



**HAL**  
open science

# Démembrement génétique des déficiences intellectuelles et compréhension des bases physiopathologiques associées, à l'ère du séquençage à haut débit

Maéva Langouët Langouët

► **To cite this version:**

Maéva Langouët Langouët. Démembrement génétique des déficiences intellectuelles et compréhension des bases physiopathologiques associées, à l'ère du séquençage à haut débit. Génétique. Université René Descartes - Paris V, 2014. Français. NNT : 2014PA05T045 . tel-01127049

**HAL Id: tel-01127049**

**<https://theses.hal.science/tel-01127049>**

Submitted on 6 Mar 2015

**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.

# Mémoire de Thèse

en vue de l'obtention du titre de

Docteur en Sciences de la Vie et de la Santé de l'université Paris Descartes

École doctorale : GC2ID/BioSPC

Spécialité : Génétique moléculaire des maladies du développement

présentée par

Maéva Langouët

Démembrement génétique des déficiences intellectuelles et  
compréhension des bases physiopathologiques associées, à l'ère du  
séquençage à haut débit

soutenue le 03 Décembre 2014  
devant le jury composé de

Professeur Valérie CORMIER-DAIRE  
Professeur Nicolas LEVY  
Docteur Janet HALL  
Docteur Fiona FRANCIS  
Docteur Gérard PIERRON  
Docteur Laurence COLLEAUX

Président  
Rapporteur  
Rapporteur  
Examinateur  
Examinateur  
Directeur de thèse



*« Ils ne savaient pas que c'était impossible, alors ils l'ont fait. »*

Mark Twain



## REMERCIEMENTS

### **Au Dr Laurence Colleaux**

J'adresse mes premiers et mes plus grands remerciements à ma directrice de thèse. Il me semble que la réussite d'une thèse ne dépend pas uniquement de l'étudiant ou uniquement du directeur, mais plutôt du binôme directeur-étudiant. Je pense avoir pris une excellente décision il y a trois ans (et trois jours) lorsque j'ai choisi de faire mon doctorat avec toi. Tu as parfaitement su m'encadrer et me faire avancer sur mes projets de thèse et au-delà et j'ai énormément apprécié travailler avec toi. Je te remercie, Laurence, pour ta rigueur et tes compétences mais aussi ta générosité et ta confiance.

### **Au Pr Arnold Munnich**

Je vous remercie de m'avoir accueillie dans votre Unité de Recherche, l'ex-U781 où j'ai effectué mon Master 2 et la plus grande partie de mon doctorat. Je vous remercie pour vos conseils et nos discussions enrichissantes. Je tiens à vous exprimer toute ma gratitude et mon profond respect.

### **Au Pr Alain Fischer et au Pr Arnold Munnich**

Je vous remercie de m'accueillir au sein de l'institut *Imagine* depuis Janvier 2014. Soyez assurés de ma reconnaissance.

### **Au Pr Nicolas Lévy et au Dr Janet Hall**

Je vous remercie d'avoir accepté d'être rapporteurs de ce travail. Soyez assurés de ma reconnaissance.

### **Au Dr Fiona Francis et au Dr Gérard Pierron**

Je vous remercie d'avoir accepté de participer à mon jury de thèse. Soyez assurés de ma reconnaissance.

### **Au Pr Valérie Cormier-Daire**

Vous me faites l'honneur de présider mon jury de thèse. Je vous remercie pour vos conseils et votre bienveillance.

### **Aux collaborateurs**

Je remercie les équipes de Dr Saadi, Dr Chaouch, Dr Stern, Pr Stoppa-Lyonnet, Pr O'Driscoll, Pr Brown, Pr Tisdale, Pr Barouki, Pr Tauber et Pr Mancini. Je tiens à remercier également Marie, Guillaume et Dennis avec qui j'ai apprécié travailler.

### **Au Dr Sylvain Hanein**

Je te remercie pour tes très nombreux conseils, tes relectures et précieux commentaires, ta générosité et tout simplement ton amitié. Tu m'as offert l'opportunité d'avoir un second mentor tout au long de ma thèse et ce de façon complètement désintéressée. Je mesure la dimension de cette faveur que tu m'as faite et l'incidence qu'elle a eue sur ma progression. Je ne te remercierai jamais assez !

### **Aux Dr Valérie Malan, Dr Jeanne Amiel et Dr Christine Bole-Feysot**

Je vous remercie pour vos précieux conseils, vos relectures et votre soutien.

### **Au Dr Vincent Cantagrel**

Je tiens à te remercier pour tes relectures et tes précieux conseils délivrés au cours de nos longues discussions concernant mon avenir !

### **Aux Dr Mathilde Doyard, Dr Anastella Pelet et Dr Ludovic Martin**

Je vous remercie pour vos relectures, vos conseils et votre amitié. Vous m'avez prêté une oreille attentive tout l'été et elle a été d'un réconfort certain.

### **Aux Dr Catherine Bui et Dr Lam-Son Nguyen**

Je vous remercie pour vos conseils pratiques au quotidien, votre amitié et toutes nos discussions autour du sujet passionnant, quoique légèrement (c'est un euphémisme !) anxigène, du choix du post-doc !

### **Au Dr Yamina Hamel**

Je te remercie d'avoir partagé ton expérience avec moi, de m'écouter et de me conseiller. Tu as cette incroyable capacité à toujours garder un regard positif, devant les bonnes et les mauvaises nouvelles. Je te remercie pour ton immense soutien, ta grande générosité et ton amitié.

### **À mon équipe**

Je remercie très chaleureusement les « neuneux » (ou les « IS » !) pour votre soutien et votre amitié. Je tiens à vous communiquer à quel point il est agréable de travailler avec des personnes telles que vous : entières. Vous avez chacun participé à la réussite et l'achèvement de cette thèse, par votre aide au quotidien, nos discussions actives et l'ambiance paisible, sérieuse et légère à la fois, qui règne dans notre groupe. Je remercie Laurence, Nathalie et Valérie. Je remercie aussi Anne, ma collègue de bureau parfaite. Je remercie les anciens, Oriane qui m'a formée pendant mon Master 2, encouragée et conseillée au tout début de ma thèse et Sébastien avec qui nous avons passé une excellente année 2012 et avec qui j'ai beaucoup apprécié travailler. En 2013, le groupe s'est agrandi et bonifié avec l'arrivée de Lam-Son plus connu sous le nom de Sonny (ou \*\*\*\*\* !) et Vincent, que j'aime appeler Mr Cell (et récemment Mr Premier de sa commission). Ensuite nous avons eu la chance de « dérober » Michaël, que nous connaissions et apprécions déjà, à l'équipe « amie » voisine. Daniel et Florine nous ont rejoint, accompagnés de leur motivation et de leur bonne humeur et très récemment l'équipe s'est encore agrandie avec l'arrivée de Hisham et Julien ! J'ai gardé pour la fin, les deux personnes qui m'ont accompagnée depuis le début, mes voisines de paillasse au 2<sup>e</sup> étage de la tour Lavoisier. Elles ont été « la base » de cette équipe pendant toute la durée de ma thèse et je tiens tout particulièrement à leur adresser mes remerciements. La brillante Giulia, merci pour ta sensibilité, ton amitié et bien sûr ton célèbre grain de folie et Karine (-Karine), qui m'a beaucoup appris ces dernières années, merci pour tout, ta générosité, ta franchise et ton soutien.

### **À tous les membres de l'ex-U781**

Je vous remercie tous très chaleureusement pour vos conseils et votre aide. J'ai été ravie de vous rencontrer et d'avoir fait ma thèse à vos côtés. Nous voilà répartis sur divers niveaux dans le tout nouveau *Imagine*, ce magnifique institut que nous attendions tous. Cependant, aujourd'hui, je ne regrette absolument pas d'avoir passé la majorité de ma thèse à l'U781. Cette proximité, parfois pénible, était aussi source d'échange et de productivité, poussait aux discussions, générait des associations, et aussi a engendré une incroyable solidarité et une amitié entre ses membres. Je ne mentionne ici aucun nom car je voudrais tous vous citer ; soyez sûrs que je garderai un excellent souvenir de ces années au cours desquelles vous êtes devenus ma « famille scientifique ».

### **À tous les membres de l'institut Imagine**

Je tiens à remercier tout *Imagine*, particulièrement les équipes des plateformes de génomique, de bio-informatique et d'imagerie, les « bisounours » du 3<sup>e</sup> étage et toute l'équipe « doc-postdoc/YR2I », pour votre énergie et votre motivation au quotidien !

### **À mes amis**

Je tiens à remercier toute mon équipe de water-polo du SCNCR avec qui j'ai passé des moments forts depuis de nombreuses années et tout au long de ma thèse, de près ou de loin. Je remercie tous les copains de

Choisy-le roi, tout particulièrement mes amies de longue date qui me suivent, m'écoutent et me soutiennent : Aude, Camille, Coline, Delphine, Dounia, Héléna, Laura, Lucile, Marion et Stéphanie. Je tiens à remercier Héléna pour ton accueil chaleureux cet été pour cette semaine de vacances-rédaction, Marion et Lucile pour m'avoir supportée lors d'une petite crise de nerf inattendue, et Aude, pour ton immense soutien au quotidien, toi qui a suivi de près, spirituellement et géographiquement, toute ma thèse. Celle-ci ne se serait peut-être pas passée aussi bien sans les débriefes chez Serge, notre QG, à mi-chemin entre ton boulot et le mien (deux cafés allongés, comme d'habitude ?).

### **À ma famille**

Je terminerai en exprimant ma reconnaissance à ma famille : mes grands parents, pour leurs relectures et leurs encouragements ; mes parents, toujours là pour moi ; et mon frère et ma sœur que j'aime plus que tout et qui doivent, depuis peu, partager la vedette avec mes deux (et bientôt trois ?) adorables petites nièces, Jena et Cléo. Je remercie également ma grande famille, localisée à divers endroits dans le monde loin d'être des plus déplaisants, pour les vacances comme pour la rédaction. On dit qu'on ne choisit pas sa famille, mais si c'était à faire, je vous choisirais ! Vous retrouver est toujours un grand moment. Je vous remercie tous pour votre accueil chaleureux et votre soutien. Je tiens à remercier particulièrement Hoël et Mathilde, mes deux piliers sur qui je peux toujours compter, pour la thèse et pour la vie. Merci pour votre aide précieuse, votre soutien démesuré et votre générosité sans limite.



