'UPD est à la coordonnée 2 947 366.

Figure 39 : profil de la patiente PSL_6146 avec disomie uniparentale segmentaire en mosaïque du 11.

AN A HAR STREET inter ein werte fangereit felder ster in der eine feldere bereite bereit les 6146 [111] (1

Bien que cette anomalie moléculaire soit considérée comme responsable du syndrome de Wiedemann-Beckwith, la patiente ayant cette anomalie dans notre série n'avait aucun des signes cliniques de ce syndrome, à savoir : macrosomie, grande taille, macroglossie, viscéromégalie, anomalies des oreilles, anomalies de la paroi abdominale, etc. Par la suite, une mutation *de novo* du gène *TUBA1A* a été retrouvée chez cette patiente, expliquant son phénotype de DI sévère avec malformations cérébrales typiques.

- Une disomie uniparentale complète et homogène du chromosome 15, avec un mix d'isodisomie et d'hétérodisomie, d'origine maternelle (figure 40).

Figure 40 : profil de la patiente PSL_8591 avec disomie uniparentale du 15



Les UPD du 15 d'origine maternelle sont responsables du syndrome de Prader-Willi, ce qui correspond bien au phénotype de la patiente PSL_8591, atteinte d'une DI-S avec obésité, épilepsie, petite taille et petites extrémités. Nous avons donc considéré que cette UPD du 15 était l'étiologie de la DI-S de cette patiente.

- deux isodisomies complètes et homogènes du chromosome 16 d'origine paternelle (figure 41).

Figure 41 : profil des patients PSL_7798 et PSL_10170 avec disomie uniparentale du 16



Etant donné qu'il n'y a pas de syndrome caractérisé provoqué par l'UPD paternelle du 16, nous avons effectué une recherche de monosomie du 16 en mosaïque dans les fibroblastes des 2 patients concernés, en FISH avec la sonde centromérique, sur 200 cellules. Cette recherche a cependant été négative.

Ces 2 patients présentent chacun un phénotype très différent. Actuellement, nous ne connaissons pas la signification, sur le plan étiologique, de ces 2 UPD du 16, que nous avons donc considérées comme des VOUS.

4.2. CNV pathogènes de petite taille (articles 1, 2 et 3)

Parmi les CNV que nous avons vérifiés (contenant des séquences codantes et non répertoriées dans DGV), 303 CNV étaient de petite taille, ne contenant que 1 ou 2 gènes.

Nous avons conclu que quatre de ces CNV étaient la cause de la DI, car ils emportaient des gènes connus pour causer une DI : *UBE3A*, *FLNA*, *FOXP1* et *ANKRD1*. Nous avons par ailleurs retrouvé une délétion du gène *DMD*, responsable de dystrophinopathie, que nous avons considérée comme pathogène, bien qu'il s'agisse d'une observation fortuite, n'expliquant pas à elle seule le phénotype du patient. En outre, nous nous sommes également intéressés à d'autres gènes délétés, qui nous ont semblé être de potentiels gènes candidats de DI (cf. 4.4).

En collaboration avec l'équipe des Docteurs Faivre et Thauvin-Robinet (CHU Dijon, France), nous avons également démontré que les réarrangements du gène *CAMTA1* constituaient une cause de DI et d'ataxie (cf. 4.4.1.1).

Au total, dans notre série, nous avons donc mis en évidence 6 petites délétions pathogènes (tableau 31).

| patient | chr | taille (kb) | début | fin | copies | gènes inclus | transmission | |
|-----------|-----|----------------|-----------|-----------|--------|--------------|------------------------|--|
| PSL_7458 | 15 | 76 | 25572823 | 25625654 | 1 | UBE3A | maternelle (empreinte) | |
| PSL_9055 | 1 | 49 | 6854452 | 6903599 | 1 | CAMTA1 | de novo | |
| PSL_9819 | х | 45 | 153540978 | 153585854 | 1 | FLNA-TKTL1 | de novo | |
| PSL_10017 | х | 269 | 31664816 | 31933609 | 0 | DMD | maternelle | |
| PSL_10406 | 3 | 234 | 71105864 | 71339355 | 1 | FOXP1 | de novo | |
| PSL_12419 | 16 | 148 | 89328478 | 89477051 | 1 | ANKRD11 | par le père atteint | |

Tableau 31 : CNV pathogènes ne contenant que 1 ou 2 gènes, retrouvés dans notre série DI

Nous avons également retrouvé une délétion n'emportant que le gène *SP3*, que nous avons considérée comme potentiellement impliquée dans la DI (cf. 4.4.1.2).

4.2.1. Patiente PSL_7458 avec délétion d'UBE3A (article 1)

Il s'agit de 2 sœurs, de 13 et 16 ans, présentant toutes les deux un syndrome d'Angelman. Leur présentation clinique est détaillée dans l'article 1 (Piard et al. 2011). Pour résumer, elles ont

toutes les deux un retard psychomoteur avec marche à 24 mois, une épilepsie, une microcéphalie, un important retard de langage (quelques mots seulement à 16 et 13 ans), une dysmorphie évocatrice avec grande bouche. La cadette a par ailleurs une ataxie et des rires immotivés.

Le bilan, avec recherche de délétion et anomalie de méthylation en 15q11.2 et un séquençage d'*UBE3A*, était négatif.

Nous avions analysé l'ADN de la cadette en ACPA dans un cadre diagnostique (PSL_7458) et n'avons initialement pas retrouvé d'anomalie. Puis les 2 patientes ont été étudiées en MLPA (kits P064 et M028) et une délétion des exons 6 à 12 a été retrouvée transmise par la mère.

Nous avons ensuite repris nos données d'ACPA pour comprendre pourquoi cette délétion n'avait pas été detectée lors de l'analyse initiale : la délétion est de petite taille (76 kb), ce qui rend sa détection difficile, aussi bien à l'œil, que par l'algorithme de segmentation cnvPartition.

Au moment de l'analyse initiale, nous établissions notre référence à partir du fichier *cluster* (cf. 3.2.4.2.5.) comme préconisé par Illumina. Puis nous avons refait notre référence par la méthode du *reclustering*, ce qui a diminué la dispersion. Et après ce *reclustering*, cnvPartition a détecté la délétion (figure 42). Depuis cette expérience, nous avons modifié l'analyse des données, en réalisant systématiquement notre référence par la méthode du *reclustering* ; nous avons également réanalysé nos données antérieures de puces afin de détecter éventuellement de petites délétions qui auraient été manquées.



Figure 42 : délétion de 76 kb d'UBE3A (patiente PSL_7458)

4.2.2. Patiente PSL_9819 avec délétion de FLNA (article 2)

La présentation clinique de cette patiente de 12 ans est détaillée dans l'article 2 (Chardon et al. 2012). En résumé, elle est atteinte d'une paraplégie spastique, et a présenté un retard psychomoteur (marche à 24 mois), une épilepsie à partir de 3 ans, un foramen ovale persistant et une bicuspidie aortique. L'IRM cérébrale a montré des hétérotopies nodulaires périventriculaires, ainsi qu'une séquelle d'accident vasculaire.

Une mutation de FLNA a été recherchée sans retrouver de variant pathogène.

L'analyse en ACPA effectuée ensuite a retrouvé une délétion Xq28 *de novo* de 45 kb, incluant les exons 30 à 48 du gène *FLNA* et les exons 7 à 13 du gène *TKTL1* (figure 43). Nous avons considéré, étant donné le phénotype, que l'haploinsuffisance de *FLNA* était responsable de la pathologie de cette patiente. Cette délétion a été rapportée en collaboration avec l'équipe de Kym Boycott (Ottawa, Canada), décrivant également une seconde délétion de *FLNA* (Chardon et al. 2012).

Figure 43 : délétion de 45 kb de FLNA (patiente PSL_9819)



4.2.3. Patiente PSL_10406 avec délétion de FOXP1

Cette patiente de 33 ans, a présenté un retard psychomoteur (marche à 24 mois) et un retard de langage. Elle a développé dans l'enfance des troubles psychotiques avec un repli, des automutilations et des hallucinations visuelles et auditives.

L'analyse en ACPA a retrouvé une délétion 3p14.1 *de novo* de 234 kb emportant les exons 3 et 4 du gène *FOXP1* (Figure 44). Etant donné que des mutations et délétions *de novo* ont été décrites dans ce gène chez des patients présentant une DI, un retard de langage et des troubles neuropsychiatriques (Hamdan et al. 2010, Horn et al. 2010), nous avons conclu à la pathogénicité de ce CNV.

Figure 44 : délétion de 234 kb de FOXP1 (patiente PSL_10406)



4.2.4. Patiente PSL_12419 avec délétion de ANKRD11

Cette patiente de 33 ans, avait une petite taille à -4DS, un surpoids, une DI légère et une dysmorphie. Son père avait également une petite taille à -4DS et la même dysmorphie, mais pas de DI.

L'ACPA a retrouvé une délétion 16q24 de 148 kb des exons 4 à 15 du gène *ANKRD11*, gène du syndrome KBG (Sirmaci et al. 2011), aussi bien chez la patiente que chez son père (figure 45). Ni l'un ni l'autre ne présentent toutefois les anomalies dentaires classiques, mais ils ont la petite taille et la dysmorphie caractéristiques. Nous avons donc conclu à la pathogénicité du CNV.

Figure 45 : délétion de 148 kb de ANKRD11 (patiente PSL_12419)



4.2.5. Patient PSL_10017 avec délétion de DMD

Ce patient âgé de 16 ans a des TSA avec stéréotypies et anxiété, ainsi qu'un retard de langage et une DI.

L'ACPA a retrouvé une délétion Xp21.1 de 269 kb des exons 48 à 55 (en phase) du gène *DMD* responsable de dystrophinopathie (figure 46). Il s'agit d'une découverte fortuite car il n'y avait pas de signe musculaire. Le bilan réalisé au décours de l'ACPA retrouve des CPK augmentées, ce qui confirme le risque de développement d'une dystrophie de Becker.

Le patient a par ailleurs une duplication 16p11.2, transmise par le père, qui est un facteur de susceptibilité au TSA et à la DI. Et l'absence de dystrophine, en particulier de l'isoforme Dp140, délétée chez ce patient est également associée à la DI (Felisari et al. 2000). Il est donc possible que la présence de ces 2 CNV, facteurs de susceptibilité, ait entraîné le phénotype de DI et de TSA, bien que cette hypothèse reste à confirmer.

Figure 46 : délétion de 269 kb de DMD (patient PSL_10017)



4.3. CNV facteurs de susceptibilité.

En dehors des résultats pathogènes, nous nous sommes également intéressés aux principaux facteurs de susceptibilité connus, retrouvés dans notre série (tableau 32).

Tableau 32 : facteurs de susceptibilité à la DI retrouvés dans notre série. Il faut signaler que nous n'avons pas pu vérifier la transmission parentale pour un certain nombre de nos patients.

| locus | CNV | patients | transmission transmission | | de novo |
|----------|-------|----------|---------------------------|------------|---------|
| | | | maternelle | paternelle | |
| 1q21.1 | perte | 1 | 1 | 0 | 0 |
| 1q21.1 | gain | 3 | 2 | 0 | 0 |
| 15q13.3 | perte | 4 | 0 | 2 | 1 |
| 16p11.2 | perte | 5 | 0 | 0 | 0 |
| 16p11.2 | gain | 7 | 1 | 2 | 2 |
| 16p13.11 | perte | 2 | 0 | 0 | 0 |
| 16p13.11 | gain | 3 | 0 | 1 | 1 |

Dans notre série, les facteurs de susceptibilité les plus fréquents sont les gains et pertes de la région 16p11.2. Nous avons inclus ces patients dans une grande étude collaborative dirigée par les équipes de Philippe Froguel (Lille, France) et Jacques Beckman (Lausanne, Suisse), dont les résultats sont présentés dans l'article 4 (Jacquemont et al. 2011).

Dans cette étude, outre la DI et les troubles neuropsychiatriques, la délétion était fortement associée à un BMI élevé, alors que la duplication était associée à un BMI diminué.

Remarquons que, dans notre série, il semble bien y avoir une tendance au surpoids parmi les 5 patients délétés : 3 patients ont un poids > $97^{\text{ème}}$ percentile, 1 patient a un poids entre le $75^{\text{ème}}$ et le $97^{\text{ème}}$ percentile, et 1 patient a un poids à la moyenne.

Pour les 7 patients dupliqués, on ne note pas particulièrement de tendance à un petit poids : 1 patient a un poids $< 3^{\text{ème}}$ percentile, 1 patient a un poids entre le $25^{\text{ème}}$ et le $50^{\text{ème}}$ percentile, 2 patients ont un poids à la moyenne, 1 patient a un poids entre le $50^{\text{ème}}$ et le $75^{\text{ème}}$ percentile et 2 patients on un poids $> 75^{\text{ème}}$ percentile.

4.4. Gènes candidats (article 4)

Selon la stratégie développée à la partie 2.4, nous avons identifié plusieurs gènes candidats à partir de l'ACPA en diagnostic, gènes que nous avons cherchés à valider en séquençage.

4.4.1. Gènes candidats identifiés par des CNV de novo.

Comme expliqué dans notre partie 2.4., dans le cas de délétions *de novo*, nous émettons l'hypothèse d'une transmission AD ; il s'agit donc le plus souvent de cas sporadiques.

4.4.1.1. Séquençage du gène *CAMTA1* (article 3)

Le travail concernant ce gène a été l'objet de l'article 3 (Thevenon*, Lopez*, Keren*, et al. 2012 ; * les 3 premiers auteurs ayant contribué à ce travail à part égale).

Nous avons retrouvé une délétion 1p36.31 de 49 kb, de formule arr[hg19] 1p36.31(6,854,452-6,903,599)x1 dn, chez une patiente (PSL_9055) de 3 ans, présentant une ataxie congénitale et un léger retard de développement. La description clinique plus détaillée de la patiente se trouve dans l'article 3 (cas index de la famille 3). La délétion emporte les exons 2 et 3 du gène *CAMTA1* et n'est pas retrouvée chez les parents (figure 47).

Figure 47 : délétion de 49 kb de CAMTA1 (patiente PSL_9055)



Chez cette patiente, même si la délétion est de petite taille, plusieurs arguments sont en faveur de la pathogénicité de ce CNV, ainsi que celle de l'haploinsuffisance de *CAMTA1* :

- la délétion est *de novo* ;

- *CAMTA1* est exprimé préférentiellement dans le cortex temporal et le cervelet (Huentelman et al. 2007), ce qui est cohérent avec les principaux signes cliniques : retard des apprentissages et ataxie ;

- une délétion *de novo* de *CAMTA1* a déjà été décrite chez un patient présentant un retard de langage, des TSA et une ataxie (patient 3 de l'article de Mikhail et al. 2011).

Afin de valider l'implication de ce gène dans le phénotype de notre patiente, nous avons séquencé les séquences codantes de *CAMTA1* (23 exons, 36 amplicons) avec la technique de pyroséquençage (Roche 454, cf. 3.5.2.) chez 47 patients présentant une ataxie congénitale avec DI (30 cas) ou sans DI (17 cas), sélectionnés parmi les patients du centre de référence des déficiences intellectuelles de causes rares. Ce séquençage n'a toutefois pas permis de retrouver de variant pathogène dans notre série.

Lors de cette étude, nous avons appris que l'équipe du Pr Faivre et du Dr Thauvin-Robinet (CHU Dijon, France) avait réuni 2 familles atteintes de DI-NS et qui présentaient chacune un réarrangement intragénique du gène *CAMTA1*, entraînant un stop prématuré (une délétion et une duplication), ségrégeant avec la pathologie (familles 1 et 2 dans l'article 3). Nous avons donc décidé de continuer ce travail en collaboration avec cette équipe. Du fait de la présence d'une ataxie chez notre patiente, leurs patients ont été neurologiquement réévalués et une ataxie a été mise en évidence. L'équipe de Dijon a par ailleurs séquencé les exons 2 à 23 chez 50 patients DI-NS et les exons 2 à 7 chez 100 patients DI-NS. Nous nous sommes chargés du séquençage de l'exon 1 de leur première cohorte de 50 patients. Aucune mutation pathogène n'a été retrouvée chez ces patients DI-NS.

4.4.1.2. Séquençage du gène *SP3*

Nous avons retrouvé une délétion 2q31.1 de 570 kb, de formule arr[hg19] 2q31.1(174,363,575-174,915,012)x1 dn, chez une patiente (PSL_11404) avec une DI-NS modérée (figure 48). Elle n'avait pas de signe dysmorphique, ni de signe neurologique, ni de malformation.

Figure 48 : délétion de 570 kb de SP3 (patiente PSL_11404)



Cette délétion ne contient qu'un seul gène : *SP3* qui est un facteur de transcription ayant un domaine de liaison de l'ADN en doigt de zinc. D'après la base de données Unigene, son expression est ubiquitaire.

Chez cette patiente, malgré le fait que ce CNV est de petite taille, plusieurs arguments sont en faveur de la pathogénicité de ce CNV ainsi que celle de l'haploinsuffisance de *SP3* :

- la délétion est *de novo* ;

- *SP3* interagit avec les histones déacétylases *HDAC1*, *HDAC2* et *HDAC4*, sachant que *HDAC4* est impliquée dans la DI (Liu et al. 2006).

En outre, le gène ne contient que 7 exons, ce qui nécessite donc relativement peu de travail de séquençage.

Nous avons donc séquencé les régions codantes de *SP3* (11 amplicons) par séquençage type Sanger (cf. 3.5.1.) chez 95 patients ayant une DI-NS, sélectionnés par le Centre de référence des déficiences intellectuelles de causes rares, afin de valider l'implication de ce gène dans le phénotype de la patiente. Mais ce séquençage n'a malheureusement pas retrouvé de variant pathogène. Ce résultat n'exclut pas que la délétion de *SP3* soit bien la cause de la DI chez la patiente ; mais il montre que des mutations de ce gène sont une cause probablement très rare de DI.

4.4.1.3. Autres gènes candidats retrouvés dans des délétions *de novo*

En dehors des délétions incluant les gènes *CAMTA1* et *SP3*, nous avons identifié une délétion 15q26.1 de 218 kb, de formule arr[hg19] 15q26.1(93,431,164-93,649,122)x1 dn, contenant les gènes *RGMA* et *CHD2* (figure 49).

Le patient délété (PSL_8467), âgé de 13 ans, a une DI modérée, avec hétéroagressivité. Il a présenté un retard de langage, mais pas de retard moteur. Il a par ailleurs une épilepsie, depuis l'âge de 3 ans, ainsi qu'une microcéphalie à -2DS.

Figure 49 : délétion de 218 kb de RGMA et CHD2 (patient PSL_8467)



Ce patient a un frère (PSL_8256) âgé de 8 ans qui, comme lui, a une DI modérée et a présenté un retard de langage sans retard moteur, mais qui en revanche n'a pas d'hétéroagressivité ni d'épilepsie. Son PC est à -1,5DS. La délétion n'a pas été retrouvée chez ce frère. Les 2 gènes délétés chez notre patient, *RGMA* et *CHD2*, sont intéressants et peuvent être considérés comme des gènes potentiellement candidats.

- *RGMA (RGM domain family, member A)* est un inhibiteur de la croissance axonale, aussi bien durant le développement qu'à l'âge adulte. Il permet de guider les axones sur des trajets spécifiques (Key et al. 2012).

- *CHD2* (*Chromodomain Helicase DNA-binding protein* 2) est un facteur de transcription ubiquitaire de la même famille que le gène *CHD7*, responsable du syndrome CHARGE. Des mutations dans ce gène ont été très récemment (juillet 2013) retrouvées chez environ 1% des patients présentant une encéphalopathie épileptique (Carvill et al. 2013).

Nous avons suspecté que cette délétion était responsable des signes présents chez le patient mais absents chez son frère, à savoir l'épilepsie et les troubles du comportement. Néanmoins, la sélection d'une cohorte de validation nous est apparue difficile, étant donné l'incertitude phénotypique. Avec les données récentes de la littérature (Carvill et al. 2013), il apparait désormais que cette délétion est au moins responsable de l'épilepsie.

4.4.2. Gènes candidats identifiés par des délétions transmises.

Dans le cas des CNV transmis contenant un gène candidat de DI, nous nous sommes placés dans l'hypothèse d'une transmission autosomique récessive avec mutation délétère sur l'allèle non délété (cf. partie 2.2.2.). Ainsi, pour 3 gènes autosomiques, nous avons effectué le séquençage de l'ensemble des régions codantes à la recherche d'une mutation sur l'autre allèle.

4.4.2.1. Séquençage du gène *STXBP6*

Nous avons retrouvé une délétion 14q12 de 161 kb, de formule arr[hg19] 14q12(25,349,607-25,510,923)x1 pat, chez un patient de 15 ans (PSL_7261) ayant une DI sévère, des TSA, des rires immotivés, et une épilepsie débutée à l'âge de 9 mois. Cette délétion inclut les exons 2 et 3 du gène *STXBP6 (Syntaxin Binding Protein 6)* et elle est transmise par le père bien portant (figure 50).

Figure 50 : délétion de 161 kb de *STXBP6* (patient PSL_7261).



Malgré le fait que la délétion est transmise par un parent sain, *STXBP6* nous est apparu comme un bon gène candidat pour plusieurs raisons :

- il est exprimé dans le cerveau (base de données Unigene) ;

- il code pour une protéine régulant le complexe SNARE qui participe à la régulation des vésicules présynaptiques (Scales et al. 2002) ;

- il est de la même famille que le gène *STXBP1* dont l'haploinsuffisance est responsable d'une DI sévère avec épilepsie à début précoce.

Nous avons donc séquencé, avec la méthode de Sanger, les régions codantes de *STXBP6* (5 exons, l'exon 1 n'étant pas traduit, soit 5 amplicons), chez notre patient (PSL_7261) afin de rechercher une mutation délétère sur l'allèle non délété. Ce séquençage n'a pas permis de retrouver de mutation délétère.

4.4.2.2. Séquençage du gène *CNTNAP4*

Nous avons retrouvé une délétion 16q23.1 de 916 kb, de formule arr[hg19] 16q23.1(75,766,089-76,682,263)x1 mat, chez une patiente de 11 ans (PSL_8059) ayant une DI-NS légère (figure 51). Cette délétion inclut le gène *CNTNAP4* (*Contactin Associated Protein-Like* 4) et elle est transmise par la mère.

Figure 51 : délétion de 916 kb incluant CNTNAP4 (patiente PSL_8059).



Bien que la délétion soit transmise par un parent bien portant, *CNTNAP4* nous a semblé être un bon gène candidat pour plusieurs raisons :

- il est exprimé préférentiellement dans le cerveau ;

- il est de la même famille que *CNTNAP2*, protéine synaptique ligand de la neurexine, dont les mutations sont impliquées dans la DI (Gregor et al. 2011) ;

Nous avons donc séquencé, en méthode de Sanger, les régions codantes de *CNTNAP4* (23 exons, 25 amplicons), chez le patient PSL_8059 mais nous n'avons pas détecté de mutation délétère sur l'allèle non délété.

4.4.2.3. Séquençage du gène *NUDT12*

Nous avons retrouvé une délétion 5q21.2 de 1,1 Mb, de formule arr[hg19] 5q21.1q21.2(102,694,113-103,794,722)x1 mat, chez une patiente de 11 ans (PSL_7699) ayant une DI-S modérée avec troubles du comportement à type d'hétéroagressivité, microcéphalie postnatale à -4DS, petite taille postnatale à -4DS et tétralogie de Fallot. Cette délétion inclut le gène *NUDT12 (nudix-type motif 12)* et elle est transmise par la mère (figure 52).

Figure 52 : délétion de 1,1 Mb incluant NUDT12 (patiente PSL_7699)



Nous avons séquencé, avec la méthode de Sanger, les régions codantes de *NUDT12* (7 amplicons) chez la patiente PSL_7699 dans l'hypothèse d'une mutation sur l'allèle non délété mais nous n'avons pas détecté de variant.

4.4.3. Gènes candidats identifiés par une délétion homozygote (ATAD3A et ATAD3B)

Nous avons identifié une délétion homozygote 1p36.33, de 10 kb, de formule arr[hg19] 1p36.33(1,404,810-1,415,099)x0 (figure 53), chez une patiente (PSL_11774) atteinte d'une DI légère (détectée lors de l'apprentissage de la lecture et de l'écriture), et présentant des troubles

de l'équilibre à type de dystonie et de syndrome cérébelleux, apparus à 13 ans et s'aggravant progressivement.

Cette délétion est retrouvée à l'état hétérozygote chez les parents bien portants, et n'est pas retrouvée chez son frère, lui aussi bien portant. Il est à remarquer que la patiente ne présente pas de grand LOH, signe de consanguinité.

Cette délétion emporte au minimum les 2 premiers exons du gène *ATAD3B* et le dernier exon du gène *ATAD3C (ATPase family, AAA domain containing 3)*. Les gènes *ATAD3A, ATAD3B* et *ATAD3C* sont des gènes paralogues, conservés chez les eucaryotes. *ATAD3A* et *ATAD3B* codent pour des protéines de la membrane mitochondriale qui stabilisent les "nucléoïdes" mitochondriaux, c'est-à-dire les complexes associant ADNmt et protéines. *ATAD3C* est un paralogue de *ATAD3A* et *ATAD3B* mais tronqué en N-terminal et en C-terminal (Li et al. 2011). *ATAD3A* et *ATAD3B* sont exprimés dans le cerveau (base de données Unigene).

Après discussion avec le Dr Fanny Mochel, (consultation des maladies métaboliques, GH Pitié-Salpêtrière, Paris) et le Dr Anne Lombes, (physiopathologie cellulaire et moléculaire des maladies mitochondriales, GH Pitié-Salpêtrière, Paris), le phénotype de la patiente est effectivement évocateur d'une maladie mitochondriale, et cette délétion est donc probablement en rapport avec la pathologie.

Nous avons alors décidé de séquencer les régions codantes de *ATAD3B* dans une série de 31 patients, sélectionnés par le Centre de référence des déficiences intellectuelles de causes rares, présentant un phénotype similaire à la patiente PSL_11774. Nous avons également séquencés *ATAD3A*, dans la même série étant donné sa fonction comparable avec *ATAD3B*.

Nous avons séquencé les exons 2 à 16 des deux gènes (15 amplicons par gène). Le séquençage des exons 1 n'a pas été réalisé étant donné la difficulté de mise au point liée à la proportion importante de GC et la forte homologie entre les 2 gènes. Ce séquençage n'a pas permis de retrouver de mutation délétère.

Figure 53 : délétion de 10 kb incluant *ATAD3B* et *ATAD3C* (patiente PSL_11774).



4.4.4. Gènes candidats identifiés par homozygosity mapping et séquençage de l'exome

L'*homozygosity mapping* permet de définir des régions candidates (2.4.4.), comprenant le plus souvent plusieurs dizaines, voire centaines de gènes. Lorsqu'un de ces gènes nous est apparu comme un très bon candidat, nous l'avons séquencé spécifiquement. Dans le cas contraire, devant l'impossibilité de séquencer, par méthode de Sanger, tous les gènes candidats, nous avons fait effectuer un séquençage de l'exome en NGS.

4.4.4.1. Famille de DI-S présentant un phénotype proche du Cockayne

Nous avons étudié le cas d'un frère et d'une soeur (PSL_9730 et PSL_9731), nés de parents consanguins, présentant la même pathologie très sévère :

- retard de croissance anténatal (RCIU), s'aggravant en post-natal ;

- difficultés alimentaires apparues à l'âge de 4 mois (les 3 premiers mois de vie ayant été sans particularité sur le plan neurologique) ;

- régression des acquisitions dans la petite enfance (perte de la station debout) ;
- microcéphalie post-natale, s'aggravant progressivement ;

- vieillissement prématuré ;
- énophtalmie ;
- lipodystrophie faciale et tronculaire ;
- neuropathie périphérique démyélinisante ;
- surdité de perception ;
- rétinopathie pigmentaire "poivre et sel" et opacités cornéennes ;
- lésions squelettiques épiphyso-métaphysaires ;
- brachydactylie et enraidissement des articulations des membres ;
- micropénis et cryptorchidie chez le garçon.

La sœur aînée est décédée à l'âge de 7 ans dans un tableau de choc septique. Le frère est âgé de 5 ans au moment de l'ACPA.

Les parents sont par ailleurs en bonne santé et ont un troisième enfant, également en bonne santé, ce qui évoque une pathologie autosomique récessive (figure 54).



Figure 54 : arbre de la famille dont le gène YWHAE a été séquencé

La première hypothèse a été celle d'un syndrome de Cockayne, malgré les lésions osseuses non classiques, l'absence de photosensibilité, l'IRM cérébrale normale et la dysmorphie un peu différente. Mais la réparation de l'ADN et le séquençage du gène CSB sont normaux. Le bilan métabolique est par ailleurs normal, ainsi que le caryotype et l'étude des gènes *LMNA* et *FACE1*.

L'étude en puce SNP n'a pas permis de mettre en évidence d'anomalie chromosomique responsable de la pathologie. Mais une étude d'homozygotie n'a retrouvé que 3 régions d'intérêt, pour un total de 12 Mb, communes aux 2 atteints, et hétérozygotes chez le frère sain, de formule chromosomique arr[hg19] 3p14.2p13(62,572,391-70,606,921)x2 hmz,11q23.3(116,139,120-118,740,864)x2 hmz,17p13.3(114,670-1,493,715)x2 (figure 55). Ces régions d'intérêt contiennent en tout 102 gènes dont aucun gène connu de Cockayne, ni de *Xeroderma Pigmentosum*.

4.4.4.1.1. Séquençage du gène YWHAE

Parmi les 102 gènes dans les régions d'intérêts, nous nous sommes intéressés à YWHAE (tyrosine 3-monooxygenase / tryptophan 5-monooxygenase activation protein, epsilon polypeptide).



Figure 55 : LOH commun aux 2 atteints, PSL_9730 et PSL_9731, contenant YWHAE

YWHAE est un gène de régulation de la transduction du signal intracellulaire. Il s'agit d'un des deux gènes critiques de la région du syndrome de Miller-Dieker : sa délétion est nécessaire,

en plus de celle de *PAFAH1B1* (qui est responsable de lissencéphalie) pour entrainer le phénotype complet de syndrome de Miller-Dieker (Cardoso et al. 2013). En outre, des délétions hétérozygotes 17p13.3 incluant, *YWHAE* mais pas *PAFAHB1*, sont responsables d'un phénotype de DI modérée avec retard de croissance staturo-pondérale (Schiff et al. 2010).

YWHAE nous a donc semblé être un gène candidat possible pour expliquer ce DI-S familial, sachant qu'on s'attend, en cas de mutation homozygote comme nous l'avons ici, à un phénotype plus sévère qu'avec des patients ayant une délétion hétérozygote comme on en trouve dans les cas rapportés jusqu'à présent.

Nous avons donc séquencé, en méthode de Sanger, les régions codantes de *YWHAE* (9 exons, 9 amplicons), chez le frère atteint (PSL_9731) mais n'avons trouvé aucune mutation délétère.

4.4.4.1.2. Séquençage de l'exome

Après avoir séquencé, sans succès, le gène *YWHAE* dans la famille précédente, nous avons fait effectuer un séquençage de l'exome chez le patient PSL_9731 au Centre National de Génotypage (CNG), en utilisant comme filtre principal, outre ceux décrits dans notre partie 2.5.3.3., une analyse limitée aux locus d'intérêt. Nous avons également éliminé les variants retrouvés chez d'autres individus du même run afin d'éviter les faux positifs.

Cette analyse a détecté 26 497 SNV (*Single Nucleotid Variants*) et 2 499 indel. Parmi ces variants, nous avons identifié une substitution nucléotidique homozygote G>A, en position 116633617, ce qui correspond à une mutation faux-sens dans l'exon 4 du gène *BUD13* (transcrit NM_032725.3 dans la base de données refseq), selon la nomenclature HGVS : cDNA : NM_032725.3:c.688C>T, ADN génomique : Chr11(GRCh37):g.116633617G>A, Protéine : p.Arg230*.

Ce variant est rare puisqu'il n'est pas répertorié dans dbSNP, 1000 genomes, HapMap ou *Exome Variant Server*.

Un séquençage selon la méthode Sanger a confirmé la mutation et déterminé que les parents et le frère sain sont hétérozygotes (figure 56).

Figure 56 : électrophorégrammes (logiciel Applied Biosystem Seqscape) montrant la mutation à l'état homozygote chez les 2 individus atteints (PSL_9730 et PSL_9731) et à l'état hétérozygote chez les parents et le frère sain (PSL_9728, PSL_9729 et PSL_10632).



BUD13 est exprimé sous la forme de deux isoformes d'après refseq (NM_032725 et NM_001159736). La mutation retrouvée dans l'exon 4 ne touche que la première isoforme, qui est l'isoforme majeure (d'après les données d'annotation Aceview).

4.4.4.1.3. Séquençage du gène *BUD13*

Afin de rechercher d'autres patients avec des mutations du gène *BUD13*, nous avons séquencé, en méthode de Sanger, les régions codantes de ce gène (10 exons, 10 amplicons), dans une série de 27 patients présentant un syndrome de Cockayne sans mutation dans le gène *ERCC6* adressés par le Pr Vincent Laugel (CHRU Strasbourg, France). Ce séquençage n'a pas détecté de mutation délétère dans cette série de patients.

4.4.4.1.4. Séquençage du gène *RBMX2*

Dans le cas de la levure, la protéine Bud13 fait partie du complexe de rétention du splicéosome, également appelé complexe RES. Ce complexe contient une autre protéine, Ist3, qui, lorsqu'elle est inactivée, produit le même phénotype sur la levure que l'inactivation de Bud13 (Dziembowski et al. 2004). Après discussion avec le Pr Bertrand Séraphin (IGBMC, Strasbourg, France), spécialisé dans l'étude de l'épissage chez la levure, nous avons donc décidé d'effectuer un séquençage du gène *RBMX2* (6 exons, 5 amplicons), qui est l'homologue humain de *Ist3* dans la série de patients avec syndrome de Cockayne sans mutation du gène ERCC6. Ce séquençage n'a pas détecté de mutation délétère.

4.4.4.2. Séquençage de l'exome dans une famille de DI-S présentant une agénésie du corps calleux (PSL_10908 et PSL_10909)

Il s'agit de 2 frères, l'un de 10 ans, l'autre de 9 mois, nés de parents consanguins, ayant une DI-S associée à une agénésie complète du corps calleux et une microcéphalie à -2DS (figure 57).

L'étude en puce SNP n'a pas permis de mettre en évidence d'anomalie chromosomique responsable de la pathologie. Mais une étude d'homozygotie n'a retrouvé que 3 régions d'intérêt pour un total de 5,7 Mb, communes aux 2 atteints, et non présentes chez la soeur et le frère sains, de formule chromosomique arr[hg19] 6q22.31(119,051,696-122,075,563)x2 hmz,

14q23.1(59,604,252-60,823,009)x2 hmz, 14q23.3q24.1(66,461,979-67,972,922)x2 hmz. Ces régions d'intérêt contiennent 26 gènes, dont aucun gène connu de DI ou d'agénésie du corps calleux.

Devant l'absence de gène particulièrement évocateur de DI, nous avons fait effectuer un séquençage de l'exome chez les patients PSL_10908 et PSL_10909 par la société Integragen, en utilisant comme filtre principal, outre ceux décrits dans la partie 2.5.3.3., une analyse limitée aux locus d'intérêt des variants présents chez les 2 frères. Nous avons également éliminé les variants ayant une fréquence > 1% dans la base de données interne d'Integragen.

Figure 57 : arbre de la famille ayant une DI-S et agénésie du corps calleux ; on a pratiqué sur les individus atteints un séquençage de l'exome



Cette analyse a détecté 30 378 SNV et 7 277 indel chez l'individu PSL_10908 ; ainsi que 60 234 SNV et 7 059 indels chez l'individu PSL_10909. Nous n'avons malheureusement retrouvé aucun variant potentiellement délétère dans les 26 gènes des régions sélectionnées. Une analyse des données d'exome sans focalisation sur les régions d'intérêt retrouve 205 variants rares, présents chez les 2 atteints : 184 compatibles avec un mode de transmission AD, 18 avec un mode de transmission AR (hétérozygotes composites) et 3 avec un mode lié à l'X. Aucun de ces variants ne provoque un stop prématuré. Pour le moment, nous n'avons donc pas identifié la mutation qui est responsable de la pathologie dans cette famille.
4.5. Caractérisation de points de cassure de translocation (article 5)

Nous avons caractérisé les points de cassure d'une translocation identifiée dans une famille avec deux sœurs atteintes de DI, épilepsie et microcéphalie. Ce travail est détaillé dans l'article 5.

Les deux sœurs, de 18 et 28 ans, présentaient à la fois une DI-S, une épilepsie, une microcéphalie à -4DS, et une petite taille à -5DS à début anténatal. Nous avons effectué une ACPA chez la plus jeune soeur (PSL_8584) et retrouvé une duplication 2p25 de 5,5 Mb, de formule arr[hg19] 2p25.3p25.2(72,185-5,571,360)x3, associée à une délétion 22q13.3 de 5 Mb, de formule arr[hg19] 22q13.31q13.33(46,094,759-51,169,045)x1 (figure 58).

L'analyse en FISH a déterminé qu'il s'agit d'un dérivé de translocation réciproque déséquilibrée, porté également par la deuxième soeur atteinte. Le père et une soeur saine âgée de 23 ans sont porteurs de la translocation réciproque sous sa forme équilibrée.

L'analyse en ACPA a révélé que le point de cassure sur le chromosome 22 coupait le gène *ATXN10*, dont l'expansion d'un quintuplet ATTCT, dans l'intron 9, est responsable de l'ataxie spino-cérébelleuse 10.

Figure 58 : duplication 2p25 de 5,5Mb associée à une délétion de 22q13.3 de 5Mb dont le point de cassure passe dans le gène *ATXN10* (patiente PSL_8584).



Il existe donc une disruption d'*ATXN10*, chez le père et la soeur porteurs de la translocation réciproque équilibrée, alors qu'ils ne présentent pas d'ataxie.

Nous avons voulu caractériser précisément les points de cassure par séquençage en méthode Sanger. Nous avons donc, dans un premier temps, réalisé des PCR dans les intervalles situés entre la dernière sonde normale et la dernière sonde anormale sur les chromosomes 2 et 22 (22 combinaisons d'amorces en tout). Puis nous avons séquencé les 2 fragments amplifiés.

Ce séquençage a permis de déterminer que le point de cassure sur le chromosome 22 était situé dans l'intron 2 de *ATXN10* à la coordonnée chr22:46087421-46087422. Sur le chromosome 2, la coordonnée du point de cassure est : chr2:5578252-5578253 (figure 59).





A noter la présence d'une insertion d'une base cytosine supplémentaire au niveau d'un des 2 points de cassure. Cette cytosine est normalement présente, sur le chromosome 2 normal et non sur le chromosome 22, or on la retrouve sur les 2 dérivés de translocation.

4.6. Résultats des patients présentant une DI et desTSA (article 6)

Parmi nos résultats, nous avons voulu analyser plus spécifiquement les données de CNV, et le travail de séquençage de gènes candidats effectué chez les patients de notre série présentant des TSA associés à la DI. Ce travail a été réalisé en collaboration avec le Dr Caroline Nava, qui travaille au Centre de recherche de l'institut du cerveau et de la moelle, UPMC/Inserm UMR-S975/CNRS UMR 7225 (GH Pitié-Salpêtrière, Paris). Les résultats de cette étude sont détaillés dans l'article 6 (Nava*, Keren* et al. 2013 ; * les 2 premiers auteurs ayant contribué à ce travail à part égale). Dans l'article, les CNV des patients ayant des TSA sans DI ou ayant été passés sur des puces Illumina Human 660W-Quad (hors diagnostic) ne font pas partie du sujet de ce travail et n'ont pas été inclus dans les statistiques présentées dans la partie 3.1.