La déficience intellectuelle (DI) est définie comme un «fonctionnement intellectuel général inférieur à la moyenne, qui s'accompagne de limitations significatives du fonctionnement adaptatif» (DSM-V). Il s'agit d'un handicap fréquent qui concerne près de 3% de la population générale. L'identification de l'étiologie d'une DI est une question primordiale car elle permet d'optimiser la prise en charge des patients sans risque de passer à côté d'une cause curable, et d'évaluer le risque de récurrence dans la famille afin d'offrir un conseil génétique pour les grossesses à venir.

Malgré les récents progrès, l'étiologie de la maladie reste inconnue dans près de 40% des cas. Le démantèlement des causes génétiques et la compréhension des bases physiopathologiques des DI constituent donc un grand défi scientifique et médical. Par ailleurs, l'identification des gènes impliqués dans les DI et le décryptage des processus cellulaires sous-jacents à ces phénotypes sont une approche de choix pour étudier le développement et la plasticité cérébrale chez l'homme d'une part et entrevoir le développement de nouvelles stratégies thérapeutiques d'autre part.

Le travail de thèse présenté dans ce manuscrit s'inscrit dans cette thématique de recherche et a porté sur l'analyse, par la méthode de *Whole Exome Sequencing* (WES), de cinq familles indépendantes dans lesquelles ségrège une forme syndromique de DI.

La première partie détaille les résultats obtenus pour l'analyse de trois familles consanguines dans lesquelles ségrège une DI autosomique récessive. La seconde partie présente l'étude de deux familles indépendantes dont les enfants atteints présentent une clinique très semblable.

Au total, ce mémoire décrit l'identification de i) deux gènes précédemment associés à la DI (*WDR62* et *AP4M1*), ii) deux gènes candidats (*RAD54B* et *HERC2*), potentiels modificateurs des symptômes observés, puis iii) la définition d'un nouveau mode d'hérédité, et enfin iv) la caractérisation de deux nouveaux gènes impliqués dans la DI (*TTI2* et *NONO*) suivie des études fonctionnelles des effets des mutations sur les cellules de patients et l'analyse d'un modèle murin *Nono<sup>gt</sup>*.

## ABSTRACT

Intellectual deficiency (ID) is characterized by a broad range of deficits in higher brain functions that result in significant limitations in adaptive and cognitive capacities required for competence in daily living, communication, social interaction and integration, self-direction, and work (DSM-V). ID affects approximately 3% of the population. Identifying ID causes is essential to improve patients' care services with no risk to miss a curable cause, but also to provide genetic counselling to the family for future pregnancies. Little is known about the biological bases of these conditions. Indeed, despite recent advances in cytogenetic and molecular genetic, the cause of the mental handicap remains unexplained in 40% of the cases. Understanding the molecular bases of these disorders is therefore an important medical challenge for the next years. Also, ID genes identification and analysis of the cellular mechanisms underlying these conditions should provide significant insight into the molecular and cellular pathways involved in cognition and may lead to new therapeutic trials aiming at improving the daily living of these patients and their families.

The PhD work presented here report on the analysis, using *Whole Exome Sequencing* (WES), of five different families presenting with syndromic ID.

The first part develops results from the analysis of three consanguineous families with an autosomal recessive form of ID. The second part presents the study of 2 unrelated male ID patients who presented the same clinical features.

Overall, this work allowed the identification of i) two genes previously associated with ID (*WDR62* and *AP4M1*), ii) two candidate genes (*RAD54B* and *HERC2*), potential modifiers of the phenotype, then iii) the definition of a novel hereditary mode, and finally iv) the characterisation of two new genes of ID (*TTI2* and *NONO*) followed by the functional analysis of mutations effects in patients' cells and the *Nono<sup>gt</sup>* mouse model.

# TABLE DES MATIÈRES

<b>Remerciements</b>	<b>7</b>
<b>Résumé</b>	<b>8</b>
<b>Abstract</b>	<b>9</b>
<b>Table des matières</b>	<b>13</b>
<b>Table des figures</b>	<b>15</b>
<b>Liste des tableaux</b>	<b>16</b>
<b>Index</b>	<b>17</b>
<b>I Introduction</b>	<b>19</b>
<b>1 Les déficiences intellectuelles (DI)</b>	<b>20</b>
1.1 Définitions des DI, classification et prévalence . . . . .	20
1.1.1 Définition . . . . .	20
1.1.2 Classification . . . . .	21
1.1.3 Prévalence . . . . .	22
1.2 L'étiologie des DI jusqu'à la révolution NGS ( <i>Next-Generation Sequencing</i> ) . . . . .	23
1.2.1 Causes environnementales . . . . .	23
1.2.2 Causes génétiques . . . . .	24
<b>2 L'apport des nouvelles technologies</b>	<b>36</b>
2.1 La révolution du séquençage à très haut débit ou <i>Next-Generation Sequencing</i> (NGS) : un bref historique . . . . .	36
2.2 Les méthodes d'enrichissement spécifique . . . . .	37
2.3 Démembrement génétique des DI par NGS, mise en place d'une approche innovante . . . . .	38

2.3.1	Preuve de concept . . . . .	38
2.3.2	Application aux DI . . . . .	39
<b>3</b>	<b>Mécanismes physiopathologiques des DI, l'état de l'art</b>	<b>41</b>
<b>4</b>	<b>Objectifs des travaux</b>	<b>44</b>
<b>II</b>	<b>Approche Whole Exome Sequencing (WES) pour l'étude de familles consanguines : du monogénique au polygénique</b>	<b>46</b>
<b>5</b>	<b>Identification d'une mutation dans un gène connu de DI -WDR62- et illustration de l'intérêt du WES</b>	<b>47</b>
5.1	La famille, les patients . . . . .	47
5.2	Stratégie . . . . .	52
5.3	Résultats . . . . .	52
5.3.1	Cartographie par autozygotie et recherche de gènes candidats . . . . .	52
5.3.2	Étude de la stabilité de la protéine RAD54B mutée . . . . .	55
5.3.3	WES sur un unique probant . . . . .	57
5.4	Évaluation de la contribution de chaque variant . . . . .	58
5.4.1	Le gène <i>WDR62</i> , données de la littérature . . . . .	58
5.5	Discussion et conclusion préliminaires . . . . .	65
<b>6</b>	<b>Combinaison de deux méthodes conduisant à la caractérisation d'un nouveau gène de DI : TT12</b>	<b>66</b>
6.1	La famille, les patients . . . . .	66
6.2	Stratégie . . . . .	66
6.3	Résultats : publication n°1, [Langouët et al., 2013] . . . . .	66
6.4	Discussion et perspectives . . . . .	80
6.4.1	Les principales caractéristiques cliniques observées dans les anomalies de la réparation de l'ADN . . . . .	80
6.4.2	Études des voies de signalisation des PIKKs . . . . .	82
6.4.3	Perspectives et conclusion préliminaires . . . . .	83
<b>7</b>	<b>Le NGS, un outil pour définir de nouveaux modes de transmission génétique</b>	<b>85</b>
7.1	La famille, les patients . . . . .	85
7.2	Stratégie . . . . .	85
7.3	Résultats : publication n°2 (en préparation) . . . . .	85
7.4	Conclusions et perspectives : étude fonctionnelle du variant c.595A>T du gène <i>AZGP1</i> . . . . .	101
7.5	Définition du concept de syndrome des mutations contiguës . . . . .	102

---

<b>III Approche WES sur des cas sporadiques cliniquement semblables : identification d'un nouveau syndrome de DI et démonstration du rôle de NONO dans la biologie des synapses GABAergiques</b>	<b>103</b>
<b>8 Approche adoptée</b>	<b>104</b>
8.1 Présentation des familles . . . . .	104
8.2 WES sur les trios . . . . .	104
<b>9 Résultats</b>	<b>105</b>
9.1 Identification des mutations du gène <i>NONO</i> : publication n°3 (soumise) . . . . .	105
9.2 Données supplémentaires : rôle de <i>NONO</i> au sein des paraspeckles . . . . .	158
<b>10 Discussion et perspectives</b>	<b>160</b>
10.1 Syndrome MCCID : un syndrome polymalformatif lié à une dérégulation transcriptionnelle globale . . . . .	160
10.2 Contribution des anomalies du rythme circadien au phénotype des patients MCCID . . . . .	160
10.3 Anomalie des synapses GABAergiques et pathologies neurodéveloppementales : des pistes thérapeutiques . . . . .	162
<b>IV Discussion et perspectives : Apport et limites du WES pour le démantèlement génétique de la DI</b>	<b>164</b>
<b>11 Bilan des travaux de thèse</b>	<b>165</b>
<b>12 Réflexions autour de l'outil WES</b>	<b>167</b>
12.1 Évolution et avantages de l'approche WES au cours des 4 dernières années . . . . .	167
12.1.1 Études familiales . . . . .	167
12.1.2 Études de cohorte de patients . . . . .	167
12.2 Limites du WES . . . . .	168
<b>13 Perspectives pour l'identification des gènes de DI ; l'approche de demain</b>	<b>169</b>
13.1 Amélioration de la démarche diagnostique . . . . .	169
13.1.1 Mise en place d'un réseau d'échange d'informations : CARID . . . . .	169
13.1.2 Vers un test diagnostique unique de la DI . . . . .	169
13.2 Exploration du génome entier : premiers pas . . . . .	170
<b>14 Validation fonctionnelle des variants : dissertation autour du choix des modèles</b>	<b>172</b>
<b>Matériels et Méthodes</b>	<b>175</b>
<b>15 Capture et séquençage de l'exome</b>	<b>176</b>

---

<b>16</b>	<b>Génotypage</b>	<b>179</b>
<b>17</b>	<b>Bio-informatique</b>	<b>180</b>
<b>18</b>	<b>Fractionnement cellulaire</b>	<b>181</b>
<b>Annexes</b>		<b>182</b>
<b>A</b>	<b>Figures</b>	<b>183</b>
<b>B</b>	<b>Tables</b>	<b>200</b>
<b>Bibliographie</b>		<b>203</b>

## TABLE DES FIGURES

1.1	Distribution du QI dans la population générale. . . . .	22
1.2	Illustration de la recombinaison homologue non-allélique, figure modifiée d'après Stankiewicz and Lupski [2002]. . . . .	26
1.3	Représentation des gènes de DILX non-syndromique (NS) et leur localisation. . . . .	29
1.4	Représentation des gènes de DILX syndromique (S) et leur localisation. . . . .	30
1.5	Principe de la méthode de cartographie par autozygotie : recherche de mutations homozygotes. . . . .	32
1.6	Principe de la méthode de cartographie par autozygotie : recherche de mutations hétérozygotes composites. . . . .	33
2.1	Résultats obtenus sur l'étude de 136 familles, d'après Najmabadi et al. [2011]. . . . .	40
3.1	Composants synaptiques impliqués dans les conditions neurodéveloppementales, d'après [Srivastava and Schwartz, 2014]. . . . .	41
3.2	Classification des fonctions moléculaires des gènes des 4 listes, d'après [Srivastava and Schwartz, 2014]. . . . .	42
3.3	Classification des processus biologiques des gènes des 4 listes, d'après [Srivastava and Schwartz, 2014]. . . . .	43
5.1	Arbre généalogique de la famille Mo. . . . .	47
5.2	Photographies des patients de la famille Mo. . . . .	48
5.3	Caractéristiques cliniques des patients de la famille Mo. . . . .	52
5.4	Carte physique de la région candidate identifiée sur le chromosome 8 (source : UCSC). . . . .	53
5.5	Électrophoregrammes représentant la mutation c.1243C>T du gène <i>RAD54B</i> chez deux enfants atteints et un contrôle. . . . .	54
5.6	Alignement des séquences protéiques des orthologues de <i>RAD54B</i> . . . . .	54
5.7	Stabilité de la protéine <i>RAD54B</i> . . . . .	55
5.8	Localisation cellulaire de la protéine <i>RAD54B</i> . . . . .	56
5.9	Intensités de fluorescence du marquage <i>RAD54B</i> . . . . .	56

5.10	Électrophoregrammes représentant la mutation c.1901T>C du gène <i>WDR62</i> chez la famille Mo. . . . .	57
5.11	Mutations <i>WDR62</i> identifiées à ce jour, figure modifiée d'après McDonell et al. [2014]. . . . .	58
5.12	Patients <i>WDR62</i> rapportées dans la littérature, inspiré de Poulton et al. [2014]. . . . .	62
5.13	IRM des patients de la famille Mo. . . . .	63
6.1	Caractéristiques cliniques des maladies liées à une anomalie de la réparation. . . . .	81
9.1	Distribution cytoplasmique vs nucléaire des ARN testés : résultats de qPCR. . . . .	158
9.2	Distribution cytoplasmique vs nucléaire des ARN testés : résultats de RT-PCR. . . . .	159
13.1	Bilan des fréquences de diagnostic positif obtenues selon l'approche utilisée, d'après Gilissen et al. [2014]. . . . .	171
18.1	Position des amorces permettant de sélectionner les ARN édités, modifié d'après Chen and Carmichael [2009]. . . . .	181
A.1	Principe de la CGH sur métaphase et sur puce à ADN (CGH array), d'après Sanlaville et al. [2005]. . . . .	183
A.2	Comparaison des principes des 3 principales technologies de NGS. . . . .	184
A.3	Ensemble des étapes des la technologie Illumina (source : Illumina). . . . .	185
A.4	Étape de fixation et d'amplification de la librairie sur la flow cell (source : Illumina). . . . .	186
A.5	Réaction de séquençage par la technique Illumina (source : Illumina). . . . .	186
A.6	Technique d'enrichissement Agilent SureSelect, modifié de la source : Protocole Agilent SureSelect Target Enrichment System. . . . .	187
A.7	La technologie RainDance : système de PCR en microgouttelettes, d'après Tewhey et al. [2009]. . . . .	188
A.8	Capture des exons avec les sondes inversées (MIPs), d'après Turner et al. [2009]. . . . .	189
A.9	Les 3 méthodes d'enrichissement spécifique : PCR ciblées (a), MIPs (b) et capture par hybridation spécifique (c), d'après Mamanova et al. [2010]. . . . .	190
A.10	Du génome à la variation, d'après Bell et al. [2011]. . . . .	191
A.11	Le mécanisme de réparation des CDB dépendant de la recombinaison homologue d'après Liu et al. [2014a]. . . . .	192
A.12	Forme des noyaux des cellules : . . . . .	193
A.13	Circularité des noyaux dans les fibroblastes du patient III.2, les cellules lymphoïdes AT163 <i>ATM</i> <sup>-/-</sup> et les contrôles correspondants (fibroblastes contrôles et cellules lymphoïdes EBV010). . . . .	194
A.14	Niveau de ROS mesuré par DCFDA dans les fibroblastes du patient III.2 et les fibroblastes contrôles. . . . .	195
A.15	Schéma de la voie de signalisation mTOR. . . . .	196
A.16	Test de l'expression du gène <i>TTI2</i> dans divers tissus humain par RT-PCR. . . . .	197
A.17	Agenda de sommeil. . . . .	199



## LISTE DES TABLEAUX

1.1	Les différents degrés de sévérité des DI. . . . .	21
1.2	Fréquences associées aux différentes causes environnementales connues. . . . .	24
1.3	Syndromes microdélétionnels à points de cassures récurrents. . . . .	27
1.4	Gènes de DIAR découverts par la méthode de cartographie par autozygotie. . . . .	34
1.5	Les syndromes associés au locus 15q11-q13. . . . .	35
5.1	Mensurations des parents. . . . .	48
5.3	Génotypage des locis connus de microcéphalie. . . . .	52
5.4	Les régions IBD candidates. . . . .	53
5.6	Fréquence des différents signes cliniques observés chez les patients Mo et la totalité des patients <i>WDR62</i> . . . . .	64
7.1	Dosage des concentrations plasmatiques de ZAG. . . . .	101
7.2	Comparaison des caractéristiques du SGC, du digénisme et du SMC. . . . .	102
11.1	Bilan de l'analyse par WES des 5 familles présentées. . . . .	165
13.1	Comparaison des outils de WES vs WGS. . . . .	171
15.1	Les différents projets de WES réalisés. . . . .	177
16.1	Les différents projets de cartographie par autozygotie réalisés. . . . .	179
B.1	Données d'intensités de fluorescence du marquage RAD54B : . . . . .	200
B.2	Bilan Immunologique du patient III.2 de la famille El. . . . .	201
B.3	Dosage des marqueurs tumoraux chez les 3 enfants atteints de la famille El. . . . .	201
B.4	Modélisation des fonctions dérégulées chez les patients MCCID. . . . .	202

A-T, 80  
A-TLD, 80  
AAIDD, 20  
ACPA, 25  
ADN, 176  
AFP, 82  
ANOVA, 56  
APA, 20  
AS, 27  
  
btw, 62  
  
Ca, 52  
CARID, 169  
CC, 62  
CD4, 82  
CD8, 82  
CDB, 55  
CDG-Ia, 31  
CGH, 183  
CHARGE, 31  
CHC, 82  
CHM, 62  
CHO, 102  
CIM-10, 20  
CNVs, 25  
cpm, 201  
CRISPR, 173  
CSF, 62  
CT, 62  
  
DCFDA, 195  
  
DE, 160  
DGS/VCF, 27  
DI, 20  
DIAD, 31  
DIAR, 31  
DILX, 28  
DS, 21  
DSBs, 81  
DSM-V, 20  
DUP, 27  
  
EA, 57  
ECG, 161  
EEG, 161  
EMG, 161  
EOG, 161  
ES, 57, 173  
EuroMRX, 28  
EVS, 168  
  
FISH, 25  
FXS, 28  
  
g/l, 201  
GABA, 162  
GO, 42  
GTC, 62  
  
HC, 62  
HR, 80  
  
IBD, 32  
IC, 56

---

ID, 9  
Ig, 201  
IGOLD, 28  
IPA, 202  
iPSCs, 172  
IRAlus, 181  
IRM, 63  
KO, 162  
  
L, 62  
LC-MS/MS, 101  
LCRs, 26  
lod-score, 33  
  
m, 62  
mat, 62  
MC, 62  
Mg, 52  
MIPs, 37  
MPS, 52  
MRI, 62  
MSG, 62  
  
N, 62  
na, 62  
NAHR, 26  
nd, 52  
NGS, 31  
NHEJ, 80  
nl, 52  
NPCs, 58  
NS, 22  
  
OMS, 20  
  
P, 52  
PAH, 31  
pat, 62  
PC, 52  
PCR, 176  
PCU, 31  
Pds, 52  
PMG, 62  
  
PWS, 35  
  
QD, 21  
QI, 21  
  
R, 62  
ROS, 82  
RPM, 21  
RT, 181  
  
S, 22  
S2D, 31  
SD, 62  
SFS, 38  
sg, 52  
SGC, 26  
SLO, 31  
SMC, 102  
SMS, 27  
SNC, 20  
SNP, 33  
SNV, 170  
SSBs, 81  
  
T, 52  
TALENS, 173  
TCL, 52  
TEMRA, 82  
TG, 52  
  
UTR, 28  
UV, 32  
  
WB, 55  
WES, 38  
WGS, 170  
  
y, 62  
  
ZFNs, 173

**Première partie**

**Introduction**

# CHAPITRE 1

## LES DÉFICIENCES INTELLECTUELLES (DI)

L'équipe de recherche dirigée par le Dr Laurence Colleaux s'intéresse à la génétique des pathologies du système nerveux central, comme les déficiences intellectuelles (DI). Ces maladies fréquentes et gravement invalidantes sont responsables de 10% des dépenses de santé dans les pays développés, loin devant le cancer ou les maladies cardiovasculaires et constituent donc un enjeu majeur de santé publique. L'absence de diagnostic moléculaire laisse les familles sans explications, or elles sont déterminantes pour aider ces familles à mieux accepter la maladie en lui donnant un sens. Ce travail aspire à mieux comprendre les bases moléculaires et physiopathologiques des DI d'une part et les processus nécessaires à la mise en place du système nerveux central (SNC) au cours du développement d'autre part. Ces résultats ont également pour vocation immédiate d'offrir une meilleure prise en charge des patients et de leurs familles par les généticiens en leur apportant un conseil génétique pour les grossesses à venir mais aussi, à plus long terme, de fournir les éléments nécessaires à la conception de nouvelles stratégies thérapeutiques.