Pour résumer, l'ACPA a retrouvé 114 CNV rares dont 41 pertes de copies et 73 gains de copies chez 160 patients DI et TSA. Parmi ces CNV, deux ont été considérés comme pathogènes en raison de leur taille, de leur caractère *de novo*, ou de leur implication connue dans les TSA et la DI : une triplication 15q11-q13 de 4,5 Mb, de formule arr[hg19] 15q11.2q13.1(21,234,149-25,713,937)x4 (PSL_8082) et une délétion 9p24 de 3,8 Mb, de formule arr[hg19] 9p24.3p24.2(317,782-4,084,321)x1 (PSL_8378).

Les différentes délétions hétérozygotes retrouvées ont permis d'identifier plusieurs gènes candidats :

- *DOCK8* et *DOCK10*, qui codent pour des protéines de signalisation cellulaire interagissant avec les Rho-GTPases, sont tous les deux exprimées dans le cerveau. Une disruption de *DOCK8*, due à une translocation réciproque équilibrée, a déjà été décrite dans une famille présentant DI et épilepsie (Griggs et al. 2007) ;

- *SMARCA2*, qui code pour un facteur de transcription appartenant au même complexe que la protéine FMRP codée par *FMR1*, qui est le gène de l'X fragile. Des mutations faux sens, n'inactivant pas totalement la fonction du gène, ont récemment été décrites chez des patients atteints du syndrome de Nicolaides-Baraitser ;

- *NYAP2*, qui code pour une protéine de signalisation neuronale (Yokoyama et al. 2011).

Dans l'hypothèse d'une pathologie de transmission AR, les patients ayant des délétions de ces gènes ont eu un séquençage de leur allèle non délété par la méthode de Sanger.

Une substitution nucléotidique G>A, faux sens, a été retrouvée dans le gène *DOCK10*, de formule HGVS cDNA : NM_014689.2:c.6460G>A, ADN génomique : Chr2(GRCh37):g.225630539C>T, protéine: p.Asp2154Asn. Ce variant est répertorié dans la base dbSNP avec une fréquence de 0,0004. Il est classé comme "*possibly damaging*" avec un score de 0,82, d'après l'algorithme de prédiction PolyPhen-2.

Aucune mutation délétère n'a été retrouvée dans les gènes DOCK8, SMARCA2 et NYAP2.

Nos effectifs dans cet article sont inférieurs à ceux présentés à la partie 4.1 car d'autres patients ont été analysés depuis.

5. Discussion

Dans cette partie, nous discuterons des résultats obtenus à partir de notre série de patients DI passés en ACPA. Nous analyserons d'abord l'intérêt que présente notre série sur le plan diagnostique, et nous verrons ensuite les nouveaux gènes candidats qu'elle nous a permis d'identifier.

5.1. Utilisation des puces SNP en diagnostic

5.1.1. Analyse globale des résultats

Notre série comporte en moyenne 1,63 patient de sexe masculin pour 1 patiente de sexe féminin, ce qui est concordant avec les études de prévalence montrant un excès de garçon (cf. 1.1.5.).

Sur les 1 187 patients pour lesquels nous avons effectué une ACPA en diagnostic, nous avons retrouvé 143 CNV pathogènes, ce qui représente environ 12% de diagnostics réalisés.

Ces résultats sont concordants avec ceux des grandes séries et revues de la littérature qui retrouvent un taux de diagnostics qui varie entre 5 à 20% (Stankiewicz et al. 2007, Keren (a), 2010, Miller et al. 2010, Cooper et al. 2011, Hochstenbach et al. 2011, Schaaf et al. 2011, Shoukier et al. 2013).

Concernant le taux diagnostique en fonction de la clinique, nous avons retrouvé significativement plus de CNV pathogènes chez les patients atteints de DI-S comparés aux patients DI-NS : 14,6% contre 6,6% ($p = 1,4.10^{-6}$ avec le test chi2). Il est classiquement admis que la probabilité de retrouver une anomalie chromosomique chez un patient DI est d'autant plus grande qu'il existe un nombre important de signes associés (De Vries et al. 2001), et nos

résultats sont en accord avec ce principe. Quant à nos patients présentant des TSA associés, ils présentent significativement moins de CNV pathogènes : 5% contre 12,9% pour nos autres patients DI (p=0,00025 avec le test chi2). Ces résultats concordent avec le taux de CNV *de novo* retrouvés dans de grandes séries de patients TSA, et qui varie de 5 à 7% (Itsara et al. 2010, Pinto et al. 2010, Levy et al. 2011, Sanders et al. 2011). Cette différence peut s'expliquer par le fait que, en dehors de l'épilepsie, les signes associés en particulier dysmorphiques et malformatifs, sont plus rares dans les TSA en général que dans la DI, retrouvés dans environ 20% des cas (Miles et al. 2005, Angkustsiri et al. 2011). Dans notre série, sur nos 256 patients présentant des TSA, seulement 8 ont été classés en DI-NS, ce qui explique probablement nos résultats.

5.1.2. Variation du nombre de CNV retrouvés en fonction du type de puce utilisé

Durant ce travail, nous avons utilisé deux types de puces SNP :

- les Illumina HumanCNV370-duo et HumanCNV370-quad (qui analysent les mêmes marqueurs SNP), de décembre 2007 à mars 2009 (153 patients) ;

- les Illumina HumanCytoSNP-12 d'avril 2009 à mai 2013 (1 034 patients);

Le nombre moyen de CNV identifiés est très différent selon qu'on utilise l'un ou l'autre de ces deux types de puces: 10,9 par patient pour les CNV370 contre 3,2 par patient pour les CytoSNP. Cette différence peut s'expliquer par le design des puces. Les CNV370 ont été conçues initialement pour les études GWAS, afin d'étudier plus particulièrement les risques associés à des variants communs (CNV ou SNP). Les CNV fréquents (polymorphiques) sont donc particulièrement bien couverts (source : *HumanCNV370 datasheet, Illumina*). Les puces HumanCytoSNP ont, au contraire, été mises au point pour le diagnostic cytogénétique afin de détecter préférentiellement les variants rares. La densité en sondes est donc plus importante sur des régions d'anomalies chromosomiques pathogènes connues et sur le chromosome X, tandis que certaines régions très polymorphiques, comme le locus des défensines en 8p ou le 9q proximal, ne sont pas couvertes (source : *InfiniumDx*TM *CytoSNP-12 Assay datasheet, Illumina*).

Par conséquent, il est logique que l'excès de CNV retrouvé avec les CNV370 soit lié à la densité en sondes au niveau des régions polymorphiques. Si nous nous intéressons uniquement aux CNV rares (non décrits dans DGV) et contenant des séquences codantes, nous en trouvons en moyenne 0,39 par patient avec les CNV370 et 0,56 CNV avec les CytoSNP. On peut donc en conclure que les CytoSNP sont plus performantes que les CNV370 pour la détection des CNV rares (p = 0,0001 avec le test chi2), malgré le fait qu'elles détectent globalement moins de CNV. Ce résultat démontre l'intérêt des CytoSNP dans une utilisation diagnostique où l'on s'intéresse prioritairement aux CNV rares et non aux CNV polymorphiques. De plus, étant donné que le nombre total de CNV retrouvés par les CytoSNP est inférieur à celui des CNV370, l'analyse et l'interprétation sont plus simples.

En ce qui concerne le nombre de diagnostics réalisés, un CNV pathogène a été retrouvé chez 9,8% des patients avec les CNV370, et 12,3% des patients avec les CytoSNP. Cette différence n'est pas significative (p = 0,36 avec le test chi2), ce qui est probablement dû au fait que la majorité des CNV pathogènes sont de grande taille, donc aussi bien détectés par les deux types de puces. Il faudrait probablement augmenter de beaucoup les effectifs pour que l'excès de petits CNV pathogènes retrouvés par les CytoSNP soit significatif.

5.1.3. Interprétation des CNV en fonction du type d'anomalie et de la taille

Lors de l'analyse des CNV, la taille et le type d'anomalie (gain ou perte) sont pris en compte pour la détermination du caractère pathogène. Plus le CNV est large, et plus la probabilité est grande que des régions génomiques de fonction critique, sensibles au nombre de copies, soient affectées, entraînant un retentissement phénotypique. Il est généralement admis que les duplications sont des altérations génomiques mieux tolérées que les délétions (Lee et al. 2007b, Mannings et al. 2010, Miller et al. 2010, South et al. 2011, Kearney et al. 2011).

Les résultats de notre série sont concordants avec ces principes : nous avons observé en effet une corrélation entre la taille du CNV, son type (gain ou perte) et la pathogénicité (Figure 60).

Nous avons retrouvé une taille moyenne de 6,5 Mb pour les délétions pathogènes, contre 11 Mb pour les duplications pathogènes, ce qui suggère qu'à taille égale une délétion a plus de chance d'être pathogène.

Figure 60 : *boxplots* de la taille des CNV en fonction de leur type et de leur interprétation. La partie gris clair représente la taille des CNV entre la médiane et le premier quartile, la partie gris foncé représente la taille des CNV entre la médiane et le troisième quartile. Les extrémités représentent la taille du CNV au 9ème percentile (en bas) et au 91ème percentile (en haut). Les CNV notés "VOUS" sont des CNV d'intérêt, contenant des séquences codantes et non présents dans DGV, pour lesquels nous n'avons pas pu vérifier le caractère transmis ou *de novo*.



Nous retrouvons cette différence lorsque nous regardons la fréquence des anomalies pathogènes en fonction de la taille minimale des CNV (figure 61).

Figure 61 : proportion de CNV pathogènes dans notre série en fonction de la taille minimale. Les pertes de copies sont en rouge et les gains de copies sont en bleu.



La figure 61 montre qu'à taille égale, la proportion de CNV pathogènes est plus importante dans les délétions. Le coefficient de détermination (en régression linéaire) est de $r^2 = 0,57$ pour les délétions et de $r^2 = 0,30$ pour les duplications.

Ces données confirment donc ce qui est classiquement admis : il y a d'autant plus de risque qu'un CNV soit pathogène qu'il s'agit d'une perte de nombre de copies et que sa taille est grande.

Quant à la taille minimale des CNV analysés, les recommandations internationales préconisent une résolution d'analyse d'au moins 400 kb (Miller et al. 2010, Kearney et al. 2011). Dans notre méthode d'analyse, nous ne fixons pas de seuil minimum de taille pour la détection des CNV : nous n'utilisons qu'un seuil de "*CNV confidence*" qui reflète la vraisemblance du CNV (cf. partie 3.2.4.2.6). Ainsi, nous avons pu détecter 6 CNV pathogènes, de taille inférieure à 400 kb, ce qui représente 0,5% de nos 1 087 patients, et nous a permis de détecter 6 CNVs pathogènes supplémentaires, soit une augmentation de notre taux diagnostique de 4,4%. Ces résultats nous encouragent donc à continuer d'obtenir la

meilleure résolution d'analyse possible sur les puces que nous utilisons, sans fixer artificiellement un seuil minimal de taille. Il faut toutefois noter qu'aucun de ces 6 CNV de petite taille ne correspond à un gain de nombre de copies, ce qui justifierait de ne continuer cette stratégie d'analyse que pour les délétions.

5.1.4. Apport des puces SNP en diagnostic du fait de la détection des disomies uniparentales

En diagnostic, les puces SNP ont comme avantage, comparées à la CGH array, qui reste la technique d'ACPA la plus utilisée, de pouvoir détecter les LOH, et donc les UPD, lorsqu'elles sont des isodisomies (les 2 chromosomes sont génotypiquement identiques).

Lorsque l'UPD survient dans une région soumise à l'empreinte, cela a pour conséquence des anomalies de l'expression d'un ou de plusieurs gènes, ce qui peut avoir des conséquences cliniques pathologiques (Butler et al. 2009).

C'est ce mécanisme qui explique le syndrome de Prader-Willi constaté chez la patiente PSL_8591 avec UPD maternelle, complète et homogène, du chromosome 15, que nous avons diagnostiqué en ACPA. La patiente, âgée de 17 ans au moment du diagnostic, présente un phénotype compatible avec ce syndrome : DI, obésité, petites extrémités, épilepsie, retard de croissance. L'interrogatoire de la famille n'avait toutefois pas permis de mettre en évidence d'hypotonie ni de difficultés d'alimentation en période néonatale, ce qui fait pourtant partie de l'histoire habituelle de la maladie. C'est la raison pour laquelle cette étiologie n'avait pas été évoquée avant l'ACPA et c'est finalement l'examen pangénomique qui a permis de poser le diagnostic. Selon notre expérience, le diagnostic de syndrome de Prader-Willi est, dans la grande majorité des cas, orienté cliniquement, et confirmé par des tests ciblés (étude de méthylation en première intention puis FISH). Depuis fin 2007, cette patiente est la seule patiente Prader-Willi diagnostiquée en ACPA dans notre laboratoire, alors que nous y avons dans le même temps diagnostiqué, en FISH, 15 patients atteints de syndrome de Prader-Willi. Le diagnostic de ce syndrome en ACPA est donc réservé aux formes atypiques, et l'excès de diagnostics permis par les puces SNP par rapport à la CGH array est réservé aux UPD, soit 25 à 30% des cas (Goldstone et al. 2004).

En ce qui concerne la patiente PSL_9378 avec une UPD maternelle, complète et homogène, du chromosome 9, nous avons suspecté un mécanisme différent d'une pathologie liée à l'empreinte. En effet, il n'existe pas de syndrome caractéristique connu causé par l'UPD maternelle du chromosome 9. Etant donné que cette UPD est composée à la fois d'isodisomie et d'hétérodisomie, nous avons posé l'hypothèse d'un mécanisme de formation par une correction de trisomie incomplète (Conlin et al. 2010). Nous avons donc effectué une recherche de mosaïque sur des fibroblastes de la patiente et retrouvé une trisomie 9 dans 4% des cellules (figure 62). Des disomies maternelles du 9 associées à une trisomie en mosaïque ont déjà été décrites en pré et postnatal (Willatt et al. 1992, Slater et al. 2000); mais dans ces cas l'UPD avait été suspectée suite à la détection de l'UPD. Sur le plan clinique, notre patiente présente un phénotype très proche du cas décrit par l'équipe de Willatt, ayant une trisomie 9 dans 6% des cellules, en particulier la dysmorphie avec un important microrétrognathisme, et une forte scoliose. Nous avons donc considéré cette anomalie chromosomique comme causale.

Figure 62 : FISH avec la sonde centromérique du chromosome 9, de la patiente PSL_9378 ayant une UPD du 9, retrouvant une trisomie 9 dans 4% des fibroblastes



Les deux précédentes UPD, du chromosome 15 et du chromosome 9, détectées en ACPA, nous ont permis de faire un diagnostic étiologique des troubles présentés par ces patients, contrairement aux trois autres UPD que nous avons détectées.

Nos deux patients porteurs d'une iUPD paternelle, complète et homogène, du chromosome 16 ont des phénotypes distincts. Tous les deux ont une DI et un retard de croissance, mais la DI est beaucoup plus sévère pour la patiente PSL_10170 que pour le patient PSL_7798. Le patient PSL_7798 a par ailleurs une ataxie cérébelleuse, une cardiomyopathie hypertrophique et une épilepsie, tandis que la patiente PSL_10170 a une arthrogrypose, une surdité de perception et un syndrome de Pierre Robin. De plus, leur dysmorphie est très différente. Les UPD du chromosome 16 sont parmi les UPD les plus fréquentes (Kotzot et al. 1999), mais elles sont d'origine maternelle dans la quasi-totalité des cas. Un seul cas d'UPD du 16 d'origine paternelle, diagnostiqué de façon fortuite, a été rapporté dans la littérature chez une enfant de 13 mois bien portante (Kohlhase et al. 2000). Nous avons conclu de ces données de la littérature, et de la constatation que nos deux patients présentaient un phénotype distinct, que leur pathologie n'était pas liée à une anomalie de l'empreinte. Etant donné que ces 2 UPD sont composées totalement d'isodisomie, nous avons posé l'hypothèse d'un mécanisme de formation par une correction incomplète de monosomie (Conlin et al. 2010). Mais la recherche, en FISH, de monosomie en mosaïque sur fibroblastes a été négative sur 200 cellules pour nos 2 patients. Au final, nous considérons actuellement ces 2 UPD du 16 comme des variants de signification inconnue. Il se peut que ces UPD aient permis l'expression de deux maladies AR différentes liées à des mutations délétères sur un chromosome 16 paternel, se retrouvant à l'état homozygote dans l'UPD. Cela nécessiterait de réaliser un séquençage exonique et d'analyser spécifiquement les gènes du chromosome 16.

Notre dernière UPD est une iUPD paternelle, segmentaire et en mosaïque, de la partie terminale du bras court du chromosome 11. Cette anomalie est connue pour être responsable du syndrome de Wiedemann-Beckwith (Romanelli et al. 2011). Or notre patiente n'a aucun signe de Wiedemann-Beckwith, pas de macrosomie, d'avance staturale, de macroglossie, de viscéromégalie, d'anomalie de morphologique des oreilles, d'anomalie de la paroi abdominale, etc. De plus, une mutation *de novo* du gène *TUBA1A* a, par la suite, été mise en évidence chez cette patiente, ce qui concorde tout à fait avec son phénotype et en particulier ses anomalies visibles à l'IRM cérébrale. Nous pensons donc que c'est cette mutation qui est responsable du phénotype et non l'UPD. L'absence de manifestation phénotypique de cette UPD peut s'expliquer par sa période d'apparition. Les anomalies en mosaïque peuvent survenir après la conception, tout au long de la vie. Et les conséquences phénotypiques sont différentes selon que l'anomalie survient durant le développement ou après le développement. Certaines études

en ACPA réalisées dans la population générale ont montré l'apparition, au cours de la vie, d'UPD en mosaïque, dont des UPD 11pter, sans que ces patients n'aient de syndrome de Wiedemann-Beckwith (Jacobs et al. 2012, Laurie et al. 2012). Nous pensons que c'est également le cas pour notre patiente ; l'apparition tardive au cours du développement de son UPD pourrait être la raison de l'absence de syndrome de Wiedemann-Beckwith. A noter que, si nous avons conclu que cette UPD n'était pas en rapport avec le phénotype de la patiente, elle a tout de même démontré que les puces SNP permettaient de diagnostiquer le syndrome de Wiedemann-Beckwith, ainsi que d'évaluer la taille et le taux de mosaïque de l'UPD, ce qui nous a conduit à collaborer à divers travaux sur les disomies uniparentales avec l'équipe d'Irène Netchine, Sylvie Rossignol et Yves Le Bouc (Hôpital Trousseau, Paris, France), (Keren (b) et al. 2013, Maupetit-Méhouas et al. 2013).

Au total, sur nos 1 187 patients, nous avons identifié 5 UPD (0,42%) dont seulement deux nous ont permis de faire un diagnostic étiologique (0,17%). Cette fréquence est environ 15 fois supérieure à celle des UPD estimée dans la population générale (Robinson et al. 2000), sachant par ailleurs que nous ne détectons pas, en puces SNP, les hétérodisomies complètes. Dans la littérature, les taux d'UPD retrouvées en puces SNP, dans des séries de patients adressés pour ACPA diagnostique, sont comparables : de 0,22 à 0,4% pour le taux d'UPD total et de 0,14 à 0,24% pour le taux d'UPD de pathogénicité certaine (tableau 33).

| Etude | effectif | UPD prouvées, | taux d'UPD, |
|-------------------------|----------|--------------------|--------------------|
| | total | totales/pathogènes | totales/pathogènes |
| Conlin et al. 2010 | 2 019 | 8/5 | 0,40% / 0,24% |
| Bruno et al. 2011 | 5 000 | 11 / 7 | 0,22% / 0,14% |
| Papenhausen et al. 2011 | 13 000 | 38 / 21 | 0,29% / 0,16% |

Dans notre série, la détection des UPD permet de faire passer le nombre de diagnostics de 143 à 145, soit une augmentation de 1,4%. Cette augmentation du taux diagnostique grâce à la détection des UPD d'une part, et d'autre part à la détection des régions homozygotes en cas de consanguinité, fait préférer les puces SNP, "lorsque cela est possible", à la CGH array, d'après les recommandations du Collège Américain de Génétique Médicale (Kearney et al. 2010). Selon notre expérience et d'après les données de la littérature, l'augmentation du taux diagnostique est toutefois relativement faible et peut ne pas justifier les efforts nécessaires au changement de plateforme pour les laboratoires utilisant la CGH array en routine et ne maîtrisant pas les puces SNP.

5.1.5. Interprétation des CNV considérés comme des facteurs de susceptibilité

Dans notre série, nous avons retrouvé 25 patients porteurs de CNV connus comme étant des facteurs de susceptibilité. L'interprétation de ce type de CNV demeure encore difficile et représente un raisonnement totalement nouveau pour les cytogénéticiens, habitués à des remaniements, soit pathogènes, soit polymorphiques.

Premièrement, il est mal aisé de classer ces remaniements selon les catégories que nous avons initialement établies dans notre base de données (pathogène, non pathogène, transmis, non vérifié) ou qui ont été établies par l'ISCA (pathogène, bénin, probablement pathogène, probablement bénin, VOUS).

Dans notre base de données, un même variant a pu être classé de façon différente, selon les patients, en fonction de son caractère *de novo* ou transmis, mais également en fonction des données de la littérature au moment de l'identification du CNV. Par exemple, nous avons identifié 7 duplications 16p11.2, dont 4 classées comme pathogènes, 1 non vérifiée (donc VOUS), et 2 transmises par un parent sain (donc "probablement bénignes"). Pourtant, l'effet physiopathologique de cette duplication n'est probablement pas différent, qu'elle soit *de novo* ou transmise ou qu'elle ait été identifiée au début ou à la fin de notre étude : il s'agit d'un facteur de susceptibilité, impliqué dans la DI, mais insuffisant à lui seul pour la provoquer.

Il se pose un autre problème qui est celui des facteurs de susceptibilité qui ne sont pas encore caractérisés comme tels dans la littérature. De ce fait, il n'est pas actuellement possible de les identifier en utilisant notre arbre décisionnel, et l'anomalie, selon son caractère *de novo* ou transmis, sera interprétée comme l'étiologie de la maladie ou comme probablement bénigne, alors que l'on est sans doute devant une cause complexe et multifactorielle. En outre, étant donné que ces facteurs de susceptibilité peuvent être portés par des individus en bonne santé, ils peuvent être répertoriés dans DGV comme variants polymorphiques et être d'emblée classés comme bénins.

En effet, le problème principal est celui de la spécificité et de la valeur prédictive de ces anomalies en diagnostic : lorsqu'on identifie un facteur de susceptibilité chez un patient DI, il nous est actuellement impossible de déterminer si ce CNV est bien en rapport avec le phénotype du patient ou s'il est le résultat d'une découverte fortuite, de sorte que le conseil génétique chez les apparentés est difficile. D'ailleurs, lorsqu'on découvre actuellement un de ces CNV en prénatal ou en néonatal, il est impossible de prévoir si cela entrainera une DI.

Il demeure un dernier problème, celui de l'information à transmettre par le cytogénéticien au prescripteur. L'intérêt clinique de ces réarrangements varie en fonction du risque : certains CNV ont un Odds Ratio > 10 comme la délétion 1q21.1 ; d'autres ont un Odds Ratio < 2 comme la délétion proximale 15q11.2 entre BP1 et BP2, ce qui n'est pas très différent de l'Odds Ratio du sexe masculin dans la DI (Malhotra et al. 2012). L'objectif serait donc de fixer un seuil à partir duquel l'augmentation du risque et de la valeur prédictive est suffisante pour que le patient et la famille en soient informés, mais il n'existe pour le moment aucune recommandation sur ce sujet. Pour notre part, nous avons choisi de ne pas rendre les CNV considérés comme polymorphiques dans DGV, même s'ils ont par ailleurs été décrits comme des facteurs de risque dans certaines études, ces CNV correspondant à des Odds Ratio faibles, telle la délétion proximale 15q11.2 (Cooper et al. 2011, Malhotra et al. 2012). Quant aux autres CNV, nous les avons considérés, dans le compte-rendu remis aux prescripteurs, comme des facteurs de risque, même si cette catégorie n'existe pas pour le moment dans notre base de données. Notre attitude sera amenée à évoluer en même temps que les connaissances sur ces facteurs de risque et d'éventuelles recommandations quant à leur interprétation.

5.2. Identification de gènes candidats de DI

Ce travail de recherche de CNV en ACPA diagnostique chez des patients atteints de DI, nous a permis d'identifier plusieurs gènes candidats (tableau 34) que nous avons cherchés à valider selon plusieurs stratégies (figure 63).

Figure 63 : stratégies d'identification et de validation des différents gènes candidats de DI de notre série.



Tableau 34 : gènes candidats identifiés, avec le processus physiologique dans lequel ils sont impliqués, ainsi que les gènes de la même famille impliqués dans la DI

| gènes | processus physiologique | commentaires |
|---------|-------------------------------|--|
| CAMTA1 | facteur de transcription | - |
| SP3 | facteur de transcription | - |
| CNTNAP4 | organisation de la densité | famille de CNTNAP2, impliqué dans la DI, les TSA |
| | présynaptique. | et le syndrome de Tourette. |
| | régulation de la | |
| NUDT12 | concentration des | - |
| | nucléotides intracellulaires. | |
| STXBP6 | fonction prósynantique | famille de STXBP1, impliqué dans une |
| | ionction presynaptique. | encéphalopathie épileptique. |

| DOCK8/DOCK10 | signalisation cellulaire. | DOCK8 impliqué dans une famille de DI et épilepsie. |
|----------------------|---|--|
| SMARCA2 | facteur de transcription. | même complexe protéique que <i>FMR1</i> , gène de l'X fragile. Impliqué dans le syndrome de Nicolaides- Baraitser. |
| NYAP2 | signalisation neuronale. | - |
| ATAD3A/ATAD3B/ATAD3C | stabilisation des nucléoïdes mitochondriaux. | - |
| YWHAE | signalisation cellulaire. | gène critique du syndrome de Miller-Dieker. |
| BUD13 | rétention des ARNm dans le splicéosome. | - |
| RBMX2 | rétention des ARNm dans le splicéosome. | même complexe protéique que BUD13. |
| ATXN10 | survie et différenciation neuronale. | gène impliqué dans l'ataxie spino-cérébelleuse 10 (SCA10). Cassure identifiée de façon fortuite dans une famille de DI, chez les non atteints. |

5.2.1. Identification de gènes candidats par les petites délétions de novo

Deux petites délétions *de novo* nous ont permis d'identifier 2 gènes candidats de DI : *CAMTA1* et *SP3* qui étaient les seuls gènes délétés. Le fait que des CNV soient *de novo* est reconnu comme un argument fort de pathogénicité (Lee et al. 2007, Kearney et al. 2011). Toutefois, ce n'est pas un argument absolu, les CNV *de novo* survenant dans la population générale avec un taux, à chaque génération, de 1,2 / 100 génome (sur des puces de 30 kb de résolution), et la plupart du temps sans conséquence pathologique (Itsara et al. 2010). Il est donc nécessaire, une fois un gène candidat identifié, de répliquer ce résultat chez d'autres patients. Afin de valider la responsabilité de ces 2 gènes dans la DI, nous avons réuni pour chacun une cohorte de réplication composée de patients atteints du même phénotype : 47 patients atteints d'ataxie congénitale +/-DI pour *CAMTA1* et 95 patients atteints de DI-NS pour *SP3*. Mais ce séquençage ne nous a pas permis de mettre en évidence de mutation délétère.

Nous considérons tout de même la délétion de *CAMTA1* comme pathogène car, du fait de notre collaboration avec l'équipe du CHU Dijon, deux autres familles ont été identifiées avec des réarrangements intragéniques de *CAMTA1* entraînant un stop prématuré, qui ségrégent avec la DI des patients. En raison du phénotype de notre patiente ayant un retard de développement associé à une ataxie congénitale, les individus atteints des deux autres familles ont été réexaminés et une ataxie a été mise secondairement en évidence. Par ailleurs un

patient ayant une délétion de *CAMTA1* a été rapportée dans la littérature, avec, lui aussi, un phénotype de DI et d'ataxie cérébelleuse (Mikhail et al. 2011). Ainsi une délétion de *CAMTA1* a été retrouvée dans quatre familles indépendantes avec un phénotype homogène de DI et d'ataxie, ce qui est de notre point de vue un argument majeur pour impliquer *CAMTA1* dans cette pathologie. De plus, d'autres arguments mineurs vont également dans le sens d'une pathogénicité de la délétion de ce gène : il est exprimé préférentiellement dans le cortex temporal et le cervelet, ce qui concorde avec le phénotype de DI et d'ataxie (Huentelman et al. 2011), et aucune mutation provoquant un stop prématuré n'est décrite dans la base *Exome Variant Server* sur plus de 6 500 individus contrôles. Au final, nous considérons donc que l'haploinsuffisance de *CAMTA1* est bien responsable d'une pathologie associant une DI et une ataxie congénitale, même si les mutations ponctuelles dans ce gène sont probablement rares.

Concernant la délétion de *SP3* que nous avons retrouvée chez une patiente DI, il est possible que l'absence de mutation délétère dans notre cohorte de réplication soit liée au fait que les mutations de ce gène représentent une cause très rare de DI. Plusieurs arguments semblent confirmer cette hypothèse : il s'agit d'un facteur de transcription interagissant avec *HDAC4*, lui même impliqué dans la DI, et aucune mutation provoquant un stop prématuré n'est décrite chez 6 500 contrôles dans *Exome Variant Server*, ce qui est en faveur de l'importance physiologique de la protéine. Pour confirmer notre hypothèse, le séquençage d'une plus grande cohorte de réplication serait nécessaire, car, en l'absence d'un deuxième cas avec haploinsuffisance de *SP3*, nous ne pouvons pas non plus exclure le fait que cette délétion pourrait être un variant polymorphique, survenu *de novo* de manière fortuite, chez cette patiente DI.

Enfin, nous avons retrouvé une délétion *de novo* emportant les gènes *RGMA* et *CHD2* chez un patient atteint de DI, troubles du comportement, épilepsie et microcéphalie à -2DS. Ces deux gènes sont de très bons candidats potentiels dans la DI, *RGMA* étant un répresseur de la guidance axonale et *CHD2* un facteur de transcription de la même famille que *CHD7*, le gène du syndrome CHARGE. Toutefois, le fait que son frère, également atteint de DI n'ait pas cette délétion pose un problème dans la ségrégation de l'anomalie avec la pathologie. Il nous a donc paru difficile de rattacher de façon claire un phénotype à cette délétion, et ainsi de recruter une cohorte de réplication pour le séquençage. En conséquence, nous ne sommes pas allés plus loin dans l'étude de ces deux gènes. Très récemment (juillet 2013), l'équipe de Heather

Mefford a retrouvé des mutations hétérozygotes délétères *de novo* du gène *CHD2* chez environ 1% des patients atteints d'encéphalopathie épileptique (Carvill et al. 2013). Cela confirme que la délétion de notre patient est bien pathogène et probablement responsable de l'épilepsie, même si elle n'explique pas la totalité du phénotype.

Les petites délétions *de novo* retrouvées en ACPA ont déjà permis d'identifier plusieurs gènes de DI comme *CHD7*, *MBD5*, ou *PCDH19*. Dans notre étude, sur trois petites délétions *de novo*, contenant un ou deux gènes antérieurement non décrits comme responsables de DI, nous avons identifié un nouveau gène de DI avec ataxie congénitale : *CAMTA1*. Le caractère *de novo* demeure donc un marqueur important de pathogénicité dans les DI sporadiques. Et il est désormais utilisé dans des études de séquençage de l'exome qui recherchent des mutations ponctuelles (SNV ou indel) *de novo* plutôt que des CNV. Vissers et al. (2010) ont séquencé l'exome de dix patients ainsi que de leurs parents afin détecter des mutations *de novo* : cette étude leur a permis d'identifier neuf variants *de novo*, dont six qu'ils ont considérés comme probablement pathogènes, en raison du type de mutation et de la fonction du gène touché. Cette méthode a également été utilisée dans d'autres pathologies neuropsychiatriques comme les TSA (O'Roak et al. 2011) ou la schizophrénie (Girard et al. 2011) et, chaque fois, elle a permis d'identifier des variants probablement pathogènes et de bons gènes candidats. Néanmoins, dans tous les cas, demeure malgré tout la difficulté de répliquer ces données dans des cohortes de patients de même phénotype, étant donné la rareté de ces variants.

5.2.2. Identification de gènes candidats par les petites délétions transmises

Les petites délétions transmises nous ont permis d'identifier 7 gènes candidats de DI : *CNTNAP4, NUDT12, STXBP6, DOCK8, DOCK10, SMARCA2, NYAP2.* Nous avons sélectionné ces gènes parce, pour leur majorité, ils interviennent sur des processus physiopathologiques déjà impliqués dans la DI (cf. 1.2.3.2.) et, pour certains, ils possèdent d'autres gènes de leur famille responsables de DI (cf. tableau 34). Pour ces gènes, nous avons posé l'hypothèse d'une possible mutation ponctuelle sur l'allèle non délété. Cela aurait pu expliquer, dans le cas d'une transmission autosomique récessive, que le parent délété soit bien portant alors que le cas index est atteint. Mais le séquençage des séquences codantes de ces gènes chez les patients délétés n'a pas permis de mettre en évidence de mutation délétère. A partir de ces résultats, nous pouvons émettre, pour chaque gène, différentes hypothèses :

- la délétion du gène est un polymorphisme rare, et n'est pas impliquée dans la pathologie du patient ;

- l'allèle restant est inactivé mais la cause de cette inactivation n'a pas été mise en évidence par le séquençage des séquences codantes (par exemple délétion ou mutation dans une séquence de régulation, ou épimutation de l'allèle restant) ;

- la délétion est bien impliquée dans la DI des patients, mais n'est pas suffisante, à elle seule, pour que la pathologie s'exprime. Les patients ont dû être soumis à d'autres facteurs, génétiques ou environnementaux, non présents chez le parent transmetteur, pour que la pathologie s'exprime. On est alors dans le cadre de facteurs de susceptibilité, de digénisme ou multigénisme. Un nombre croissant de pathologies digéniques, n'obéissant pas à une hérédité mendélienne monogénique, est désormais décrit, et dans ce cas, apporter la preuve de la pathogénicité d'une mutation est très complexe (Schaffer et al. 2013).

Etant donné que nous continuons d'effectuer en prospectif de l'ACPA diagnostique chez des patients DI, nous espérons identifier d'autres patients ayant une délétion de ces gènes afin d'avoir des arguments supplémentaires quant à leur pathogénicité. A terme, une base de données nationale des CNV retrouvés en ACPA va être mise en place par le réseau national d'analyse chromosomique sur puce ADN (réseau Achro-Puce). Cette base de données pourra nous aider à identifier certains patients analysés dans d'autres laboratoires, porteurs des CNV candidats que nous avons identifiés.

5.2.3. Identification de gènes candidats par les petites délétions homozygotes

Nous avons identifié une délétion homozygote des gènes *ATAD3B* et *ATAD3C* chez une patiente atteinte de DI légère et d'un syndrome cérébelleux s'aggravant progressivement.

ATAD3A est un gène très conservé au cours de l'évolution : en effet, il existe chez les plantes et les nématodes. Il s'agit d'une ATPase située dans la membrane mitochondriale et ayant un rôle critique dans l'organisation des nucléoïdes mitochondriaux, c'est-à-dire des structures

associant protéines et ADNmt. Les gènes *ATAD3B* et *ATAD3C* sont des paralogues d'*ATAD3A* apparus chez les primates, au cours de l'évolution (Li et al. 2011). *ATAD3B* a 93% d'homologie avec *ATAD3A*, et comme ce dernier, il est exprimé dans le cerveau et semble avoir une fonction comparable. *ATAD3C* code pour une protéine tronquée en N-terminal, et dont la séquence C-terminale est également différente ; sa fonction n'est pas connue, et il n'existe aucune donnée sur son expression tissulaire (base de données Unigene).

Le phénotype de la patiente nous est apparu comme compatible avec une maladie mitochondriale ; nous avons donc posé l'hypothèse de la pathogénicité de cette délétion homozygote, avec un mode de transmission AR. Et comme pour *SP3* et *CAMTA1*, nous avons voulu répliquer ce résultat en recherchant des mutations dans une cohorte de patients de même phénotype. Nous n'avons séquencé qu'*ATAD3A* et *ATAD3B*, étant donné l'absence de données sur *ATAD3C*, et n'avons trouvé aucune mutation chez 24 patients ayant une DI et une ataxie progressive.

Dans le même temps, le Dr Lombes (Physiopathologie cellulaire et moléculaire des maladies mitochondriales, GH Pitié-Salpêtrière, Paris) a réalisé des études morphologiques sur les fibroblastes du patient et a retrouvé un compartiment mitochondrial très anormal, pauvre et fragmenté (figure 64).

Figure 64 : étude du compartiment mitochondrial avec Ac anti MtCO2 en vert. Ce compartiment apparait beaucoup plus pauvre et fragmenté chez le patient que chez le contrôle (images fournies par le Dr A. Lombes).





Etant donné ces résultats, nous avons toujours une forte suspicion de l'implication de cette délétion dans le phénotype du patient, mais en l'absence de réplication chez d'autres patients, nous ne l'avons pas démontrée. Il est possible que le phénotype soit lié à l'absence d'*ATAD3C*, ou à l'absence concomitante d'*ATAD3B* et *ATAD3C* ; dans ce cas, la probabilité de retrouver des mutations ponctuelles responsables de la pathologie est faible. Il est toutefois également possible que cette délétion soit responsable des anomalies morphologiques mais pas du phénotype clinique. Désormais, pour avancer dans l'étude de la pathogénicité de cette anomalie, notre objectif est d'identifier, en prospectif, un second patient délété pour ces 2 gènes, dans le cadre de l'ACPA diagnostique effectuée en routine dans le laboratoire.

5.2.4. Identification de gènes candidats par homozygosity mapping

L'*homozygosity mapping* réalisé avec les puces SNP nous a permis de définir des locus d'intérêt, où se trouve probablement la mutation homozygote causale de 2 familles consanguines, dans l'hypothèse d'une transmission AR.

Dans la première famille, le séquençage de l'exome des 2 individus atteints, nous a permis d'identifier une mutation homozygote du gène *BUD13* chez les 2 atteints, présente à l'état hétérozygote chez les parents et le frère sains.

Plusieurs arguments nous portent à croire que cette mutation est délétère :

- il s'agit d'une mutation non-sens, entraînant probablement une dégradation du transcrit par le système NMD et donc une absence d'expression de la protéine ou une expression très réduite de la protéine tronquée ;

- cette mutation est privée, non répertoriée, y compris à l'état hétérozygote, dans les bases de données dbSNP, HapMap, 1000 genomes et *Exome Variant Server* ;

-

- aucune mutation entraînant un codon stop prématuré dans *BUD13* n'est répertoriée dans la base de données *Exome Variant Server* chez 6 500 individus contrôles.

BUD13 est un gène très conservé depuis la levure, donc ayant probablement un rôle physiologique important ;

Aucune étude fonctionnelle de *BUD13* n'a été réalisée chez l'Homme, sa fonction est donc mal connue. Mais son homologue chez la levure est une protéine du complexe du splicéosome, et donc impliqué dans l'épissage des préARN messagers (Dziembowski et al. 2004).