### 1.1 Définitions des DI, classification et prévalence

#### 1.1.1 Définition

Les définitions de la DI de l'*American Association on Intellectual and Developmental Disabilities* (AAIDD) et de la Classification Internationale des Maladies (CIM-10) de l'Organisation Mondiale de la Santé (OMS) sont relativement proches de la définition de l'*American Psychiatric Association* (APA), la plus largement utilisée dans la littérature scientifique, rédigée dans la cinquième édition du manuel « *Diagnostic and Statistical Manual of Mental Disorders* » abrégé DSM-V. Selon cette source, la DI, diagnostiquée pendant la période de développement (avant l'âge de 18 ans), correspond à une faiblesse généralisée des capacités mentales avec un impact sur le fonctionnement adaptatif des trois domaines suivants : le domaine conceptuel (langage, lecture, écriture, mathématiques, raisonnement, connaissance, mémoire, temps, espace, argent...), le domaine social (empathie, responsabilité sociale, jugement social, compétences de communication interpersonnelle, méfiance/naïveté appropriée, capacité à créer et maintenir des relations amicales, capacité à résoudre les problèmes sociaux...) et le domaine pratique (autonomie : hygiène, utilisation de moyens de

transport, responsabilités professionnelles, gestion du budget, gestion de l'emploi du temps, organisation du travail...).

### 1.1.2 Classification

La classification de ces affections repose sur le degré de sévérité de la DI. Le fonctionnement intellectuel général est apprécié selon divers tests cognitifs standardisés et adaptés à l'âge et au niveau de développement du patient. Pour les patients les plus jeunes, les cliniciens utilisent des tests psychométriques qui permettent d'estimer le Quotient de Développement (QD) chez l'enfant, comme le test de Denver ou l'échelle de Brunet-Lézine. Si un déficit des acquisitions est observé, il est alors nommé Retard Psychomoteur (RPM). Avec l'apparition du langage, les cliniciens sont ensuite en mesure d'évaluer le Quotient Intellectuel (QI). Celui-ci, composé des sous-catégories QI verbal et QI de performance, représente la mesure la plus fréquemment utilisée pour classer les DI.

Dans une population donnée, les scores du QI suivent une distribution normale dont la moyenne, fixée arbitrairement, est de 100 et la déviation standard (DS) de 15. Par conséquent, 95% des individus ont un QI entre -2 DS et +2 DS (approximation de 1,96) correspondant à des scores entre 70 et 130. Par définition, la DI correspond à un score du QI inférieur à 70 (soit -2 DS). Cette notion de seuil est perdue dans le DSM-V, dernière version du DSM publiée le 18 Mai 2013, et n'existe aujourd'hui que dans la CIM-10. Les seuils indiqués dans la table 1.1 se réfèrent donc au DSM-IV tandis que les caractéristiques développementales des quatre degrés de gravité, de la DI légère à la DI profonde, sont définies par le DSM-V [Daily et al., 2000]. Richardson [1989] propose une classification simplifiée avec uniquement deux degrés de sévérité : la DI légère et la DI sévère (table 1.1).

DI légère	DI modérée	DI sévère	DI profonde
85% des cas 50<QI<70	10% des cas 35<QI<49	3 à 4% des cas 20<QI<34	1 à 2% des cas QI<20
Les patients sont autonomes, peuvent avoir un parcours scolaire avec soutien et nécessitent une aide pour gérer leurs finances, les transports et leur santé.	Les patients 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.	Les facultés communicatives sont basiques, la prise en charge nécessite une aide quotidienne.	Les patients sont totalement dépendants d'une aide extérieure et nécessitent une prise en charge permanente.
DI légère	DI sévère		

**Table 1.1** : Les différents degrés de sévérité des DI.

La présence éventuelle d'autres signes cliniques associés aux DI amène un autre niveau de classification. Lorsque la DI est le seul signe clinique observé chez le patient, on nomme celle-ci une DI isolée ou non-syndromique (NS) ou encore non-spécifique. À l'inverse, lorsqu'elle est associée à d'autres symptômes (malformations, dysmorphie faciale, anomalies biochimiques ou radiologiques), on l'appelle DI syndromique (S).

### 1.1.3 Prévalence

On estime que les DI affectent 2,5% de la population générale et la prévalence théorique des DI sévères (-3 DS) serait de 0,23%. Diverses études montrent que les DI touchent 1 à 8% de la population générale [Ropers, 2008, 2010]. Ces estimations varient car elles sont biaisées par la sous-population étudiée et les tests de QI, souvent culture-dépendant. L'ensemble des études épidémiologiques s'accordent tout de même sur une prévalence de la DI de 2 à 3% dans la population générale. Par ailleurs, la prévalence des DI sévères observée est plus élevée que celle estimée (de l'ordre de 0,3 à 0,5%) [Cans et al., 1999, Ropers, 2010]. L'excédent de sujets avec un QI inférieur à 50 comparé au taux attendu suggère une distribution bimodale des scores de QI, comme indiqué sur la figure 1.1.

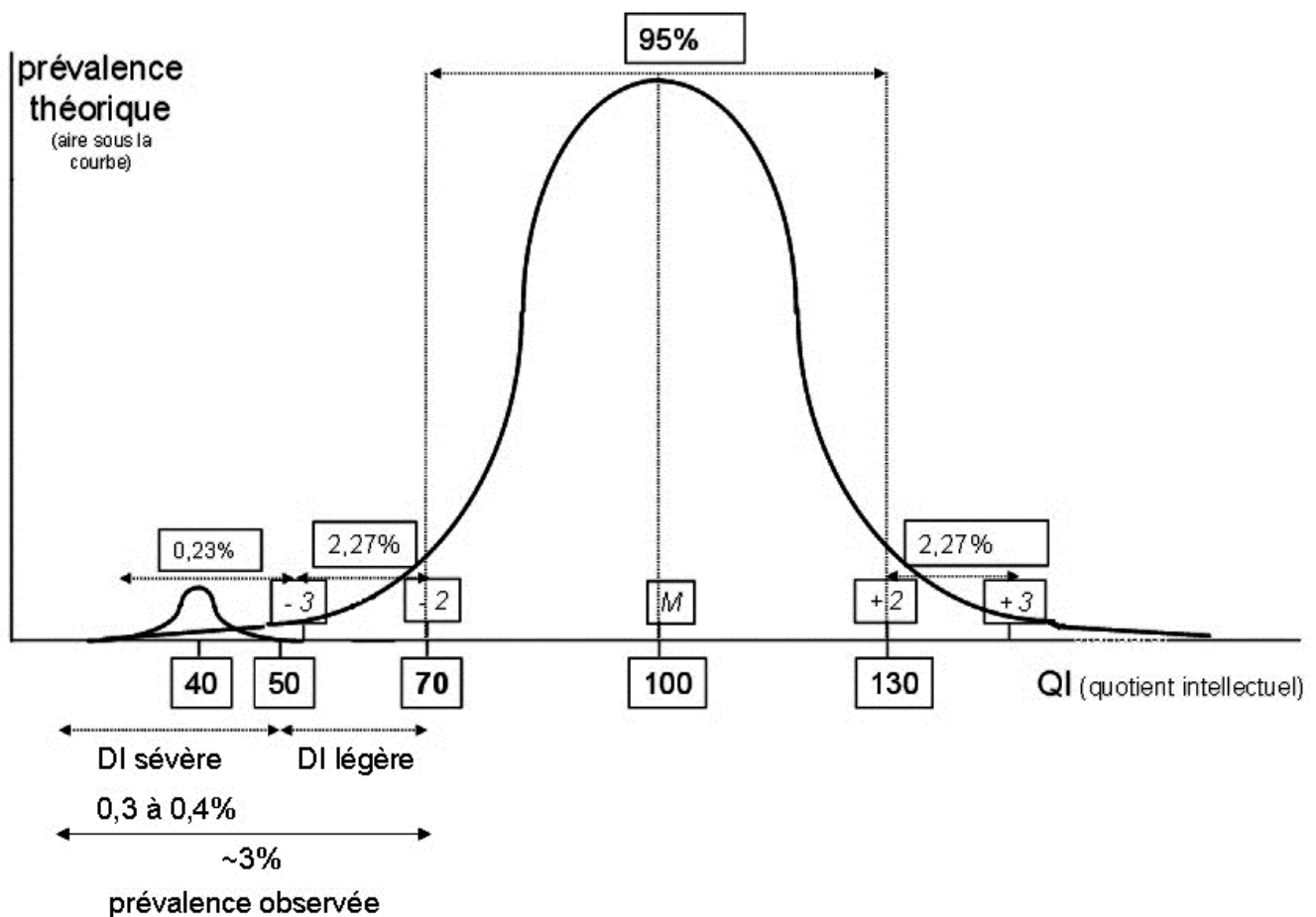


Figure 1.1 : Distribution du QI dans la population générale.

Enfin, on observe un sex-ratio déséquilibré (entre 1,3 et 1,4) avec une majorité de sujets mâles parmi les patients atteints de DI [Bahi-Buisson et al., 2006, Ropers, 2006].

---

## 1.2 L'étiologie des DI jusqu'à la révolution NGS (Next-Generation Sequencing)

Les DI sont des pathologies neurodéveloppementales dont la cause se manifeste pendant la période de gestation, à la naissance ou pendant les premières années de vie, alors que le système nerveux est encore en développement. En effet, après la naissance, une phase de développement rapide avec maturation se déroule durant les premières années de l'enfance. Une deuxième phase, plus lente, lui succède jusqu'à la fin de l'adolescence [collective INSERM, 2004].

### 1.2.1 Causes environnementales

Les déficiences intellectuelles dont la cause est environnementale représentent 15 à 20% des cas de DI [collective INSERM, 2004]. Ces causes sont citées ci-dessous en fonction de la période de survenue.

#### i) Période prénatale

Elle correspond à la période comprise entre le premier jour des dernières règles et le début du travail.

Les infections fœtales (cytomégalovirus CMV, toxoplasmose, herpès, rubéole, etc...) sont responsables d'environ 5% des DI [Benoist et al., 2008, Mwaniki et al., 2012].

De même, l'exposition *in utero* à des agents exogènes tels que l'alcool, certaines drogues (cocaïne), le tabac et certains médicaments neurotoxiques, peut altérer le développement du système nerveux central. Ces intoxications seraient responsables d'environ 2% des DI [Clifford et al., 2012].

Behrooz et al. [2011] et Bath et al. [2013] rapportent une plus forte prévalence de malformations congénitales et de déficiences intellectuelles engendrées par certaines carences nutritionnelles durant la période périconceptionnelle, telles que le déficit en folates, l'hyperphénylalaninémie maternelle ou l'hypothyroïdisme et la carence en iode.

#### ii) Périodes péri-natales et post-natales

La période périnatale s'étend du début du travail jusqu'au septième jour de vie. Pendant cette période les accidents tels que l'anoxie cérébrale, l'hémorragie cérébrale et l'ictère nucléaire (atteinte cérébrale suite à l'accumulation de bilirubine) peuvent entraîner une DI. Comme reporté dans un précédent rapport d'expertise de l'INSERM [collective INSERM, 2004], la grande prématurité augmente aussi de façon importante le risque de troubles cognitifs.

La période qui suit est dite post-natale. Une DI peut alors provenir de maladies infectieuses de la petite enfance telles que la méningite, l'encéphalite, les encéphalopathies post-vaccinales. Les intoxications peuvent également favoriser l'apparition d'un déficit intellectuel. L'effet neurotoxique d'une exposition à des métaux lourds tels que le plomb a été largement démontré [Mendola et al., 2002, Min et al., 2009, McDermott et al., 2011]. De même, le traumatisme crânien, principale cause de handicap et première cause de mortalité chez les moins de 15 ans, provoqué par les accidents de la voie publique (60% des cas), des chutes (25%) ou des accidents de sport (15%) peut générer de multiples séquelles cognitives dont la gravité est corrélée à l'ampleur du traumatisme crânien [Anderson et al., 2012].



Enfin, les facteurs socio-économiques tels que la malnutrition, les abus ou les carences émotionnelles représentent également un facteur de risque de DI. En effet, un soutien émotionnel ainsi qu'une stimulation cognitive provenant de l'entourage sont nécessaires au développement d'un individu. Cependant, la part des causes environnementales dans la prévalence des DI diffère selon le niveau de développement des pays. En effet, l'information, la formation et la prévention des risques auprès des femmes enceintes et du personnel médical sont plus ou moins riches et engendrent une qualité différente de la prise en charge des patients selon les États [Ropers, 2010]. Il est donc difficile d'évaluer la part des facteurs socio-environnementaux des pathologies induites par les conditions de vie. Néanmoins, des études cas-témoins suggèrent un impact distinct de ces facteurs sur l'apparition d'un déficit intellectuel. Par exemple, une étude finlandaise a ainsi établi une corrélation positive entre le niveau d'instruction des parents et leur statut socio-économique et la prévalence de la DI [Heikura et al., 2008]. De même, une étude américaine a montré que des enfants nés de mères avec moins de 12 années de scolarisation ont un risque 7 fois plus élevé de développer un retard cognitif comparé à ceux nés de mères ayant suivi des études supérieures [Chapman et al., 2002]. Contrairement aux facteurs organiques, les facteurs psychosociaux apparaissent d'autant plus importants qu'on se situe dans le cadre de la déficience légère. En effet, dans les DI légères (85% des DI), les facteurs organiques sont plus rarement retrouvés. Par exemple des anomalies chromosomiques ne sont retrouvées que dans 4 à 8% des cas. En revanche, des facteurs organiques sont mis en évidence dans 75% des cas des DI sévères (QI < 50) et la prévalence est la même quel que soit le milieu socio-économique de l'enfant [Ropers, 2008, 2010]. Aussi, pour les DI de cause organique déterminée, les conditions environnementales psychologiques et sociales jouent un rôle dans le sens de l'aggravation ou de l'atténuation des symptômes. La table 1.2 ci-dessous, résume les fréquences associées aux différentes causes environnementales citées précédemment.

Causes		Fréquences (%)
Environnementales	Traumatisme	5
	Infection	5
	Intoxication	2
	Prématurité	5
	Autres	1
	Total	18

**Table 1.2 :** Fréquences associées aux différentes causes environnementales connues.

### 1.2.2 Causes génétiques

Les causes génétiques de la DI sont nombreuses et diverses. Il est difficile d'en dresser une liste exhaustive étant donné la fréquence à laquelle de nouvelles étiologies sont décrites aujourd'hui. Parmi les 25 000 gènes du génome humain, on estimait en 1995, à environ 1/3 ceux exprimés au niveau du cerveau [Takahashi et al., 1995]. Récemment, du fait de l'amélioration de l'annotation du génome humain d'une part et des technologies d'autre part, cette estimation a été revue à la hausse : 84% des gènes ont été observés exprimés dans au moins une des régions du cerveau, où ils jouent un rôle dans le développement et le fonctionnement cérébral,

---

permettant vraisemblablement d'expliquer l'extrême hétérogénéité génétique des DI [Hawrylycz et al., 2012, Shen et al., 2012].

### **i) Anomalies chromosomiques**

Les anomalies chromosomiques sont la cause la plus fréquente de DI. Il s'agit d'anomalies chromosomiques de nombre (aneuploïdies) ou de structures telles que les délétions, duplications, translocations ou inversions.

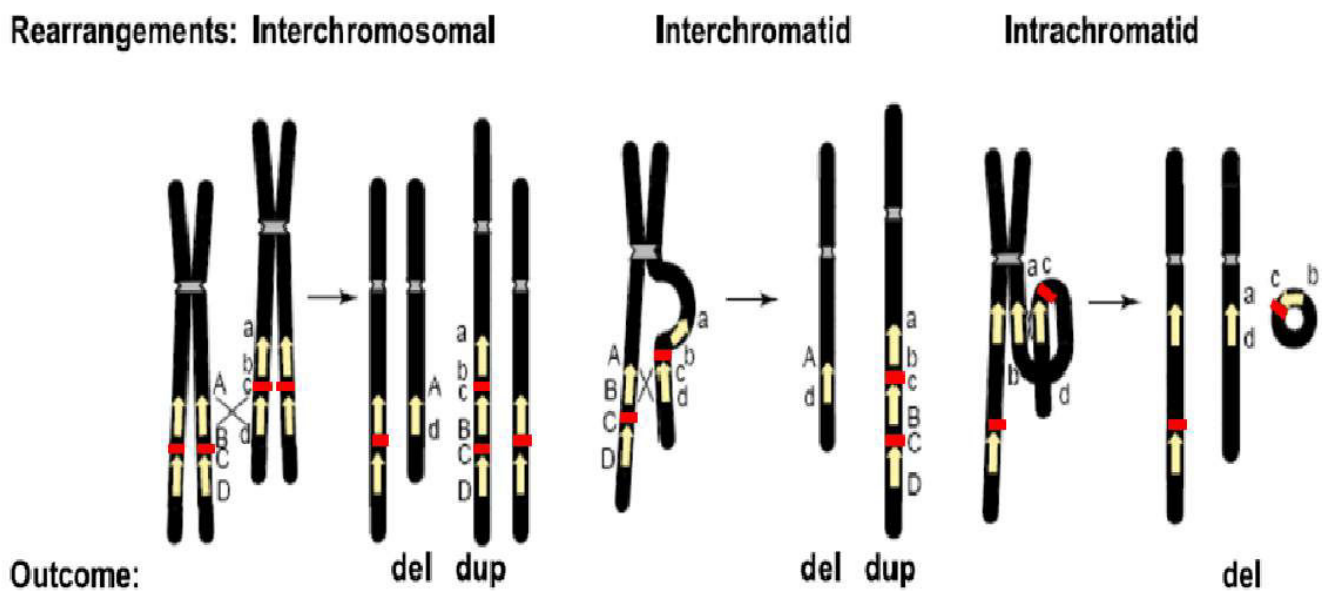
L'aneuploïdie, correspondant à la perte ou au gain d'un chromosome entier, est un phénomène fréquent à la conception [Regan, 2010]. Les aneuploïdies peuvent être homogènes (dans toutes les cellules de l'organisme) ou en mosaïque (dans certains types cellulaires uniquement). Les aneuploïdies homogènes résultent le plus souvent d'une non-disjonction méiotique et peuvent se traduire par une trisomie (présence d'un chromosome normal surnuméraire) ou une monosomie (perte d'un chromosome). Le syndrome de Down ou trisomie 21, avec une prévalence de 1/750 ou 1/800, est la forme génétique la plus fréquente de DI. Il correspond à la première anomalie chromosomique décrite associée à une pathologie et identifiée sur caryotype marqué en Giemsa [Lejeune et al., 1959].

Les anomalies chromosomiques de structure apparaissent à la suite de cassures chromosomiques suivies par un ou plusieurs ré-assemblages inattendus et affectent un ou plusieurs chromosomes, homologues ou non. Elles peuvent être équilibrées ou non. De manière générale, les anomalies équilibrées n'entraînent pas de déséquilibre du matériel chromosomique et ne sont donc pas associées à un phénotype. Exception faite des cas des remaniements chromosomiques équilibrés dont un des points de cassure interrompt un gène ou altère son expression. Aussi, lors de la méiose, les anomalies équilibrées peuvent entraîner la formation de gamètes déséquilibrés générant des zygotes anormaux après fécondation conduisant ultérieurement à la nécessité d'interrompre la grossesse ou la naissance d'enfants porteurs d'anomalies congénitales. Ainsi, les anomalies non-équilibrées surviennent *de novo* ou à la suite d'un remaniement parental équilibré.

Les premiers outils de caryotype ne permettaient que le diagnostic des anomalies de nombre ou de structure de grande taille (supérieures à 25 Mb). Puis, dans les années 1990, l'essor de la cytogénétique moléculaire avec les sondes, ces nouveaux outils de biologie moléculaire utilisés dans la technique de FISH (Fluorescent *In Situ* Hybridization), rendait possible la caractérisation d'anomalies submicroscopiques (inférieures à 4Mb) d'une région d'intérêt mais ne permettait pas l'étude globale du génome. Cette technique est donc utilisée lorsqu'une région chromosomique candidate est proposée à la suite d'un examen clinique détaillé mettant en évidence un tableau clinique évocateur.