Le splicéosome est un complexe nucléaire formé de ribonucléoprotéines (snRNP), qui sont composées de protéines associées à de petits ARN (U1, U2, U4, U5 et U6), et de protéines non snRNP (Jurica et al. 2003). Le splicéosome s'assemble sur les préARNm pour effectuer l'épissage, comprenant l'excision des séquences non codantes (introns) et la ligation des séquences codantes (exons), il en résulte un ARNm mature. Le splicéosome est localisé dans le noyau, au niveau des granules interchromatiniennes (ou *"speckles"* en anglais).

Chez la levure, *BUD13* code pour une protéine non snRNP du splicéosome qui appartient au complexe de rétention (complexe RES pour *REtention and Splicing*) qui permet de maintenir les préARNm dans le splicéosome. Son inactivation entraîne un défaut global de l'épissage avec accumulation des préARNm et diminution des ARNm matures. Les levures, dont notre gène d'intérêt est inactivé, sont viables mais ont un phénotype anormal avec croissance diminuée. Par ailleurs, les mutations responsables du syndrome de Cockayne entraînent également une perturbation de la transcription des ARNm, ce qui pourrait expliquer certains signes cliniques communs à nos deux patients (Cleaver et al. 2013).

Quelques pathologies humaines constitutionnelles sont connues pour être causées par des mutations dans des gènes codant pour des protéines du splicéosome. Il s'agit de certaines rétinopathies pigmentaires (rappelons que nos patients ont également une rétinopathie pigmentaire) : *RP9, RP11, RP13, RP18* et *RP33* (Linder et al. 2011) ; et d'une hypotrichose liée à des mutations du gène *SNRPE* (Pasternack et al. 2012). Dans ces derniers cas, contrairement à la famille que nous étudions, le mode de transmission est de type autosomique dominant, et les mutations sont de type faux-sens, n'altérant que de façon partielle la fonction des protéines. Seule une altération partielle de l'épissage de gènes exprimés dans la rétine a été retrouvée lorsque certains des gènes de rétinite pigmentaire sont inactivés. Le nanisme microcéphalique ostéodysplasique primaire de type I (MOPD1) est

également dû à des mutations dans un gène impliqué dans l'épissage. Il existe d'ailleurs des points communs entre le phénotype des patients de notre famille et celui des patients MOPD1 : RCIU, microcéphalie, difficultés alimentaires, aspect vieilli de la peau, brachydactylie et limitations articulaires. Mais le tableau est plus sévère dès la période néonatale dans le MOPD1 et la dysmorphie est très différente de celle de nos patients. Cette pathologie est causée par des mutations homozygotes ou hétérozygotes composites faux-sens du gène *RNU4ATAC* (Edery et al. 2011), dont la protéine appartient au complexe mineur du splicéosome (responsable de l'épissage d'un type d'intron particulier : les introns de type U12). Nous avons vérifié l'absence de mutation de ce gène chez nos patients sur nos données d'exome. Enfin, un rôle dans la régulation de l'épissage est suspecté pour *FMR2*, gène responsable du retard mental lié au locus FRAXE. Ce gène est localisé dans les granules interchromatiniennes et son inactivation *in vitro* réduit l'activité des ESE (*exonic splicing enhancer*), ce qui pourrait provoquer des défauts de l'épissage alternatif (Bensaid et al. 2009).

Comme pour *CAMTA1*, *SP3* et *ATAD3*, une réplication de ces résultats est nécessaire pour démontrer l'implication de *BUD13* dans la pathologie de nos patients. Nous avons donc séquencé *BUD13* dans une cohorte de 27 patients Cockayne, non mutés dans *ERCC6*, étant donné certaines similitudes phénotypiques de nos patients avec ce syndrome (régression des acquisitions, vieillissement prématuré, microcéphalie, énophtalmie, rétinite pigmentaire et opacités cornéennes), bien qu'ils aient également quelques discordances (anomalies squelettiques, dysmorphie différente, absence de photosensibilité). Nous avons séquencé aussi dans cette cohorte le gène *RBMX2* qui est l'homologue d'un gène de la levure, *Ist3*, qui appartient également au complexe RES et dont l'inactivation a pour conséquence le même phénotype que l'inactivation de *BUD13*, c'est-à-dire une croissance diminuée. Mais nous n'avons retrouvé aucune mutation délétère, ni dans *BUD13*, ni dans *RBMX2*.

Au final, nous avons donc cette forte suspicion de pathogénicité pour cette mutation non-sens de *BUD13*, sans avoir toutefois réussi à la démontrer. Afin d'avancer, nous proposons :

- d'inclure d'autres patients ayant le même phénotype que ceux de nos individus atteints (et non plus un phénotype de Cockayne) et de séquencer *BUD13* et *RMBX2*. Dans ce but, une présentation des patients sera faite lors d'une réunion nationale de génétique clinique, par la clinicienne qui les suit, le Dr Odile Boute (CHRU Lille, France) ;

- d'étudier en immunofluorescence l'expression cellulaire de *BUD13* pour voir s'il est bien localisé, comme chez la levure, au niveau du splicéosome (c'est-à-dire dans les granules interchromatiniennes). Nous étudierons également en immunofluorescence la morphologie des granules interchromatiniennes de nos patients ;

- de rechercher des anomalies de l'épissage chez nos deux patients en puce d'expression, et rechercher les mêmes conséquences que l'inactivation de *BUD13* chez la levure, à savoir une augmentation du rapport PréARNm / ARNm mature.

Concernant la deuxième famille consanguine, dont les deux frères atteints ont une DI et une agénésie du corps calleux, et pour laquelle nous avons fait un *homozygosity mapping*, l'étude en exome, n'a pas permis de retrouver de mutation délétère dans les régions d'intérêt. Nous voyons plusieurs hypothèses pour l'expliquer :

- la mutation causale a été manquée par l'analyse d'exome (région non présente dans le kit de capture : intron, région régulatrice, ou à profondeur de séquençage trop faible...) ;

- la transmission de la pathologie n'est pas AR mais AD avec mosaïque germinale (ce qui expliquerait la récurrence alors que les parents sont bien portants). Pour le vérifier, il faudrait rechercher une mutation hétérozygote présente chez les deux atteints, absente (ou retrouvée avec peu de lectures) chez les parents ;

- la transmission de la maladie est AR mais avec une mutation hétérozygote composite non liée à la consanguinité ;

- la transmission de la maladie est récessive liée à l'X.

Nous avons au total 205 variants rares dans cette famille qui seraient compatibles avec l'une des trois dernières hypothèses ; aucun ne résulte en un codon stop prématuré. Nous n'avons donc pas réussi à mettre en évidence la mutation pathogène de cette famille à ce jour. Notre objectif actuel est de réaliser un exome chez d'autres patients atteints de DI et agénésie du corps calleux, et d'utiliser les données de notre famille afin de rechercher des mutations survenant de façon récurrente dans un ou plusieurs gènes. Cela sera fait dans le cadre du

PHRC national ACCREM que coordonne le Dr Héron (Centre de référence national des déficiences intellectuelles de causes rares, GH Pitié-Salpêtrière, Paris, France).

5.2.5. Identification d'un gène candidat par caractérisation de translocation équilibrée

Nous avons identifié une disruption du gène de l'*ATXN10*, en 22q13.31, causée par une translocation familiale 2;22, présente sous sa forme déséquilibrée chez deux sœurs atteintes de DI-S, et sous sa forme équilibrée chez une troisième soeur et le père. Le point de cassure a pu être caractérisé par une ACPA chez une des deux sœurs atteintes de DI-S porteuses de la translocation sous sa forme déséquilibrée, puis par un séquençage du point de cassure chez le père.

Après le séquençage du point de cassure, une petite insertion d'une cytosine a été détectée au niveau du point de cassure. Cette cytosine n'est normalement présente que sur le chromosome 2 normal, mais se retrouve au niveau des deux dérivés de translocation. Cette insertion correspond possiblement à une *information scar*, liée à un mécanisme de réparation par NHEJ (figure 65).

Figure 65 : séquences au niveau des points de cassure sur les chromosomes 2 et 22 et les dérivés de translocation. En rouge : régions de microhomologie, présentes à la fois sur le chromosome 2 et le chromosome 22. En rouge et gras : la cytosine, présente uniquement sur le chromosome 2 normal, et sur les deux dérivés de translocation.



ATXN10, qui code pour l'ataxine 10, est le gène de l'ataxie spino-cérébelleuse 10 (SCA10, OMIM#603516) qui est une maladie neurodégénérative autosomique dominante, caractérisée par un syndrome cérébelleux apparaissant entre 20 et 50 ans et, dans quelques cas, par une épilepsie. Une expansion en tandem d'un pentanucléotide (ATTCT)n est présente dans l'intron 9. La taille de cette expansion varie de 10 à 29 pour les allèles normaux et de 800 à 4 500 pour les allèles pathologiques (Matsura et al. 2000).

Une étude neurologique poussée des deux individus porteurs de la translocation équilibrée ne montre aucun signe d'ataxie ni d'atteinte neurologique. Une apparition plus tardive de l'ataxie ne peut être exclue chez la fille âgée de 23 ans, mais est par contre très improbable chez le père âgé de 57 ans, étant donné que la maladie survient avant 50 ans dans les cas décrits.

ATXN10 est un gène exprimé de façon ubiquitaire, dont la fonction exacte de la protéine est actuellement inconnue. Des études réalisées chez le rat ont montré qu'un *knock-down* par ARN interférence provoquait une dégénérescence des neurones cérébelleux, ce qui serait en faveur d'un mécanisme de type perte de fonction chez les patients SCA10 (Marz et al. 2004). A l'inverse, des études chez des souris *null* hétérozygotes n'ont montré aucun phénotype clinique et anatomo-pathologique, ce qui serait en défaveur d'un mécanisme de type perte de fonction chez les patients SCA10 (Wakamiya et al. 2006). Les deux études insistent sur les différences biologiques entre le modèle animal et l'Homme, et donc sur la nécessité d'étudier des échantillons de patients pour évaluer au mieux l'effet pathologique de l'expansion. Dans notre famille, le père et la deuxième fille ont une déficience de *ATXN10* à l'état hétérozygote mais ont un examen neurologique normal. Cette observation est en accord avec l'étude chez la souris et est donc en défaveur d'une haploinsuffisance de l'ataxine 10 humaine provoquée par l'expansion du pentanucléotide (ATTCT)n chez les patients SCA10.

Des expériences récentes sur des lignées cellulaires hybrides somatiques ont montré qu'il n'y a aucune interférence sur le niveau de la transcription ou de l'ARNm (Wakamiya et al. 2006). Toutes ces données sont en défaveur d'un simple mécanisme de gain ou de perte de fonction, ce qui est soutenu par des études se concentrant sur l'effet génomique de la mutation : l'expansion du pentanucléotide provoquerait un mésappariement de l'ADN chez la levure (Potaman et al. 2003) et dans les lignées cellulaires humaines (Liu et al. 2007), ce qui crée un déplacement des origines de réplication de l'ADN et une instabilité génomique (Edwards et al. 2009), entraînant une probable mort cellulaire.

Ainsi, l'ataxie spino-cérébelleuse de type 10 pourrait être liée à des anomalies de structure de la chromatine causées par l'expansion du pentanucléotide et non pas à une expression anormale du gène *ATXN10*. L'haploinsuffisance de l'*ATXN10*, observée chez nos patients sans phénotype "SCA10", est en accord avec cette hypothèse.

Cette translocation, retrouvée dans notre cohorte d'individus DI, est une illustration de la possibilité de mieux caractériser un gène sans rapport avec la pathologie initialement étudiée.

Conclusion

Nos résultats montrent l'intérêt d'étudier une cohorte de patients, adressés dans un cadre diagnostique, en puces SNP, afin d'identifier de nouveaux gènes candidats de DI.

Sur notre série de 1 187 patients, nous obtenons un taux diagnostique de 12,3% et nous avons identifié dix gènes candidats de DI. Par ailleurs, nous avons également mieux caractérisé, de façon fortuite, le rôle physiopathologique d'un gène, *ATXN10*, sans rapport avec la DI.

Toutefois, ce travail montre également les limites actuelles de notre stratégie. D'abord, un grand nombre des CNV que nous avons retrouvés sont classés VOUS ou "probablement bénins", ce qui implique une incertitude quant à leur signification clinique. Il est probable qu'un petit nombre d'entre eux soient pathogènes, en association éventuelle avec d'autres facteurs génétiques ou environnementaux. Cette limite pourra être surmontée dans l'avenir grâce à une meilleure connaissance du génome humain : les publications de la littérature et les bases de données publiques, permettre de connaitre de mieux en mieux l'effet des CNV, bénins ou pathogènes, d'un nombre grandissant de régions.

La deuxième limite concerne la validation des gènes candidats de DI, qui repose encore sur des techniques de séquençage ciblées, lourdes en termes de charge de travail et coûts réactif. Cela nous a imposé, premièrement de limiter notre travail à un nombre restreint de gènes candidats, sélectionnés sur des critères stringents, et deuxièmement, de valider ces gènes sur des petites cohortes de patients. Or, la grande majorité des gènes de DI ne concernent qu'une proportion très restreinte de patients, < 1%. En conséquence, les anomalies que nous avons identifiées sur nos gènes candidats, probablement responsables de pathologies très rares, n'ont pas pu être répliquées chez d'autres patients DI. A l'avenir, deux évolutions permettront de surmonter cette limite : d'abord, l'évolution très rapide des techniques de séquençage, qui deviennent de moins en moins coûteuses en personnel et en consommables ; ensuite, la mise en place de bases de données de ce type existe déjà : *exome variant server*, incluant des variants retrouvés en exome chez des patients présentant des pathologies cardiaques et

respiratoires. Il est probable qu'une ou plusieurs bases de données publiques répertoriant des variants retrouvés chez des patients DI voient également le jour dans un futur proche, ce qui permettra de faciliter grandement la validation des gènes candidats. On peut par exemple citer l'IRDiRC (*International Rare Disease Research Consortium*) qui a, entre autres, pour objectif d'harmoniser et de rendre publique un grand nombre de données, ceci afin d'accélérer la caractérisation clinique et moléculaires des maladies rares.

Bibliographie

Albayram O, Alferink J, Pitsch J, et al. Role of CB1 cannabinoid receptors on GABAergic neurons in brain aging. *Proc Natl Acad Sci U S A*. 2011;108:11256-61.

Altug-Teber O, Dufke A, Poths S, et al. A rapid microarray based whole genome analysis for detection of uniparental disomy. *Hum Mutat.* 2005;26:153-9.

Angkustsiri K, Krakowiak P, Moghaddam B, et al. Minor physical anomalies in children with autism spectrum disorders. *Autism*. 2011;15:746-60.

Ball MP, Thakuria JV, Zaranek AW, et al. A public resource facilitating clinical use of genomes. *Proc Natl Acad Sci U S A*. 2012;109:11920-7.

Bensaid M, Melko M, Bechara EG, et al. FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res*. 2009;37:1269-79.

Bentley DR. Whole-genome re-sequencing. Curr Opin Genet Dev. 2006;16:545-52.

Beirne-Smith M, Patton JR, Kim SH. Mental retardation: An introduction to intellectual disabilities. 7th ed. Upper Saddle River (USA): Pearson; 2006

Bessa C, Lopes F, Maciel P. Molecular Genetics of Intellectual Disability. Latest Findings in Intellectual and Developmental Disabilities. Ed : Tan U. Rijeka (Croatia): Intech Research; 2012:149-76.

Bhalla K, Luo Y, Buchan T, Beachem MA, et al. Alterations in CDH15 and KIRREL3 in patients with mild to severe intellectual disability. *Am J Hum Genet*. 2008;83:703-13.

Bruno DL, White SM, Ganesamoorthy D, et al. Pathogenic aberrations revealed exclusively by single nucleotide polymorphism (SNP) genotyping data in 5000 samples tested by molecular karyotyping. *J Med Genet*. 2011;48:831-9.

Bryson SE, Bradley EA, Thompson A, Wainwright A. Prevalence of autism among adolescents with intellectual disabilities. *Can J Psychiatry*. 2008;53:449-59.

Butler MG. Genomic imprinting disorders in humans: a mini-review. J Assist Reprod Genet. 2009;26:477-86.

Cardoso C, Leventer RJ, Ward HL, et al. Refinement of a 400-kb critical region allows genotypic differentiation between isolated lissencephaly, Miller-Dieker syndrome, and other phenotypes secondary to deletions of 17p13.3. *Am J Hum Genet*. 2003;72:918-30.

Carvill GL, Heavin SB, Yendle SC, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet. 2013;45:825-30.

Chardon JW, Mignot C, Aradhya S, Keren B, et al. Deletion of filamin A in two female patients with periventricular nodular heterotopia. *Am J Med Genet A*. 2012;158A:1512-6.

Chelly J, Khelfaoui M, Francis F, et al. Genetics and pathophysiology of mental retardation. *Eur J Hum Genet*. 2006;14:701-13.

Chen JM, Cooper DN, Férec C, et al. Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol*. 2010;20:222-33.

Chen X, Shen Y, Zhang F, et al. Molecular analysis of a deletion hotspot in the NRXN1 region reveals the involvement of short inverted repeats in deletion CNVs. *Am J Hum Genet*. 2013;92:375-86.

Cleaver JE, Bezrookove V, Revet I, Huang EJ. Conceptual developments in the causes of Cockayne syndrome. *Mech Ageing Dev.* 2013;134:284-90.

Cohen GA, Doze VA, Madison DV. Opioid inhibition of GABA release from presynaptic terminals of rat hippocampal interneurons. *Neuron*. 1992;9:325-35.

Conlin LK, Thiel BD, Bonnemann CG, et al. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet*. 2010;19:1263-75.

Cooper GM, Coe BP, Girirajan S, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011;43:838-46.

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45:984-94.

Curry CJ, Stevenson RE, Aughton D, et al. Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics. *Am J Med Genet.* 1997;72:468-77.

Daily DK, Ardinger HH, Holmes GE. Identification and evaluation of mental retardation. *Am Fam Physician*. 2000;61:1059-67.

Decouflé P, Boyle CA, Paulozzi LJ, Lary JM. Increased risk for developmental disabilities in children who have major birth defects: a population-based study. *Pediatrics*. 2001;108:728-34.

de Leeuw N, Hehir-Kwa JY, Simons A, et al. SNP array analysis in constitutional and cancer genome diagnostics--copy number variants, genotyping and quality control. *Cytogenet Genome Res.* 2011;135:212-21.

de Ligt J, Willemsen MH, van Bon BW, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med.* 2012;367:1921-9.

de Vries BB, White SM, Knight SJL, et al. Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. *J Med Genet*. 2001;38:145-50.

Depienne C, Bouteiller D, Keren B, et al. Sporadic infantile epileptic encephalopathy caused by mutations in PCDH19 resembles Dravet syndrome but mainly affects females. PLoS Genet. 2009;5:e1000381.

des Portes V, Boddaert N, Sacco S, et al. Specific clinical and brain MRI features in mentally retarded patients with mutations in the Oligophrenin-1 gene. *Am J Med Genet A*. 2004;124A:364-71.

Down JLH. Observations on an Ethnic Classification of Idiots. *London Hospital Reports*. 1866;3:259-262.

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Arlington (USA): American Psychiatric Publishing; 2013.

Dziembowski A, Ventura AP, Rutz B, et al. Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J.* 2004;23:4847-56.

Dulac C. Brain function and chromatin plasticity. Nature. 2010;465:728-35.

Edery P, Marcaillou C, Sahbatou M, et al. Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science*. 20118;332:240-3.

Edwards SF, Hashem VI, Klysik EA, Sinden RR. Genetic instabilities of (CCTG).(CAGG) and (ATTCT).(AGAAT) disease-associated repeats reveal multiple pathways for repeat deletion. *Mol Carcinog*. 2009 Apr;48(4):336-49.

Eiben B, Bartels I, Bähr-Porsch S, et al. Cytogenetic analysis of 750 spontaneous abortions with the direct-preparation method of chorionic villi and its implications for studying genetic causes of pregnancy wastage. *Am J Hum Genet*. 1990;47:656-63.

Ellison JW, Rosenfeld JA, Shaffer LG. Genetic basis of intellectual disability. *Annu Rev Med.* 2013;64:441-50.

Emerson E. Poverty and people with intellectual disabilities. *Ment Retard Dev Disabil Res Rev.* 2007;13:107-13.

Fan JB, Gunderson KL, Bibikova M, et al. Illumina universal bead arrays. *Methods Enzymol.* 2006;410:57-73.

Felisari G, Martinelli Boneschi F, et al. Loss of Dp140 dystrophin isoform and intellectual impairment in Duchenne dystrophy. *Neurology*. 2000;55:559-64.

Field M, Tarpey P, Boyle J, et al. Mutations in the RSK2(RPS6KA3) gene cause Coffin-Lowry syndrome and nonsyndromic X-linked mental retardation. *Clin Genet*. 2006;70:509-15.

Fombonne E. Epidemiological surveys of autism and other pervasive developmental disorders: an update. *J Autism Dev Disord*. 2003;33:365-82.

Frints SG, Froyen G, Marynen P, Fryns JP. X-linked mental retardation: vanishing boundaries between non-specific (MRX) and syndromic (MRXS) forms. *Clin Genet*. 2002;62:423-32.

Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol.* 2009;27:182-9.

Goldstone AP. Prader-Willi syndrome: advances in genetics, pathophysiology and treatment. *Trends Endocrinol Metab.* 2004;15:12-20.

Gregor A, Albrecht B, Bader I, et al. Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. *BMC Med Genet*. 2011;12:106.

Griggs BL, Ladd S, Saul RA, et al. Dedicator of cytokinesis 8 is disrupted in two patients with mental retardation and developmental disabilities. *Genomics*. 2008;91:195-202.

Grouchy J, Turleau C. Clinical Atlas of Human Chromosomes 2nd ed. New York (USA): Wiley;1984.

Guerrini R, Shanahan JL, Carrozzo R, et al. A nonsense mutation of the ATRX gene causing mild mental retardation and epilepsy. *Ann Neurol.* 2000;47:117-21.

Gunderson KL, Kruglyak S, Graige MS, et al. Decoding randomly ordered DNA arrays. *Genome Res.* 2004;14:870-7.

Hochstenbach R, Buizer-Voskamp JE, Vorstman JA, Ophoff RA. Genome arrays for the detection of copy number variations in idiopathic mental retardation, idiopathic generalized epilepsy and neuropsychiatric disorders: lessons for diagnostic workflow and research. *Cytogenet Genome Res.* 2011;135:174-202.

Hosoki K, Kagami M, Tanaka T, et al. Maternal uniparental disomy 14 syndrome demonstrates prader-willi syndrome-like phenotype. *J Pediatr*. 2009;155:900-903.e1.

Hoyer J, Ekici AB, Endele S, et al. Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am J Hum Genet*. 2012;90:565-72.

Filges I, Röthlisberger B, Noppen C, et al. Familial 14.5 Mb interstitial deletion 13q21.1-13q21.33: clinical and array-CGH study of a benign phenotype in a three-generation family. *Am J Med Genet A*. 2009;149A:237-41.

Forsgren L, Edvinsson SO, Nyström L, Blomquist HK. Influence of epilepsy on mortality in mental retardation: an epidemiologic study. *Epilepsia*. 1996;37:956-63.

Fujihara Y, Satouh Y, Inoue N, et al. SPACA1-deficient male mice are infertile with abnormally shaped sperm heads reminiscent of globozoospermia. *Development*. 2012;139:3583-9.

Girard SL, Gauthier J, Noreau A, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet.* 2011;43:860-3.

Hamdan FF, Daoud H, Rochefort D, et al. De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *Am J Hum Genet*. 2010;87:671-8.

Harbour CK. History of Intellectual Disability, CIRRIE 2010, http://cirrie.buffalo.edu/
Harris JC. Intellectual disability: understanding its development, causes, classification, evaluation, and treatment. New York (USA): Oxford University Press; 2006:2-98.

Hildebrandt F, Heeringa SF, Rüschendorf F, et al. A systematic approach to mapping recessive disease genes in individuals from outbred populations. *PLoS Genet*. 2009;5:e1000353.

Horn D, Kapeller J, Rivera-Brugués N, et al. Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. *Hum Mutat.* 2010;31:E1851-60.

Huentelman MJ, Papassotiropoulos A, Craig DW, et al. Calmodulin-binding transcription activator 1 (CAMTA1) alleles predispose human episodic memory performance. *Hum Mol Genet*. 2007;16:1469-77.

ISCN2013. An International System for Human Cytogenetic Nomenclature. Ed: Shaffer L, McGowan-Jordan J, Schmid M. Basel (Switzerland): Karger; 2013.

Itsara A, Wu H, Smith JD, et al. De novo rates and selection of large copy number variation. *Genome Res.* 2010;20:1469-81.

Jacquemont S, Reymond A, Zufferey F, [...], Keren B, [...], Froguel P. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature*. 2011;478:97-102.

Jacobs KB, Yeager M, Zhou W, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet*. 2012;44:651-8.

Jelliffe-Pawlowski LL, Shaw GM, Nelson V, Harris JA. Risk of mental retardation among children born with birth defects. *Arch Pediatr Adolesc Med.* 2003;157:545-50.

Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*. 2003;12:5-14.

Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*. 1987;236:1567-70.

Liu F, Pore N, Kim M, Voong KR, et al. Regulation of histone deacetylase 4 expression by the SP family of transcription factors. *Mol Biol Cell*. 2006;17:585-97.

Kaminsky EB, Kaul V, Paschall J, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med.* 2011;13:777-84.

Kaufman L, Ayub M, Vincent JB. The genetic basis of non-syndromic intellectual disability: a review. *J Neurodev Disord*. 2010;2:182-209.

Kearney HM, Thorland EC, Brown KK, et al. Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med.* 2011;13:680-5.

Key B, Lah GJ. Repulsive guidance molecule A (RGMa): a molecule for all seasons. *Cell Adh Migr*. 2012;6:85-90.

Keren B (a), Schluth-Bolard C, Egea G, Sanlaville D. New technologies for the human genome exploration. *Arch Pediatr.* 2010;17:1605-8.

Keren B (b), Jacquette A, Depienne C, et al. Evidence against haploinsuffiency of human ataxin 10 as a cause of spinocerebellar ataxia type 10. *Neurogenetics*. 2010;11:273-4.

Keren B (a), Moncla A, Sanlaville D. Les syndromes microdélétionnels. Syndromes dysmorphiques. Ed: Lacombe D, Philip N. Progrès en Pédiatrie. Rueil-Malmaison (France): Doin. 2013, en cours de publication.

Keren B (b), Chantot-Bastaraud S, Brioude F, et al. SNP arrays in Beckwith-Wiedemann syndrome: an improved diagnostic strategy. *Eur J Med Genet*. 2013, en cours de publication.

Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature*. 2008;453:56-64.

Kohlhase J, Janssen B, Weidenauer K, et al. First confirmed case with paternal uniparental disomy of chromosome 16. *Am J Med Genet*. 2000;91:190-1.

Kotzot D. Abnormal phenotypes in uniparental disomy (UPD): fundamental aspects and a critical review with bibliography of UPD other than 15. *Am J Med Genet*. 1999;82:265-74.

Krantz ID, McCallum J, DeScipio C, et al. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. *Nat Genet*. 2004;36:631-5.

Kurotaki N, Imaizumi K, Harada N, et al. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet*. 2002;30:365-6.

Larson SA, Lakin KC, Anderson L, et al. Prevalence of mental retardation and developmental disabilities: estimates from the 1994/1995 National Health Interview Survey Disability Supplements. *Am J Ment Retard*. 2001;106:231-52.

Laurie CC, Laurie CA, Rice K, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. 2012;44:642-50.

Lee JA (a), Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*. 2007;131:1235-47.

Lee C (b), Iafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet*. 2007;39S:S48-54.

Lehrke R. Theory of X-linkage of major intellectual traits. *Am J Ment Defic*. 1972;76:611-9.

Lejeune J, Gautier M, Turpin R. Étude des chromosomes somatiques de neuf enfants mongoliens . *C R Hebd Seances Acad Sci.* 1959;248:1721-2.

Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev.* 2002;8:117-34.

Levy D, Ronemus M, Yamrom B, et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*. 2011;70:886-97.

Li S, Rousseau D. ATAD3, a vital membrane-bound mitochondrial ATPase involved in tumor progression. *Med Sci.* 2011;27:1089-95.

Lieber MR, Ma Y, Pannicke U, Schwarz K. Mechanism and regulation of human nonhomologous DNA end-joining. *Nat Rev Mol Cell Biol.* 2003;4:712-20.

Linder B, Dill H, Hirmer A, et al. Systemic splicing factor deficiency causes tissuespecific defects: a zebrafish model for retinitis pigmentosa. *Hum Mol Genet*. 2011;20:368-77.

Liu G, Bissler JJ, Sinden RR, Leffak M. Unstable spinocerebellar ataxia type 10 (ATTCT*(AGAAT) repeats are associated with aberrant replication at the ATX10 locus

and replication origin-dependent expansion at an ectopic site in human cells. *Mol Cell Biol.* 2007;27:7828-38.

Lubs HA, Stevenson RE, Schwartz CE. Fragile X and X-linked intellectual disability: four decades of discovery. *Am J Hum Genet*. 2012;90:579-90.

Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci USA*. 2010;107:961-8.

Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell.* 2012;148:1223-41.

Manning M, Hudgins L; Professional Practice and Guidelines Committee. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med.* 2010;12:742-5.

Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated highdensity picolitre reactors. *Nature*. 2005;437:376-80.

Martin JP, Bell J. A pedigree of mental defect showing sex-linkage. *J Neurol Psychiatry*. 1943;6:154-7.

März P, Probst A, Lang S, et al. Ataxin-10, the spinocerebellar ataxia type 10 neurodegenerative disorder protein, is essential for survival of cerebellar neurons. *J Biol Chem.* 2004;279:35542-50.

Matsuura T, Yamagata T, Burgess DL, et al. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat Genet*. 2000;26:191-4.

Mefford HC, Batshaw ML, Hoffman EP. Genomics, intellectual disability, and autism. *N Engl J Med.* 2012;366:733-43.

Mikhail FM, Lose EJ, Robin NH, 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:2386-96.

Miles JH, Takahashi TN, Bagby S, Sahota PK, Vaslow DF, Wang CH, Hillman RE, Farmer JE. Essential versus complex autism: definition of fundamental prognostic subtypes. *Am J Med Genet A*. 2005;135:171-80.

Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet*. 2010;86:749-64.

Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet*. 2010;19:R131-6.

Oberlé I, Rousseau F, Heitz D, et al. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*. 1991;252:1097-102.

Oeseburg B, Dijkstra GJ, Groothoff JW, et al. Prevalence of chronic health conditions in children with intellectual disability: a systematic literature review. *Intellect Dev Disabil.* 2011;49:59-85.

Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004;5:557-72.

O'Roak BJ, Deriziotis P, Lee C, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet*. 2011;43:585-9.

Papenhausen P, Schwartz S, Risheg H, et al. UPD detection using homozygosity profiling with a SNP genotyping microarray. *Am J Med Genet A*. 2011;155A:757-68.

Pasternack SM, Refke M, Paknia E, et al. Mutations in SNRPE, which encodes a core protein of the spliceosome, cause autosomal-dominant hypotrichosis simplex. *Am J Hum Genet.* 2013;92:81-7.

Patja K, Mölsä P, Iivanainen M. Cause-specific mortality of people with intellectual disability in a population-based, 35-year follow-up study. *J Intellect Disabil Res.* 2001;45:30-40.

Patja K, Iivanainen M, Vesala H, et al. Life expectancy of people with intellectual disability: a 35-year follow-up study. *J Intellect Disabil Res.* 2000;44:591-9.

Petterson B, Bourke J, Leonard H, et al. Co-occurrence of birth defects and intellectual disability. *Paediatr Perinat Epidemiol.* 2007;21:65-75.

Piard J, Depienne C, Keren B, et al. Intragenic deletion of UBE3A gene in 2 sisters with Angelman syndrome detected by MLPA. *Am J Med Genet A*. 2011;155A:3170-3.

Pinto D, Pagnamenta AT, Klei L, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010;466:368-72.

Poirier K, Lacombe D, Gilbert-Dussardier B, et al. Screening of ARX in mental retardation families: Consequences for the strategy of molecular diagnosis. *Neurogenetics*. 2006;7:39-46.

Ramocki MB, Zoghbi HY. Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature*. 2008;455:912-8.

Rauch A, Hoyer J, Guth S, et al. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A*. 2006;140:2063-74.

Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*. 2012;380:1674-82.

Raymond FL. Monogenic Causes of Mental Retardation. Genetics of Mental Retardation. Ed : Knight SJL. Monogr Hum Genet. Basel (Switzerland): Karger; 2010;18:89-100.

Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444-54.

Regan R, Willatt L. Mental Retardation: Definition, Classification and Etiology. Genetics of Mental Retardation. Ed : Knight SJL. Monogr Hum Genet. Basel (Switzerland): Karger; 2010;18:16-30

Roeleveld N, Zielhuis GA, Gabreëls F. The prevalence of mental retardation: a critical review of recent literature. *Dev Med Child Neurol*. 1997;39:125-32.

Romanelli V, Meneses HN, Fernández L, et al. Beckwith-Wiedemann syndrome and uniparental disomy 11p: fine mapping of the recombination breakpoints and evaluation of several techniques. *Eur J Hum Genet*. 2011;19:416-21.

Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. *Science*. 1998;281:363-365.

Rowe LR, Lee JY, Rector L, et al. U-type exchange is the most frequent mechanism for inverted duplication with terminal deletion rearrangements. *J Med Genet*. 2009;46:694-702

Salvador-Carulla L, Reed GM, Vaez-Azizi LM, et al. Intellectual developmental disorders: towards a new name, definition and framework for "mental retardation/intellectual disability" in ICD-11. *World Psychiatry*. 2011;10:175-80.

Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*. 2011;70:863-85.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74:5463-7.

Scales SJ, Hesser BA, Masuda ES, Scheller RH. Amisyn, a novel syntaxin-binding protein that may regulate SNARE complex assembly. *J Biol Chem.* 2002;277:28271-9.

Schaaf CP, Wiszniewska J, Beaudet AL. Copy number and SNP arrays in clinical diagnostics. *Annu Rev Genomics Hum Genet*. 2011;12:25-51.

Schäffer AA. Digenic inheritance in medical genetics. *J Med Genet.* 2013. En cours de publication.

Schalock RL, Luckasson RA, Shogren KA, et al. The Renaming of Mental Retardation: Understanding the Change to the Term Intellectual Disability. *Intellect Dev Disabil.* 2007;45:116-24.

Schalock RL, Borthwick-Duffy SA, Bradley VJ, et al. Intellectual Disability: Definition, Classification, and Systems of Supports (11h edition). Annapolis Junction (USA): American Association on Intellectual and Developmental Disabilities; 2010.

Scheerenberger RC. A History of Mental Retardation. Baltimore: Brookes Publishing Company; 1983

Schiff M, Delahaye A, Andrieux J, et al. Further delineation of the 17p13.3 microdeletion involving YWHAE but distal to PAFAH1B1: four additional patients. *Eur J Med Genet*. 2010;53:303-8.

Sharp AJ, Locke DP, McGrath SD, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*. 2005;77:78-88.

Shoukier M, Klein N, Auber B, et al. Array CGH in patients with developmental delay or intellectual disability: are there phenotypic clues to pathogenic copy number variants? *Clin Genet.* 2013;83:53-65.

Sirmaci A, Spiliopoulos M, Brancati F, et al. Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. *Am J Hum Genet*. 2011;89:289-94.

Slater HR, Ralph A, Daniel A, et al. A case of maternal uniparental disomy of chromosome 9 diagnosed prenatally and the related problem of residual trisomy. *Prenat Diagn.* 2000;20:930-2.

Smith CE, Llorente B, Symington LS. Template switching during break-induced replication. *Nature*. 2007;447:102-5.

South ST, Brothman AR. Clinical laboratory implementation of cytogenomic microarrays. *Cytogenet Genome Res.* 2011;135:203-11.

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

Stankiewicz P, Beaudet AL. Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. *Curr Opin Genet Dev.* 2007;17:182-92.

Steemers FJ, Chang W, Lee G, et al. Whole-genome genotyping with the single-base extension assay. *Nat Methods*. 2006;3:31-3.

Steinberg G, Stromsborg K, Thomas L, et al. Strategies for covalent attachment of DNA to beads. *Biopolymers*. 2004;73:597-605.

Stephenson MD, Awartani KA, Robinson WP. Cytogenetic analysis of miscarriages from couples with recurrent miscarriage: a case-control study. *Hum Reprod.* 2002;17:446-51.

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;89:551-63.

Tarpey PS, Smith R, Pleasance E, et al. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet*. 2009;41:535-43.

Tyrer F, McGrother C. Cause-specific mortality and death certificate reporting in adults with moderate to profound intellectual disability. *J Intellect Disabil Res.* 2009;53:898-904.

van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. Annu Rev Genet. 2011;45:81-104.

van Houdt JK, Nowakowska BA, Sousa SB, et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat Genet*. 2012;44:445-9.

van Karnebeek CD, Scheper FY, Abeling NG, et al. Etiology of mental retardation in children referred to a tertiary care center: a prospective study. *Am J Ment Retard*. 2005;110:253-67.

Verloes A, Héron D, Billette de Villemeur T, et al. Diagnostic investigations for an unexplained developmental disability. *Arch Pediatr.* 2012;19:194-207.

Verrotti A, Spalice A, Ursitti F, et al. New trends in neuronal migration disorders. *Eur J Paediatr Neurol.* 2010;14:1-12.

Vissers LE, van Ravenswaaij CM, Admiraal R, et al. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet*. 2004;36:955-7.

Vissers LE, de Ligt J, Gilissen C, et al. A de novo paradigm for mental retardation. *Nat Genet*. 2010;42:1109-12.

Wakamiya M, Matsuura T, Liu Y, et al. The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10. *Neurology*. 2006;67:607-13.

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;81:768-79.

Weise A, Mrasek K, Klein E, et al. Microdeletion and microduplication syndromes. J *Histochem Cytochem*. 2012;60:346-58.

Willatt LR, Davison BC, Goudie D, et al. A male with trisomy 9 mosaicism and maternal uniparental disomy for chromosome 9 in the euploid cell line. *J Med Genet*. 1992;29:742-4.

Yamazawa K, Ogata T, Ferguson-Smith AC. Uniparental disomy and human disease: an overview. *Am J Med Genet C Semin Med Genet*. 2010;154C:329-34.

Yokoyama K, Tezuka T, Kotani M, et al. NYAP: a phosphoprotein family that links PI3K to WAVE1 signalling in neurons. *EMBO J.* 2011;30:4739-54.

Zoghbi HY, Bear MF. Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. *Cold Spring Harb Perspect Biol.* 2012;4(3).

Références Internet

1000 genomes : http://www.1000genomes.org/ AAIDD : http://aaidd.org/ Aceview : http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/ CIM10: http://apps.who.int/classifications/icd10/browse/2010/en ISCA/dbvar : http://www.ncbi.nlm.nih.gov/dbvar/ dbSNP : https://www.ncbi.nlm.nih.gov/SNP/ DECIPHER : http://decipher.sanger.ac.uk/ DGV: http://projects.tcag.ca/variation/ Exome Variant Server : http://evs.gs.washington.edu/EVS/ Genereviews : http://www.ncbi.nlm.nih.gov/books/NBK1116 HapMap : http://hapmap.ncbi.nlm.nih.gov/ Illumina : https://www.illumina.com/ IRDiRC : http://www.irdirc.org/ MutPred : http://mutpred.mutdb.org/ OMIM : http://www.ncbi.nlm.nih.gov/omim Polyphen : http://genetics.bwh.harvard.edu/pph2/ SIFT : http://sift.jcvi.org/ Unigene : http://www.ncbi.nlm.nih.gov/unigene

Keren Boris – Thèse de doctorat - Année 2013

Article 1

RESEARCH LETTER



Intragenic Deletion of UBE3A Gene in 2 Sisters With Angelman Syndrome Detected by MLPA

Juliette Piard,^{1,2}* Christel Depienne,^{3,4} Boris Keren,⁵ Estelle Fédirko,³ Oriane Trouillard,³ Perrine Charles,¹ and Delphine Heron¹

¹Unité Fonctionnelle de Génétique Médicale AP-HP, Département de Génétique et Cytogénétique, Centre de Référence «Déficiences Intellectuelles de Causes Rares», Groupe Hospitalier CRicm, UMR-S975 (Ex-U679), Pitié-Salpêtrière, Paris, France

²Centre de Génétique Humaine, Université de Franche-Comté, CHU Besançon, Besançon, France

³Unité Fonctionnelle de Neurogénétique Moléculaire et Cellulaire, AP-HP, Département de Génétique et Cytogénétique, Centre de Génétique Moléculaire, Hôpital de la Pitié-Salpêtrière, Paris, France

⁴CRicm, UMR-S975 (Ex-U679), Université Pierre et Marie Curie-Paris 6, Paris, France

⁵Unité Fonctionnelle de Cytogénétique, AP-HP, Département de Génétique et Cytogénétique, Fédération de Génétique, Hôpital de la Pitié-Salpêtrière, Paris, France

Received 18 August 2011; Accepted 2 September 2011

TO THE EDITOR:

Angelman syndrome (AS) is characterized by severe intellectual disability with major speech impairment, ataxia and behavioral uniqueness including a combination of frequent laughter and smiling, apparent happy demeanor, excitable personality, and hypermotor behavior [Williams, 2010]. Characteristic morphological facial features are common but may be subtle and include broad mouth, wide spaced teeth, prognathia, protruding tongue, strabismus, and deep-set eyes [Smith et al., 1996; Sandanam et al., 1997; Pelc et al., 2008a,b,c]. The underlying molecular deficit is the failure of expression of the maternal copy of the imprinted UBE3A gene due to maternal deletions, paternal uniparental disomy (UPD), imprinting center defects or point mutations in the maternally inherited UBE3A gene [Clayton-Smith and Laan, 2003; Jedele, 2007]. Classical methods used to confirm the clinical diagnosis comprise methylation studies and sequencing of the UBE3A gene [Clayton-Smith and Laan, 2003].

In this report we present two sisters with AS associated with atypical clinical manifestations. A heterozygous partial microdeletion within the *UBE3A* gene, undetectable by direct sequencing, was identified using multiplex ligation-dependent probe amplification (MLPA).

The affected family was European and included two sisters from healthy unrelated parents. Patient 1 was born at term following an uneventful pregnancy. Birth parameters were within normal ranges. Strabismus was noted at 4 months. Developmental milestones were delayed. She sat unsupported after 9 months and walked independently at 24 months with unsteady, wide-based gait and flexed arms. There was no speech development. By the age of 30 months, she developed tonic seizures controlled with Sodium Valproate. The EEG showed generalized slow wave activity with paroxysmal spike-waves. Then, she presented with absence and myoclonic seizures when she stopped the anticonvulsant therapy at

How to Cite this Article:

Piard J, Depienne C, Keren B, Fédirko E, Trouillard O, Charles P, Heron D. 2011. Intragenic deletion of UBE3A gene in two sisters with Angelman syndrome detected by MLPA.

Am J Med Genet Part A 155:3170-3173.

the age of 6. An improvement in epilepsy was obtained with Ethosuximide. As a child, she had sleep disturbances, stereotypies and a happy disposition with frequent smiling and laughing. However, no inappropriate paroxysmal laughter was noted. She attended a school that provided special needs education from the age of 6. At age 16, she could only say three words. She became more anxious over time and her demeanor appeared less happy. She still had stereotypies and was on anticonvulsant treatment with seizures remission. Her height was 152 cm (-2 standard deviation (SD)), her weight was 63 kg (+2 SD), and her head circumference was 52 cm (-2 SD). Mild obesity with abnormal fat distribution and subtle dysmorphic facial features were noted. These included a large mouth, widely spaced teeth, deep-set eyes, and upslanting palpebral fissures (Fig. 1). She was also noted to have marked lymphedema of the lower limbs, which was treated by lymph drainage (Fig. 2a,b). Clinical investigations revealed normal biochemical markers,

*Correspondence to:

Dr. Juliette Piard, Service de Génétique Médicale, CHU Besançon, 2 place Saint Jacques, 25000 Besançon, France. E-mail: juliette.piard@gmail.com Published online 7 November 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/ajmg.a.34334

© 2011 Wiley Periodicals, Inc.

PIARD ET AL.



FIG. 1. Morphological facial features of Patient 1. Note large mouth, wide spaced teeth, deep set eyes, strabismus. [Color figure can be seen in the online version of this article, available at http:// onlinelibrary.wiley.com/journal/10.1002/[ISSN]1552-4833]

steatorrhea and a peripheral interstitial syndrome associated with primary lymphatic disease on thoracic CT-scan.

Patient 2 is the younger sister of Patient 1. She was born at term after uneventful pregnancy. Birth parameters were within normal ranges. Stereotypies of rocking type were reported at 2 months by the parents. She sat unsupported at a normal age and walked independently at 24 months. As with her sister, she had no speech development. She presented with tonic and myoclonic seizures at 5 years successfully controlled with Sodium Valproate. She attended a special needs school at age 6. She developed gastrointestinal dysfunction with constipation, severe gastroesophageal reflux disease and swallowing difficulties leading to recurrent aspiration pneumonia. At the age of 8, she had a pulmonary abscess in her right lower lobe. A thoracic CT-scan performed at this time revealed a coincidental cardiac malformation with a right aortic arch, a Kommerell's diverticulum and a retroesophageal sub-clavian left artery causing mild tracheal obstruction. Ataxic gait was noted by age 10. At 13 years and 5 months she could only speak a few words and her comprehension seemed to lag behind that of her sister. She had a happy demeanor with smiling and laughing and hyperactive behavior. Her height was 152 cm (-0.5 SD), her weight was 42 kg(M), and her head circumference was 50.5 cm (-2.5 SD). Clinical examination revealed dysmorphic features (Fig. 3a) with macrostomia, widely spaced teeth, strabismus, deep-set eyes, and upslanting palpebral fissures. In addition she developed orthopedic complications including hyperlordosis, scoliosis, and contractures in her ankles and thumbs (Fig. 3b).

Karyotype, methylation analysis of the 15q11-q13 region and FISH using DNA probe specific for the Prader–Willi/Angelman critical region were normal in both sisters. Direct sequencing of the



FIG. 2. a,b: Lower limb lymphedema in Patient 1. [Color figure can be seen in the online version of this article, available at http:// onlinelibrary.wiley.com/journal/10.1002/[ISSN)1552-4833]



FIG. 3. a: Morphological facial features of Patient 2. Note large smiling mouth with wide teeth, deep-set eyes, strabismus, and upslanting palpebral fissures. b: Hyperlordosis and contractures observed in Patient 2. [Color figure can be seen in the online version of this article, available at http://onlinelibrary.wiley.com/ journal/10.1002/[ISSN]1552-4833]



coding regions of UBE3A failed to identify a point mutation in the gene and analysis using either a Genosystem array CGH (containing 490 BAC clones) or an Integrachip array CGH (containing 3,172 clones) did not detect pathogenic micro-rearrangements. Other negative investigations included a broad metabolic screen, analysis of sub-telomeric rearrangements by MLPA and direct sequencing of the coding sequence of MECP2. Finally, MLPA using the P064 Mental Retardation-1 kit (MCR-Holland) revealed a partial heterozygous deletion of UBE3A in Patient 1. The deletion was confirmed by MLPA using the ME028 kit (designed to explore the 15q11-q13 interval) and was shown to encompass at least exons 6-12. The same deletion was detected in the affected sister (Patient 2) and their healthy mother but not in the father (Fig. 4). We . We used the Illumina HumanCytoSNP-12 array to more precisely characterize the size of the deletion. This analysis demonstrated that the deletion encompassed exons 6-14 and spanned approximately 70 Kb.

The patients presented in this report have classical phenotypic manifestations of AS including severe intellectual disability with major speech impairment, moderate ataxia, microcephaly, seizures, and behavioral disorders with happy dispositions, frequent smiling, and sleep disturbance. However, neurological and behavioral problems were less severe than have been described previously [Smith et al., 1996; Sandanam et al., 1997; Pelc et al., 2008a,b,c]. Medical complications such as obesity, constipation, and esophageal reflux are commonly seen in AS, however unusual clinical findings were noted in our two patients. Patient 1 had marked lower limb lymphedema and Patient 2 had a right aortic arch. To our knowledge, this is the first report describing such anomalies in AS patients. We cannot conclude, however, whether these clinical findings are related to AS or are incidental.

A partial micro-deletion in *UBE3A* was identified in the two affected sisters using MLPA. This deletion was inherited by the mother, who did not show any symptoms of AS, suggesting that she had received it from her father. Unfortunately, this hypothesis could not be confirmed since the maternal grandfather was deceased at the time of the study. This sub-microscopic deletion was much smaller than the typical AS deletion and was neither detectable by methylation testing nor by FISH analysis using classical probes nor by BAC-array. In the literature, five other reports have described a micro-deletion within the *UBE3A* gene causing AS. Different techniques were used to emphasize the submicroscopic molecular defect including allelic loss at micro-satellite loci [Bürger et al., 2002], semi-quantitative dosage PCR technique [Boyes et al., 2006], CGH-array [Lawson-Yuen et al., 2006], and MLPA [Cali et al., 2010; Beleza-Meireles et al., 2011]. All these patients presented with typical AS and were not distinguishable from AS patients with classical deletion, UPD or point mutations in *UBE3A*. No lymphedema, right aortic arch, or other malformations were described in those patients.

Early diagnosis of AS, detection of the underlying defect and mode of inheritance is crucial in terms of genetic counseling especially if the methylation study is normal, which theoretically increases the risk of recurrence. Indeed, AS is caused by different genetic defects, which have similar consequences in terms of neurological development and behavior, but which are associated with completely different risks of recurrence in the family of the affected child. In our case, the risk of recurrence was 50%.

Our study supports previous reports and highlights that the presence of micro-deletions within the *UBE3A* gene, undetectable by DNA methylation studies and direct sequencing, can be responsible for typical AS. We have also confirmed that MLPA analysis can successfully identify this molecular defect. Geneticists should be aware that different techniques are know available to detect micro-deletions in the *UBE3A* gene and should absolutely be considered after traditional testing in patients with a high clinical suspicion of AS.

REFERENCES

- Beleza-Meireles A, Cerqueira R, Sousa SB, Palmeiro A, Ramos L. 2011. Novel deletion encompassing exons 5–12 of the UBE3A gene in a girl with Angelman syndrome. Eur J Med Genet, DOI: 10.1016/ j.ejmg.2011.02.010.
- Boyes L, Wallace AJ, Krajewska-Walasek M, Chrzanowska KH, Clayton-Smith J, Ramsden S. 2006. Detection of a deletion of exons 8–16 of the

3173

PIARD ET AL.

UBE3A gene in familial Angelman syndrome using a semi-quantitative dosage PCR based assay. Eur J Med Genet 49:472–480.

Bürger J, Horn D, Tönnies H, Neitzel H, Reis A. 2002. Familial interstitial 570 kbp deletion of the UBE3A gene region causing Angelman syndrome but not Prader-Willi syndrome. Am J Med Genet 111:233–237.

Cali F, Ragalmuto A, Chiavetta V, Calabrese G, Fichera M, Vinci M, Ruggeri G, Schinocca P, Sturnio M, Romano S, Romano V, Elia M. 2010. Novel deletion of the E3A ubiquitin protein ligase gene detected by multiplex ligation-dependant probe amplification in a patient with Angelman syndrome. Exp Mol Med 42:842–848.

- Clayton-Smith J, Laan L. 2003. Angelman syndrome: A review of the clinical and genetic aspects. J Med Genet 40:87–95.
- Jedele K. 2007. The overlapping spectrum of Rett and Angelman syndromes: A clinical review. Semin Pediatr Neurol 14:108–117.
- Lawson-Yuen A, Wu BL, Lip V, Sahoo T, Kimonis V. 2006. Atypical cases of Angelman syndrome. Am J Med Genet Part A 140A:2361–2364.

Pelc K, Boyd S, Cheron GA, Dan B. 2008a. Epilepsy in Angelman syndrome. Seizure 17:211–221.

- Pelc K, Cheron G, Boyd S, Dan B. 2008b. Are there distinctive sleep problems in Angelman syndrome? Sleep Med 9:434–441.
- Pelc K, Cheron GA, Dan B. 2008c. Behavior and neuropsychiatric manifestations in Angelman syndrome. Neuropsychiatr Dis Treat 4(3):577–584.
- Sandanam Tls, Beange H, Robson L, Woolnough H, Buchholz T, Smith A. 1997. Manifestations in institutionalised adults with Angelman syndrome due to deletion. Am J Med Genet 70:415–420.
- Smith A, Wiles C, McGill J, Wallace G, Dixon J, Selby R, Colley A, Marks R, Trent R. 1996. Clinical features in 27 patients with Angelman syndrome resulting from DNA deletion. J Med Genet 33:107–112.
- Williams CA. 2010. The behavioral phenotype of the Angelman syndrome. Am J Med Genet Part C 154C:432–437.

Article 2

RESEARCH LETTER



Deletion of Filamin A in Two Female Patients With Periventricular Nodular Heterotopia

Jodi Warman Chardon,¹* Cyril Mignot,² Swaroop Aradhya,³ Boris Keren,² Alexandra Afenjar,² Anna Kaminska,⁴ Cherif Beldjord,⁵ Delphine Héron,² and Kym M. Boycott¹

¹Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada

²Département de Génétique, Groupe Hospitalier Pitié Salpétrière, APHP, Paris, France

³GeneDx, Gaithersburg, Maryland

⁴Unité Clinique d'Electrophysologie, Hôpital Necker, APHP, Paris, France

⁵Service de Biologie Moléculaire et Génétique, Hôpital Cochin, APHP, Paris, France

Manuscript Received: 21 January 2012; Manuscript Accepted: 14 March 2012

TO THE EDITOR:

Periventricular nodular heterotopia (PNH) is a neuronal migration disorder characterized by uncalcified macroscopic gray matter nodules adjacent to lateral ventricles [Dobyns et al., 1996]. PNH has a wide clinical spectrum, ranging from asymptomatic to intractable seizures with severe cognitive impairment [Parrini et al., 2006]. PNH can be isolated or observed in association with other brain, skeletal, or visceral abnormalities such as microcephaly, polymicrogyria, cerebellar hypoplasia, frontonasal dysplasia, ambiguous genitalia, limb anomalies [Parrini et al., 2006], cardiac malformations [Jefferies et al., 2010] and features of Ehlers-Danlos syndrome [Sheen et al., 2005]. The etiology of PNH is heterogeneous; the most common cause of isolated PNH is the X-linked form (OMIM 30004) caused by presumed loss-of-function mutations of the filamin A gene (FLNA, locus Xq28) [Ekşioğlu et al., 1996; Fox et al., 1998]. Phenotypes associated with probable gain-of-function mutations in FLNA cause the otopalatodigital spectrum of disorders [Robertson et al., 2003; Clark et al., 2009]. The FLNA gene is immediately adjacent to the emerin gene (EMD) which is mutated in X-linked Emery-Dreifuss muscular dystrophy (EDMD; OMIM 300384) [Bione et al., 1994], a neuromuscular condition characterized by contractures, humeroperoneal muscle weakness, and cardiac involvement [Emery, 2000; Muchir and Worman, 2007].

While most genetic causes of PNH are due to mutations in *FLNA*, other genetic contributions have been identified including mutations in *ARFGEF2* localized at 20q13.13, which causes an autosomal recessive form of PNH (OMIM 608097) [Sheen et al., 2004]. In addition, chromosome abnormalities have been found in association with PNH phenotypes [Andrade, 2009], including deletions (5q14.3–q15 [Cardoso et al., 2009]; 7q11 [Ferland et al., 2006]), and duplications (5p15.1 [Sheen et al., 2003]). We describe two female patients with PNH and normal *FLNA* sequencing who were subsequently analyzed by microarray to assess for chromosome abnormalities that could be contributing to their clinical presentations; deletions involving *FLNA* were identified in both patients, which have implications for diagnostic testing strategies and counseling.

How to Cite this Article:

Warman Chardon J, Mignot C, Aradhya S, Keren B, Afenjar A, Kaminska A, Beldjord C, Héron D, Boycott KM. 2012. Deletion of filamin A in two female patients with periventricular nodular heterotopia. Am J Med Genet Part A 158A:1512–1516.

Patient 1 is a 42-year-old female seen for neurogenetics consultation regarding the etiology of her seizure disorder. She is the eldest of two children born to nonconsanguineous parents; the family history was unremarkable. The patient experienced a traumatic brain injury secondary to a motor vehicle collision at 5 years of age. Cognitive impairment and seizures were noted after the accident. Prior to the accident, she had no seizures and was developing normally. She is being treated with anti-epileptic therapy with good seizure control. The patient is nulliparous and past medical history also includes dysfunctional uterine bleeding due to uterine fibroids, migraines and right hip replacement secondary to pelvic and right hip injuries sustained in the motor vehicle collision. Neurological examination was remarkable for a mild right hemiparesis, spasticity, and hyperreflexia likely secondary to the accident. Muscle bulk and tone was within normal limits, and no contractures were present. There was no evidence of joint hyperextensibility or bruising. Cardiorespiratory examination and EKG were unremarkable and a transthoracic echocardiogram was nor-

DOI 10.1002/ajmg.a.35409

© 2012 Wiley Periodicals, Inc.

^{*}Correspondence to:

Jodi Warman Chardon, MD, MSc, Department of Genetics, Children's Hospital of Eastern Ontario, 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada. E-mail: jwarman@cheo.on.ca Article first published online in Wiley Online Library (wileyonlinelibrary.com): 20 April 2012

WARMAN CHARDON ET AL.

mal with no valvular or aortic abnormality. On laboratory examination, there were no hematological abnormalities. Bilateral ventricular heterotopia was present on MRI (Fig. 1A), in addition to probable post-traumatic changes in the left temporal lobe. EEG demonstrated mild to moderate diffuse disturbance of cerebral activity with some possible epileptiform activity in the left temporal lobe. Based on the clinical context and diagnostic imaging, there was a high index of suspicion for a mutation in FLNA but complete gene sequencing was unrevealing. Whole-genome array CGH was performed using a custom-designed oligonucleotide array, with a probe placed, on average, every 35 kb and every 5-20 kb in more than 150 targeted regions of clinical significance. Analysis of DNA from this patient revealed an apparently mosaic deletion with a minimum size of 34.5 kb (chrX:153576890-153611431) and a maximum size of 67.6 kb (chrX:153560505-153628073), based on the GRCh37 genome build, and included both the entire FLNA and EMD genes (Fig. 1B). The log2 ratios for probes in FLNA showed an approximate value of -0.5, indicating a mosaic deletion. Taqman-based quantitative PCR analysis with a probe targeted to FLNA confirmed a copy number of more than one but less than two, suggesting the presence of a mosaic deletion. This result was also confirmed with a targeted array that contained probes in every exon of FLNA. Patient 1 had no clinical features suggestive of EDMD consistent with this being an X-linked disorder and that most female carriers are generally asymptomatic [Emery, 2000].

Patient 2 is a 12 year-old girl born to unrelated healthy parents and was referred for intellectual disability, spastic hemiparesis, and epilepsy associated with brain MRI features of unilateral perisylvian stroke (Fig. 2A) and PNH (Fig. 2B). Her mother experienced five spontaneous miscarriages. She was the product of a twin pregnancy characterized by a fetal loss and was delivered by cesarean section for fetal distress. Growth parameters were normal. Spastic hemiplegia was diagnosed at 9 months. Psychomotor development was delayed, with sitting achieved at 12 months and walking at 27 months of age. Seizures occurred at 3 years of age and included absence seizures and clonic seizures. Her IQ score (WISC-IV) was 47 with homogeneous cognitive functioning and she demonstrated behavioral difficulties. Clinical examination showed mild left-sided hemiparesis, normal OFC, multiple limb bruises, and absence of cardiac abnormalities or features of Ehlers-Danlos syndrome. Echocardiography revealed a patent foramen ovale and a bicuspid aortic valve. Laboratory testing indicated persistent macrothrombocytopenia (platelet count 130-150 giga/l, average volume 11.8 fl). Brain MRI showed atrophy of the right cerebral hemisphere compatible with a diagnosis of perinatal arterial stroke and multiple PNH in both hemispheres. Sequencing of the FLNA gene identified a heterozygous c.1286C>T (p.T429M) variant inherited from her father and thus considered a benign polymorphism. Chromosome microarray analysis using the Illumina HumanCytoSNP-12 v2.1 chip revealed a deletion with a minimum size of 44.9 kb (chrX:153540977-153585854) and a maximum size of 56.2 kb (chrX:153535981-153592201), based on the GRCh37 genome build. This deletion included the 3' coding and non-coding region of FLNA and the 3' coding and non-coding region of the adjacent gene TKTL1.

Filamin A is a large ubiquitous actin-binding protein implicated in cytoskeletal organization [Robertson, 2005]. Mutations in *FLNA* explain virtually all X-linked familial PNH and 26% of sporadic cases in females with classic bilateral PNH [Parrini et al., 2006]. Deletions and duplications of *FLNA* have been exceptionally reported as causing PNH; each on a single occasion [Fink et al., 1997; Nurden et al., 2011]. Somatic mosaicism in patients with *FLNA* intragenic mutations has been reported to cause PNH and may result in milder disease phenotypes [Parrini et al., 2004; Masurel-Paulet et al., 2011]. We report an affected individual (Patient 1) with a mosaic deletion of the *FLNA* and *EMD* genes. The Xq28 cytogenetic band contains several loci at which localized segmental duplications exist and predispose to disease-causing rearrangements [Naylor et al., 1996; Aradhya et al., 2001]. The



FIG. 1. Imaging and microarray findings for patient 1. A: MRI FLAIR imaging demonstrating bilateral contiguous periventricular nodular heterotopia. B: Mosaic deletion detected by array CGH. The arrow represents the oligonucleotide probes positioned over the FLNA-EMD region and show a deviation in log2 ratio compared to the adjacent black probes that are aligned on the baseline and represent normal log2 ratio or copy number.

1513

AMERICAN JOURNAL OF MEDICAL GENETICS PART A



FIG. 2. Imaging findings for patient 2. A: T2-weighted brain MRI showing atrophy of the right hemisphere compatible with perinatal middle cerebral artery stroke. B: Bilateral periventricular nodular heterotopia (arrows) demonstrated on T2-weighted brain MRI.

complete deleted region described here is flanked by 11.3-kb inverted repeat sequences that are >99% homologous [Small et al., 1997]. These homologous sequences have been previously characterized and shown to cause a complex rearrangement leading to a unique deletion of the EMD gene in a patient with EDMD as well as a partial duplication of the FLNA [Small et al., 1997]. A second report described two additional deletions at this locus that disrupted EMD only and not FLNA [Small and Warren, 1998]. Our report is the first to describe a deletion that removes both genes due to what appears to be illegitimate non-allelic homologous recombination between the repeat sequences. Although the repeat sequences are in opposing orientation, and are therefore expected to confer susceptibility to inversions rather than deletions and reciprocal duplications, precedence exists for deletion mutations to arise from recombination between local inverted repeat sequences (e.g., deletion of F8 in cases of hemophilia A [Kaufman et al., 2006] and of IKBKG in some patients with incontinentia pigmenti [Fusco et al., 2009]). Directly oriented sequence microhomology within inverted repeats might account for this phenomenon.

The occurrence of a perinatal ischemic stroke in patient 2 may be related to mutation in *FLNA*. Middle cerebral artery strokes have been reported in three female patients with *FLNA* mutation and a PNH phenotype [Fox et al., 1998] and in one patient with a Melnick–Needles phenotype [Foley et al., 2010], and rupture of carotid aneurysm with intracerebral hemorrhage has been reported in an older female patient [de Wit et al., 2009]. Furthermore, structural anomalies of large and small vessels, including in the brain, are found in FLNA null-mice [Feng et al., 2006]. Cardiac valvular malformations and patent foramen ovale are possible factors having contributed to the stroke in patient 2, however, data from other patients with *FLNA* mutations and FLNA-nullmice suggest abnormal vasculature may have played a role. Vascular anomalies in FLNA-related disorders may not be restricted to large and middle vessels and malformed brain microvessels may be associated with the development of PNH [Kakita et al., 2002]. In this respect, brain damage caused by a head trauma in patient 1 and in a previously reported patient [Kakita et al., 2002] could possibly result from fragility of the brain microvasculature, a hypothesis that will require further evaluation.

Features of Ehlers-Danlos syndrome, including aortic dilatation, bruising, and joint hypermobility, have been reported in patients with PNH [Sheen et al., 2005] but were absent in the patients reported here. Cardiovascular abnormalities have been associated with FLNA mutations, including dysplastic pulmonary valve and clefting of the mitral valve [Jefferies et al., 2010], bicuspid aortic valve (also identified in patient 2), and patent ductus arteriosus [Fox et al., 1998]. Thus, patients with PNH may require routine structural cardiac or aortic imaging to screen for cardiovascular abnormalities associated with PNH, particularly if presenting with Ehlers-Danlos phenotype. In addition, patient 2 demonstrated macrothrombocytopenia, and multiple bruises. Platelet abnormalities and bleeding diatheses have been reported with patients with FLNA mutations, presenting solely with bleeding disorders [Nurden et al., 2011]. Therefore, detailed hematological investigations should be considered for patients presenting with bruising or other manifestations of platelet dysfunction such as epistaxis.

We report on two patients with bilateral PNH, seizures, and intellectual disability with novel deletions involving the *FLNA* gene supporting loss-of-function of Filamin A as the pathogenesis of PNH. To our knowledge, these two patients are the second and third individuals reported with this disease mechanism. Our observations suggest that evaluation of *FLNA* copy number should be considered when assessing patients with PNH with negative *FLNA* gene sequencing.

ACKNOWLEDGMENTS

We thank the families for their participation. K.M.B. is supported by a Clinical Investigatorship Award from the Canadian Institutes of Health Research - Institute of Genetics.

REFERENCES

- Andrade D. 2009. Genetic basis in epilepsies caused by malformations of cortical development and in those with structurally normal brain. Hum Genet 126:173–193.
- Aradhya S, Bardaro T, Galgóczy P, Yamagata T, Esposito T, Patlan H, Ciccodicola A, Munnich A, Kenwrick S, Platzer M, D'Urso M, Nelson DL. 2001. Multiple pathogenic and benign genomic rearrangements occur at a 35kb duplication involving the NEMO and LAGE2 genes. Hum Mol Genet 10:2557–2567.
- Bione S, Maestrini E, Rivella S, Mancini M, Regis S, Romeo G, Toniolo D. 1994. Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. Nat Genet 8:323–327.
- Cardoso C, Boys A, Parrini E, Mignon-Ravix C, McMahon JM, Khantane S, Bertini E, Pallesi E, Missirian C, Zuffardi O, Novara F, Villard L, Giglio S, Chabrol B, Slater HR, Moncla A, Scheffer IE, Guerrini R. 2009. Periventricular heterotopia, mental retardation, and epilepsy associated with 5q14.3–q15 deletion. Neurology 72:784–792.
- Clark AR, Sawyer GM, Robertson SP, Sutherland-Smith AJ. 2009. Skeletal dysplasias due to filamin A mutations result from a gain-of-function mechanism distinct from allelic neurological disorders. Hum Mol Genet 18:4791–4800.
- de Wit MC, Kros JM, Halley DJ, de Coo IF, Verdijk R, Jacobs BC, Mancini GM. 2009. Filamin A mutation, a common cause for periventricular heterotopia, aneurysms and cardiac defects. JNNP 80:426–428.
- Dobyns WB, Andermann E, Andermann F, Czapansky-Beilman D, Dubeau F, Dulac O, Guerrini R, Hirsch B, Ledbetter DH, Lee NS, Motte J, Pinard JM, Radtke RA, Ross ME, Tampieri D, Walsh CA, Truwit CL. 1996. X-Linked malformations of neuronal migration. Neurology 47: 331–339.
- Ekşioğlu YZ, Scheffer IE, Cardenas P, Knoll J, DiMario F, Ramsby G, Berg M, Kamuro K, Berkovic SF, Duyk GM, Parisi J, Huttenlocher PR, Walsh CA. 1996. Periventricular heterotopia: An X-linked dominant epilepsy locus causing aberrant cerebral cortical development. Neuron 16:77–87.
- Emery AE. 2000. Emery-Dreifuss muscular dystrophy A 40 year retrospective. Neuromuscul Disord 10:228–232.
- Ferland RJ, Gaitanis JN, Apse K, Tantravahi U, Walsh CA, Sheen VL. 2006. Periventricular nodular heterotopia and Williams syndrome. Am J Med Genet Part A 140A:1305–1311.
- Feng Y, Chen MH, Moskowitz IP, Mendonza AM, Vidali L, Nakamura F, Kwiatkowski DJ, Walsh CA. 2006. Filamin A (FLNA) is required for cell-cell contact in vascular development and cardiac morphogenesis. Proc Natl Acad Sci USA 103:19836–19841.
- Fink JM, Dobyns WB, Guerrini R, Hirsch BA. 1997. Identification of a duplication of Xq28 associated with bilateral periventricular nodular heterotopia. Am J Hum Genet 61:379–387.
- Foley C, Roberts K, Tchrakian N, Morgan T, Fryer A, Robertson SP, Tubridy N. 2010. Expansion of the spectrum of FLNA mutations associated with Melnick-Needles syndrome. Mol Syndromol 1:121– 126.
- Fox JW, Lamperti ED, Ekşioğlu YZ, Hong SE, Feng Y, Graham DA, Scheffer IE, Dobyns WB, Hirsch BA, Radtke RA, Berkovic SF, Huttenlocher PR, Walsh CA. 1998. Mutations in filamin 1 prevent migration of cerebral

cortical neurons in human periventricular heterotopia. Neuron 21: 1315-1325.

- Fusco F, Paciolla M, Pescatore A, Brigida Lioi M, Ayuso C, Faravelli F, Gentile M, Zollino M, D'Urso M, Giuseppina Miano M, Valeria Ursini M. 2009. Microdeletion/Duplication at the Xq28 IP locus causes a de novo IKBKG/NEMO/IKKgamma exon4_10 deletion in families with Incontinentia Pigmenti. Hum Mutat 30:1284–1291.
- Jefferies JL, Taylor MD, Rossano J, Belmont JW, Craigen WJ. 2010. Novel cardiac findings in periventricular nodular heterotopia. Am J Med Genet 152:165–168.
- Kakita A, Hayashi S, Moro F, Guerrini R, Ozawa T, Ono K, Kameyama S, Walsh CA, Takahashi H. 2002. Bilateral periventricular nodular heterotopia due to filamin 1 gene mutation: Widespread glomeruloid microvascular anomaly and dysplastic cytoarchitecture in the cerebral cortex. Acta Neuropathol 104:649–657.
- Kaufman RJ, Antonarakis SE, Fay PJ. 2006. Factor VIII and hemophilia A. In: Colman RW, Hirsh J, Marder VJ, Clowes AW, George JN, editors. Hemostasis and thrombosis: Basic principles and clinical practice, 5th edition. Philadelphia, PA: Lippincott - Raven. pp 151–175.
- Masurel-Paulet A, Haan E, Thompson EM, Goizet C, Thauvin-Robinet C, Tai A, Kennedy D, Smith G, Yee Khong T, Solé G, Guerineau E, Coupry I, Huet F, Robertson S, Faivre L. 2011. Lung disease associated with periventricular nodular heterotopia and a FLNA mutation. Eur J Med Gen 54:25–28.
- Muchir A, Worman H. 2007. Emery-Dreifuss muscular dystrophy. Curr Neurol Neurosci Rep 7:78–83.
- Naylor JA, Nicholson P, Goodeve Anne, Hassock S, Peake I, Gianelli F. 1996. A novel DNA inversion causing severe hemophilia A. Blood 87: 3255–3261.
- Nurden P, Debili N, Coupry I, Bryckaert M, Youlyouz-Marfak I, Solé G, Pons AC, Berrou E, Adam F, Kauskot A, Lamazière JM, Rameau P, Fergelot P, Rooryck C, Cailley D, Arveiler B, Lacombe D, Vainchenker W, Nurden A, Goizet C. 2011. Thrombocytopenia resulting from mutations in filamin A can be expressed as an isolated syndrome. Blood 118: 5928–5937.
- Parrini E, Mei D, Wright M, Dorn T, Guerrini R. 2004. Mosaic mutations of the FLN1 gene cause a mild phenotype in patients with periventricular heterotopia. Neurogenetics 5:191–196.
- Parrini E, Ramazzotti A, Dobyns WB, Mei D, Moro F, Veggiotti P, Marini C, Brilstra EH, Dalla Bernardina B, Goodwin L, Bodell A, Jones MC, Nangeroni M, Palmeri S, Said E, Sander JW, Striano P, Takahashi Y, Van Maldergem L, Leonardi G, Wright M, Walsh CA, Guerrini R. 2006. Periventricular heterotopia: Phenotypic heterogeneity and correlation with filamin A mutations. Brain 129:1892–1906.
- Robertson SP. 2005. Filamin A: Phenotypic diversity. Curr Opin Genet Dev 3:301–307.
- Robertson SP, Twigg SR, Sutherland-Smith AJ, Biancalana V, Gorlin RJ, Horn D, Kenwrick SJ, Kim CA, Morava E, Newbury-Ecob R, Orstavik KH, Quarrell OW, Schwartz CE, Shears DJ, Suri M, Kendrick-Jones J, Wilkie AO. 2003. Localized mutations in the gene encoding the cytoskeletal protein filamin A cause diverse malformations in humans. Nat Genet 33:487–491.
- Sheen VL, Wheless JW, Bodell A, Braverman E, Cotter PD, Rauen KA, Glenn O, Weisiger K, Packman S, Walsh CA, Sherr EH. 2003. Periventricular heterotopia associated with chromosome 5p anomalies. Neurology 60:1033–1066.
- Sheen VL, Ganesh VS, Topcu M, Sebire G, Bodell A, Hill RS, Grant PE, Shugart YY, Imitola J, Khoury SJ, Guerrini R, Walsh CA. 2004. Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex. Nat Genet 36:69–76.
- Sheen VL, Jansen A, Chen MH, Chen MH, Parrini E, Morgan T, Ravenscroft R, Ganesh V, Underwood T, Wiley J, Leventer R,

AMERICAN JOURNAL OF MEDICAL GENETICS PART A

Vaid RR, Ruiz DE, Hutchins GM, Menasha J, Willner J, Geng Y, Gripp KW, Nicholson L, Berry-Kravis E, Bodell A, Apse K, Hill RS, Dubeau F, Andermann F, Barkovich J, Andermann E, Shugart YY, Thomas P, Viri M, Veggiotti P, Robertson S, Guerrini R, Walsh CA. 2005. Filamin A mutations cause periventricular heterotopia with Ehlers–Danlos syndrome. Neurology 64:254–262.

Small K, Iber J, Warren ST. 1997. Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. Nat Genet 16:96–99.

Small K, Warren ST. 1998. Emerin deletions occurring on both Xq28 inversion backgrounds. Hum Mol Genet 7:135–139.

Keren Boris – Thèse de doctorat - Année 2013

Article 3

Downloaded from jmg.bmj.com on June 28, 2013 - Published by group.bmj.com

Chromosomal rearrangements

ORIGINAL ARTICLE

ABSTRACT

Intragenic CAMTA1 rearrangements cause non-progressive congenital ataxia with or without intellectual disability

Julien Thevenon,^{1,2} Estelle Lopez,² Boris Keren,^{3,4} Delphine Heron,⁵ Cyril Mignot,⁵ Cecilia Altuzarra,⁶ Mylène Béri-Dexheimer,⁷ Céline Bonnet,⁷ Eloi Magnin,⁸ Lydie Burglen,⁹ Delphine Minot,¹ Jacqueline Vigneron,¹⁰ Sophie Morle,⁶ Mathieu Anheim,⁵ Perrine Charles,⁵ Alexis Brice,^{3,4} Louise Gallagher,¹¹ Jeanne Amiel,¹² Emmanuel Haffen,⁸ Corinne Mach,³ Christel Depienne,^{3,4} Diane Doummar,¹³ Marlène Bonnet,¹⁴ Laurence Duplomb,² Virginie Carmignac,² Patrick Callier,^{2,15} Nathalie Marle,¹⁵ Anne-Laure Mosca-Boidron,¹⁵ Virginie Roze,¹⁶ Bernard Aral,¹⁷ Ferechte Razavi,¹⁸ Philippe Jonveaux,⁸ Laurence Faivre,^{1,2} Christel Thauvin-Robinet^{1,2}

For numbered affiliations see end of article

Correspondence to Dr Christel Thauvin-Robinet, Hôpital d'Enfants, Centre de Génétique, 10 Bd Maréchal de Lattre de Tassigny, Dijon 21079, France; christel.thauvin@chu-dijon.fr

Received 28 February 2012 Bevised 10 April 2013 Accepted 12 April 2012

Background Non-progressive congenital ataxias (NPCA) with or without intellectual disability (ID) are clinically and genetically heterogeneous conditions. As a consequence, the identification of the genes responsible for these phenotypes remained limited. Objective Identification of a new gene responsible for

NPCA and ID. Methods Following the discovery of three familial or sporadic cases with an intragenic calmodulin-binding transcription activator 1 (CAMTA1) rearrangement identified by an array-CGH and recruited from a national collaboration, the authors defined the clinical and molecular characteristics of such rearrangements, and searched for patients with point mutations by direct sequencing. Results Intragenic copy number variations of CAMTA1 were all located in the CG-1 domain of the gene. It segregated with autosomal dominant ID with nonprogressive congenital cerebellar ataxia (NPCA) in two unrelated families, and was de novo deletion located in the same domain in a child presenting with NPCA. In the patients with ID, the deletion led to a frameshift, producing a truncated protein, while this was not the case for the patient with isolated childhood ataxia. Brain MBI of the patients revealed a pattern of progressive atrophy of cerebellum medium lobes and superior vermis, parietal lobes and hippocampi. DNA sequencing of the CG-1 domain in 197 patients with sporadic or familial nonsyndromic intellectual deficiency, extended to full DNA sequencing in 50 patients with ID and 47 additional patients with childhood ataxia, identified no pathogenic mutation

Conclusion The authors have evidence that loss-offunction of CAMTA1, a brain-specific calcium responsive transcription factor, is responsible for NPCA with or without ID.

Accession numbers CAMTA1 reference sequence used was ENST00000303635. Protein sequence was ENSP00000306522

INTRODUCTION

Clinically and genetically heterogeneous conditions such as childhood ataxia or intellectual disability (ID) hindered the elucidation of those groups of pathologies. Childhood ataxias comprise a group of neurological signs secondary to a dysfunction of the cerebellum and its connections (cerebellar ataxia) and/or to problems in the posterior column of the spinal cord or peripheral nerves (sensory ataxia). Non-progressive congenital cerebellar ataxia (NPCA) manifests itself in infancy as hypotonia, delayed motor development and ataxia with or without dysarthria.1 ID and spasticity can also be associated with NPCA. NPCA is presumably monogenic in over half of the cases with a mainly autosomal recessive mode of inheritance. Structural abnormalities of the cerebellum are not always present on neuroimaging.