Dans les années 2000, l'avènement de la technique de l'analyse comparative sur puce à ADN (ACPA) rendait possible la mise en évidence de remaniements submicroscopiques entre 5 et 10 Mb, non visibles sur caryotype, appelés CNVs (pour Copy Number Variants). Aujourd'hui, la résolution de cette technique ayant considérablement progressée, elle offre la possibilité d'une étude pangénomique avec une résolution encore supérieure, de quelques Kb (voir figure A.1 en annexe). La recherche systématique d'anomalies chromosomiques chez les patients atteints de DI idiopathique a permis, par leur analyse rétrospective, de décrire de nouveaux syndromes associés à des microdélétions ou des microduplications. En effet, classiquement, la clinique des syndromes était initialement décrite (comme la trisomie 21) et les bases génétiques élucidées par la suite. L'ACPA a permis d'appliquer la démarche inverse ; l'identification de l'anomalie chromosomique précédant la description phénotypique.

Plusieurs anomalies chromosomiques de petite taille sont récurrentes car un même mécanisme en est à l'origine et l'architecture du génome explique ce constat. Les régions télomériques, riches en séquences minisatellites hypervariables (séquences répétées en tandem de 10 à 60 paires de bases ou pb), sont des zones singulièrement sensibles aux cassures et de nombreux syndromes cliniquement reconnaissables, dus à des remaniements dans ces zones, ont été identifiés : 1p36, 2q37, 3pter, 9q34, 18qter et 22q13. D'autre part, des séquences d'ADN répétées de 10 à 400 kb, appelées duplicons ou LCRs (pour Low Copy Repeats) sont présentes en plusieurs copies dans le génome humain, espacées de 500 kb à 4 Mb et présentent plus de 90% d'identité. L'homologie de ces séquences favorise les recombinaisons homologues non-alléliques (NAHR pour non-allelic homologous recombination) lors de la méiose entraînant la duplication ou la délétion des régions comprises entre deux duplicons de même orientation, comme illustré dans la figure 1.2, adaptée d'après l'article de Gu et al. [2008].



**Figure 1.2 :** Illustration de la recombinaison homologue non-allélique, figure modifiée d'après Stankiewicz and Lupski [2002].

Une recombinaison entre LCR (en jaune) de même orientation, entre chromosomes homologues (premier cas) ou entre chromatides sœurs (second cas), entraîne une délétion ou une duplication du segment intermédiaire (en rouge) tandis qu'un échange entre LCR de même orientation sur une même chromatide conduit à la délétion du segment intermédiaire (dernier cas).

Un dosage inapproprié des gènes contenus dans le fragment génomique remanié engendre vraisemblablement le phénotype clinique associé. Selon la taille du fragment génomique impliqué, un dosage génique inadapté peut se manifester sous la forme de syndrome des gènes contigus (SGC) ou de maladie mendélienne monogénique.

La notion de syndrome des gènes contigus a été introduite par Schmickel [1986] afin d'expliquer l'ensemble des signes cliniques caractéristiques d'une microdélétion hémizygote. L'auteur propose que le phénotype observé chez ces patients soit la conséquence de l'haploinsuffisance de l'ensemble des gènes contenus dans le

segment délété. Le syndrome de WAGR (Wilms tumor, Aniridia, Genitourinary anomalies, mental Retardation) est un exemple de SGC et associe un risque de survenue de tumeur de Wilms à tout âge, une aniridie totale ou partielle, des anomalies génito-urinaires allant de l'ambiguïté sexuelle à l'ectopie testiculaire, ainsi qu'une DI de degré variable. La délétion du gène *PAX6* (Paired box gene 6) est responsable de l'aniridie alors que l'haploinsuffisance du gène *WT1* (Wilms' tumor suppressor gene) est associée au risque de survenue de tumeur et des anomalies uro-génitales [Fischbach et al., 2005]. Cette théorie est valable pour la très grande majorité des cas mais elle a toutefois été démentie pour certains syndromes, comme les syndromes d'Angelman (AS), de DiGeorge/vélo-cardio-facial (DGS/VCF) et de Smith-Magenis (SMS), pour lesquels l'anomalie d'un seul gène explique la très grande majorité du phénotype, respectivement *UBE3A*, *TBX1* et *RAI1* [Kishino et al., 1997, Baldini, 2005, Williams et al., 2012]. Dans ces cas précis, on appelle ce gène, « le gène majeur ». La table 1.3 ci-dessous, résume les principaux syndromes de DI et/ou malformations congénitales reconnaissables cliniquement, causés par des CNVs récurrents [Stankiewicz and Lupski, 2010].

Syndrome	OMIM	Locus	Réarrangement	Gène(s)	Incidence
Williams-Beuren (WBS)	194050	7q11.23	délétion	SGC incl. <i>ELN</i>	1/20000
dup(7)q11.23	609757	7q11.23	duplication	SGC incl. <i>ELN</i>	1/7500
Angelman (AS)	105830	15q11-q12	mat dél, DUP15 pat	<i>UBE3A</i>	1/10000 - 1/20000
Prader-Willi (PWS)	176270	15q11-q12	dél pat, DUP15 mat	SGC	1-9/100000
Miller-Dieker (MDLS)	247200	17p13.3	délétion	SGC incl. <i>LIS1</i>	1/100000
Smith-Magenis (SMS)	182290	17p11.2	délétion	<i>RAI1</i>	1/15000 - 1/25000
Potocki-Lupski (PTLS)	610883	17p11.2	duplication	SGC	1/25000
del(17)q21	610443	17q21.31	délétion	SGC incl. <i>KANSL1</i>	1/16000
DiGeorge/vélo-cardio-facial (DGS/VCFS)	188400	22q11.2	délétion	<i>TBX1</i>	1/5000
dup(22)q11	608363	22q11.2	duplication	SGC incl. <i>TBX1</i>	1/4000

**Table 1.3 :** Syndromes microdélétionnels à points de cassures récurrents.

Cependant, si ces syndromes microdélétionnels sont dits récurrents car les points de cassures observés sont invariablement les mêmes impliquant une prévalence relativement élevée (1/4000 pour la duplication 22q11.2 par exemple), il existe de nombreux autres remaniements, rares et donc difficiles à diagnostiquer et à interpréter.

Enfin, il faut citer les disomies uniparentales (DUP) ou unidisomies, qui correspondent à la présence chez un individu de deux chromosomes d'une même paire provenant d'un seul de ses parents. Ainsi, ces phénomènes sont susceptibles de révéler des maladies récessives.

---

## ii) Formes monogéniques de DI

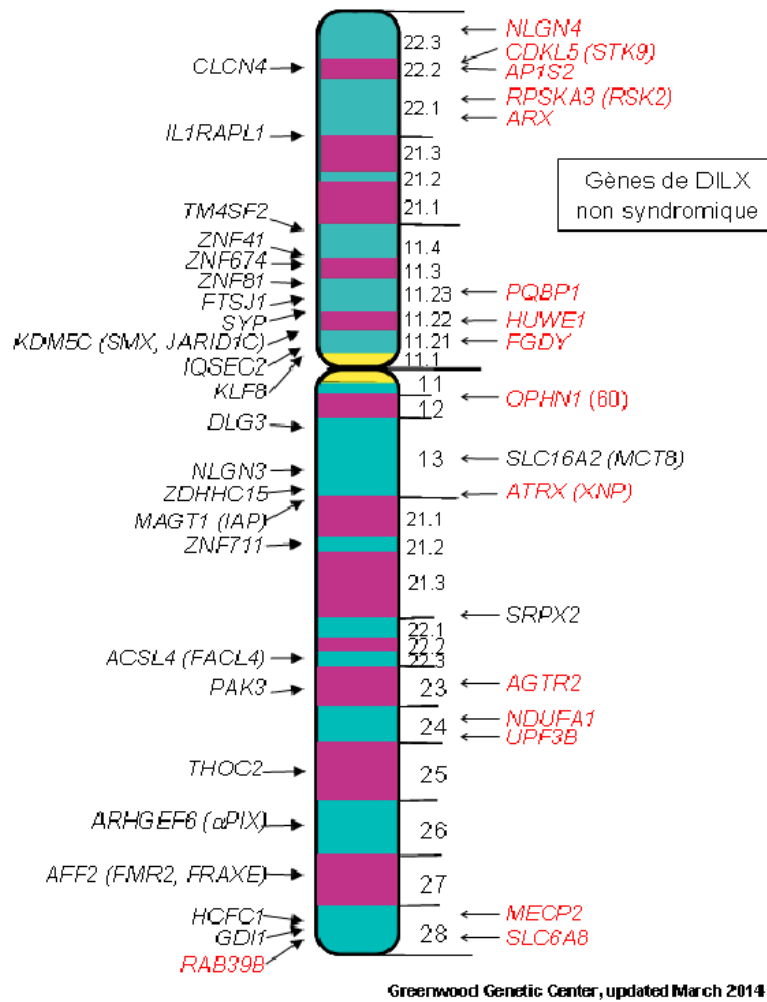
Les formes monogéniques des DI résultent de l'anomalie d'un seul gène. Elles sont présentées ci-dessous selon les trois modes de transmission mendéliens : lié au chromosome X, autosomique dominant et autosomique récessif.

### Les formes liées au chromosome X (DILX) :

L'excès de garçons parmi les patients ayant une DI et l'observation de familles avec une DI ségrégeant selon un mode lié au chromosome X suggéraient que les anomalies des gènes du chromosome X contribuaient de façon importante à l'étiologie de ces affections. Les recherches concernant l'implication du chromosome X dans les DI ont donc débuté dans les années 1990, plus précocement que pour les formes liées aux autosomes. Le syndrome de l'X fragile (FXS), correspondant à un déficit intellectuel de léger à sévère qui peut être associé à des troubles du comportement et à des signes physiques caractéristiques, a été la première anomalie génétique identifiée dans le groupe des DILX et correspond à la cause la plus fréquente de DI d'origine monogénique. Elle est due à un nombre de répétitions supérieure à 200 d'un triplet CGG dans la région 5'UTR (UnTranslated Region) du gène *FMR1* (Xq27.3), entraînant une répression de son expression [Verkerk et al., 1991, Mandel and Heitz, 1992]. La constitution du Consortium européen « EuroMRX » en 1995, s'appuyant sur des réseaux de cliniciens et de chercheurs, a permis de regrouper les données cliniques d'une large cohorte de 600 familles multiplex (composées de plusieurs enfants atteints) dans lesquelles ségrége une DILX. L'analyse des données de cartographie génétique et de clonage positionnel de ces familles a largement accéléré la découverte de gènes de DILX. Le projet « EuroMRX » s'est poursuivi par un séquençage systématique de 90 gènes de DILX, connus ou candidats permettant l'identification des mutations causales dans 42% des familles [de Brouwer et al., 2007].