ID is the most prevalent severe handicap in children, affecting 1%-3% of the population.3 ID is grouped into syndromic ID (S-ID) and non-syndromic ID (NS-ID). ID can be caused by environmental and/or genetic factors. The high degree of heritability of ID has been highlighted, in particular with the estimation that up to 50% of severe cases are caused by genetic abnormalities.⁴ Genetic causes include various chromosomal rearrangements diagnosed by conventional or array-based techniques, with a broad range of prevalence,5 as well as mutations of numerous autosomal or X linked genes.8 It has recently been shown that de novo mutations or intragenic rearrangements were a major cause of sporadic ID.⁹ To date, 15 genes have been identified in autosomal dominant NS-ID by candidate genes strategies (SYNGAP1, KIF1A, GRIN1, EPB41L1 and CACNG2),¹⁰¹¹ by mapping breakpoints of chromosomal rearrangement (DOCK8, CDH15, KIRREL3 and ARID1B)^{12 13}, by microarrays indicating de novo copy number varia-tions (CNVs) (*MBD5* and *SHANK2*)¹⁴ ¹⁵ and by exome sequencing (DYNC1H1, YY1, DEAF1, CIC).

400

Chromosomal rearrangements

Their encoded proteins exert highly specific functions in neuronal connectivity and plasticity. $^{\!\!8}$ 16

The calmodulin-binding transcription activator 1 (CAMTA1 gene) is a member of the family of calmodulin-binding transcriptional activators (CAMTAs), and has been shown to be conserved in all organisms from multicellular eukaryotes to humans. The *CAMTA1* gene, first identified in a human brain-specific CDNA cloning study, belongs to the human CAMTA family¹⁷ defined by four conserved protein domains, including the family-specific DNA-binding domain termed CG-1, which contains a bipartite nuclear localisation signal with transcriptional regulation properties, a transcription factor immunoglobulin-like DNA-binding domain), calmodulin-binding IQ motifs (IQ domain) and ankyrin repeats necessary for oligomerisation (ANK domain).^{18–23}

In this article, we report on three independent families with NPCA with or without ID, secondary to intragenic *CAMTA1* rearrangements disrupting the CG-1 domain.

SUBJECTS

Based on the identification of an intragenic rearrangement of the *CAMTA1* gene in family 1 identified by an array-CGH, a national call for collaboration was performed to reinforce the hypothesis of the implication of this gene in ID. Among the 23 000 array-CGH analyses performed in France for the work-up of developmental delay and/or intellectual deficiency, two additional families were recruited. Family pedigrees and patients' pictures are presented in figure 1. The patients presented with dysmorphic facial features that are detailed in figure 1. Major clinical and neurological findings are summarised in table 1. A specific consent for research investigations was accepted by each patient or legal representative. The research protocol was approved by the local ethic committee.

Family 1

Patient III-6 was referred to the genetics centre for delayed psychomotor development. She was the second child of nonconsanguineous parents with a family history of ID. She had

Figure 1 Pedigrees and photographs of families 1–3. Intellectual disability is represented by black boxes, and cosegregates with the family. Note some common facial features including a long face with a pointed chin, bulbous nose with anteverted nostrils, a long philtrum and a thick lower lip with down-turned corners of the mouth. speech delay with normal social interactions. Standard chromosomal analysis revealed 47,XXX aneuploidy. The brain CTscan showed no abnormality. Her four-and-a-half-year-old sister (III-7) presented with right eye strabismus, unsteady gait and behavioural difficulties. She was unable to speak. Her 2-year-old brother (III-8) presented with neonatal global hypotonia followed by developmental delay. He could walk unaided but with an ataxic gait at 22 months of age. Upper arm myoclonic epilepsy was noticed by the mother. The EEG and brain CT-scan were normal.

The mother (II-6) also presented with ID without behavioural abnormalities and went to schools for students with special needs. She currently works in a job reserved for the disabled. Clinical examination revealed an ataxic gait and mild intentional tremor. There were no lateralised falls, Romberg signs or nystagmus. Neuropsychological evaluation revealed mild ID with homogeneous performances in verbal and non-verbal processes, and poor abstract verbal reasoning. The PRI score was 66-73 revealing borderline intellectual abilities (poor spatial perception, visual abstract processing, non-verbal abstract problem solving and non-verbal reasoning). Low-speed processing and problems with selective attention were noticed (speed processing index =61; normal =100) with an impact on concentration when mental mathematical problems were considered. There was dissociation between memory processes in the auditory-verbal and visual modality, indicating a lack of rehearsal processing in episodic visual memory. Results showed a working memory index at 50, indicating short-term memory disabilities predominantly on auditory memory modality. Brain MRI showed mild hippocampus atrophy, simplified gyration of the dentate gyri, and bilateral posterior cortical atrophy including the parietal lobes and the precuneus and cuneus. There was a cerebellar atrophy of the medium lobes and of the superior vermis (figure 2A). An 18FDG PET scan revealed a hypometabolism of the left temporal region and temporal medial lobe, biparietal lobe and the precuneus/cuneus (figure 2B).

Patient II-4, a half-sister of patient II-6, attended a special needs school and worked as a non-qualified operator. The clinical



J Med Genet 2012;49:400-408. doi:10.1136/jmedgenet-2012-100856

| | Family 1 | | | | | | | Famiry 2 | | | |
|---|---------------------------------|---------------|----------------------|------------------------------|--------------|------------------------------|---------------|---------------------|---------------|-------------|---------------------------------------|
| | 9-111 | 11-7 | 8-111 | 11-6 | II-4 | III-3 | 111-4 | E | II-2 | 2 | Family 3 |
| Development history | | | | | | | | | | | |
| Term of gestation of birth | 38 | 40 | 39 | NR | NR | Term | 36 + 5 | 38 | 38 | NR | |
| Measurements at birth | Z | z | z | NR | z | z | z | N | Z | NR | IUGR |
| Neonatal hypotonia | No | No | Yes | NR | NR | No | No | N | Yes | NR | Yes |
| Sitting age (months) | 6 | 11 | 10 | NR | NR | 10 | NR | NR | NR | NR | 6 |
| Walking age (months) | 18 | 24 | 20 (with assistance) | NR | NR | 18 | 22 | 24 | 18 | Delayed | 21 |
| First word association (months) | 48 | Absent | Absent | NR | NR | Delayed | Delayed | 36 | 36 | NR | Delayed |
| Clinical examination | | | | | | | | | | | |
| Age at clinical examination (years old) | 9 | 4 | 2 | 41 | 38 | 18 | 14 | 16 | 12 | 42 | 4 |
| Growth parameters | Z | z | z | z | z | z | Z | Macrocephaly | z | z | z |
| Facial dysmorphism | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Neurological examination | | | | | | | | | | | |
| Cerebellar symptoms | | | | | | | | | | | |
| Ataxic gait | Yes | No | Yes | No | No | Yes | No | Yes | No | Yes | Yes |
| Dysmetry | Mild | No | Yes | No | No | DiiM | No | Mild | No | Mild | Yes |
| Instability | Yes | Yes | Yes | Mild | Mild | Mild | Mild | Yes | Mild | Yes | Yes |
| Dysarthria/abnormal pronunciation | Yes | NR | NR | Yes | Yes | No | Yes | No | No | No | Yes |
| SARA score (/56) | 6 | 6 | 6 | 9 | 9 | 8 | 4 | 6 | 8 | 9 | 6 |
| Standard examination | | | | | | | | | | | |
| Nerve sensitivity | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| Tendon reflexes | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Polykinetic | Normal | Normal | Normal |
| Pyramidal symptoms | No | No | No | No | No | No | No | Yes | No | No | No |
| Tremor | No | No | No | No | No | No | No | No | No | No | No |
| Vestibular symptoms | | | | | | | | | | | |
| Nystagmus | No | No | No | No | No | No | No | No | No | No | No |
| Romberg sign | No | No | No | No | No | No | No | No | No | No | No |
| Convulsions | No | No | 2 febrile crisis | No | No | No | No | No | No | No | No |
| Strabismus | Yes | No | No | No | No | No | No | No | No | No | No |
| Neuropsychological evaluation (IQ, VIQ, PIQ) | 58, 52, 57 | 67, 78, 61 | NR | 58, 59, 58 | 65, 66, 68 | 51 | 61, 76, 65 | 40, 50, 50 | 53, 55, 56 | 54, 52, 60 | Dissociation between verbal and |
| | | | - | | | | | - | | - | non-verbal skills |
| Scholarship/job | Special needs | Special needs | NN | Job reserved for disabled | Not working | Job reserved for disabled | Special needs | Special needs | Special needs | Not working | Special needs |
| Behavioural troubles | Stereotypies, aggressiveness | No | No | No | No | Solitary | No | ASD, aggressiveness | No | No | No |
| Brain MRI | NR | NR | NR | Pathological | Pathological | NR | Abnormal | NR | NR | NR | Normal |

Keren Boris – Thèse de doctorat - Année 2013

Downloaded from jmg.bmj.com on June 28, 2013 - Published by group.bmj.com

Ch

402

J Med Genet 2012;49:400-408. doi:10.1136/jmedgenet-2012-100856

Chromosomal rearrangements

Figure 2 Brain imaging. (A) MRI examination of patient II-4 (48-yearsold), II-6 (41-years-old), III-4 (14-yearsold) from family 1 and patient from family 3 (4-years-old). A progressive pattern could be identified showing mild progressive cerebellar atrophy of the superior vermis and medium cerebellar lobes (star), marked bilateral posterior cortical atrophy including the parietal lobes (*), precuneus and cuneus (#). Mild bilateral hippocampal atrophy (Scheltens's medial temporal lobe atrophy scale: grade 2) and simplified gyration of the dentate gyri were noticed (arrows). No white matter abnormality was identified. (B) PET scan (radiotracer: FDG) showing the left temporal area including the left temporal medial lobe the biparietal lobe and the precuneus/cuneus hypometabolism (arrows).



examination revealed static instability associated with mild ataxic gait, with no evidence of vestibular dysfunction. Brain MRI showed a similar pattern of mild hippocampus atrophy, simplified gyration of the dentate gyri, and bilateral posterior cortical atrophy including the parietal lobes and the precuneus and cuneus. Cerebellum showed atrophy of the medium lobes and of the superior vermis (figure 2A). Her elder son (patient III-3), a 23-year-old boy, was described as hypotonic from birth with psychomotor delay but no behavioural troubles although he is rather solitary. He attended a special needs school and now works at a supported employment. He acquired reading, writing and lives independently. He presented with left eye strabismus and an ataxic gait. Her younger son (patient III-4) also presented with psychomotor delay. He attended a special needs school from 10 years of age. He followed speech and language therapy for delayed receptive language. Brain MRI showed no supratentorial atrophy, but a mild atrophy of the superior vermis (figure 2A).

The maternal grandmother (patient I-5) had limited schooling and took up early employment in manual jobs.

J Med Genet 2012;49:400-408. doi:10.1136/jmedgenet-2012-100856

Family 2

The proband (II-1), a 16-year-old male, was the first child of nonconsanguineous parents (figure 1). He had delayed motor skills (walked at 24 months) and speech delay (first words at 36 months). He attended a special needs school, learnt to write and read simple sentences. He is now following a vocational course to work in a job for people with disabilities. He shows neuropsychological diagnosed mild ID, associated with autistic spectrum manifestations, poor interactive skills and bursts of aggressiveness. The clinical examination noted a large forehead, palpebral oedema, enophtalmia, a wide flat nose, short ears, a small mouth and abnormally implanted teeth. His younger sister (II-2) presented with neonatal global hypotonia and mild psychomotor delay. She walked at 14 months with an ataxic gait and has normal behaviour. Speech was delayed, with the first words at 24 months and poor elocution. Neuropsychological evaluation revealed mild ID. She attended a special educational institution. She had similar facial features to her brother.

Their mother (I-1) also presented with mild ID, attended a special needs school and is currently unemployed. The clinical

403

Chromosomal rearrangements

examination revealed an ataxic gait with static instability and mild dysmetry. No vestibular signs were noticed. Neuropsychological examination revealed a mild ID. She comes from a large family of 14 brothers and sisters with no other case of ID. Her parents were deceased.

Family 3

The proband was the first child of healthy unrelated parents. Hypotonia and motor delay were noted from birth. Ataxia was observed at 17 months; she walked independently at 21 months. The clinical evaluation at 22 months showed an interactive child with good understanding abilities. She was able to build a stack of three cubes, scribbled, played symbolic games and spoke about 10 words. She had neither dynamic cerebellar signs nor motor deficit but walked with a wide-based gait with occasional falls. The ataxia improved and became visible only during tandem walking at 4 years. She was able to draw a square, presented mild speech difficulties but spoke well-structured sentences. This was confirmed by the neuropsychological evaluations performed at 4 years and 2 months: (1) WPPSI-III testing showed normal non-verbal skills (blocks design 10:19, matrix reasoning 12/19, pictures concept 10:19) and deficient verbal skills due to defective articulation and fluency and (2) Nepsy testing showed normal sensorimotor but mildly altered visuospatial (design copying) functions.

Cohorts of patients with sporadic or autosomal dominant $\ensuremath{\text{ID}}\xspace$ childhood ataxia

Two cohorts of patients were assembled. Group 1 included 150 male or female patients with NS-ID, including 25 (17%) with a family history of ID compatible with an autosomal dominant mode of inheritance. All patients had normal standard chromosomal analysis and half of the series had a normal array-CGH. When performed, cerebral MRI as well as plasma amino acids and urinary organic acid chromatography showed no abnormalities. Group 2 included 47 patients with non-progressive and early-onset childhood ataxia, with (30 patients) or without ID (17 patients). Informed consent was obtained for all tested patients.

METHODS Array-CGH/SNP

DNA was extracted from peripheral blood lymphocytes. Patients issued from families 1 and 2 were studied using the Human Genome CGH Microarray 105K (Agilent Technologies, Santa Clara, California, USA). Data were processed with feature extraction (V.9.1) software and the results were analysed with CGH analytics (V.4.0) software (Agilent®) in the Hg19 genome assembly. The CNVs identified through an array-CGH were confirmed by quantitative PCR or FISH analysis. For family 3, chromosome microarray analysis was performed on the proband and her parents using Illumina HumanCytoSNP-12 v2.1 chips. All of the procedures were carried out according to the manufacturer's protocol. Raw data were analysed using Illumina GenomeStudio v2011.1 software. To determine the log ratios and the B allele frequencies of every SNP, we generated our reference by using the GeneTrain 2.0 cluster algorithm with 95 other samples processed in the same run. CNVs were called with the cnvPartition v2.4.4 algorithm (Genecall Threshold of 35), and also by visual analysis on the GenomeStudio Genome Viewer. Mapping data were analysed on the human genome sequence using ensemble website, with Hg18 rebuild. CNVs were assessed in the Database of Genomic Variants.

404

Molecular impact of the CAMTA1 deletion Skin fibroblasts

After informed consent was obtained, primary fibroblasts were harvested from skin biopsy from the patient II-6 (family 1) and two healthy controls. The cells were cultured in DMEM (HyClone) containing 10% fetal bovine serum (HyClone) and 1% penicillin/streptavidin (PromoCell). For each sample, total RNA was extracted from three independent culture wells using the RNeasy mini kit (Qiagen) following the manufacturer's recommendations and stored at -80° C.

CAMTA1 cDNA sequencing analysis and quantification

From the extracted mRNA, two experiments were realised. Total cDNA was synthetised from 1 μg of RNA using the Superscript III RT-PCR kit (Invitrogen) using usual procedures.

First, a reverse transcription with specific *CAMTA1* primers was performed to produce cDNA. PCR primers were chosen in exons 3 and 7 using PrimerBlast online software (PCR primers available on request). A standard Touchdown PCR protocol was used. The different amplified cDNA were sequenced to evidence the frameshift secondary to the exon 4 deletion.

The latter experiment was realised in order to quantify the transcription of the mutated allele. *CAMTA1* primers were designed using PrimerBlast online software (PCR primers and conditions available on request). *GAPDH* was used as the reference gene. Quantitative reverse transcriptase PCR (qRT-PCR) data were run on a LightCycler 480 (Roche) using the QuantiTect SYBR green PCR kit (Qiagen). Relative quantification was calculated using the Light Cycler 480 SW1.5 software.

In silico analysis

In patients with unavailable fibroblast cultures, in silico mRNA prediction was used. Deleted exons diagnosed by an array-CGH were removed from the reference sequence. Translation prediction was performed with a prediction tool. Alignments were performed with the clustalw2 program.

Genomic CAMTA1 sequencing analysis

Group 1

Twenty-three coding exons and intronic flanking regions of the CAMTA4 gene were amplified using a touchdown protocol in 50 patients from group 1 (n=150). Exons encoding the CG-1 region (exons 2–7) were sequenced for 100 additional patients (PCR primers and conditions available on request). PCR fragments were purified using the multiscreen Vacuum Manifold system (Millipore). Sequencing was performed using an ABI BigDye Terminator Cycle Sequencing kit (v3.1) (Applied Biosystems) in an ABI 3130 sequencer. Sequence data were analysed with SeqScape v2.7 (Applied Biosystems) and Sequencer v4.1.

Group 2

Thirty-six amplicons were designed using the GS FLX Titanium Fusion Primers tool from Integrated DNA Technology (primers available on request) (n=47). The library was prepared according to the Roche Amplicon Library Preparation Manual. GC Rich buffer 1X from the Roche Fastart DNA Polymerase kit was added for amplicon 1. Amplicons were pooled and sequenced on the Roche GS Junior in a single run, according to the Roche emPCR Amplification Method protocol followed by the GS Junior Sequencing Method protocol. The resulting sequences were analysed using GS AVA software.

Cerebral expression of CAMTA1 in mouse embryos

Cerebral expression of CAMTA1 was obtained from the brain map of the developing mouse referenced on the Allen institute

J Med Genet 2012;49:400-408. doi:10.1136/jmedgenet-2012-100856

Chromosomal rearrangements

website, and provided as free open resources for the global scientific community. $^{\rm 27}$

RESULTS

Array-CGH/SNP results

All of the rearrangements are reported in Hg18 and represented in figure 3. In family 2, the rearrangement was proved to be a tandem duplication by FISH analysis (probe RP11-334N17).

Molecular impact of the CAMTA1 rearrangements

In family 1, mRNAs obtained from cultured skin fibroblasts allowed a direct analysis of the transcriptional consequences of the intragenic rearrangement. Standard PCR amplification allowed the identification of four PCR products and cDNA fragments were sequenced separately (figure 3B). They included: (i) the native cDNA (named CAMTA1), (ii) an alternative spliced cDNA lacking exon 6 (named CAMTA1 Δ 6), (iii) a cDNA lacking exon 4 (named CAMTA1 Δ 4 Δ 6). The absence of exon 4

led to a frameshift (+1) and a subsequent premature stop codon at position 256 (figure 3B,C). In silico prediction of the cDNA sequences revealed that in family 3, the deletion of exons 2 and 3 was predicted not to alter the frame (figure 3C).

Using the fibroblasts from the CAMTA1 patient II-6 (family 1), we were able to quantify the normal, unmutated mRNA only by qRT-PCR. The quantity of mutated mRNA was lower than the sensitivity of the experiment could quantify (data not shown).

Genomic CAMTA1 sequencing analysis in the NS-ID/childhood ataxia cohort

No pathogenic *CAMTA1* mutation was identified in the coding sequence. One exonic missense sequence variation was identified. Prediction tools were in favour of a benign sequence variation. All of the other variants were intronic. All of the identified variations were previously known polymorphisms listed in control databases or synonymous mutations. No mutations in the exon—intron boundaries were found.



Figure 3 Molecular impact of the intragenic *CAMTA1* rearrangements in reported patients. (A) All the reported rearrangements implicated the CG-1 domain. An 81 Kb deletion at 1p36.31p36.23 encompassing exon 4 (arrCGH 1p36.31p36.23 (7119268–7200395)x1) was identified in family 1 and cosegregated with the disease. A 539 kb duplication at 1p36.31p36.23 (arrCGH 1p36.31p36.23 (6882372–7422115)x3) was identified in the two affected siblings of family 2, inherited from the mother (also affected). The rest of the family was not accessible for study. (A) A de novo 49 kb deletion at the 1p36.31 locus (arrSNP 1p36.31 (6777038–6826186)x1 dn) encompassing exons 2 and 3 was diagnosed in the proband of family 3 (figure 3A). (B) Analysis of the cDNA in fibroblasts revealed that the deletion of exon 4 caused a frameshift. (C) Protein products from the intragenic rearrangements of *CAMTA1* demonstrating a truncation (families 1 and 2), contrasting with the patient with childhood ataxia (family 3). The deletion of exons 3 and 4 (patient form Pinto *et al*⁷; family 1) both led to the production of a truncated protein and a premature termination codon respectively in positions 48 and 256.

J Med Genet 2012;49:400-408. doi:10.1136/jmedgenet-2012-100856

Chromosomal rearrangements

Cerebral expression of CAMTA1 during mouse development

Embryonic mice show a global brain expression of the CAMTA1 protein that became restricted to the cerebellum, hippocampi and olfactive bulbs later on in development (figure 4). At E15.5. CAMTA1 is expressed in the whole brain. In the forebrain, positivity is strong in the ventricular zone of the cerebral hemisphere and the olfactory recessus, fulfilled at this stage with migrating neuroblasts. In the midbrain, positivity is stronger in the dorsal wall. In the hindbrain, the first rhombomere shows positivity as well as the ventral neuroepithelium (figure 4A). At E18.5, positivity is marked in the cortical plate and reduced in the ventricular zone. The olfactory system and the cerebellar anlage express also positivity (figure 4B). At P4, positivity is found in the olfactory system and is weakly present in the superficial layer of the cortical plate and in the gyrus dentate of the hippocampus. In the cerebellum, the vermian external granular layer shows positivity (figure 4C). At P14, positivity is restricted to the olfactory system and also found in the internal granular layer of the vermis (figure 4C).

DISCUSSION

We report on three independent NPCA \pm ID families with an intragenic rearrangement affecting the CG-1 domain of the *CAMTA1* gene, arising de novo or segregating with the disease making *CAMTA1* a novel gene for S-ID. DNA sequencing analysis of CAMTA1 in a cohort of 197 patients with ID and/or NPCA failed to identify further mutations, indicating that micro-rearrangements are the most frequent mutations encountered in this gene.

All of the rearrangements affected the CG-1 domain of the protein and two-thirds were predicted to create a frameshift and a truncated protein (figure 3). There was no truncating mutation identified in more than 7000 exomes realised on patients without ID (NHLB1 Exome Sequencing Project data). By interrogating the Database of Genomics Variants, we found one intragenic CNV affecting the *CAMTA1* gene identified in a healthy individual.²⁸ The Q-PCR control of the deletion finally

defined that the CNV was purely intronic (S Scherer, personal communication). Apart from the identification of intragenic CAMTA1 deletions/duplication, other evidence supports the disease-causing role of the gene as: (i) the CAMTA1 gene appears to be predominantly expressed in the human brain¹⁷ and its integrity is necessary for visual response in the fruit fly, which suggests that it has a superior function in mammalians; (ii) the cellular functions of genes being under CAMTA1 transcriptional control suggest a putative role of CAMTA1 in promoting neuronal differentiation and are responsible for ID \pm ASD when genes such as SHANK3, CHRNA7, NRG1 and 2, CAMK4 and CASK are mutated; 30 31 and (iii) a replicated association between a CAMTA1 allele (rs4908449; p=0.0002) and differential performance in normal episodic memory, which adds CAMTA1 to the list of calcium-responsive proteins responsible for memory processes in humans.

Physiopathological mechanisms by which the loss of function of CAMTA1 leads to NPCA and ID remain to be determined. The function of the CAMTA proteins has been studied in various models.¹⁹ In Arabidopsis, CAMTAs located in the nucleus and interacted with double-stranded DNA and calmodulin and activate transcription in response to various cellular stresses.^{20 21} The CG-1 domain of *CAMTA1* is crucial for nuclear localisation and further transcription regulation in neuronal cells.^{23 31} The analysis of the CAMTA1 cDNA from cultured fibroblasts (family 1, patient II-6) allowed the identification of different transcriptional products (figure 3). The detection of the mutated cDNA by sequencing does not exclude the possibility of a residual truncated protein. However, in the patient's fibroblasts, the qRT-PCR was unable to quantify the mutated cDNA. This non-quantifiable cDNA suggested that nonsense-mediated mRNA decay could be activated and is in favour of haploinsufficiency causing the phenotype. This argument is reinforced by a transcriptomic analysis suggesting that haploinsufficiency of the *CAMTA1* gene could lead to a transcriptional misregulation of known genes implicated in neuronal differentiation and proliferation.³⁰ A dominant

Figure 4 CAMTA1 expression during mouse brain development. The inset shows the expression of CAMTA1 in the murine brain as detected by in situ hybridisation. Embryonic mice show a global brain expression of the CAMTA1 protein that became restricted to the cerebellum, hippocampi and olfactive bulbs later on in development. At E15.5, CAMTA1 is expressed in the forebrain (prosencephalon), midbrain (mesencephalon) and hindbrain (rhombencephalon). In the forebrain, positivity is strong in the ventricular zone of the cerebral hemisphere and the olfactory recessus (white arrow), fulfilled at this stage with migrating neuroblasts. In the midbrain, positivity is stronger in the dorsal wall (#). In the hindbrain, the first rhombomere shows positivity as well as the ventral neuroepithelium (star). At E18.5, marked positivity in the cortical plate, and reduced in the ventricular zone. The olfactory system (white arrow) and the



cerebellar anlage (star) also express positivity. At P4, positivity is found in the olfactory system (white arrow) and is weakly present in the superficial layer of the cortical plate and in the gyrus dentate of the hippocampus (black arrow). In the cerebellum, the vermian external granular layer shows positivity (star). At P14, positivity is restricted to the olfactory system (white arrow) and also found in the internal granular layer of the vermis (star).

J Med Genet 2012;49:400-408. doi:10.1136/jmedgenet-2012-100856

207

406

Chromosomal rearrangements

negative model could not be ruled out since oligomerisation of CAMTA1 is needed for the proper function of nuclear local-isation and transcriptional regulation.^{22 23 29}

Although ID was the leading feature in the majority of patients of our cohort, ataxia appeared as a key feature in the cases with CAMTA1 rearrangements. From the patients we report, ataxia is consistent and of early onset although it can be mild and easily missed. Interestingly, one patient with a rearrangement of the CAMTA1 gene including the promoter and the first three exons has been reported in a series of patients with ID and ASD screened for intragenic rearrangements by high resolution array-CGH. The patient presented with developmental delay, ID, attention deficit hyperactivity disorder and pervasive developmental disorder associated with an ataxic gait and ³⁴ One speech disorders such as velopharyngeal deficiency. further patient with ID and autism was identified in the context of a genome wide CNV study in autism, but no clinical data are provided.7 CAMTA1 mutations/rearrangements are likely to be extremely rare, since we found none in a series of 196 cases and the patient above was the only one in a cohort of 996 patients with ASD (both with or without ID) detected by an array-CGH and found to harbour an intragenic CAMTA1 rearrangement." Congenital non-progressive ataxia appeared isolated in the family with the inframe deletion (family 3) but associated with ID in families with truncating rearrangements (family 1). Thus, both ID and ataxia will be associated with CAMTA1 loss of function. Cerebellar symptoms, such as ataxia, are compatible with the cerebellar expression of CAMTA1 in mouse at late stages of development (figure 4).

As expected, considering the functions of CAMTA1 in learning and memory processes we identified a number of manifestations in the adult patients we investigated through careful neuropsychological evaluation in family 1. Members of the family demonstrated a lack of rehearsal processing in episodic visual memory. This clinical finding was found in association with the presence of mild bilateral hippocampal atrophy, simplified gyration of the dentate gyri and severe bilateral posterior cortical atrophy including the parietal lobes, the precuneus and the cuneus on neuroimaging (figure 2). A similar pattern of mild parietal atrophy was identified in the two adult patients, although not specific. These findings converge with the results of functional MRI of patients with differential episodic memory performance as described by Huentelman

Web resources

Primer blast designing tool (http://www.ncbi.nlm.nih.gov/tools/ primer-blast/), translating prediction tool (http://web.expasy.org/ translate/), reference genomic sequence (http://www.ensembl.org), protein alignment sequences (http://www.genome.jp/tools/clustalw/ website), Database of Genomic Variants: (http://projects.tcag. ca/variation/), Decipher database (http://decipher.sanger.ac.uk/), Control DNA SNP databases on Hapmap (http://www.hapmap.org); dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/), NHLB1 Exome Sequencing Project (ESP) Exome Variant Server (http://evs. gs.washington.edu/EVS/), Polyphen 2 (http://genetics.bwh.harvard. edu/pph2/), SIFT (http://sift.jcvi.org/), OMIM (http://omim.org/). Allen brain institute, mouse brain CAMTA1 images: http://developingmouse.brain-map.org/data/compare/ivt.html?ispopup=true &include=100046547,100046440,100046759,100057301, 100083105,100042096,73716734,69540333,100057147.

J Med Genet 2012;49:400-408. doi:10.1136/jmedgenet-2012-100856

and colleagues.³² In this study, there was an asymmetrical activation of the mediotemporal lobe in two comparison groups on functional brain imaging. Here, PET scan analysis revealed asymmetrical hypometabolism of regions involved in memory processes (figure 2). Those neuropsychological data support the role of CAMTA1 in the visual and sound perception network in human left cerebrum. A bio computational analysis of STAT2 mediated transcriptional control in human versus chimpanzee study underlined the expression of CAMTA1 in this network.35

In conclusion, the identification of three novel independent families with intragenic rearrangements combined with their specific cerebral expression and the functions of the calciumdependent transcription factor adds CAMTA1 to the list of genes responsible for NPCA with ID.

Author affiliations

Centre de Génétique et Centre de Référence «Anomalies du Développement et Syndromes Malformatifs», Hôpital d'Enfants, CHU Dijon, Dijon, France ²Equipe Génétique et Anomalies du Développement, Université de Bourgogne, Faculté

de Médecine, Dijon, France ³INSERM, U975 and CNRS 7225, CRICM, Hôpital Pitié-Salpêtrière, Paris, France AP-HP, Département de Génétique et de Cytogénétique, Centre de Génétique
 Moléculaire et Chromosomique, Hôpital Pitié-Salpêtrière, Paris, France
 ⁵Département de Génétique et Centre de Référence "Déficiences intellectuelles de

causes rares", Groupe Hospitalier Pitié-Salpêtrière et CRICM, INSERM UMRS975, La Pitié Salpêtrière, Paris, France

^{el}Service de Neuropédiatrie, CHU de Besançon, Besançon, France ²Laboratoire de Génétique et EA 4368, CHU de Nancy, Nancy, France

⁸Neurologie, CHU de Besançon, Besançon, France ⁹Centre de référence des malformations et maladies congénitales du cervelet et Service de Génétique Médicale, APHP, Hôpital Trousseau, et INSERM U676, Hôpital

 Service de Generative Metalale, Arthr, Hohiai Housseau, et Hoschwoord, Hohiai Robert Derör, Paris, France
 ¹⁰U.F. de Génétique Clinique, Maternité Régionale, Nancy, France
 ¹¹Department of Psychiatry, Trinity College Dublin, Dublin, Ireland
 ¹²Département de Génétique, Hôpital Necker Enfant Malade, AP-HP et Université René Descartes, Paris, France

³Département de Neuropédiatrie, Hôpital A.Trousseau, AP-HP, Paris, France ¹⁴Centre de Référence des Troubles du Langage et des apprentissages, Pédiatrie, CHU

Dijon, Dijon, France ¹⁵Laboratoire de Cytogénétique, Plateau Technique de Biologie, CHU de Dijon, Dijon,

France ⁶Laboratoire de Cytogénétique, CHU de Besançon, Besançon, France ¹⁷Laboratoire de Génétique Moléculaire, Plateau Technique de Biologie, CHU de Dijon,

 Dijon, France
 ¹⁸Service d'Histologie-Embryologie-Cytogénétique, Hôpital Necker-Enfants Malades, Paris, France

Acknowledgements The authors thank the families for their contribution and Coralie Rastel for neuropsychological evaluation of the proband of family 3. They also thank the Allen Brain Institute for providing the data, and Andy Pang and Dr Stephen Scherer for assessing controls and sharing unpublished data in DGV

Contributors Patient recruitment and phenotype: CT-R, DH, CM, CA, DM, MA, PC, JV, SM, LG, CD, LB, DD, MB, JT, LF, JA, EM, EH. Experimental analysis: EL, CM, LD, VC, BA, BK, MB, CB, AB, PC, NM, A-LM, VR, PJ, FR. All the authors have participated in writing and reviewing the manuscript. JT, EL and BK contributed equally to this participated of T. P. are Microsoft Mc and T. P. are Microsoft Mc and T. P. are Microsoft Mc and T. P. and Mc and Mc and T. P. and Mc and T. And T work. LF and CT-R equally directed the project.

Funding Funding provided by the Regional Council of Burgundy.

Competing interests None

Patient consent Obtained

Ethics approval Ethics approval was provided by the local ethics committee.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- Steinlin M. Non-progressive congenital ataxias. Brain Dev 1998;20:199–208. Steinlin M, Zangger B, Boltshauser E. Non-progressive congenital ataxia with or without cerebellar hypoplasia: a review of 34 subjects. Dev Med Child Neurol 1998-40-148-54
- 1990,400,146–34.
 Chelly J, Kheffaoui M, Francis F, Chérif B, Bienvenu T. Genetics and pathophysiology of mental retardation. *Eur J Hum Genet* 2006;14:701–13.
 Inlow JK, Restfo LL. Molecular and comparative genetics of mental retardation.
- Genetics 2004;16:835-81.

Chromosomal rearrangements

- de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM, Reijmersdal S, Nillesen WM, Huys EH, Leeuw N, Smeets D, Sistermans EA, Feuth T, van RavenswaaijArts CM, van Kessel AG, Schoemmakers EF, Brunner HG, Veltman JA. Diagnostic genome profiling in mental retardation. *Am J Hum Genet* 5. 2005;77:606-16.
- 2005;77:606-16.
 Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, Zenker M, Hüffmeier U, Thiel C, Rischendorf F, Nürnberg P, Reis A, Trautmann U. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. Am J Med Genet A 2006;140:2063-74.
 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Balley AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Prina L, Bruego SE, Carsen AE, Casell L, Cherey L, Chume BH, Cochrong L, Corsello C, 6
- TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farar P, Fernandez BA, Folstein SE, Fornbonne E, Freitag CM, Gilbert J, Gilberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussi J, Benshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Satl J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Waldace S, Wang K, Wasajank TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook H, Coon H, Louccaro ML, Devilin S, Betancur C, Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010;**46**:6368–72.
- 8.
- Sutcliffer JS, Betancur C, Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010;**466**:368–72. **Kaufman L**, Ayub M, Vincent JB. The genetic basis of non-syndromic intellectual disability: a review. *J Neurodev Disord* 2011;**2**:182–209. **Vissers LE**, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van BOn BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA. A de novo paradigm for mental retardation. *Nat Genet* 2010;**42**:1109–12. **Janden EE** Courtier, L Sciencelman D, Norcey A, Yang Y, Pellerin S, Debranierica, 9
- 2010;42:1109—12.
 Hamdan FF, Gauthier J, Spiegelman D, Noreau A, Yang Y, Pellerin S, Dobrzeniecka S, Côté M, Perreau-Linck E, Carmant L, D'Anjou G, Fombonne E, Addington AM, Rapoport JL, Delisi LE, Krebs MO, Mouaffak F, Joober R, Mottron L, Drapeau P, Marineau C, Lafrenière RG, Lacaille JC, Rouleau GA, Michaud JL; Synapse to Disease Group. Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. N Engl J Med 2009;360:599—605. 10.
- J Med 2009;360:599–605. Hamdan FF, Gauthier J, Araki Y, Lin DT, Yoshizawa Y, Higashi K, Park AR, Spiegelman D, Dobrzeniecka S, Piton A, Tomitori H, Daoud H, Massicotte C, Henrion E, Diallo O; S2D Group, Shekarabi M, Marineau C, Shevell M, Maranda B, Mitchell G, Nadeau A, D'Anjou G, Vanasse M, Srour M, Lafrenière RG, Drapeau P, Lacaille JC, Kim E, Lee JR, Igarashi K, Huganir RL, Rouleau GA, Michaud JL. Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am J Hum Genet* 2011;88:306–16. Maranetaller L Scrappenge S, Lorgen Denoiseure P Karmiscrack B, Nathreth M, Wahl 11.
- Wagenstaller J, Spranger S, Lorenz-Depierux B, Karmierzak B, Nathrath M, Wahl D, Heye B, Glaser D, Liebscher V, Meitinger T, Strom TM. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with 12.
- Bhalla K, Luo Y, Buchan T, Beachern MA, Guzauskas GF, Ladd S, Bratcher SJ, Schroer RJ, Balsamo J, DuPont BR, Lilien J, Srivastava AK. Alterations in CDH15 and 13. KIRREL3 in patients with mild to severe intellectual disability. Am J Hum Gener 2008-83-703-13
- Griggs BL, Ladd S, Saul RA, DuPont BR, Srivastava AK. Dedicator of cytokinesis 8 is disrupted in two patients with mental retardation and developmental disabilities Genomics 2008;91:195-202.

- Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, Endris V, Roberts W, Szatmari P, Pinto D, Bonin M, Riess A, Engels H, Sprengel R, Scherer SW, Rappold GA. Mutations in the SHANKZ synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat Genet* 2010;42:499–91.
 Krab LC, Goorden SM, Elgersma Y. Oncogenes on my mind: ERK and MTOR signaling in cognitive diseases. *Trends Genet* 2008;24:499–510.
 Nagase T, Ishikawa K, Suyama M, Kikuno R, Hirosawa M, Miyajima N, Tanaka A, Kotani H, Nomura N, Ohara O. Prediction of the coding sequences of unidentified human genese. Yll The commentate sequences of 100 new CIMA clones from brain 15.
- 16.
- 17. human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res* 1998;**5**:355–64. **Bouché N**, Scharlat A, Snedden W, Bouchez D, Fromm H. A novel family of
- 18. calmodulin-binding transcription activators in multicellular organisms. J Biol Chem
- 2002·277·21851-61
- 2002;277:21851–61. Finkler A, Ashery-Padan R, Fromm H. CAMTAs: calmodulin-binding transcription activators from plants to human. *FEBS Lett* 2007;**581**:3893–8. Doherty CJ, Van Buskirk HA, Myers SJ, Thomashow MF. Roles for Arabidopsis CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. *Plant Cell* 2009;**21**:972–84. 20
- 21
- Galon Y, Snin Cell 2009 21:912–94.
 Galon Y, Snin C, Fromm H. How calmodulin binding transcription activators (CAMTAs) mediate auxin responses. *Plant Signal Behav* 2010;5:1311–14.
 Han J, Gong P, Reddig K, Mitra M, Guo P, Li HS. The fly CAMTA transcription factor potentiates deactivation of rhodopsin, a 6 protein-coupled light receptor. *Cell* 22.
- 2006 127 847-58
- 2006; 127:847–58. Gong P, Han J, Beddig K, Li HS. A potential dimerization region of dCAMTA is critical for termination of fly visual response. J Biol Chem 2007; 282:21253–8. Sattler JM, Dumont R. Assessment of Children: WISC-IV and WPPSI-III Supplement. San Diego, Mk: J.M. Sattler Publisher, 2004. Kaufman AS, Lichtenberger E. Assessing Adolescent and Adult Intelligence. 3rd edn.
- 24
- 25.
- Kaurman AS, Lichtenberger E. Assessing Adolescent and Aduit Intelligence. 3rd edn. Hoboken, NJ: Wiley, 2006.
 Schmitz-Hübsch T, du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, Giunti P, Globas C, Infante J, Kang JS, Kremer B, Mariotti C, Melegh B, Pandolfo M, Rakowicz M, Ribai P, Rola R, Schöls L, Szymanski S, van de Warrenburg BP, Dürr A, Klockgether T, Fancellu R. Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology 2006;66:1717-20.
 Allen Developing Mouse Brain Atlas [Internet]. Seattle, WA: Allen institute for brain Science. 2000. http://developingmouse.brain.man.org
- 27.
- Alter Developing Woulds Data Addates, Jeaues, VA: Alter Institute for Hamiltonian Science, 2009, http://developingmouse.brain-map.org Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, Park H, Hurles ME, Lee C, Venter JC, Kirkness EF, Levy S, Feuk L, Scherer SVV. Towards a comprehensive structural variation map of an individual human genome. *Genome* Biol 2010;11:R52
- Han J, Reddig K, Li HS. Prolonged G(q) activity triggers fly rhodopsin endocytosis 29.
- and degradation, and reduces photoreceptor sensitivity. *EMBO J* 2007;26:4966—73.
 Henrich KO, Bauer T, Schulte J, Ehemann V, Deubzer H, Gogolin S, Muth D, Fischer M, Benner A, König R, Schwab M, Westermann F. CAMTA1, a 1p36 tumor 30. suppressor candidate, inhibits growth and activates differentiation programs in
- suppressor candidate, inhibits growth and activates differentiation programs in neuroblastoma cells. Cancer Res 2011;**71**:3142–51. Schraivogel D, Weinmann L, Beier D, Tabatabai G, Eichner A, Zhu JY, Anton M, Sixt M, Weller M, Beier CP, Meister G, CAMTA1 is a novel tumour suppressor regulated by miR-9/9(*) in glioblastoma stem cells. *EMBO J* 2011;**30**:4309–22. **HuenteIman MJ**, Papassotropolos A, Craig DW, Hoerndli FJ, Pearson JV, Huynh KD, Corneveaux J, Hänggi J, Mondadori CR, Buchmann A, Reiman EM, Henke K, de Quervain DJ, Stephan DA. Calmodulin-binding transcription activator 1 (CAMTA1) alleas profilease human encicit momonu cerforcargen. *Hum MAI* Genet 32.
- Uderviain DJ, Stepnan DA, Cairnoounn-binding transcription activator 1 (CAWIA1) alleles predispose human episodic memory performance. *Hum Mol Genet* 2007;16:1469–77.
 Need AC, Attix DK, McEvoy JM, Cirulli ET, Linney KL, Hunt P, Ge D, Heinzen EL, Maia JM, Shianna KV, Weale ME, Cherkas LF, Clement G, Spector TD, Gibson G, Goldstein DB. A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum Mol Genet* 2009;18:4650–61. 33
- Mikhail FM, Lose EJ, Robin NH, Descartes MD, Nutledge KD, Rutledge SL, Korf BR, Carroll AJ. 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* 2011;**15**:2386–36. 34
- Wang L, Huang J, Jiang M, Lin H, Signal Transducer and Activator of Transcription 2 (STAT2) Metabolism Coupling Postmitotic Outgrowth to visual and sound perception network in human left cerebrum by Biocomputation. J Mol Neurosci. Published Online 35 First: 5 January 2012

J Med Genet 2012;49:400-408. doi:10.1136/imedaenet-2012-100856

Keren Boris – Thèse de doctorat - Année 2013

Article 4

LETTER

doi:10.1038/nature10406

Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus

A list of authors and their affiliations appears at the end of the paper

Both obesity and being underweight have been associated with increased mortality^{1,2}. Underweight, defined as a body mass index (BMI) \leq 18.5 kg per m^2 in adults and ≤-2 standard deviations from the mean in children, is the main sign of a series of heterogeneous clinical conditions including failure to thrive³⁻⁵, feeding and eating disorder and/or anorexia nervosa^{6,7}. In contrast to obesity, few genetic variants underlying these clinical conditions have been reported^{8,9}. We previously showed that hemizygosity of a ~600-kilobase (kb) region on the short arm of chromosome 16 causes a highly penetrant form of obesity that is often associated with hyperphagia and intellectual disabilities¹⁰. Here we show that the corresponding reciprocal duplication is associated with being underweight. We identified 138 duplication carriers (including 132 novel cases and 108 unrelated carriers) from individuals clinically referred for developmental or intellectual disabilities (DD/ID) or psychiatric disorders, or recruited from populationbased cohorts. These carriers show significantly reduced postnatal weight and BMI. Half of the boys younger than five years are underweight with a probable diagnosis of failure to thrive, whereas adult duplication carriers have an 8.3-fold increased risk of being clinically underweight. We observe a trend towards increased severity in males, as well as a depletion of male carriers among non-medically ascertained cases. These features are associated with an unusually high frequency of selective and restrictive eating behaviours and a significant reduction in head circumference. Each of the observed phenotypes is the converse of one reported in carriers of deletions at this locus. The phenotypes correlate with changes in transcript levels for genes mapping within the duplication but not in flanking regions. The reciprocal impact of these 16p11.2 copy-number variants indicates that severe obesity and being underweight could have mirror aetiologies, possibly through contrasting effects on energy balance.

Copy-number variants (CNVs) at the 16p11.2 locus have been associated with cognitive disorders including autism (deletions) and schizo-phrenia (duplications)¹¹⁻¹³, conditions that have been suggested to lie at opposite ends of a single spectrum of psychiatric phenotypes¹⁴. We and others have reported that a deletion of this region spanning 28 genes (Supplementary Table 1) increases the risk of morbid obesity 43-fold (Supplementary Fig. 1) 10,15 . We hypothesized that the reciprocal duplication, with its resulting increase in gene dosage, may influence BMI in a converse manner. The duplication was identified in 73 out of 31,424 patients with DD/ID, a frequency consistent with previous reports¹³ (Table 1). Four additional cases were identified among 1,080 patients affected by bipolar disease or schizophrenia. Compared to its prevalence in seven European population-based genome-wide asso-ciation study (GWAS) cohorts¹⁶⁻¹⁸ (31 out of 58,635 individuals), the ($P = 4.23 \times 10^{-13}$; odds ratio = 4.4, 95% confidence interval = 2.9-6.9) and the psychiatric cohorts ($P = 3.6 \times 10^{-3}$; odds ratio = 7.0, 95% confidence interval = 1.8-19.9) (Table 1), strengthening previous reports of similar associations^{12,13}. Our data do not support a two-hit model¹⁹ for the effects of 16p11.2 duplications or deletions (Supplementary Text and Supplementary Table 2).

We compared available data on weight, height and BMI for 106 independent duplication carriers (including published cases) to data for reference populations matched for gender, age and geographical location (Table 2, Methods and Supplementary Tables 3 and 4). The duplication was strongly associated with lower weight (mean *Z*-score -0.56; $P = 4.4 \times 10^{-4}$) and lower BMI (mean *Z*-score -0.67; $P = 2.0 \times 10^{-3}$) (Table 2 and Supplementary Table 5). Birth parameters (n = 48) were normal, indicating a postnatal effect. Adults carrying the duplication had a relative risk of being clinically underweight (BMI <18.5) of 8.3 (95% confidence interval = 4.4 - 15.9, $P = 1.53 \times 10^{-10}$) (see Methods). Concordantly, none of the 3,544 patients in our obesity cohorts^{10,15} carried the duplication (Table 1).

To investigate these associations further, we carried out separate analyses of carrier patients (DD/ID and psychiatric) and non-medically ascertained carriers (population-based cohorts, plus 11 transmitting parents and three other affected first-degree relatives for whom data were available) (Table 2). Each category had significantly lower weight and BMI, with similar effect sizes. However, the proportion of underweight cases (BMI ≤ -2 s.d.) was higher in the first group than in the second group (17 out of 76 compared to 2 out of 40; P = 0.017). Note that the impact of the duplication on underweight status might be underestimated here owing to prescription of antipsychotic treatments that are often associated with weight gain²⁰ (Supplementary Table 6).

Having demonstrated an association of the duplication with being underweight, we investigated the implications of gender for the resulting phenotypes (Fig. 1, Supplementary Fig. 2 and Supplementary Table 7, In DD/ID patients, the impact of the duplication on being underweight is stronger in males; the effect in females is in the same direction, but is smaller and not statistically significant (Table 2). A similar and significant difference (P = 0.0168) was observed in adult carriers (all groups combined): the relative risk of being underweight for males is 23.2 (95% confidence interval = 9.1–59.3, $P = 4.6 \times 10^{-11}$); for females it is only 4.7 (95% confidence interval = 1.9–11.8, $P = 9.9 \times 10^{-4}$). A gender bias was also observed in the ascertainment of DD/ID duplication carriers, in which we have an excess of males (51 males:33 females, P = 0.044). By contrast, carriers from the general population showed a strong overrepresentation of females (10 males:21 females, P = 0.035) (Supplementary Text). A similar bias was observed among transmitting parents (7 males:23 females, $P = 5.53 \times 10^{-4}$). Thus, there is an overrepresentation of males in the medically ascertained group, and a depletion in the non-medically ascertained one. We suggest that males may be more likely than females to present severe phenotypes, and that this may account for the gender bias because severely affected males may be less likely to be recruited to adult population cohorts or to be reproductively successful.

As previously reported²¹, the duplication was also associated with reduced head circumference (mean Z-score -0.89; $P = 7.8 \times 10^{-6}$) (Fig. 1), 26.7% presenting with microcephaly (head circumference ≤ -2 s.d.), whereas carriers of the reciprocal deletion had an increased head circumference (mean Z-score +0.57; $P = 1.79 \times 10^{-5}$) (Supplementary Fig. 3 and Supplementary Table 8): an additional instance of a mirror phenotype associated with reciprocal changes in copy number at this locus. Notably, head circumference Z-score score score late

6 OCTOBER 2011 | VOL 478 | NATURE | 97

©2011 Macmillan Publishers Limited. All rights reserved



Table 1 | 16p11.2 rearrangements in cases and controls

| Ascertainment | Cohorts | Duplicat | ion | Deletion | Total | |
|-------------------|--|-------------------|------------------------|-------------------|------------------------|--------|
| | | п | P† | n | P† | |
| Neuro- | Unspecified DD/ID* from 28 cytogenetic centres | 72 | | 113 | | 30,323 |
| developmental | ADHD1, deCODE | 0 | | 1 | | 591 |
| disorders | Childhood autism [‡] , deCODE | 0 | | 2 | | 159 |
| | Childhood autism spectrum disorder‡, deCODE | 1 | | 3 | | 351 |
| | TOTAL | 73 | 4.23×10^{-13} | 119 | 5.43×10^{-32} | 31,424 |
| | Rearrangement frequency (95% CI) | 0.23% (0.18-0.29) | | 0.38% (0.31-0.45) | | |
| Family history | First-degree relatives of probands | 30 | | 35 | | 43/62 |
| Adult psychiatric | Schizophrenia, deCODE | 0 | | 1 | | 657 |
| symptoms | Bipolar disease, Rouen | 1 | | 0 | | 156 |
| | Schizophrenia, schizo-affective, Rouen | 3 | 5 | 0 | | 267 |
| | TOTAL | 4 | 3.57×10^{-3} | 1 | 3.78×10^{-1} | 1,080 |
| | Rearrangement frequency (95% CI) | 0.37% (0.01-0.73) | | 0.09% (0-0.27) | | |
| Underweight | Eating disorder, Spain | 1§ | | 0 | | 441 |
| Obesity | Obesity, Spain | 0 | | 2 | | 653 |
| | Adult obesity, France | 0 | | 4 | | 705 |
| | Childhood obesity, France & UK | 0 | | 7 | | 1,574 |
| | Obesity bariatric surgery, France | 0 | | 2 | | 141 |
| | Obesity discordant siblings, Sweden | 0 | | 2 | | 159 |
| | Obesity and cognitive delay, France & UK | 0 | | 9 | | 312 |
| | TOTAL | 0 | 4.21×10^{-1} | 26 | 2.52×10^{-19} | 3,544 |
| | Rearrangement frequency (95% Cl) | 0 | | 0.73% (0.45-1.01) | | |
| Population-based | NFBC1966 Finnish | 4 | | 3 | | 5,319 |
| cohorts | CoLaus Swiss | 5 | | 0 | | 5,612 |
| | EGCUT Estonian | 2 | | 1 | | 2,994 |
| | deCODE Iceland | 17 | | 18 | | 36,601 |
| | SHIP Germany | 1 | | 2 | | 4,070 |
| | KORA F3+F4 Germany | 2 | | 1 | | 3,458 |
| | Paediatric family study | 0 | | 0 | | 581 |
| | TOTAL | 31 | | 25 | | 58,635 |
| | Rearrangement frequency (95% Cl) | 0.05% (0.03-0.07) | | 0.04% (0.03-0.06) | | |

CI, Confidence interval; ADHD, attention-deficit hyperactivity disorder. *Not a disease-specific cohort. Detailed distribution is provided in the online methods. †Fisher's exact test, compared to the combined frequency in general population groups. There was no overlap between these 3 cohorts. Advpical duplication (see Supplementary Fig. 5). II Total number of parental pairs tested for duplication/deletion. 13 out of 43 duplications and 27 out of 62 deletion cases were de novo.

positively with those of BMI in carriers of both the duplication (rho = 0.37; $P = 2.65 \times 10^{-3}$) and the deletion (rho = 0.42; $P = 1.9 \times 10^{-5}$) (Supplementary Methods). This indicates that head circumference and BMI may be regulated by a common pathway, or that a causal relationship exists between these two traits in these patients. Alternatively, the two phenotypes may arise from distinct genes and pathways. A full list of malformations and secondary phenotypes reported in duplication carriers ascertained for DD/ID is available in Supplementary Table 9.

In view of the importance of modified eating behaviours in obesity and being underweight, the clinical reports of duplication carriers were screened for evidence of such modified behaviours. In 11 out of 77 clinically ascertained cases, clinicians had spontaneously reported low food intake and selective and restrictive eating behaviour, again mirroring one of the phenotypes—hyperphagia—seen in deletion carriers¹⁰ (Supplementary Table 6) and indicating that the duplication may increase the risk of eating disorders. Consequently, we carried out multiplex ligation-dependent probe amplification (MLPA, Supplementary Table 10) to screen for 16p11.2 rearrangements in 441 patients diagnosed with eating disorders, including anorexia nervosa, bulimia and binge eating disorder (Table 1 and Supplementary Text). No duplications of the entire region were identified, but one out of 109 anorexia nervosa patients carried an atypical 136-kb duplication that encompasses the sialophorin (*SPN*) and quinolinate phosphoribosyltransferase (*QPRT*) genes (Supplementary Fig. 4). This single, smaller duplication does not allow us to draw any firm conclusions, but together with other atypical rearrangements, it may, in the future, be essential for establishing the roles of the 28 genes within the region.

Large genomic structural variants are known to affect the expression of genes not only within the affected region but also at a distance²²⁻²⁵. Therefore, it is possible that the phenotypes observed in 16p11.2 deletion and duplication individuals are due to effects on the expression of genes mapping outside the rearranged interval, rather than to gene dosage within the 600-kb deletion or duplication. We measured

Table 2 Comparisons of the height, weight and BMI distributions in duplication carriers and controls.

| | Strata | Combined† | | | DD/ID or psychiatric† | | | Non-medically ascertained: | | |
|--------|--------|--------------|----------------------|-----|-----------------------|----------------------|----|----------------------------|----------------------|----|
| | | Mean Z-score | P-value | n* | Mean Z-score | P-value | n* | Mean Z-score | P-value | n* |
| BMI | All | -0.47 | 2.0×10^{-3} | 102 | -0.56 | 4.1×10^{-3} | 76 | -0.45 | 6.0×10^{-3} | 40 |
| | Male | -0.54 | 2.1×10^{-2} | 52 | -0.71 | 1.3×10^{-2} | 43 | -0.31 | 2.0×10^{-1} | 14 |
| | Female | -0.4 | 1.8×10^{-2} | 50 | -0.37 | 8.3×10^{-2} | 33 | -0.52 | 4.2×10^{-3} | 26 |
| Weight | All | -0.56 | 4.4×10^{-4} | 104 | -0.65 | 1.3×10^{-3} | 78 | -0.61 | 3.0×10^{-3} | 40 |
| 0 | Male | -0.64 | 5.8×10^{-3} | 53 | -0.79 | 4.4×10^{-3} | 44 | -0.57 | 8.8×10^{-2} | 14 |
| | Female | -0.47 | 1.7×10^{-2} | 51 | -0.47 | 6.5×10^{-2} | 34 | -0.63 | 8.6×10^{-3} | 26 |
| Height | All | -0.24 | 4.8×10^{-2} | 103 | -0.33 | 3.6×10^{-2} | 77 | -0.15 | 1.8×10^{-1} | 40 |
| | Male | -0.34 | 4.5×10^{-2} | 52 | -0.4 | 4.6×10^{-2} | 43 | -0.29 | 1.2×10^{-1} | 14 |
| | Female | -0.14 | 2.6×10^{-1} | 51 | -0.24 | 2.1×10^{-1} | 34 | -0.07 | $3.7 	imes 10^{-1}$ | 26 |

The available BMI, weight and height data for duplication carriers were transformed to Z-scores using gender- and age-matched reference populations, and one-tailed t-tests were carried out to determine whether the mean Z-scores deviated from zero. Significant differences were identified by reference to cutoffs controlling the false discovery rate at 5% (see Methods): BMI, 0.022; weight, 0.032; height, 0.025. Significant results are indicated in bold. Data were not available for all subjects. "Relatives of probands were excluded as required, to avoid including more than one member of the same family in a single analysis. Hincluding 24 cases from the literature Gupplementary Table 3), iPopulation-based cases and first-degree relatives of probands.

98 | NATURE | VOL 478 | 6 OCTOBER 2011

©2011 Macmillan Publishers Limited. All rights reserved





Figure 1 | Effect of the chromosome 16p11.2 duplication on BMI and head circumference. Z-score values of BMI and head circumference in carriers of the 16p11.2 duplication, stratified by gender and age group. The most severe effect is observed in children at 0–5 years of age. Boxplots represent the fifth, twenty-fifth, median, seventy-fifth and ninety-fifth percentile for each age group. Light grey and dark grey backgrounds represent ≤ -2 and ≤ -3 s.d., respectively, corresponding to the WHO definition of moderately and severely underweight³⁰. BMI is decreased in adolescent and adult females.

relative transcript levels of 27 genes mapping within or near to the rearrangement, using lymphoblastoid cell lines (Supplementary Tables 1 and 11): six from deletion carriers, five from duplication carriers and ten from gender- and age-matched controls (Supplementary Table 12). Expression levels correlated positively with gene dosage for all genes in the CNV region (Fig. 2), consistent with published partial results from adipose tissue¹⁰. Mean relative transcript levels in deletion and duplication carriers were, respectively, 67% and 214% of the levels measured in controls (Supplementary Table 13). Although genes proximal (centromeric) to the rearrangement interval

showed no significant variation in relative transcript levels between patients and controls (Fig. 2), distal (telomeric) genes showed a marked alteration in relative expression. However, their expression levels, including that of *SH2B1* (for which gene dosage and a nearby single nucleotide polymorphism (SNP) have been associated with obesity^{15,26}), were similarly upregulated in cell lines of both deletion and duplication carriers, showing no apparent correlation between transcript level and either copy number or phenotype (Fig. 2). Although lymphoblastoid cells may not recapitulate obesity-relevant tissues, previous experiments have shown a high degree of correlation between expression levels in different tissues and cell lines²², indicating that the same pathways may be similarly disrupted in different cell lineages. Thus, any involvement of these distal genes in the control of BMI in these subjects seems unlikely.

Our study demonstrates the power of very large screens (>95,000 samples: to our knowledge the largest of its kind so far) to characterize the clinical and molecular correlates of a rare functional genomic variant. We demonstrate unambiguously that carrying the 16p11.2 duplication confers a high risk of being clinically underweight, and show that reciprocal changes in gene dosage at this locus result in several mirror phenotypes. As in the schizophrenia/autism14 and microcephaly/ macrocephaly²¹ dualisms, abnormal eating behaviours, such as hyperphagia and anorexia, could represent opposite pathological manifestations of a common energy-balance mechanism, although the precise relationships between these mirror phenotypes remain to be determined. We speculate that head circumference (which correlates with brain volume²⁷), and thus neuronal circuitry, may affect cognitive function and energy balance in patients with 16p11.2 rearrangements (possibly through eating behaviour). Consistent with this are previous reports that a subgroup of children with microcephaly show a concomitant reduction in weight percentile28. Our findings also support the observation that severe overweight and underweight phenotypes correlate with lower cognitive functioning4.29. Thus, abnormal food intake may be a direct result of particular neurodevelopmental disorders. Although it is possible that the 16p11.2 region encodes distinct genes specific for each trait, a more parsimonious hypothesis is that these clinical manifestations of dysfunction of the central nervous system are all secondary to the disruption of a single neurodevelopmental step that is sensitive to gene dosage. Further resolution of this issue may



©2011 Macmillan Publishers Limited. All rights reserved

Figure 2 Transcript levels for genes within and near to the 16p11.2 rearrangements. a, Relative expression levels of 27 genes mapping to 16p11.2 in deletion (n = 6) and duplication (n = 5) carriers (red and green, respectively), and in control cell lines (n = 10, blue). Grey lines denote the extent of the 16p11.2 CNV (29.5–30.1 megabases (Mb)). Complete lists of genes mapping within the rearranged interval, and of the quantitative PCR assays, are in Supplementary Tables 1 and 11, respectively. For the possible relevance of each of these genes to obesity/leanness and/or developmental delay/cognitive deficits, see ref. 10. b, Rank comparison (Kruskal–Wallis test)

between the expression of 27 genes mapping to 16p11.2 in deletion and duplication carriers (red and green, respectively) and in control cell lines (blue). Genes are labelled as telomeric, centromeric or within the rearranged interval (CNV). Dots correspond to the mean group rank and bars indicate the comparison interval. Groups with non-overlapping intervals are significantly different (*P*-values were adjusted for multiple testing issues using a Bonferroni correction, where the number of tests is the number of pairwise comparisons; the resulting adjusted *P*-value was less than 0.05).

6 OCTOBER 2011 | VOL 478 | NATURE | 99



require the identification of additional patients with rare atypical rearrangements in this region.

METHODS SUMMARY

Underweight is defined in adults as $BMI \leq 18.5$. In individuals younger than 18 years of age, it is defined as a Z-score ≤ -2 . Statistics. Two-tailed Fisher's exact test was used to compare frequencies of the

rearrangement in patients and controls. Z-scores were computed for all data using gender-, age- and geographically-matched reference populations. One-tailed Student's t-test was performed to test BMI, height, weight and head circumference in duplication carriers for Z-scores of less than zero. We used Kruskal-Wallis tests for differences in gene expression patterns. P-values were adjusted using a Bonferroni correction, considering the number of pairwise comparisons; the result-ing adjusted *P*-value was less than 0.05. The relative risk of being underweight was calculated as the ratio of the fraction of underweight individuals among duplication

carriers versus our control group. Discovery of CNVs. Carriers of 16p11.2 duplication and deletion were identified through various procedures: (1) comparative genomic hybridization with Agilent 44K, 60K, 105K, 180K, 244K arrays; (2) Illumina Human317, Human370, HumanHap550, Human610 and 1M BeadChips; (3) Affymetrix 6.0, 500K geno-typing arrays; (4) quantitative multiplex PCR of short fluorescent fragments (QMPSF); (5) fluorescent in situ hybridization (FISH); (6) MLPA. CNV analyses of GWAS data were carried out using cnvHap, a moving-window average-intensity procedure, a Gaussian mixture model, circular binary segmentation, QuantiSNP, PennCNV, BeadStudio GT module and Birdseed. At least two independent algo rithms were used for each cohort.

Expression analyses. Lymphoblastoid cell lines were established from carriers and controls. SYBR Green quantitative PCR was performed to assess relative expression of genes

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 February; accepted 29 July 2011.

Published online 31 August 2011.

- Berrington de Gonzalez, A. et al. Body-mass index and mortality among 1.46 million white adults. N. Engl. J. Med. **363**, 2211–2219 (2010). Flegal, K. M., Graubard, B. I., Williamson, D. F. & Gail, M. H. Cause-specific excess 2 eaths associated with underweight, overweight, and obesity. J. Am. Med. Assoc.
- 298, 2028-2037 (2007). 298, 2028-2037 (2007). Olsen, E. M. et al. Failure to thrive: the prevalence and concurrence of anthropometric criteria in a general infant population. Arch. Dis. Child. 92, 100-114 (2007). 3 109-114 (2007).
- Corbett, S. S. & Drewett, R. F. To what extent is failure to thrive in infancy associated 4. nitive development? A review and meta-analysis. J. Child Psychol. with poore
- 5
- with poorer cognitive development? A review and meta-analysis. J. Comp. 1 system Psychiatry 45, 641–654 (2004). Rudolf, M. C. & Logan, S. What is the long term outcome for children who fail to thrive? A systematic review. Arch. Dis. Child. **90**, 925–931 (2005). Bravender, T. et al. Classification of eating disturbance in children and adolescents: proposed changes for the DSM-V. Eur. Eat. Disord. Rev. **18**, 79–89 6. (2010)
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR 39–134 and 583–596 (Am. Psychiatric Assoc., 2000). Scherag, S., Hebebrand, J. & Hinney, A. Eating disorders: the current status of molecular genetic research. *Eur. Child Adolesc. Psychiatry* 19, 211–226 (2010) 7
- 8.
- Wang, K. et al. A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. *Mol. Psychiatry* doi:10.1038/mp.2010.107 (2010).
 Walters, R. G. et al. A new highly penetrant form of obesity due to deletions on chromosome 16p112. *Nature* 463, 671–675 (2010).
 Marshall, C. R. et al. Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* 82, 477–488 (2008).
 Morchus, S. E. et al. Winerdunipications of 16p112 are associated with

- Marshan, C. R. *et al.* Structure **1 42**, 477-**48**8 (2008).
 McCarthy, S. E. *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genet* **42**, 1223-1227 (2009).
 Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667-675 (2008).
 Crespi, B., Stead, P. & Elliot, M. Evolution in health and medicine Sackler colloquium: comparative genomics of autism and schizophrenia. *Proc. Natl Acad. Sci. USA* **107** (Supp1), 1736-1741 (2010).
 Bochukova, E. G. *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* **463**, 666-670 (2010).
 Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **8**, 6 (2008).
 Sabatti, C. *et al.* Genenic structure of Europeans: a view from the north-east. *PLoS ONE* **4**, e5472 (2009).
 Ririrajan, S. & Eichler, E. E. Phenotypic variability and genetic susceptibility to

- Girrajan, S. & Eichler, E. E. Phenotypic variability and genetic susceptibility to genomic disorders. *Hum. Mol. Genet.* 19, R176–R187 (2010).

100 | NATURE | VOL 478 | 6 OCTOBER 2011 ©2011 Macmillan Publishers Limited. All rights reserved

- 20. Pramyothin, P. & Khaodhiar, L. Metabolic syndrome with the atypica
- 21. ociated with
- Pramyotrin, P. & Khaodniar, L. Metabolic syndrome with the atypical antipsychotics. *Curr. Opin. Endocrinol. Diabetes Obes.* **17**, 460–466 (2010). Shinawi, M. *et al.* Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioral problems, dysmorphism, epilepsy, and abnormal head size. *J. Med. Genet.* **5**, 332–341 (2009). Merla, G. *et al.* Submicroscopic deletion in patients with Williams–Beuren
- Mera, G. et al. Submicroscopic deletion in patients with Williams-beuren syndrome influences expression levels of the nonhemizygous flanking genes. *Am.* J. Hum. Genet. **79**, 332–341 (2006). Stranger, B. E. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007). Henrichsen, C. N. et al. Segmental copy number variation shapes tissue transcriptomes. *Nature Genet.* **41**, 424–429 (2009). 23.
- 24
- Ricard, G. *et al.* Phenotypic consequences of copy number variation: insights from Smith–Magenis and Potocki–Lupski syndrome mouse models. *PLoS Biol.* **8**, e1000543 (2010).
- BIO00543 (2010).
 Willer, C. J. et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genet.* **41**, 25–34 (2009).
 Courchesne, E., Carper, R. & Akshoomoff, N. Evidence of brain overgrowth in the first year of life in autism. *J. Am. Med. Assoc.* **290**, 337–344 (2003).
 Baxter, P. S., Rigby, A. S., Rotsaert, M. H. & Wright, I. Acquired microcephaly: causes, 26
- 27.
- 28. patterns, motor and IQ effects, and associated growth changes. Pediatrics 124, 590-595 (2009).
- 590–595 (2009). Li, Y., Dai, Q., Jackson, J. C. & Zhang, J. Overweight is associated with decreased cognitive functioning among school-age children and adolescents. *Obesity (Silver Spring)* **16**, 1809–1815 (2008). de Onis, M., Blossner, M., Borghi, E., Frongillo, E. A. & Morris, R. Estimates of global prevalence of childhood underweight in 1990 and 2015. *J. Am. Med. Assoc.* **291**, 2600–2606 (2004). 29
- 30 2600-2606 (2004).

Supplementary Information is linked to the online version of the paper at .nature.cor n/nati

Foundation (A Reymond), the Telethon Action Suisse Foundation (A.Reymond), the Swiss National Science Foundation (A.Reymond, J.S.B., S.B. and S.R.J., a Swiss National Science Foundation Sinergia grant (S.J., D.M., S.B., J.S.B. and A.Reymond), the European Commission anEUploidy Integrated Project grant 037627 (A.Reymond, S.B., X.E., H.G.B. and S.E.A), the Ludwig Institute for Cancer Research (AV), the Swiss Institute of Bioinformatics (S.B. and Z.K.), an Imperial College Department of Medicine PhD studentship (J.S.E.-S.M), the Comprehensive Biomedical Research Centre, Imperial College Healthcare NHS Trust, and the National Institute for Health Research (P.E.) the Meloroma Trust and the Medical Research Council (A IE R. and P.E.) the Imperial College Healthcare NHS Trust, and the National Institute for Health Research (P.E.), the Wellcome Trust and the Medical Research Council (ALF.B. and P.F.), the Instituto de Salud Carlos III (SCIII)-FIS, the German Mental Retardation Network funded through a grant of the German Federal Ministry of Education and Research (NGFNplus 01GS08160) (A.Reis), European Union-FEDER (PI081714, PS09/01778) (F.F.A., M.G. and X.E.), SAF2008-02278 (C.R.), the Belgian National Fund for Scientific Research, Flanders (N.V.A and R.F.K.), the Dutch Organisation for Health Research and Development (ZON-MW grant 917-86-319) and Hersenstichting Nederland (B.B.A.d.V.), grant 81000346 from the Chinese National Natural Science Foundation (Y.G.Y.), the Simons Foundation Autism Research Initiative, Autism Speaks and NIH grant GM061354 (J.F.G.), and Oestpreichische Nationalbank (OENB) grant 13059 (A.K.-B.), Y.S. holds a Young Investigator Award from the Children's Turnor Foundation and a Catalyst Award from Harvard Medical School. B.L.W. Nolds a Fudan Scholar Research Award from Tudan University, a grant from Chinese National '973' project on and a Catalyst Award from Harvard Medical School. B.L.W. holds a Fudan Scholar Research Award from Fudan University, a grant from Chinese National '973' project on Population and Health (2010CB529601) and a grant from the Science and Technology Council of Shanghai (09JC1402400). E.R.S. and S.L., recipients of the Michael Smith Foundation for Health Research Scholara raward, acknowledge the CIHR MOP 74502 operational grant. The Estonian Genome Center of the University of Tartu (EGCUT) received support from the EU Centre of Excellence in Genomics and FP7 grants 201413 and 245536, and from Estonian Government SF0180142508, SF0180026s09 and SF0180027510 (A.M., K.M. and A.K.). D.S. thanks the Direction Générale de l'Organisation des Soins from the French Ministry of Health for their support in the development of several array-CGH plaforms, and the Centres Labellisés Anomalies du Development in France. The Helmholtz Zentrum Munich and the State of Bavaria financed the KORK study, also supported by the German National Genome Anomalies du Development in France. The Helmholtz Zentrum Munich and the State of Bavaria financed the KORA study, also supported by the German National Genome Research Network (NGFN-2 and NGFNPUIs: 01GS0823), the German Federal Ministry of Education and Research (BMBP), and the Munich Center of Health Sciences (MC Health, LMUnnovatiy), CIBEROBN and CIBERESP are initiatives of ISCIII (Spain). SWS. holds the GlaxoSmithKline-Canadian Institutes of Health Chair in Genetics and Genomics at the University of Toronto and the Mospital for Sick Children, and is supported by Genome Canada and the McLaughlin Centre. Funding for deCODE came in part from NIH grant MH071425 (K.S.), EU grant HEALTH-2007-2.2.1-10-223423 (Project PsychCNV) and EU grant IMI-Un-VewMeds. NFEC1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 114194, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), NHLBI grant SR01HL087679-02 through the STAMPEED program (TRL1MH083268-01), NHZ Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), NHLBI gra 5R01HL087679-02 through the STAMPEED program (IRL1MH083268-01), NIH/ NIMH (5R01MH63706:02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council, UK (60500539, G0600705, PrevMeS1yn/SALVE). The DNA extractions, sample quality controls, biobank up-keeping and aliquoting was performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki. We thank M. Hass, Z. Jaros, M. Jussila, M. Koiranen,

P. Rantakallio, M. C. Rudolf, V. Soo, O. Tornwall, S. Vaara, T. Ylitalo and the French DHOS national CGH network for their help, as well as all participating patients and clinicians. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions S.J. A.Reymond, P.F. and J.S.B. wrote the manuscript with Author Contributions S.J., A Reymond, P.F. and J.S.B. wrote the manuscript with contributions from F.Z., L.H., R.G.W., N.D.B., Z.K., A.I.F.B. and A.V. L.H., A.V. and A.Reymond produced and analyzed the expression data. Z.K., A.V., R.G.W. and N.D.B. conducted the statistical analyses, guided by S.J., A.Reymond, P.F. and J.S.B. S.J., A.Reymond, F.Z., L.H., D.M., Y.S., G.T., M.S., S.B., D.C., N.L., B.B.A.d.V., B.A.F., F.F.A., M.G., A.G., J.H., A.K., C.L.C., K.M., O.S.P. D.S., M.M.V.H., S.V.G., A.T.V.-V.S., F.W., B.-L.W., Y.Y., J.A., X.E., J.F.G., A.M., S.W.S., K.S., U.T., A.I.F.B., J.S.B., P.F. and all other authors phenotyped and/or genotyped patients and/or individuals of the general population. S.J., A.Reymond and J.S.B. designed the study. All authors commented on and approved the manuscrit. manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. J.S.B. and P.F. are equally contributing senior authors and F.Z., L.H., R.G.W. and Z.K. are equally contributing second authors. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.S.B. (Jacques.Beckmann@chuv.ch).