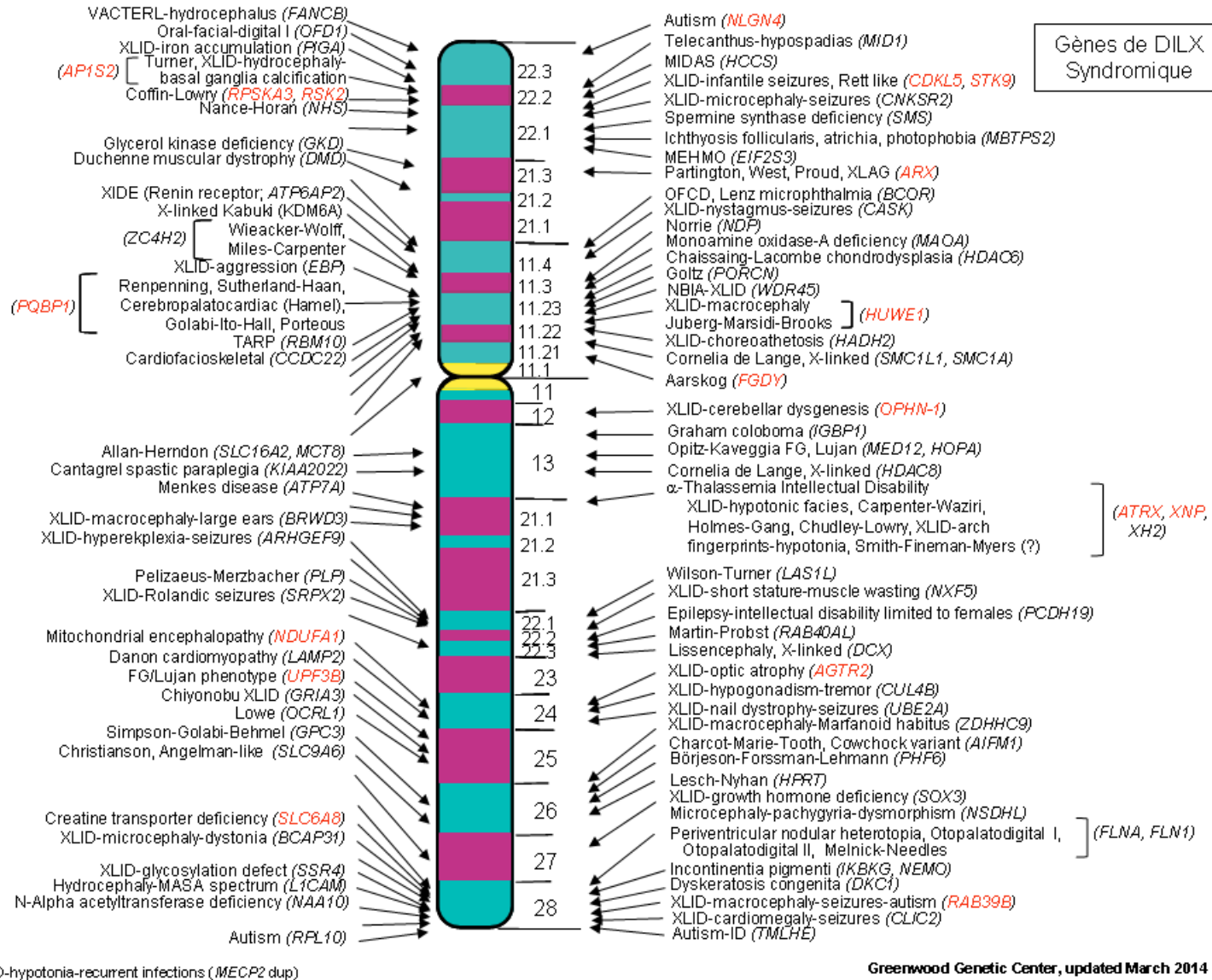
En 2010, l'étude IGOLD (International Genetics of Learning Disability), bénéficiant de l'amélioration des techniques de séquençage, a permis une avancée considérable dans le démantèlement génétique des DILX. En effet, ce premier projet de séquençage à haut débit sur une cohorte de 208 familles de DILX a ciblé les exons codants de 718 gènes connus, soit 1Mb de séquence correspondant à 75% des régions codantes du chromosome X [Raymond et al., 2009, Tarpey et al., 2009, Shoubridge et al., 2010]. Cette étude remarquait pour la première fois les besoins informatiques pour analyser et stocker d'importantes quantités de données mais aussi l'abondance de variants potentiellement pathogéniques découverts avec une approche de séquençage à haut débit soulevant les premières questions de méthodes d'interprétation et d'exclusion de variants candidats.

À ce jour, 102 gènes responsables de désordres liés au chromosome X dans lesquels la DI apparaît comme la caractéristique première ont été décrits. Ils sont rapportés dans les illustrations ci-dessous (figure 1.3 et figure 1.4) d'après le site en ligne <http://www.ggc.org/research/molecular-studies/xlid.html> lequel centralise toutes les informations actualisées sur les DILX.



**Figure 1.3 :** Représentation des gènes de DILX non-syndromique (NS) et leur localisation.

Les gènes notés en noir sont exclusivement associés à des DILX NS alors que ceux notés en rouge ont été associés à des DILX soit syndromiques, soit non-syndromiques.



**Figure 1.4 :** Représentation des gènes de DILX syndromique (S) et leur localisation.

Les gènes notés en noir sont exclusivement associés à des DILX S, ceux notés en rouge ont été associés à des DILX soit S, soit NS.

---

### **Les formes autosomiques dominantes (DIAD) :**

Certains réarrangements chromosomiques ont été des pistes pour identifier des gènes responsables de formes dominantes de DI. En effet, la détection de remaniements chromosomiques déséquilibrés chez des patients avec une pathologie commune permet d'identifier un locus d'intérêt. Les gènes contenus dans la région aneusomique sont autant de candidats intéressants à étudier chez d'autres patients exprimant la même maladie, mais sans déséquilibre génomique. Les gènes responsables des syndromes de CHARGE (MIM 214800) et de Pitt-Hopkins (MIM 602272), respectivement *CHDC7* et *TCF4* ont, par exemple, été identifiés à l'aide de cette approche [Visser et al., 2004, Amiel et al., 2007, Zweier et al., 2007, Berkel et al., 2010, Hoyer et al., 2012]. D'autre part, l'analyse des points de cassures de translocations apparemment équilibrées à phénotype anormal a également permis de mettre en évidence des loci impliqués dans les DI autosomiques dominantes (DIAD) [Daoud et al., 2009]. La cytogénétique est donc la discipline qui a permis l'identification des premiers gènes de DIAD. Néanmoins, devant l'extrême hétérogénéité clinique et vraisemblablement génétique de ces syndromes, peu d'études ont été entreprises pour tenter de déterminer leurs bases génétiques. Le projet canadien « *Synapse to Disease* » (S2D), débuté en 2006, a été le premier à utiliser une stratégie de séquençage systématique de gènes candidats. Il portait sur l'analyse de 500 gènes codant des protéines synaptiques chez une cohorte de 371 patients avec DI NS, autisme ou schizophrénie et 190 individus contrôles et a permis l'identification de mutations délétères *de novo* touchant des gènes codant des protéines du système glutamatergique (*SYNGAP1*, *STXBP1*, *SHANK3*, *KIF1A*, *GRIN1*, *CACNG2* et *EPB41L1*) [Hamdan et al., 2009]. Il faudra ensuite attendre la révolution NGS (New Generation Sequencing) pour voir se multiplier le nombre de nouveaux gènes de DIAD découverts.

### **Les formes autosomiques récessives (DIAR) :**

C'est en 1986 que le tout premier gène de DI fut décrit, le gène *PAH* dont les mutations sont responsables d'une forme de DIAR ; la phénylcétonurie (PCU). La PCU est due au déficit d'une enzyme hépatique : la phénylalanine-hydroxylase (PAH), impliquée dans le métabolisme de la phénylalanine [DiLella et al., 1986].

Toutefois, si l'identification des gènes responsables de DIAR a connu un essor considérable dans les années 1990-2000, l'extrême hétérogénéité génétique et clinique des formes autosomiques récessives et la taille souvent très restreinte des fratries ont, en revanche, longtemps freiné l'identification des gènes associés, pourtant considérées comme plus fréquentes que les DIAR ou DIAD [Ropers, 2007].

En effet, les études de liaisons classiques, basées sur le regroupement de nombreuses familles, sont performantes dans le cas de pathologies cliniquement homogènes, et avec des bases moléculaires peu hétérogènes. C'est ainsi que le gène *PMM2* dont les mutations sont responsables d'une anomalie congénitale de la glycosylation, la forme CDG-Ia (pour Congenital Disorders of Glycosylation) a pu être identifiée [Bjursell et al., 1998], dans une région préalablement délimitée par une étude de liaison réalisée sur 25 familles avec CDG-I [Martinsson et al., 1994].

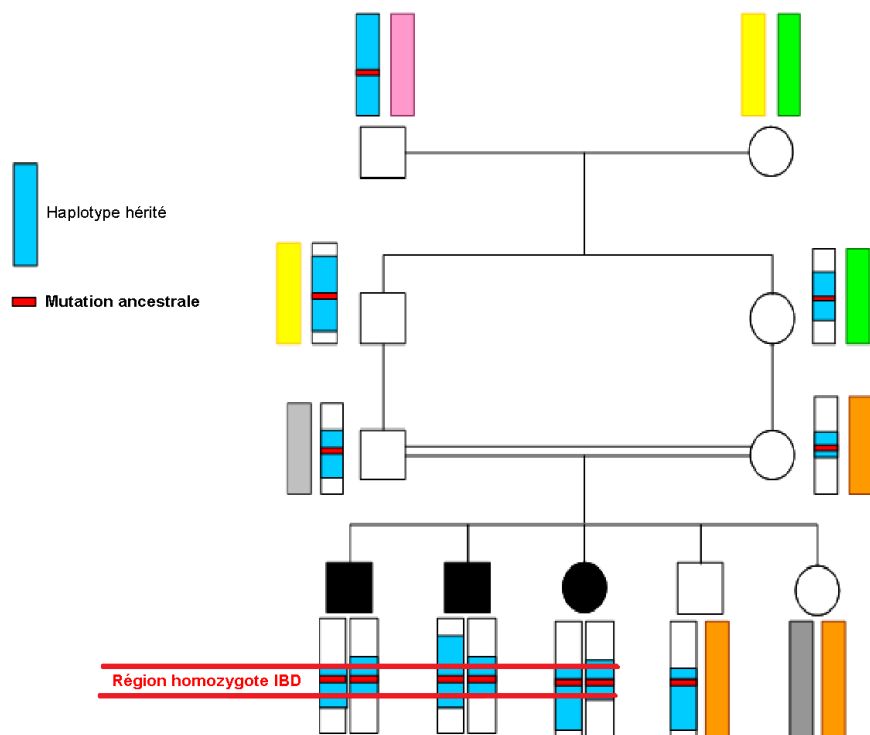
Pour s'affranchir de ces limites, plusieurs méthodes ont été développées.