Sébastien Jacquemont¹*, Alexandre Reymond²*, Flore Zufferey¹, Louise Harewood², Robin G. Walters³, Zoltán Kutalik^{4,5}, Danielle Martinet¹, Yiping Shen^{6,7}, Armand Valsesia^{4,5,5}, Noam D. Beckmann¹, Gudmar Thorfeifsson⁹, Marco Belfiore¹, Sonia Bouquillon¹⁰, Dominique Campion¹¹, Nicole de Leeuw¹⁴, Bert B. A. de Vries¹², Tonia Esko^{13,14}, Toidget A. Fernandes¹⁵, Freiando Fernández-Aranda¹⁵, José Manuel Fernández-Real¹⁷, Monica Gratacòs¹⁸, Audrey Guilmatre¹¹, Juliane Hoyer¹³, Mario-Ritta Jarvein^{17,021/22}, R. Frank Kooy³³, Ants Kurg¹³, Catto Kahana, S. Patté, Damien Sanlaville²⁵, Mieke M. Van Haelst^{3,26}, Sergi Villatoro Gomez¹⁴, Faida Walha², Bal-lin Wu^{1,27}, Yonguo Yu^{1,25}, Azzedine Aboura²⁴, Marie Claude Addor¹, Yves Alemblik⁰, Stylianos E. Antonarakis¹¹, Benoit Arveile^{22,33}, Magalie Barth³⁴, Nathalie Bednarek³⁵, Frédérique Béna³¹, Sven Bergmann⁴⁵, Mylène Berl³⁶, Laura Bernardin¹³⁷, Betha Burneiser²³, Dorninique Boneau³⁴, Armand Bottani³¹, Odile Boute³⁸, Han G. Brunner¹², Dorothée Cailley³³, Patrick Callier³⁹, Jean Christophe Cuvellie⁴², Albert David⁴⁷, Bénédicte de Freminville⁴³, Bruno ⁶, Delote⁴⁴, Marie-Ange Delrue^{32,33}, Génédicte Demeer⁴⁵, Dominique Descamps⁴⁶, Jacet David⁴⁷, Bénédicte Demeer⁴⁵, Jacet David⁴⁷, Bénédicte Demeer⁴⁵, Jacet Leoce, Faure⁴⁵, Anne Fauet⁴⁵, Florence Fellmann¹, Alessandra Ferrarini¹, Richard Fishe⁴⁵, Elisabeth Flori³⁰, Lukas Fore⁴⁷, Josin Guilla⁴¹, Jaaina Kutsane⁴¹, Jetse Antona Kasone⁴⁸, Ante Fauet⁴⁵, Broige the Hund^{42,25}, Suege Hippolyte¹, Muriel Holde⁷⁸, Gera dino Haliso^{42,4}, Sylvie Jailad^{25,65}, Caneta Kasone⁴⁸, Anse Fauet⁴⁵, Florence Fellmann , Alessandra Ferrarini¹, Richard Fishe²⁵, Satu Kaksone⁶⁸, Boris Keren⁵, Anina Kose Brande⁵³, Junge Hande³⁵, Satu Kaksone⁶⁸, Boris Keren⁵, Anina Kose Brandstatter⁵⁷, Nine V. A. Knoer²⁶, Satu Kaksone⁶⁸, Boris Keren⁵⁵, Jinia Kose Brandstatte⁷⁷, Nine V. A. M. Knoer⁴⁵, Sa

¹Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne. ¹Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. ²Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. ³Department of Genomics of Common Disease, Imperial College London, London W12 ONN, UK. ⁴Department of Medical Genetics, University of Lausanne, 1015 Lausanne, Switzerland. ⁵Swiss Institute of Bioinformatics, University of Lausanne, 1015 Lausanne, Switzerland. ⁵Navis Institute of Bioinformatics, University of Lausanne, 1015 Lausanne, Switzerland. ⁵Navis Institute of Bioinformatics, University of Lausanne, 1015 Lausanne, Switzerland. ⁵Navis Centrof or Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁸Ludwig Institute for Cancer Research, University of Lausanne, 1015 Lausanne, Switzerland. ³deCODE Genetics, Sturlugata 8, IS-101 Reykjavik, Iceland. ¹⁰Laboratoire de Génétique Médicale, Hopital Jeanne de Flandre, CHRU de Lille, 59000 Lille, France. ¹¹INSERM U614, University of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, 6500 HB

Nijmegen, The Netherlands, ¹³Institute of Molecular and Cell Biology, University of Tartu, Estonia. ¹⁴Discipline of Genetics and Medicine, Memorial University of Newtoundland, St John's A1B 3V6, Newfoundland, Canada. ¹⁵Department of Psychiatry, University Hospital of Belvitge-IDBELL, Ciber Fisiopatologia Obesidad y Nutricion (CIBEROBN), 08907 Barcelona, Spain. ¹³Department of Diabetes, Endocrinology, and Nutrition, Hospital of Belvitge-IDBELL, Ciber Fisiopatologia Obesidad y Nutricio (CIBEROBN), 08907 Barcelona, Spain. ¹³Department of Diabetes, Endocrinology, and Nutrition, Hospital Universitari de Girona Dr. Josep Trueta, Institut d'Investigació Biomédica de Girona, Ciber Fisiopatologia Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 17007 Girona, Spain. ¹³Genes and Disease Prueta, Institut (DIBERESP), 08003 Barcelona, Catalonia, Spain. ¹⁹Institute of Human Genetics, Friedrich-Alexander University Erlangen-Nuremberg, 91054 Erlangen, Germany. ²⁰Department of Epidemiology and Biostatistics, Imperial College London, School of Public Health, London W2 1PG, UK. ²¹Department of Child and Adolescent Health, National Institute for Health and Welfare, ²⁰Department of Child and Adolescent Health, Sciences, University of Oulu, and Biocenter Oulu, University of Oulu, Box 5000, 90014 University of Oulu, Finland. ²⁰Department of Medical Genetics, University and University of Oulu, and Biocenter Oulu, University and University of Oulu, Finland. ²⁰Department of Medical Genetics, University and University of Oulu, Jand ²⁰Department of Medical Genetics, University and University of Oulu, Jand ²⁰Department of Medical Genetics, University and University of Oulu, Jand ²⁰Department of Medical Genetics, University and University of Oulu, Jand ²⁰Department of Medical Constitution Bioteces, University Adog Nantes, France. ²⁰Service de Cytogénétique Constitutionnelle, Hospices Civils de Lyon, CHU de Juon and ²⁰Penetice Condecid Constitution Construction Constructi <text> Genetics, Technical University Munich, 85764 Neuherberg, Germany. ⁸³Medical Genetics Unit, IRCCS Casa Sollievo della Sofferenza Hospital, 71013 San Giovanni

6 OCTOBER 2011 | VOL 478 | NATURE | 101

©2011 Macmillan Publishers Limited. All rights reserved




Rotondo, Italy, ⁸⁴Genetics, GlaxoSmithKline R&D, 720 Swedeland Road, King of Prussia, Pennsylvania 19406, USA. ⁸⁵Institute of Clinical Chemistry and Laboratory Medicine, Ernst-Moritz-Arndt University Greifswald, D-17475 Greifswald, Germany. ⁸⁶Cardiovascular Genetics Research Unit, EA4373, Université Henri Poincaré, 54000 Nancy, France, ⁸⁷Department of Molecular Medicine and Surgery, Karolinska Institutet, 171 76 Stockholm, Sweden. ⁸⁶Service de Génétique-CLAD Ouest, 35203 Rennes cedex 2, France, ⁸⁹Service de Génétique, CHU Clémenceau, 14033 Caen, France. ⁹⁰Department of Pathology, University of British Columbia and Child and Family Research Institute, Vancouver, British Columbia V52 4H4, Canada. ⁹¹Division of Pediatrics, Ospedale San Giovanni, 6500 Bellinzona, Switzerland. ⁹³Department of Pediatrics and CEBR, G. Gaslin Institute, University of Genova, 16126 Genova, Italy. ⁹⁶Department of Internal Medicine, University of Genova, 16126 Genova, 1849. ⁹⁶Department of Internal Medicine, University of Genova, 16126 Genova, 1849. ⁹⁶Department of Internal Medicine, University of Genova, 16126 Genova, 1849. ⁹⁶Department of Internal Medicine, University of Genova, 16126 Genova, 1849. ⁹⁶Department of Internal Medicine, University of Genova, 16126 Genova, 1849. ⁹⁶Department of Paediatrics 3010 Bern, Switzerland. ⁹⁷The Autism Research Unit, Hospital for Sick Children, Toronto, Ontario MSG 1X8, Canada. ⁹⁷State Diagnostic and Counseling Center, 200 Kopavogur, Iceland. ⁹⁸University of Iceland and Landspitali University Hospital, 101 Reykjavik,

Iceland, ⁹⁹Department of Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ¹⁰⁰Genetic Services, Rinnekoti Research Foundation, Kumputie 1, SF-02980 Espoa, Finland. ¹⁰¹Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland. ¹⁰²Department of Endocrinology and Nutrition, Clinic Hospital of Virgen de la Victoria, Ciber Fisiopatologia Obesidad y Nutrición (CIBROBN), Instituto Salud Carlos III, 29010 Malaga, Spain. ¹⁰³Centre de Maladies Rares, Anomalies du Developpement Nord de France, CH Arras-CHRU Lille, 59000 Lille, France. ¹⁰⁴Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. ¹⁰⁵Institute for Community Medicine, Ernst-Moritz-Arndt University Greifswald, D-17475 Greifswald, Germany. ¹⁰⁶Department of Medical Genetics, Haartman Institute, University of Helsinki and Folkhälsan Institute of Genetics, 00251 Helsinki, Finland. ¹⁰⁷Department of Psychiatry, Oslo University Hospital I, N-0407 Oslo, Norway. ¹⁰⁵The Hospital for Sick Children, University of Toronto, Toronto, Ontario M5G L17, Canada. ¹⁰⁹CNRS 8090-Institute of Biology, Pasteur Institute, 59800 Lille, France. ¹¹⁰UOC Genetica Medica, Azienda Ospedaliera Universitä Nünchen, Klinikum rechts der Isar, 81675 München, Germany. *These authors contributed equally to this work.

102 | NATURE | VOL 478 | 6 OCTOBER 2011

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Study cohorts. For the description of these cohorts, refer to Supplementary Information.

CNV detection. Cases ascertained for intellectual disabilities and developmental delay were identified through standard medical diagnostic procedures. CNV analyses of GWAS data were variously carried out using cnvHap³¹; a moving-window average-intensity procedure; a Gaussian mixture model (Valsesia *et al.*, submitted); circular binary segmentation^{32,33}; QuantiSNP³⁴; PennCNV³⁵; BeadStudio GT module (Illumina Inc.); and Birdseed³⁶ (see below). At least two independent algorithms were used for each cohort.

Patients referred for intellectual disabilities and developmental delay. All diagnostic procedures (CGH, quantitative PCR and/or quantitative multiplex PCR of short fluorescent fragments) were carried out according to the relevant guidelines of good clinical laboratory practice for the respective countries. All rearrangements in probands were confirmed by a second independent method and karyotyping was performed in all cases to exclude a complex rearrangement. Northern Finland 1966 birth cohort (NFBC). CNV calling has been previously described¹⁰. In brief, data were normalized using Illumina BeadStudio, then GC effects on ratios were removed by regressing on GC and GC2, and wave effects were removed by fitting a Loess function³⁷. CNV analysis was done using cnvHap³¹. All called 16p11.2 duplications were validated by direct analysis of log₂ ratios. Data for each probe were normalized by first subtracting the median value across all samples (so that the distribution of ratios for each probe was centred on zero), and then dividing by the variance across all samples (to correct for variation in the sensitivity of different probes to copy-number variation). All CNV calls were confirmed by MLPA.

deCODE genetics. Illumina Human317, Human370, HumanHap550, Human610 and 1M BeadChips were used for CNV analysis. BeadStudio (version 2.0) was used to call genotypes, normalize the signal intensity data and establish the log R ratio (LRR) and Balled frequency (BAF) at every SNP according to standard Illumina protocols. All samples passed a standard SNP-based quality control procedure with a SNP call rate greater than 0.97. PennCNV³⁵, a free, open-source tool, was used for detection of CNVs. The input data for PennCNV are LRR, a normalized measure of the total signal intensity for the two alleles of the SNP, and BAF, a normalized measure of the allelic intensity ratio of the two alleles. These values are derived with the help of control genotype clusters (HapMap samples), using the Illumina BeadStudio software. PennCNV employs a hidden Markov model to analyse the LRR and BAF values across the genome. CNV calls are made on the basis of the probability of a given copy state at the current marker, as well as on the probability of observing a copy-state change from the previous marker to the current one. PennCNV uses a built-in correction model for GC content⁸⁹.

Cohorte Lausannoise (CoLaus). Data normalization and CNV calling have been previously described¹⁰. Data normalization included allelic cross-talk calibration^{39,40}, intensity summarization using robust median average, and correction for any PCR amplification bias. Wave effects were corrected by fitting a Loess function³⁷. CNV calling was done using a Gaussian mixture model (Valsesia *et al.*, submitted) that fits four components (deletion, copy-neutral, one additional copy and two additional copies) to copy-number ratios. The final copy number at each probe location is determined as the expected (dosage) copy number. The method has been validated by comparing test data sets with results from the CNAT⁴¹ and CBS^{32,33} algorithms, and by replicating a subset of CoLaus subjects on Illumina arrays. Only duplications found by both Gaussian mixture model and CBS were considered. **Estonian genome center of the University of Tartu (EGCUT)**. Genotypes were

Estonian genome center of the University of Tartu (EGCUT). Genotypes were called by BeadStudio software GT module v3.1 or GenomeStudio GT v1.6. (Illumina Inc.). Values for LRR and BAF produced by BeadStudio were formatted for further CNV analysis and break-point mapping with Hidden-Markov-Modelbased softwares QuantiSNP (ver.1.1)³⁴ and PennCNV⁴² or CNVPartition 2.4.4 (Illumina Inc.). All analyses were carried out using the recommended settings, except changing EMiters to 25 and L to 1,000,000 in QuantiSNP. For PennCNV, the Estonian-population-specific SNP allele frequency data was used. All detected duplications were confirmed by quantitative PCR.

Study of health in Pomerania (SHIP). Raw intensities were normalized using Affymetrix power tools (Affymetrix); CNV analysis was done using Birdseye from the Birdsuite software package³⁶ and PennCNV³⁵. PennCNV predictions with confidence scores less than 10 were removed. Birdsuite predictions were filtered as in ref. 15: CNV's were kept if their linkage disequilibrium (LOD) score was >10, length >1 kb, number of probes ≥5 and size per number of probes <10,000.

Kooperative Gesundheitsforschung in der Region Augsburg (KORA) F3 and F4. Genotyping for KORA F3 was performed using the Affymetrix 500K array set, consisting of two chips (Styl and Nspl). The KORA F4 samples were genotyped with the Affymetrix human SNP array 6.0. For both studies, genomic DNA from blood samples was used for analysis. Hybridization of genomic DNA was done in accordance with the manufacturer's standard recommendations. Genotyping was done in the Genome Analysis Centre of the Helmholtz Centre Munich. Genotypes were determined using BRLMM clustering algorithm (Affymetrix 500K array set) and Birdseed2 clustering algorithm (Affymetrix array 6.0). For quality control purposes, we applied a positive control and a negative control DNA every 48 samples (KORA F3) or 96 samples (KORA F4). On the chip level, only subjects with overall genotyping efficiencies of at least 93% were included. In addition, the called gender had to agree with the gender in the KORA study database. After exclusions, 1,644 individuals remained in KORA F3 and 1,814 in KORA F4 for further analysis.

MLPA analysis. We used MLPA to determine changes in the copy number of a region of about 2 Mb on chromosome 16p11.2. Briefly, we designed, using hg18, nine probes within the targeted region, one control probe outside the rearranged region and seven control probes targeting unique position in the genome (Supplementary Table 10). Assays were performed with MRC-Holland reagents according to the manufacturer's protocol⁴⁵. The analysis of the amplification products was performed by capillary electrophoresis in the DNA Analyser 3730XL and using the GeneMapper software v3.7 (Applied Biosystems). The calculations were performed independently for each experimental wer first normalized the MLPA data to minimize the amount of experimental wariation, summing all signal values of each control probe for each sample, and then dividing each signal values of each control probe for each same experiment, dividing the normalized signal values by the average calculated from all the same experiment, dividing the same experiment. The product of this calculation is termed dosage quotient (DQ). A DQ value of less than 0.65 or more than 1.25 was considered as copy-number loss or gain, respectively, as previously described⁴¹⁻⁴⁶.

Custom array-CGH for the short arm of chromosome 16. DNA samples were labelled with Cy3 and cohybridized to custom-made Nimblegen arrays with Cy5labelled DNA from the CEPH cell line GM12042. These arrays contained 71,000 probes spread across the short arm of chromosome 16 from 22.0 Mb to 32.7 Mb (at a median space of 45 bp between 27.5 Mb and 31.0 Mb), and 1,000 control probes situated in invariable regions of the X chromosome. DNA labelling, hybridization and washing were performed according to Nimblegen protocols. Scanning was performed using an Agilent G2565BA microarray scanner. Image processing, quality control and data extraction were performed using the Nimblescan software v.2.5.

Defining underweight. Underweight was defined throughout the study as $BMI \le 18.5 \text{ kg per m}^2$ in adults and $\le -2 \text{ s.d. in children}^{30,47,48}$.

Weight, height, BMI and head circumference Z-scores as a function of age. For paediatric cases, weight, height, BMI and head circumference Z-scores were determined for paediatric cases (0–18 years of age) using clinical growth charts specific to the country of origin. Children were ascertained from nine different countries. If charts were only available in percentiles, those measures were transformed into Z-scores using gender-, age- and geographically-matched reference populations (see Statistics).

For the USA and Canada, data from the Center for Disease Control and National Center for Health Statistics (CDC/NCHS) were used to calculate Z-scores⁴⁹.

For the French paediatric population, we used French national growth charts^{50,51}. For the Swiss paediatric population, we used Swiss national growth charts⁵². For Dutch participants, Dutch national growth charts were used³³. For Italian, German, Finnish and Austrian cases (n = 6), height, weight and BMI Z-scores were estimated using WHO growth charts⁵⁴.

To check for discrepancies generated by the use of different growth charts, height, weight and BMI Z-scores were recalculated using WHO growth charts for all cases under five years of age, regardless of origin (http://www.who.int/ childgrowth/standards/enf/54). Z-scores obtained using the WHO data were not significantly different. These growth standards, developed by the World Health Organization multicentre growth reference study, describe normal child growth from birth to 5 years under optimal environmental conditions. These standards can be applied to all children everywhere, regardless of ethnicity, socioeconomic status and type of feeding^{55,6}.

If necessary, percentile values were transformed to Z-scores by the inversenormal density function. When growth charts were unavailable, we used reported LMS parameters (median (M), generalized coefficient of variation (S) and skewness (L)) to obtain Z-scores via the formula:

Z-score =
$$\begin{cases} \frac{(X/M)^L - 1}{L + S}, & L > 0\\ \ln(X/M)/S L = 0, \end{cases}$$

in which X is the observed value.

In adults (>18 years of age), we estimated LMS parameters when these were unavailable from the available sex-, age- and origin-matched Swiss (CoLaus),

©2011 Macmillan Publishers Limited. All rights reserved





Estonian or French control populations. For cases identified from population-based cohorts, Z-scores were directly inferred from the cohort.

Gene expression. We established lymphoblastoid cell lines from deletion and duplication carriers, as well as from controls (Supplementary Table 12), by transforming peripheral blood mononuclear cells with Epstein-Barr virus. Patients and controls were enrolled after obtaining appropriate informed consent via the physi-cians in charge, and approval by the ethics committee of the University of Lausanne. More control cell lines were obtained from Coriell Institute for Medical Research (http://www.coriell.org/) (Supplementary Table 12). SYBR Green real-time quant-itative PCR (RT–PCR) was performed as previously described^{22,57}. Briefly, 1 µg of total RNA from lymphoblastoid cell lines was converted to complementary DNA using Superscript VILO (Invitrogen) primed with a mixture of oligo(dT) and random hexamers. Oligos were designed using the PrimerExpress program (Applied Biosystems) with default parameters (Supplementary Table 11). Non-intronspanning assays were tested for genomic contamination in standard \pm reverse transcriptase reactions. The amplification efficiency of each primer pair was tested in a cDNA dilution series, as previously described⁵⁶. A full list of genes mapping in the rearranged interval, and exclusion criteria, are presented in Supplementary Table 1. All RT-PCR reactions were performed in a 10-µl final volume and triplicates per sample. The setup in a 384-well plate format was performed using a Freedom EVO robot (TECAN) and assays were run in an ABI 7900 sequence detection system (Applied Biosystems) with the following amplification conditions: 50 $^\circ\rm C$ for 2 min, 95 $^\circ\rm C$ for 10 min, and 45 cycles of 95 $^\circ\rm C$ 15 s, then 60 $^\circ\rm C$ for 1 min. A final incubation of 95 $^\circ$ C for 15 s followed by 60 $^\circ$ C for 15 s was carried out to establish a dissociation curve. Each plate included the appropriate normalization to genes to control for an variability between plate runs. Raw threshold cycles (Ct) values were obtained using SDS2.4 (Applied Biosystems). To calculate the normalized relative expression ratio of individuals carrying the CNV and of controls, we used Biogazelle qBase Plus software⁵⁹ including geNorm⁶⁰. This program identified appropriate normalization genes (*EEF1A1*, *RPL13*, *GUSB* and *TBP*) having a gene-stability measure of M = 0.25. We note that one gene, *LAT*, showed a very high expression profile in one of the duplication samples (DASYL, Supplementary Table 13), reaching a relative expression value of 27.3 (s.e.m. = 1.37), compared to an average expression for other duplications of 1.89 (s.e.m. = 0.51). We cannot exclude that this finding is genuine (and confirmed it in a second experiment), but it was removed from further analyses as an outlier to give a more accurate overview of expression profiles for these genes.

In silico analysis was performed to check for brain, and specifically hypothalamus, expression of genes in the rearranged 16p11.2 interval (Supplementary Table 1). This was done using Allen Brain Atlas Resources, available from http://www.brain-map.

Cases with major neurological signs. Major neurological signs were defined by moderate to severe hypotonia, hypertonia, ataxia, spasticity, hypereflexia, hyporeflexia and/or extra-pyramidal signs, and by the presence of epilepsy.

Statistics. Student's t-test: one-tailed t-tests were performed to test whether duplication carriers have Z-score values lower than zero for BMI, height and weight. We found this analysis more suitable than linear regression analysis, correcting for confounding factors such as sex and age, because these anthro-pometric traits have a highly nonlinear dependence on these factors, as can be observed in control populations. Kruskal–Wallis test: this was used to test differences in the gene expression

pattern between deletion and duplication carriers and control individuals. Because expression values are not necessarily normally distributed, this test is more adequate than a classical one-way analysis of variance. To test pairwise differences, we computed the difference in mean group rank with its 95% confidence interval (as provided by the multcompare function in Matlab). Correction for multiple testing was done using a Bonferroni adjustment. Multiple testing: we determined false-discovery-rate-based thresholds for asso-

ciation *P*-values for each phenotype, to correct for multiple testing. For each phenotype, we replaced the observed *Z*-scores with numbers randomly drawn from a standard normal distribution and performed the same t-tests for the same strata. The procedure was repeated 1,000 times. For various P-value thresholds, we asked how many tests would be declared significant for the null set on average (over the 1,000 random draws). The false discovery rate was estimated as the ratio of this number and the actual number obtained for the observed Z-scores. Thus, we controlled the dependence between nested tests. Relative risk: among adults, we defined underweight as a BMI <18.5 (WHO

criteria). The estimated relative risk is the ratio of the fraction of underweight individuals among duplication carriers versus our control group. The standard error of log(relative risk) and its significance were calculated as previously described⁶¹. In our control group (population-based cohorts), the frequency of being underweight is 1.9% (38 males and 148 females out of 9,470). Owing to the fact that being underweight decreases with age in the general population, we resampled our control group to ensure precise age-matching

- Coin, L. J. *et al.* cnvHap: an integrative population and haplotype-based multiplatform model of SNPs and CNVs. *Nature Methods* 7, 541–546 (2010).
 Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 7, 572 (2004).
- Segmentation for the analysis of an arg-based by Copy Namber Gata. *Biostatistic 5*, 557–5572 (2004).
 Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 23, 657–653 (2007).
 Colella, S. *et al.* QuantiSNP: an objective Bayes Hidden-Markov model to detect 33.
- Coleria, S. et al. Qualitistics an objective Bayes inducti-Markov model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 35, 2013–2025 (2007). Wang, K. et al. PenncTV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17, 1665–1674 (2007). 35.
- cata. genome res. 17, 1659–1674 (2007). Korn, J. M. et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nature Genet. 40, 1253–1260 (2008). Marioni, J. C. et al. Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. Genome *Biol* 8, 828 (2007). Biol. 8, R228 (2007).
- Diskin, S. J. et al. Adjustment of genomic waves in signal intensities from whole
- Diskin, S. Stear. Adjustment of genome area in signal interfaces rolling genome SNP genotyping platforms. Nucleic Acids Res. 36, e126 (2008). Bengtsson, H., Irizarry, R., Carvalho, B. & Speed, T. P. Estimation and assessment c raw copy numbers at the single locus level. *Bioinformatics* 24, 759–767 (2008). Bengtsson, H., Ray, A., Spellman, P. & Speed, T. P. A single-sample method for 40.
- 41
- Derigsson, A., Ay, A., Speinnan, F. & Speed, T. F. A Single-sample riteriou for normalizing and combining full-resolution copy numbers from multiple platforms, labs and analysis methods. *Bioinformatics* **25**, 861–867 (2009). Huang, J. *et al.* Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum. Genomics* **1**, 287–299 (2004). Wang, B. *et al.* Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* **316**, 1194–1198 (2007).
- 43. Schouten, J. P. et al. Relative quantification of 40 nucleic acid sequences ultiplex ligation-dependent probe amplification. Nucleic Acids Res. 30, e57 2002).
- Bunyan, D. J. et al. Dosage analysis of cancer predisposition genes by multipl ligation-dependent probe amplification. Br. J. Cancer **91**, 1155–1159 (2004). 44. Itiplex
- Iligation-dependent probe amplification. *Br. J. Cancer* **91**, 1155–1159 (2004). Fernández, Let *J.* Comparative study of three diagnostic approaches (FISH, STRs and MLPA) in 30 patients with 22q11.2 deletion syndrome. *Clin. Genet.* **68**, 373–378 (2005). Slater, H. R. *et al.* Rapid, high throughput prenatal detection of aneuploidy using a novel quantitative method (MLPA). *J. Med. Genet.* **40**, 907–912 (2003). 45.
- 47
- Hoter quantitative interford (MLFA), 3. Med. Genet. 40, 907–912 (2003).
 Mei, Z. et al. Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents.
 Am. J. Clin. Nutr. 75, 978–985 (2002).
 Physical. status: the use and interpretation of anthropometry. Report of a WHO expert committee. World Health Organ. Tech. Rep. Ser. 854, 1–452 (1995).
 Kuczmarski, R. J. et al. CDC growth charts: United States. Adv. Data 314, 1–27
- 49. (2000)
- 50.
- (2000) Sempé, M., Pedron, G. & Roy-Pernot, M. P. Auxologie, Méthode et Séquences. (Théraplix, 1979). Rolland-Cachera, M. F. et al. Body mass index variations: centiles from birth to 87 years. *Eur. J. Clin. Nutr.* **45**, 13–21 (1991). 51.
- years. *Lur. J. Clin. Nutr.* **45**, 13–21 (1991). Prader, A., Largo, R. H., Molinari, L. & Issler, C. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. *Heiv. Paediatr. Acta,* Suppl. 52, 1–125 (1989). Fredriks, M. Growth Diagrams, 1997. Fourth Dutch Nation-wide Survey. 233–242 (Bohn Stafleu Van Loghum, 1997). 52.
- 53.
- de Onis, M. et al. Development of a WHO growth reference for school-aged children 54.
- George Ching, M. et al. Development for a Who grown Testereneo for Schoorage or limiter and adolescents. Bull. World Health Organ. 85, 660-667 (2007).
 de Onis, M., Garza, C., Onyango, A. W. & Borghi, E. Comparison of the WHO child growth standards and the CDC 2000 growth charts. J. Nutr. 137, 144–148 (2007).
 Mei, Z., Ogden, C. L., Flegal, K. M. & Grummer-Strawn, L. M. Comparison of the
- prevalence of shortness, underweight, and overweight among US children aged 0 prevalence of shorthess, under weight, and of weight and of the shorthess of the short of the
- (2008).
- Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25, 402–408 58. (2001)
- Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* 8, R19 (2007).
- analysis of real-time quantitative PCR data. Genome Biol. **8**, N19 (2007).
 Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. **3**, 1–11 (2002).
 Morris, J. A.& Gardner, M. J. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. Br. Med. J. (Clin. Res. Ed.) **296**,
- 1313-1316 (1988).

©2011 Macmillan Publishers Limited. All rights reserved

Article 5

Neurogenetics (2010) 11:273–274 DOI 10.1007/s10048-009-0227-8

LETTER

Evidence against haploinsuffiency of human ataxin 10 as a cause of spinocerebellar ataxia type 10

Boris Keren • Aurélia Jacquette • Christel Depienne • Patricia Leite • Alexandra Durr • Wassila Carpentier • Baya Benyahia • Gerard Ponsot • Florent Soubrier • Alexis Brice • Delphine Héron

Received: 21 October 2009 / Accepted: 23 October 2009 / Published online: 20 November 2009 © Springer-Verlag 2009

Spinocerebellar ataxia type 10 (SCA10) is an autosomal dominant neurodegenerative disorder characterized by ataxia and cerebellar atrophy and seizures in some patients. A heterozygous unstable tandem expansion of a pentanucleotide ATTCT is responsible for the disease. This mutation lies within intron 9 of the ataxin 10 gene (ATXN10), localized on 22q13.31 [1]. The size of the repeat ranges from 10 to 29 in normal alleles and from 800 to 4,500 in SCA10. The exact pathological mechanism of that expansion in triggering the disease is not yet understood.

Here, we report on a familial translocation t(2;22) (p25.3;q13.31) in three sisters and their father. The first and third sisters have severe mental retardation caused by

B. Keren (⊠) · A. Jacquette · C. Depienne · P. Leite · A. Durr ·
B. Benyahia · F. Soubrier · A. Brice · D. Héron
Département de Génétique et Cytogénétique, CGMC,
GH Pitié-Salpêtrière, AP-HP,
Paris, France
e-mail: boris.keren@psl.aphp.fr

B. Keren · C. Depienne · A. Durr · A. Brice Université Pierre et Marie Curie-Paris 6, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière, Inserm, U975, Cnrs, UMR 7225, Paris, France

W. Carpentier · F. Soubrier Plateforme Post-génomique de la Pitié-Salpêtrière, Inserm, UPMC Paris 6, GH Pitié-Salpêtrière, Paris, France

G. Ponsot

Service de Neuropédiatrie et Pathologie du Développement, Hôpital Armand Trousseau, APHP, Paris, France the unbalanced form of the translocation with monosomy 22gter and trisomy 2pter. Besides, their father and the second sister bear the balanced form of the translocation. That anomaly has been precisely characterized, first by analyzing the DNA from the sisters with the unbalanced translocation on microarray (Illumina® HumanCytoSNP-12 and Roche Nimblegen® Human CGH 2.1 M) then by direct sequencing of the breakpoints of the balanced translocation. The exact localization of those breakpoints lies at genomic positions chr2:5495703-5495704 and chr22:44466085-44466086 (hg18 assembly). On the balanced translocation, there is no gene disruption on chromosome 2, but the breakpoints on chromosome 22 fall in the intron 2 of ATXN10: Only the first 2 exons remain on the derivative chromosome 22, whereas the last 10 exons have been translocated on the derivative chromosome 2 (Fig. 1).

Careful neurologic examination of both the patients with the balanced translocation showed neither ataxia nor any other neurologic signs. Neuroimaging of the father showed a normal cerebellum (Fig. 1). A later onset of ataxia cannot be excluded in the 23-year-old daughter but is very unlikely in the 57-year-old father, since the first signs of the disease appear between 20 and 50 years of age in the previously reported cases. Therefore it can be assumed that both the father and the second sister are not affected by SCA10.

ATXN10 is a broadly expressed gene of unknown function. Animal models have shown that a knockdown of the protein provokes degeneration of rat cerebellar neurons in culture which would argue for a loss of function mechanism [2]. On the contrary, other experiments have shown that SCA10-null heterozygous mutant mice have normal brains and no ataxia, which does not support the loss of function mechanism [3]. Both studies insisted on the



Fig. 1 a Brain sagittal T2-weighted MRI showing a normal cerebellum in the father with the balanced translocation. b View of the 22q13.3 breakpoint with Illumina[®] HumanCytoSNP-12 and

c Roche Nimblegen[®] Human CGH 2.1 M DNA chips. d Localization of the 22q13.3 breakpoint in the intron 2 of ATXN10 (using BLAT in UCSC genome browser)

biological differences between animal models and humans and therefore on the necessity to study patients' samples in order to definitely assess the pathological effect of the expansion. In our report, both the father and the daughter with the balanced translocation have a heterozygous disrupted ATXN10 and no sign of SCA10, which is consistent with the latter study on mice and confirms that haploinsufficiency is not responsible for SCA10 in humans.

Moreover, recent experiments on somatic hybrid cell lines have shown that there is no interference with the level of transcription or RNA processing [3]. All those data are against a simple gain or loss of function mechanism, which is supported by recent studies focusing on the genomic effect of the mutation: Expansion of the pentanucleotide repeat causes DNA unpairing in yeast [4] and in human cell lines [5], and such unpairing creates ectopic DNA replication origins [5] and genomic instability [6], which may lead to cell degeneration. Thus, SCA10 could be related to the chromatin structure abnormalities caused by the expansion and not to an abnormal or abnormally expressed ATXN10. ATXN10 haploinsufficency observed in our patients without SCA10 is consistent with this hypothesis.

References

- Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, Khajavi M, McCall AE, Davis CF, Zu L, Achari M, Pulst SM, Alonso E, Noebels JL, Nelson DL, Zoghbi HY, Ashizawa T (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet 26:191–194
- März P, Probst A, Lang S, Schwager M, Rose-John S, Otten U, Ozbek S (2004) Ataxin-10, the spinocerebellar ataxia type 10 neurodegenerative disorder protein, is essential for survival of cerebellar neurons. J Biol Chem 279:35542–35550
- Wakamiya M, Matsuura T, Liu Y, Schuster GC, Gao R, Xu W, Sarkar PS, Lin X, Ashizawa T (2006) The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10. Neurology 67:607–613
 Potaman VN, Bissler JJ, Hashem VI, Oussatcheva EA, Lu L,
- Potaman VN, Bissler JJ, Hashem VI, Oussatcheva EA, Lu L, Shlyakhtenko LS, Lyubchenko YL, Matsuura T, Ashizawa T, Leffak M, Benham CI, Sinden RR (2003) Unpaired structures in SCA10 (ATTCT)n. (AGAAT)n repeats. J Mol Biol 326:1095–1111
- Liu G, Bissler JJ, Sinden RR, Leffak M (2007) Unstable spinocerebellar ataxia type 10 (ATTCT*(AGAAT) repeats are associated with aberrant replication at the ATX10 locus and replication origin-dependent expansion at an ectopic site in human cells. Mol Cell Biol 27:7828–7838
- Edwards SF, Hashem VI, Klysik EA, Sinden RR (2009) Genetic instabilities of (CCTG).(CAGG) and (ATTCT).(AGAAT) diseaseassociated repeats reveal multiple pathways for repeat deletion. Mol Carcinog 48(4):336–349

D Springer

npg

Article 6

European Journal of Human Genetics (2013), 1–8 © 2013 Macmillan Publishers Limited All rights reserved 1018-4813/13 www.nature.com/ejhg

ARTICLE

Prospective diagnostic analysis of copy number variants using SNP microarrays in individuals with autism spectrum disorders

Caroline Nava^{1,2,3,4,19}, Boris Keren^{5,19}, Cyril Mignot^{4,6,7,8}, Agnès Rastetter^{1,2,3}, Sandra Chantot-Bastaraud⁹, Anne Faudet⁴, Eric Fonteneau⁵, Claire Amiet¹⁰, Claudine Laurent^{1,2,10}, Aurélia Jacquette^{4,7,8}, Sandra Whalen⁴, Alexandra Afenjar^{4,6,7,8,11}, Didier Périsse^{10,12}, Diane Doummar⁶, Nathalie Dorison^{6,11}, Marion Leboyer^{13,14,15,16}, Jean-Pierre Siffroi⁹, David Cohen^{10,17}, Alexis Brice^{*,1,2,3,4,18}, Delphine Héron^{4,6,7,8} and Christel Depienne^{*,1,2,3,4,18}

Copy number variants (CNVs) have repeatedly been found to cause or predispose to autism spectrum disorders (ASDs). For diagnostic purposes, we screened 194 individuals with ASDs for CNVs using Illumina SNP arrays. In several probands, we also analyzed candidate genes located in inherited deletions to unmask autosomal recessive variants. Three CNVs, a *de novo* triplication of chromosome 15q11–q12 of paternal origin, a deletion on chromosome 9p24 and a *de novo* 3q29 deletion, were identified as the cause of the disorder in one individual each. An autosomal recessive cause was considered possible in two patients: a homozygous 1p31.1 deletion encompassing *PTGER3* and a deletion of the entire *DOCK10* gene associated with a rare hemizygous missense variant. We also identified multiple private or recurrent CNVs, the majority of which were inherited from asymptomatic parents. Although highly penetrant CNVs or variants inherited in an autosomal recessive manner were detected in rare cases, our results mainly support the hypothesis that most CNVs contribute to ASDs in association with other CNVs or point variants located elsewhere in the genome. Identification of these genetic interactions in individuals with ASDs constitutes a formidable challenge.

European Journal of Human Genetics advance online publication, 1 May 2013; doi:10.1038/ejhg.2013.88

Keywords: copy number variants; autism spectrum disorders; 15q11-q12 triplication; autosomal recessive inheritance; genetic interactions

INTRODUCTION

Autism is a neuropsychiatric disorder characterized by impaired social interactions and communication, as well as restricted, repetitive and/ or stereotyped behaviors. The term autism spectrum disorders (ASDs) is now commonly used to designate this group of highly heterogeneous and complex developmental diseases. Autism is associated with intellectual disability (ID) in ~75% in the patients and with epilepsy or EEG abnormalities in ~15–25%. ASDs are more common in male patients than in female patients, in a ratio of 4:1 that can reach 10:1 in forms with normal cognitive abilities.¹

As indicated by the increased risk of recurrence in families and concordance in twin pairs, ASDs are largely genetically determined.^{1,2} Recent studies have demonstrated that ASDs can be caused by rare,

highly penetrant point mutations, deletions, duplications and larger chromosomal abnormalities that can either arise *de novo* or be inherited.^{3–9} Known monogenic disorders account for 2–5% of syndromic cases; fragile X syndrome is usually the most common cause, followed by PTEN macrocephaly syndrome and tuberous sclerosis, each accounting for <1% of individuals with ASD.^{1,10} Large copy number variants (CNVs) are found in 5–10% of autistic patients, especially in those with syndromic ASDs.^{5,11,12} Several CNVs have been repeatedly identified in individuals with ASD, such as deletions or duplications on chromosomes 15q11–q13, 16p11.2, 7q11.23 and 22q13.^{4,5,13} Private CNVs in individuals with autism often disrupt genes encoding synaptic proteins, such as neuroligins, neurexins and SHANK proteins, or neuronal cell-adhesion proteins,

¹INSERM, U975 (CRICM), Institut du cerveau et de la moelle épinière (ICM), Hôpital Pitié-Salpêtrière, Paris, France; ²CNRS 7225 (CRICM), Hôpital Pitié-Salpêtrière, Paris, France; ³Université Pierre et Marie Curie-Paris-6 (UPMC), UMR_S 975, Paris, France; ⁴AP-HP, Hôpital Pitié-Salpêtrière, Département de Génétique et de Cytogénétique, Unité fonctionnelle de cytogénétique, Paris, France; ⁵AP-HP, Hôpital Pitié-Salpêtrière, Paris, France; ⁵AP-HP, Hôpital Pitié-Salpêtrière, Paris, France; ⁴AP-HP, Hôpital Pitié-Salpêtrière, Département de Génétique et de Cytogénétique, Unité fonctionnelle de cytogénétique, Paris, France; ⁵AP-HP, Hôpital Pitié-Salpêtrière, Département de Génétique et de Cytogénétique, Unité fonctionnelle de cytogénétique, Paris, France; ⁵AP-HP, Hôpital Pitié-Salpêtrière, Paris, France; ⁶Corupe de Recherche Clinique (GRC) 'déficience intellectuelle et autisme' UPMC, Paris, France; ⁵AP-HP, Hôpital Armand Trousseau, Service de Neuropédiatrie, Paris, France; ⁵AP-HP, Hôpital Pitié-Salpêtrière, Bojtal Armand Trousseau, Paris, France; ¹⁴Centre Diagnostic Autisme, Pitié-Salpêtrière de l'afforta de l'adolescent, Paris, France; ¹¹Centre de Référence 'déficience des anomalies du développement et syndromes malformatifs, Hôpital TrussERM, U955, Créteil, France; ¹⁴Centre Diagnostic Autisme, Pitié-Salpêtrière, Prios, France; ¹⁴Université Paris Est, Faculté de médecine, Créteil, France; ¹²Centre Diagnostic Autisme, Pitié-Salpêtrière, Pioteil, France; ¹⁶Fondation FondaMental, Créteil, France; ¹⁷Institut des Systèmes Intelligents et Robotiques, CNRS UMR 7222, UPMC-Paris-6, Paris, France; ¹⁸AP-HP, Hôpital Pitié-Salpêtrière, Département de Génétique et de Cytogénétique et de Cytogénétique undéculaire et cellulaire, Paris, France; ¹⁸AP-HP, Hôpital Pitié-Salpêtrière, Département de Génétique et de Cytogénétique, Unité fonctionnelle de neurogénétique et et curvagénétique et et de Cytogénétique Unité Rotonnelle de neurogénétique et dellulaire, Paris, France; ¹⁸AP-HP, Hôpital Pit

^{*}Correspondence: Dr C Depienne or Professor A Brice, INSERM, U975 (CRICM), Institut du cerveau et de la moelle épinière (ICM), Hôpital Pítié-Salpêtrière, Paris 75013, France. Tel: + 33 1 57 27 46 82; Fax: + 33 1 57 27 47 95; E-mail: christel.depienne@upmc.fr or alexis.brice@upmc.fr Received 11 December 2012; revised 23 February 2013; accepted 28 March 2013

CNV analysis in ASD C Nava et al

which have important roles in neurodevelopment and/or neurotransmission^{3,14–17} However, most CNVs and variants in genes involved in autism so far are also found in other neurodevelopmental disorders such as ID without autistic features and/or schizophrenia.^{12,18} Recent studies point to polygenic or oligogenic inheritance.^{6,17,19} Indeed, most of the abnormalities identified have been associated with highly variable phenotypes and seem insufficient to cause ASDs on their own. It is probable that genetic interactions (epistasis) between rare variants have an important role in the etiology of ASDs.

In this study, we screened 194 subjects with ASDs using SNP microarrays to assess genomic causes and establish genetic syndrome diagnoses. Three heterozygous CNVs, a *de novo* triplication of chromosome 15q11–q12 of paternal origin, a distal deletion on chromosome 9p24 and a *de novo* 3q29 deletion, and one homozygous deletion on chromosome 1p31.1 encompassing *PTGER3* were considered to probably cause the disorder in four individuals. In addition, we screened candidate genes located in inherited deletions to unmask autosomal recessive variants in three patients, and found a rare missense variant in *DOCK10* associated with an inherited deletion on chromosome 2q in one individual.

MATERIALS AND METHODS

Cohort description

This study first included 200 subjects with ASDs (168 males and 32 females) who were recruited in the 'Centre de Référence déficiences intellectuelles de causes rares', the 'Centre Diagnostic Autisme', Pitié-Salpêtrière Hospital (Paris, France) or the 'Fondation Lejeuné' over a period of 3 years (2009-2011). Index cases were assessed with the Autism Diagnostic Interview-Revised (ADI-R): 167 index cases (83.5%) had autism with ID and 33 (16.5%) had Asperger syndrome or high-functioning autism based on the DSM IV-TR criteria; 85% (170/200) of the patients were sporadic cases. Preliminary systematic screening included karyotyping, screening for fragile X syndrome and targeted metabolic or genetic analyses, including fluorescence in situ hybridization (FISH) analysis of regions involved in Williams-Beuren and Smith-Magenis syndromes and screening of the RAI1 gene for selected patients presenting compatible clinical features. These analyses allowed a clinical diagnosis in six males, one with fragile X syndrome, two with a supernumerary chromosome 15 of maternal origin, one with a 7a11.23 deletion, one with a 17p11.2 deletion and one with a RAI1 mutation. These six individuals were excluded from further analyses. CNV analysis was therefore performed on 194 subjects. Informed written consent was obtained from each individual or his/her parents before blood sampling. All experiments were performed in accordance with French guidelines and legislation.

High-density SNP arrays

Genomic DNA was extracted from blood cells using standard phenolchloroform or saline procedures. Index cases and the parents of individuals with possibly pathogenic CNVs were genotyped using either 370CNV-Quad (n=20), 660W-Quad (n=27) or cytoSNP-12 (n=147) microarrays (Illumina, San Diego, CA, USA). Automated Illumina microarray experiments were performed according to the manufacturer's specifications. Image acquisition was performed using an iScan System (Illumina). Image analysis and automated CNV calling was performed using GenomeStudio v.2011.1 and CNVPartition v.3.1.6 with the default confidence threshold of 35. We also used a size threshold of 20 kb for individuals analyzed with 660W-Quad microarrays, as CNVs <20 kb detected by these arrays proved to be false positives. Identified CNVs were systematically compared with those present in the Database of Genomic Variants (DGV), excluding BAC-based studies, using an in-house bioinformatics pipeline, to assess their frequency in control populations. Only CNVs identical (breakpoints and copy number) to or totally included in those described in the DGV were considered; microrearrangements partially overlapping CNVs described in DGV or with a discordant copy number were considered to be novel. CNVs with a minor allele frequency \geq 1% in at least one study comprising \geq 30 controls were excluded from further analysis. CNVs encompassing coding regions, and with a frequency <1%, were considered as possibly deleterious and were searched for in the parents of the patients when DNA was available (Figure 1). CNV frequencies were compared with the Mann-Whitney or unilateral Fisher's exact tests. Candidate genes present in the CNVs were compared with those present in AutDB (http://www.mindspec.org/autdb.html).²⁰

Analysis of candidate genes by Sanger sequencing

Specific primer pairs (available on request) were designed to screen the coding regions of DOCK10, DOCK8, NYAP2, SMARCA2 and ATP2C2 in a research context. Forward and reverse sequence reactions were performed with the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA, USA). G50-purified sequence products were run on an ABI 3730 automated sequencer (Applied Biosystems) and data were analyzed with Seqscape v.2.6 software (Applied Biosystems). Missense variants were assessed *in silico* for possible pathogenicity using PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), SIFT (http://sift.biia-star.edu.sg/) and SNPs&GO (http:// snps-and-go.biocomp.unibio.it/).



Figure 1 Strategy used for the selection of rare, possibly deleterious CNVs. The criteria included the presence of coding regions in the CNV, a minor allele frequency (MAF) <1% in the DGV, a *de novo* occurrence of the CNV, a previous association with ASDs, the recurrence of rare CNVs in our cohort and the presence of genes of interest (ie expressed in the brain and with a function related to that of genes previously involved in ASDs). The CNVs highlighted in this toty are indicated in italics for each subcategory.

European Journal of Human Genetics

npg 2

CNV analysis in ASD C Nava et al

Microsatellite marker analysis

Three microsatellite markers in the 15q11-q13 region (D15S128, D15S122 and D15S822) were used to follow the transmission of the alleles, as already described.¹³

Fluorescence in situ hybridization

To confirm array results, we performed FISH analyses on peripheral blood lymphocytes with the subtelomere 9p, UBE3A probes (Cytocell, Cambridge, UK) and RP1-196F4 in 3q29, according to the supplier's recommendations.

Methylation analysis

Whole methylation analysis was performed for the subject with the 15q11–q12 triplication (8082), a patient with Angelman syndrome presenting with autistic features, previously reported,¹³ and two young male controls by Integragen SA (Evry France). Bisulfite conversion of DNA was carried out to convert unmethylated cytosine nucleotides to uracil. Converted DNA was hybridized on Infinium HumanMethylation450 BeadChips (Illumina), according to the manufacturer's specifications. Microarrays were scanned using an Illumina BeadArray Reader, and the data were analyzed using GenomeStudio (Illumina).

RESULTS

A total of 130 CNVs encompassing at least one exon and absent or rare (minor allele frequency <1%) in the DGV were found in 94 subjects with ASD (Supplementary Table S1). These CNVs included one triplication, 82 duplications, 46 heterozygous deletions and 1 homozygous deletion. The size of the CNVs ranged from 5 to 4480 kb, with a median at 138 kb. DNA from both parents of 48 patients was available for investigation of the transmission of the CNV. Three CNVs, one triplication of 15q11-q12, one deletion on chromosome 3q29 and one duplication of 16p11.2, were absent from both parents, demonstrating that they occurred de novo in the index case (Table 1 and Supplementary Clinical Data). Seventy-four CNVs (26 deletions and 48 duplications) were inherited from an asymptomatic parent. A homozygous deletion on chromosome 1p31.1, encompassing the PTGER3 gene, was found in one patient; this deletion was heterozygous in both parents, who were both from The Comoro Islands.

Three CNVs, all confirmed by FISH analysis, were considered to be the probable cause of ASD, based on their large size, absence from healthy individuals, de novo occurrence and/or presence in previously reported syndromes associated with ID and ASD: a de novo triplication of the 15q11-q12 region in a male subject (family 772), a deletion of the 9p24 region in a female (family 15) and a deletion of 3q29 in a male (family 1315) (Figure 2 and Supplementary Figure S3). The de novo 15q11-q12 triplication spanned 4.48 Mb and included the Angelman/Prader-Willi critical region between breakpoints BP2 and BP3. Analysis of microsatellite markers in the critical region revealed that the triplication was of paternal origin and resulted from an interchromosomal recombination since it included two different paternal alleles. FISH analysis indicated that the three paternal copies were located on the same chromosome at the 15q11-q12 locus. Methylation analysis of genes located within the triplicated region confirmed increased methylation values for UBE3A and decreased methylation values for SNRPN. Interestingly, this analysis also revealed abnormal methylation values for non-imprinted genes in the region including OCA2 and GABRG3. The de novo deletion on chromosome 3q29, found in an individual with high-functioning autism, spanned 1.6 Mb and encompassed 22 genes. The deletion on chromosome 9p24 spanned 3.8 Mb and encompassed 13 genes, including SMARCA2, which encodes a protein belonging to the FMRP complex, and DOCK8. Only the DNA of the mother, who did not have the deletion, was available.

| A P | oband \$ | | | | | | | | | | | Age at | Heign (SU)/weight (SU)/ | | | |
|-----|----------|--------------------------|---|-----------|---------------------------------|------------------------------------|------------------|-----------|---------|-------------|-----------------------|-------------|--|---|-------------|---|
| | | SNP-arrays | Start-end | Size (kb) | Inheritance | Diagnosis | Sex Origi | 4 11 | valking | Langage | Epilepsy | examination | head circumference (SD) | Dysmorphic features | examination | Other |
| 00 | 182 | Tripli 15q11q12 | 21 234 149- 25 713 937 | 4480 | De novo (pat.allele) | ASD + ID (moderate) | M Asia | | 27m | Yes (delay) | No | 18y6m | 155 (-3)/NA/55 (-0.5) | Bulbous nose, nar- rowed palpebral | z | Polyembolokoilomania, delayed puberty |
| 00 | 378 [| Del 9p24 | 317782- 4084321 | 3767 | Unknown, absent in mother | ASD + ID (mod- erate to severe) | F Euro | e | 37 | Few words | Yes, West syndrome | 15y9m | 167 (+1)/82 (+5)/56 () | M) Synophris, upper lip anteverted, short phil- rum, large mouth, | z | Hyperventilation, dental agenesis, self- injurious behavior |
| 4 | 33 [| Del3q29 | 197 222 803- | 1594 | De novo | ASD without ID | M Euro | pe/Asia 1 | .em | Yes (delay) | No | 4y8m | 105 (M)/16.5 (M)/51 (M) | No | z | anu aggression |
| T | 0785 [| Oup 16p11.2 | 29541713- | 694 | Paternal | ASD + ID (mod- | M South | , , | <2y | Yes (delay) | Yes, | 19y7m | 170 (-0.5)/70 (+1)/58 | No | Hands | Gastroesophageal |
| 6 | 388 [| Dup 16p11.2 | 29 554 843- | 552 | De novo | ASD without ID | M Euro | pe 1 | 5 | Yes (delay) | Yes, | 10y2m | 141 (+1)/40 (+3)/54.5 | Epicanthus | N | VIIII |
| ÷. | 0017 [| Dup 16p11.2 | 30 107 306 29 554 843- 30 235 818 | 681 | Paternal | ASD+ID (mild) | M Euro | [ad | 1 | Normal | temporal No | 16y3m | (+1) 172 (+0.5)/58 (-0.5)/5 (+1.5) | 7 Enophtalmia, ante- verted nostrils, bulbous | z | OCD |
| - | 1385 [| Dup 16p11.2 | 29541713- 30107306 | 566 | Unknown | ASD + ID (moderate) | M Asia | , | <2.5y | Yes (delay) | No | 7y6m | 121.5 (+1)/19.2 (-2)/ 49.2 (-2) | No | z | |
| 1 | 374 1 | Del2q36 (DOCK10) | 225186974- 227064219 | 1877 | Mother | ASD + ID (moderate) | M Sub- Africi | Saharan J | .9m | Yes | Yes, tem- poral, | 14y10m | 167 (M)/47 (-1)/55 (M) | Prognatism, short phil- trum, small mouth, | z | Marfanoid |
| 1 | 39 | Hmz del1p31.1 PTGER3) | 71 187 451- 71 230 750 | 43 | Htz del in both parents | ASD + ID (moderate) | F Com | oros | mg. | Few words | No | 5y | 117 (+2.5)/21.5 (+2)/ (+2) | everteu tower rip | z | |

European Journal of Human Genetics



European Journal of Human Genetics

The duplication of the pericentromeric 16p11.2 region, found in four subjects out of 194 (2%), was the most recurrent CNV (Supplementary Figure S4). Parental DNA was available for three patients: the duplication was inherited from the father in two patients (families 845 and 1122) and occurred de novo in one (family 885). Seven CNVs common to at least two patients or five different CNVs altering a common gene were found in 25 additional cases. These CNVs included a duplication of APOO on chromosome X in two patients, a deletion of C2ORF63 in two patients, a deletion and a duplication of CTNNA3 in one patient each, which were all absent from the DGV, and a duplication encompassing the DNAH5 and TRIO genes on chromosome 5, found in two patients. Furthermore, rare CNVs, such or as maternally inherited deletions on chromosome X that alter DMD, ASMT and DDX53 (upstream of PTCHD1), or CNVs encompassing PARK2, RB1CC1, ICA1 and NXPH1 were previously reported in ASD.18,20

Twenty-nine patients had more than one rare, potentially pathogenic CNV. Two of these individuals, in which inheritance could be assessed, had inherited a CNV from each parent absent from the DGV, suggesting that the combination of the CNVs could be pathogenic. Interestingly, multiple rare CNVs were observed in 28

Table 2 Comparison of the number of CNVs identified in patients with ASD + ID and HF-ASD

| | ASD+ID | HF-ASD | |
|--------------------------------------|-------------------------|----------|--------------------|
| | $(N = 160 + 4^{a})$ | (N=33) | P-value |
| Deletions | $41 + 2^{a}$ | 6 | |
| Gains of copies | $73 + 2^{a}$ | 10 | |
| Total CNVs | $114 + 4^{a}$ | 16 | 0.3 ^b |
| Patients with one deletion | 19+2 ^a (13%) | 5 (15%) | 0.53 ^c |
| Patients with one gain of copies | 32+2 ^a (21%) | 9 (27%) | 0.34 ^c |
| Patients with multiple CNVs | 28 (17%) | 1 (3%) | 0.025 ^c |
| del/dup | 13 | 1 | |
| 2 del | 4 | 0 | |
| ≥2 dup | 11 | 0 | |
| Total patients with CNVs | 79+4 ^a (51%) | 15 (45%) | 0.36 ^c |
| Total patients with novel CNVs | 60+4 ^a (39%) | 7 (21%) | 0.037 ^c |
| Patients with novel multiple CNVs | 18 (11%) | 0 (0%) | 0.031 ^c |

Abbreviations: ASD, autism spectrum disorder; ASD + ID, low-functioning ASDs; CNV, copy number variant; HF, high-functioning; HF-ASD, high-functioning ASDs; ID, intellectual

number variant; HF, high-tunctioning; HF-ASU, Ingrisubasining, 1999, 199 deletion, one with 17p11.2 deletion and two with a supernumerary chromosome 15) previously detected by FISH analysis and karyotyping.

^bMann–Whitney test. ^cUnilateral exact Fisher's test.

CNV analysis in ASD C Nava et al

individuals with ASD and ID (17%), but only one individual with Asperger syndrome out of 33 (3%, P = 0.025; Table 2). This result suggests that combinations of CNVs are pathogenic because of additive or epistatic effects and that a burden of rare CNVs more frequently leads to low-functioning autism.

Hemizygous deletions inherited from asymptomatic parents could unmask autosomal recessive variants located on the other allele.21 The coding regions of three patients with deletions encompassing one or more candidate genes, selected on the basis of their expression in the brain and their function, were further analyzed. No pathogenic variants were detected by sequencing of SMARCA2 and DOCK8 on the non-deleted allele of the girl with the 9p24 deletion. Similarly, sequencing of ATP2C2, a gene in which variants have been reported to modulate linguistic abilities,²² detected non-pathogenic variants in an adopted female with Asperger syndrome, who had a 86-kb deletion on chromosome 16 containing only ATP2C2 and KIAA1609 (family 725). In contrast, sequencing of DOCK10 and NYAP2 in a male patient with a deletion of 1.9 Mb on chromosome 2q inherited from his asymptomatic mother, which contained only these two genes (family 625), identified a rare missense variant at the hemizygous state in DOCK10 (c.6460G>A/p.Asp2154Asn, rs111356042, minor allele frequency = 0.0004 in the European population and 0.0369 in the African-American population). It was predicted by Polyphen-2 to be possibly damaging. The patient, who was the only affected child of the couple, had both the deletion and the missense variant, unlike his siblings, suggesting that the association of the two variants could lead to an autosomal recessive disorder (Figure 3).

Although our analysis first focused on CNVs containing coding regions, we also detected six rare CNVs, in six patients, who were either intronic or intergenic but located within or close to genes encoding contactins or neuroligins: an intronic duplication in NLGN1 was detected in three patients, a duplication between CNTN4 and CNTN6 was present in three patients and intronic deletions in CNTN4 and CNTN5 were identified in two unrelated individuals. Furthermore, five intronic CNVs absent from databases and encompassing genes involved in neurotransmission or development of the nervous system (AGBL4, LRRTM4, UNC5C, CADM1, GPHN) were identified in five individuals (Supplementary Table S2). These variants could contribute to ASDs by altering regulatory regions of the genes, although this hypothesis is difficult to prove in the absence of expression data.

DISCUSSION

Many studies have recently demonstrated the important contribution of rare CNVs to the genetic puzzle of ASDs. However, the clinical significance of CNVs in individuals with ASD has been more rarely investigated.^{23,24} In this study, we report the results of a diagnostic

4

Figure 2 Identification of three pathogenic CNVs: 15g11g12 triplication (family 772, patient 8082), 9p24 deletion (family 15, patient 8378) and 3g29 deletion (family 1315, patient 433). (a) SNP array profiles of the patient with the triplication (8082): the Y axis indicates the log R (above) and the B allele frequency (below); the X axis indicates the position on chromosome 15. The red line (log R ratio profile) corresponds to the median smoothing series (Genomestudio). (b) Confirmation of the triplication by FISH analysis with a probe specific to UBE3A (in red) on peripheral blood cell metaphases and nucleus. The arrows point the triplication (c) Details of the genes included in the triplication. (d) Microsatellite (D15S822) analysis showing two alleles inherited from the father in the proband. (e) Methylation indexes of genes included in the triplicated region showing abnormal methylation of UBE3A and SNRPN in the patient 8082 and comparison with a patient with Angelman syndrome. (f) SNP array profiles of the patient with the 9p24 deletion: the Y axis indicates the log R (above) and the B allele frequency (below); the X axis indicates the position on chromosome 9. (g) Confirmation of the deletion (shown by the arrow) by FISH analysis with a probe specific to the 9p subtelomeric region (green) on peripheral blood cell metaphases. (h) Details of the genes included in the deletion. (i) SNP array profiles of the patient with the 3q29 deletion: the Y axis indicates the log R (above) and the B allele frequency (below); the X axis indicates the position on chromosome 3. (j) Confirmation of the deletion (shown by the arrow) by FISH analysis with a probe specific to the 3q29 region (RP1-196F4, green) on peripheral blood cell metaphases and nuclei. (k) Details of the genes included in the deletion.

European Journal of Human Genetics

npg 5



Figure 3 Identification of a deletion encompassing *DOCK10* associated with the c.6460G>A/p.Asp2154Asn in a male patient. (a) SNP array profiles of the patients with the deletion and details of the genes included the deletion: the Y axis indicates the log R (above) and the B allele frequency (below); the X axis indicates the position on chromosome 2. (b) Pedigree of family 625 and segregation analysis of the deletion and the c.6460G>A/p.Asp2154Asn variant. (c) Sequence electropherograms showing the c.6460G>A/p.Asp2154Asn variant in the hemizygous state in his father. (d) Alignment of the region flanking c.6460G>A/p.Asp2154Asn in orthologous proteins showing the conservation of the altered amino acids.

analysis using Illumina SNP microarrays in 194 subjects with ASDs who had been recruited prospectively. The cohort initially included 200 patients, but six individuals were excluded from the CNV analysis because a diagnostic was established by preliminary screenings. CNVs encompassing at least one exon and rare in the DGV were found in 94 subjects with ASD (48%). Yet, this analysis allowed us to make a probable diagnosis in only 3 of the 194 patients (1.5%). However, if we consider the six patients with Smith–Magenis syndrome (n=2), Williams–Beuren syndrome (n=1), supernumerary chromosome 15 (n=2) and fragile X syndrome (n=1) previously detected by FISH analysis, karyotyping, *RA11* sequencing and southern blotting, and excluded from the microarray analysis, the number of unambiguous diagnoses reaches 4.5% (9/200).

Our results confirm the frequent implication of gains of copies in the 15q11–q12 region in ASD, found in 3 out of 200 patients (1.5%). The male subject with the 15q11–q12 triplication from paternal origin had a very mild phenotype, compared to other patients with triplications of this region, including relatively good preservation of language and acquisition of reading and writing. On the contrary, the two patients with supernumerary chromosome 15 (detected before the microarray analysis) had a maternal origin and had severe ID and epilepsy in addition to autism. Duplications and triplications of this

European Journal of Human Genetics

imprinted region in ASD individuals are usually of maternal origin. However, duplications of paternal origin have occasionally been reported.^{13,25} Our findings indicate that 15q11–q12 duplications and triplications of paternal origin might be less penetrant or associated with a milder phenotype than those of maternal origin. This suggests that dysregulation of both imprinted genes (including *UBE3A*) and non-imprinted genes in the 15q11–q13 region contribute to the development of ASDs. Interestingly, both imprinted (*UBE3A* and *SRNPN*) and non-imprinted genes (*OCA2* and *GABRG3*) located in the triplicated region had abnormal methylation values in the patient with the 15q11–q12 triplication. Epigenetic dysregulations affecting the genes encoding GABA_A receptors in the brains of individuals with 15q11–q13 copy number abnormalities have been reported previously.²⁶

The second structural anomaly that was considered as probably causative is a deletion of the 9p24 region. Deletions encompassing this region, which are usually *de novo* but larger in size (Supplementary Figure S3), are known to cause the 9p deletion syndrome, a disorder characterized by ID, trigonocephaly, facial dysmorphism and sexual reversion in males. Interestingly, autism was reported in 4 out of 100 patients with this syndrome.²⁷ suggesting that this region contains genes contributing to autism in certain conditions. Moreover,

deletions on chromosome 9p24.3 encompassing C9orf66, CBWD1, DOCK8 and FOXD4 were identified recently in two individuals with ASD (frequency estimated to 0.05%).5 The deletion found in our female patient encompassed 13 genes, including DOCK8 and SMARCA2. It is likely that this deletion has occurred de novo in the proband, who had severe language impairment and West syndrome in addition to autism, as were most deletions encompassing this region in the literature, although this could not be formally demonstrated due to the unavailability of the DNA of the father. SMARCA2 encodes a protein that belongs to the FMRP complex, deficient in fragile X syndrome and is involved in a large network of interactions with other ASD related proteins such as TSC1 and NLGN3.28,29 To test the hypothesis that additional variants in SMARCA2 and/or DOCK8 contribute to autism, we sequenced the exons of these two genes in the patient with the 9p deletion, but no rare, possibly pathogenic variants were identified. Recently, heterozygous nonsynonymous mutations or partial deletions in SMARCA2 were identified as the cause of Nicolaides–Baraitser syndrome, an autosomal dominant condition with syndromic $\mathrm{ID}.^{30}$

The third probably causal abnormality was a *de novo* deletion on chromosome 3q29. The 3q29 microdeletion syndrome has been recently described as a new syndrome, probably caused by nonallelic homologous recombination.³¹ It was usually associated with dysmorphism, mild-to-moderate ID, variable congenital malformations and ASD in 27% of cases.³² This CNV was also recently found in schizophrenic individuals.³³ Our patient with this deletion had not the classical dysmorphism and had a normal intelligence. Only two patients with normal intelligence and 3q29 deletion were previously described.^{34,35} Among the 22 genes included in the 1.6 Mb deletion, several genes including *FBXO45*, *DLG1*, *BDH1* or *PAK2* (involved in brain development or synaptic transmission) and *NCBP2* (involved in RNA processes such as splicing, translation regulation, nonsense-mediated mRNA decay, RNA-mediated gene silencing by microRNAs and mRNA export) represent candidate genes possibly contributing to ASD.

Unlike these three microrearrangements, the contribution of the identified CNVs to ASD remained unclear in most cases, such as the 16p11.2 duplications identified in four individuals. Deletions and duplications involving a pericentromeric region of chromosome 16 that spans ~600 kb and contains 28–35 genes are recurrent and reciprocal because of highly homologous segmental duplications. Although the duplication was repeatedly associated with autism and schizophrenia, 5^{-36-38} it is characterized by great phenotypic variability, ranging from normal phenotypes to severe congenital malformations, and does not segregate perfectly with autism in multiplex families. These observations suggest that the 16p11.2 duplication is associated with autism with low penetrance or that other unidentified factors interact with the duplication to determine its phenotypic expression.

The CNV was inherited from asymptomatic parents in 96% of the cases in which inheritance could be assessed. Inherited CNVs possibly have a role in the etiology of ASDs in combination with other genetic factors in the deletion/duplication or elsewhere in the genome. To identify recessive mutations located on the undeleted allele, we selected three patients with rare deletions encompassing a small number of candidate genes selected on the basis of their function and expression in brain. This approach allowed us to detect a rare, possibly deleterious missense variant in *DOCK10 in trans* with a whole gene deletion. This variant and the deletion were associated in the index case but not his healthy brother and sister, which is compatible with autosomal recessive inheritance. CNVs involving *DOCK10* are extremely rare; they were not found in the DGV or in 8329 controls, and were present in only 2 out of 15767 patients.³⁸

CNV analysis in ASD C Nava et al

CNVs encompassing other members of the DOCK (Dedicator of CytoKinesis) family, such as DOCK4 and DOCK8 (located in the 9p24 deletion), have been reported in ASDs.^{5,39} However, additional cases of ASD with mutations in this gene need to be identified to confirm that autosomal recessive mutations in DOCK10 can cause ASD.

Another possible case of autosomal recessive ASD was a female with a homozygous deletion on chromosome 1p31.1, encompassing alternatively spliced exons of the PTGER3 gene. This homozygous deletion spanned 43 kb and was located in a larger (15 Mb) homozygous region. PTGER3 encodes EP3, one of four receptors for prostaglandin E2, which has many biological functions, including inhibition of gastric acid secretion, inhibition of sodium and water reabsorption in kidney, uterine contractions, fever in response to exogenous and endogenous stimuli, and modulation of neurotransmitter release in central and peripheral neurons.⁴⁰ Mice lacking Ep3 have a mild, complex phenotype, including an impaired febrile response,⁴¹ an impaired response of adrenocorticotropic hormone to bacterial endotoxin,⁴² exaggerated allergic inflammation,⁴³ obesity, increased motor activity⁴⁴ and increased survival to bacterial infection.45 PTGER3 is expressed as at least nine multiple splicing variants that have identical ligand binding properties but interact with different second messengers.46 The homozygous PTGER3 deletion identified in our patient alters six of the nine isoforms (Supplementary Figure S5). This patient, who was 5 years old when last examined, had typical autism with a mild delay. She was very agitated, tall and not obese; her history of infection and fever was unremarkable.

Altogether, our results show that genetic syndromes associated with genomic deletions or duplications are a rare cause of ASDs. The search for CNVs in individuals with ASD should be pursued, however, as it can lead to a diagnosis and appropriate genetic counseling in a small number of cases. Association of CNV analysis with exome sequencing to systematically search for autosomal recessive causes of ASD could increase the proportion of positive diagnoses. However, oligogenic or multihit models, in which the association of rare inherited CNV with point mutations contributes to the disorder, are probable in many cases of autism. Two of our subjects support this hypothesis; each had two novel, private CNVs inherited from asymptomatic parents, suggesting that one of the CNVs was not sufficient to cause ASD but that a combination of the two rearrangements was deleterious. The observation of multiple rare CNVs in 17% of the subjects with ASD and ID, but only one with Asperger syndrome (3%), also suggests that the effects of CNVs are additive or that some combinations are pathogenic. Two recent studies further support the theory of the 'two-hit model' in individuals with ASDs.^{19,47} The next challenge in ASDs is therefore to understand the interactions between rare CNVs and other rare variants in individual patients and to take into account environmental interactions that may also modulate the risk for developing ASD.48

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the P3S platform, the 'Plateau technique Mutualisé du GHU Est', and the genotyping and sequencing platform of the ICM for technical assistance and the DNA and cell bank of CRICM for DNA extraction and cell culture. We also thank Dr Merle Ruberg for critical reading of the manuscript and Sophie Rivaud-Pechoux for her help with the statistical analysis. This study was financially supported by AP-HP (DHOS), Fondation de France, ERA-NET NEURON EUHFAUTISM and INSERM.

European Journal of Human Genetics

CNV analysis in ASD C Nava et al

- 1 Miles JH: Autism spectrum disorders a genetics review. Genet Med 2011; 13:

npg

8

- Miles JFI: Autom spectrum and the second seco
- 2009; 5: e1000536.
 5 Sanders SJ, Ercan-Sencicek AG, Hus V et al: Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 2011; **70**: 863–885.
 6 O'Roak BJ, Vives L, Girrajan S et al: Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012; **485**: 246–250.
 7 Sanders SJ, Murtha MT, Gupta AR et al: De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012; **485**: 237–241.

- sequencing are strongly associated with aurism. *Nature* 2012; 489: 237–241.
 8 Neale BM, Kou Y, Liu L *et al*. Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* 2012; 485: 242–245.
 9 Szatmari P, Paterson AD, Zwaigenbaum L *et al*: Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 2007; 39: 319–328.
 10 Schaaf CP, Zoghbi HY: Solving the autism puzzle a few pieces at a time. *Neuron* 2011; 270,0000 cm²
- 70: 806-808 11 Levy D. Ronemus M, Yamrom B et al: Rare de novo and transmitted copy-number
- Levy D, Ronemus M, Yamrom B *et al*: Rare *de novo* and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 2011; **70**: 886–897.
 Guilmatre A, Dubourg C, Mosca AL *et al*: Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry* 2009; **66**: 947–956.
 Depienne C, Moreno-De-Luca D, Heron D *et al*: Screening for genomic rearrangements
- and methylation abnormalities of the 15q11-q13 region in autism spectrum disorders. Biol Psychiatry 2009: 66: 349-359.
- Biol Fsycinatity 2009; bc: 349–359.
 Bi Gal X, Sub HM, Perin JC et al: Rare structural variation of synapse and neurotransmission genes in autism. *Mol Psychiatry* 2012; 17: 402–411.
 Jamain S, Quach H, Betancur C et al: Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 2003; 34:
- 27-29 16 Durand CM, Betancur C, Boeckers TM et al: Mutations in the gene encoding the
- Suraptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* 2007; **39**: 25–27. Leblond CS, Heinrich J, Delorme R *et al*: Genetic and functional analyses of SHANK2
- mutations suggest a multiple hit model of autism spectrum disorders. PLoS Genet 2012: 8: e1002521.
- 18 Betancur C: Etiological heterogeneity in autism spectrum disorders: more than 100

- Betancur C: Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* 2011; **1380**: 42–77.
 Schaaf CP, Sabo A, Sakai Y *et al*: Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum Mol Genet* 2011; **20**: 3366–3375.
 Basu SN, Kollu R, Banerjee-Basu S: AutDB: a gene reference resource for autism research. *Nucleic Acids Res* 2009; **37**: 0832–0836.
 Hochstenbach R, Poot M, Nijman J *et al*: Discovery of variants unmasked by hemizygous deletions. *Eur J Hum Genet* 2012; **20**: 748–753.
 Newbury DF, Winchester L, Addis L *et al*: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet* 2009; **85**: 264–272.
 Salyakina D, Cukier HN, Lee JM *et al*: Copy number variants in extended autism spectrum disorder families reveal candidates potentially involved in autism risk. *PLoS* spectrum disorder families reveal candidates potentially involved in autism risk. PLoS
- Van Daalen E, Kemmer C, Verbeek NE et al: Social Responsiveness Scale-aided analysis of the clinical impact of copy number variations in autism. *Neurogenetics* 2011; 12: 315–323. One 2011; 6: e26049. 24 van Da
- 25 Veltman MW, Thompson RJ, Craig EE et al: A paternally inherited duplication in the Prader-Willi/Angelman syndrome critical region: a case and family study. J Autism Dev Disord 2005: 35: 117-127.

- 26 Hogart A, Nagarajan RP, Patzel KA, Yasui DH, Lasalle JM: 15q11-13 GABAA receptor genes are normally biallelically expressed in brain yet are subject to epigenetic dysregulation in autism-spectrum disorders. Hum Mol Genet 2007; 16: 691-703
- Vinci G, Chantot-Bastaraud S, El Houate B, Lortat-Jacob S, Brauner R, McElreavey K: Association of deletion 9p, 46,XY gonadal dysgenesis and autistic spectrum disorder. *Mol Hum Reprod* 2007; **13**: 685–689. 27 28 Sakai Y, Shaw CA, Dawson BC et al: Protein interactome reveals converging molecular
- Sakai Y, Shaw CA, Dawson BC *et al*: Protein interactome reveals converging molecular pathways among autism disorders. *SCI Transl Med* 2011; **3**: 86ra49.
 Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D: Rare *de novo* variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 2011; **70**: 898–907.
 Van Houdt JK, Nowakowska BA, Sousa SB *et al*: Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat Genet* 2012; **44**: 445–449,
- S441.

- S441.
 Willatt L, Cox J, Barber J *et al*: 3q29 Microdeletion syndrome: clinical and molecular characterization of a new syndrome. *Am J Hum Genet* 2005; **77**: 154–160.
 Quintero-Rivera F, Sharifi-Hannauer P, Martinez-Agosto JA. Autistic and psychiatric findings associated with the 3q29 microdeletion syndrome: case report and review. *Am J Med Genet A* 2010; **152**A: 2459–2467.
 Mulle JG, Dodd AF, McGrath JA *et al*: Microdeletions of 3q29 confer high risk for schizophenia. *Am J Hum Genet* 2010; **87**: 229–236.
 Cobb W, Anderson A, Turner C, Hoffman RD, Schonberg S, Levin SW: 1.3 Mb *de novo* deletion in chromosome band 3q29 associated with normal intelligence in a child. *Eur J Med Genet* 2010; **53**: 415–418.
 Li F, Lisi EC, Wohler ES, Hamosh A, Batista DA: 3q29 interstitial microdeletion syndrome: an inherited case associated with cardiac defect and normal cognition.
- me: an inherited case associated with cardiac defect and normal cognition.
- Syndroline: an interact case associated with carried detect and normal cogmutant. Eur J Med Genet 2009; 52: 349–352.
 Weiss LA, Shen Y, Korn JM et al: Association between microdeletion and microduplication at 16p11.2 and autism. N. Engl J Med 2008; 358: 667–675.
 Sahoo T, Theisen A, Rosenfeld JA et al: Copy number variants of schizophrenia 36 W
- 37 Sahoo T, susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems, Genet Med 2011: 13: 868-880.
- and behavior problems. Genet Med 2011; 13: 868–880.
 38 Cooper GM, Coe BP, Girriajan S et al: A copy number variation morbidity map of developmental delay. Nat Genet 2011; 43: 838–846.
 39 Maestrini E, Pagnamenta AT, Lamb JA et al: High-density SNP association study and copy number variation analysis of the AUTS1 and AUTS5 loc limplicate the IMMP2L-DOCK4 gene region in autism susceptibility. Mol Psychiatry 2010; 15: 954–968.
 40 Momiyama T, Todo N, Sugimoto Y, Ichikawa A, Narumiya S: Membrane depolarization the certification of more transformed in Francement FPB subtract of durating restongering neurons.
- Momiyama T, Todo N, Sugimoto Y, Ichikawa A, Narumiya S: Membrane depolarization by activation of prostaglandin E receptor EP3 subtype of putative serotonergic neurons in the dorsal raphe nucleus of the rat. Nauryn Schmiedebergs Arch Pharmacol 1996; 353: 377–381.
 Ushikubi F, Segi E, Sugimoto Y et al: Impaired febrile response in mice lacking the prostaglandin E receptor subtype EP3. Nature 1998; 395: 281–284.
 Matsuoka Y, Furuyashiki T, Bito H et al: Impaired adrenocorticotropic hormone response to bacterial endotoxin in mice deficient in prostaglandin E receptor EP1 and EP3 subtypes. Proc Natl Acad Sci USA 2003; 100: 4132–4137.
 Kunikata T, Yamane H, Segi E et al: Suppression of allergic inflammation by the prostaglandin E receptor subtype EP3. Natl memunol 2005; 6: 524–531.
 Sanchez-Alavez M, Klein I, Brownell SE et al: Night eating and obesity in the EP3R-deficient mouse. Proc Natl Acad Sci USA 2007; 104: 3003–3014.
 Kornov M, Lewis C, Serezari CH et al: E-prostancid 3 receptor deletion improves

- 11613-11617 47 Gau SS, Liao HM, Hong CC, Chien WH, Chen CH: Identification of two inherited copy
- with autism supports two-hit and compound heterozygosity number variants in a m
- number variants in a male with autism supports two-nit and compound neterozygosity models of autism. Am J Med Genet B 2012; 1598; 710–717.
 48 Guinchat V, Thorsen P, Laurent C, Cans C, Bodeau N, Cohen D: Pre-, peri- and neonatal risk factors for autism. Acta Obstet Gynecol Scand 2012; 91: 287–300.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)

European Journal of Human Genetics

Keren Boris – Thèse de doctorat - Année 2013

Résumé (français) :

L'analyse chromosomique sur puce ADN (ACPA) tend à devenir le principal examen diagnostique dans la déficience intellectuelle (DI). Parmi les techniques d'ACPA, les puces SNP ont l'intérêt de pouvoir détecter les pertes d'hétérozygotie, et par conséquent d'identifier les isodisomies uniparentales (iUPD) et les zones d'identité liées à la consanguinité. Nous avons étudié une cohorte de 1187 patients atteints de DI, dans un cadre diagnostique, sur puces SNP. Nous avons réalisé, par cette étude, 145 diagnostics (12%) dont 2 iUPD et 6 délétions n'incluant qu'un seul. De plus, nous avons détecté 639 CNV rares non décrits chez des sujets contrôles et incluant des séquences codantes, ce qui nous a permis d'identifier 11 gènes candidats dans la DI : CAMTA1, SP3, CNTNAP4, NUDT12, STXBP6, DOCK8, DOCK10, SMARCA2, NYAP2, ATAD3A et ATAD3B. Nous avons tenté de valider l'implication de ces gènes par séquençage, mais n'avons trouvé de seconde mutation pour aucun d'entre eux. Toutefois, des réarrangements de CAMTA1 ont été également retrouvés dans 2 autres familles avec un phénotype homogène (DI et ataxie congénitale) ce qui nous fait considérer avec certitude qu'il s'agit d'un gène de DI. Par ailleurs, l'homozygosity mapping, réalisé en puces SNP, a permis d'identifier, par séquençage whole exome, une mutation non-sens homozygote du gène BUD13 dans une famille de DI syndromique. Enfin, de façon fortuite, nous avons caractérisé en ACPA une translocation familiale entraînant une disruption d'un gène d'ataxie spino-cérébelleuse, ATXN10, ce qui a permis de mieux comprendre la physiopathologie de cette maladie. Au total, notre étude démontre l'intérêt des puces SNP dans la DI, d'une part en diagnostic et d'autre part pour l'identification de gènes responsables de DI.

Title: Intellectual disability: from microarray diagnosis to the identification of candidate genes

Abstract:

Chromosomal Microarray Analysis (CMA) has become the main diagnostic test in the field of intellectual disability (ID). Among CMA techniques, SNP arrays have the advantage of identifying losses of heterozygosity in addition to Copy Number Variants (CNVs). Therefore they can detect uniparental isodisomies (iUPD) and regions of identity by descent. We screened a cohort of 1,187 patients with ID, in diagnostic setting, by CMA using SNP arrays. Causal abnormalities, including 2 iUPDs and 6 deletions comprising only one gene, were detected in 145 patients (12%). Moreover we found 639 rare CNVs, absent from control individuals, which included coding sequences. Our results allowed us to identify 11 genes possibly involved in or contributing to ID: CAMTA1, SP3, CNTNAP4, NUDT12, STXBP6, DOCK8, DOCK10, SMARCA2, NYAP2, ATAD3A and ATAD3B. We then screened additional patients with similar phenotypes in order to find a second mutation, but no second mutation was identified in any of them. Besides, CAMTA1 rearrangements were also found among two other families with homogeneous phenotypes (ID and congenital ataxia), which confirms that this gene is involved in ID. Furthermore, thanks to homozygosity mapping made possible by SNP arrays combined with exome sequencing, we identified a homozygous nonsense mutation in the BUD13 gene, in a family with syndromic ID. Finally we incidentally characterized a familial translocation resulting in the disruption of ATXN10, a gene responsible for spinocerebellar ataxia. This translocation has helped to better understand the pathophysiology of the disease. Overall, our study shows the importance of SNP arrays for the molecular diagnosis of ID and as a tool to identify genes responsible for ID.

Mots clés (français) : déficience intellectuelle, puces SNP, séquençage de l'ADN.

Keywords: intellectual disability, SNP arrays, DNA sequencing.