L'approche « gène candidat » a été utilisée lorsque des signes caractéristiques tels que des anomalies biochimiques ou métaboliques permettent d'orienter la stratégie de séquençage. Cela a été le cas pour la découverte des mutations du gène *DHRC7*, codant la dernière enzyme de la voie de biosynthèse du cholestérol, responsables du syndrome Smith-Lemli-Opitz ou SLO (MIM270400). Ces patients, souffrant de DI de sévérité variable, présentent également un taux faible de cholestérol accompagné d'une accumulation de son précurseur [Irons et al., 1993, Fitzky et al., 1998, Waterham et al., 1998, Wassif et al., 1998].



Une approche de complémentation fonctionnelle a aussi fait ses preuves lorsqu'il existe un phénotype cellulaire associé à une pathologie. Par exemple, l'hypersensibilité aux rayons ultraviolets (UV) des fibroblastes de peau des patients atteints du syndrome de Cockayne (MIM 2164000, CS) a permis la découverte des bases moléculaire associées [Troelstra et al., 1992, de Lonlay, 2003].

Dans tous les autres cas, une stratégie de choix est la méthode de cartographie par autozygotie (ou homozygotie par filiation). Cette technique, décrite en 1987 par Lander and Botstein [1987], consiste à étudier des familles consanguines multiplex afin de localiser des gènes responsables de maladies autosomiques récessives rares. La consanguinité augmente de manière significative le risque de transmission de deux allèles morbides à la génération suivante : on passe en effet d'un risque de donner naissance à un enfant avec DI de 0,012 pour une union aléatoire à 0,062 pour une union de cousins germains [Bashi, 1977, Morton, 1978]. Dans ces familles, la situation la plus fréquente est la suivante : la mutation, survenue sur un des allèles chez l'ancêtre commun, est portée par chacun des deux parents, à l'état hétérozygote. Les enfants atteints héritent de deux copies identiques de l'allèle muté. Les sujets malades sont homozygotes par descendance pour l'allèle porteur de la maladie mais également pour le fragment de chromosome flanquant le locus morbide, comme illustré dans la figure 1.5.



**Figure 1.5 :** Principe de la méthode de cartographie par autozygotie : recherche de mutations homozygotes.

Dans les arbres généalogiques comme celui-ci, les symboles pleins correspondent aux individus malades. Les hommes sont représentés par un carré et les femmes par un cercle. Les doubles barres représentent les unions consanguines et les losanges correspondent aux enfants à naître, de sexe inconnu.

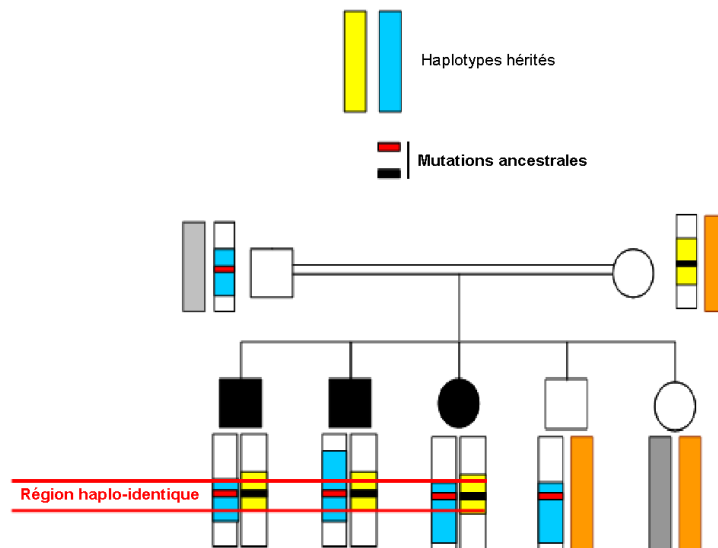
On nomme communément ces fragments, les régions IBD (pour Identical By Descent). D'autres régions sont également homozygotes par descendance, mais sont différentes d'un individu atteint à l'autre. La

méthode consiste donc à identifier des régions homozygotes communes chez l'ensemble des individus malades, hétérozygotes chez leurs deux parents et génodifférentes chez les frères et sœurs sains. On considère alors qu'il y a une probabilité de liaison génétique entre l'haplotype commun aux enfants atteints et la maladie. L'analyse de liaison est quantifiée grâce au calcul du lod (logarithm of an odds ratio)-score [Morton, 1955]. Il correspond au logarithme du rapport de vraisemblance entre deux hypothèses ; l'hypothèse de liaison génétique entre le marqueur (position connue) et le gène de la maladie (position inconnue) pour une distance génétique donnée, représentée par le taux de recombinaison  $\theta$  ( $0 \leq \theta \leq 0.5$ ) et celle d'indépendance génétique entre ce marqueur et le gène recherché ( $\theta=0.5$ ) :

$$Z(\theta) = \log_{10}[L(\theta)/L(0.5)]$$

Les régions identifiées avec un lod-score significatif, constituent des régions candidates dans lesquelles des anomalies génétiques sont alors recherchées. Cette méthode est d'application simple depuis les efforts déployés pour réaliser une carte génétique du génome humain à l'aide de marqueurs polymorphes de type microsatellites (séquences répétées en tandem <10pb) et single nucléotide polymorphism (SNP). Deux paramètres importants affectent l'efficacité de cette méthode, le nombre total d'enfants dans la famille et le degré de consanguinité : plus celui-ci est éloigné, plus le nombre de recombinaisons augmente et plus la région suspecte est réduite. Enfin, sa puissance est bien supérieure à celle de la méthode d'analyse de liaison classique qui nécessite un grand nombre d'enfants atteints.

Dans le cas de familles multiplex avec ou sans consanguinité, il est également possible d'analyser les données du génotype à la recherche de régions haplo-identiques chez tous les individus atteints (figure 1.6). L'hypothèse est alors la présence de deux mutations hétérozygotes composites, chaque mutation étant portée par un ancêtre commun à l'état hétérozygote. Les deux mutations auraient ségrégré au cours des générations pour enfin se retrouver chez la descendance à l'état hétérozygote composite, comme illustré dans la figure 1.6.



**Figure 1.6** : Principe de la méthode de cartographie par autozygotie : recherche de mutations hétérozygotes composites.

Les régions candidates définies par ces techniques comprennent un nombre très variable de gènes i.e de

quelques dizaines à plusieurs centaines selon la richesse génétique de la région d'intérêt. Jusqu'en 2009, une approche gène candidat était appliquée une fois ces régions révélées. Parmi l'ensemble des gènes connus, présents dans l'intervalle critique, seuls les gènes dont le dysfonctionnement pourrait rendre compte du phénotype observé étaient alors séquencés pour rechercher la présence d'éventuelles mutations. La fonction des produits de ces gènes, leur expression et leur éventuelle implication dans des pathologies étaient considérées. C'est ainsi que *PRSS12* a été le premier gène de DIAR NS identifié en 2002 dans l'équipe du Dr Laurence Colleaux : au moyen de l'application de la méthode de cartographie par autozygotie, la même délétion de 4 pb était observée chez les patients de 3 familles algériennes [Molinari et al., 2002]. Une dizaine de gènes ont ensuite été identifiés par la même stratégie (table 1.4)[Raymond and Tarpey, 2006, Basel-Vanagaite, 2007, Ropers, 2010, Ellison and Rosenfeld, 2013].

Gènes découverts par la méthode de cartographie par autozygotie	Position chromosomique	Taille	Référence(s)
<i>PRSS12</i>	4q26	72,7kb	[Molinari et al., 2002]
<i>CRBN</i>	3p26.2	30kb	[Higgins et al., 2004]
<i>CC2D1A</i>	19p13.12	24,7kb	[Basel-Vanagaite et al., 2006]
<i>GRIK2</i>	6q16.3	671kb	[Motazacker et al., 2007]
<i>TUSC3</i>	8p22	23,1kb	[Molinari et al., 2008, Garshasbi et al., 2008]
<i>TRAPPC9</i>	8q24.3	728,2kb	[Philippe et al., 2009, Mochida et al., 2009, Mir et al., 2009]
<i>SOBP</i>	6q21	171,2kb	[Birk et al., 2010]
<i>MED23</i>	6q23.2	54,3kb	[Hashimoto et al., 2011]
<i>ZC3H14</i>	14q31.3	50,6kb	[Pak et al., 2011]

**Table 1.4 :** Gènes de DIAR découverts par la méthode de cartographie par autozygotie.

### iii) Mode de transmission non mendélien

Parmi les modes de transmission non-mendéliens, on peut citer les pathologies liées à un défaut dans le génome mitochondrial. Chez les mammifères, la cellule-œuf reçoit uniquement les mitochondries issues du gamète femelle. Par conséquent, une anomalie génétique d'origine mitochondriale est transmise uniquement par les femmes, à toute sa descendance, quel que soit le sexe. On peut citer comme exemple l'atrophie optique de Leber, une maladie mitochondriale neurodégénérative touchant principalement le nerf optique ([www.mitomap.org](http://www.mitomap.org)). À ce jour, aucune DI de transmission mitochondriale n'a été rapportée dans la littérature. Toutefois, ce mode de transmission est parfois suspecté lorsque d'autres hypothèses n'ont pas permis de mettre en évidence une cause génétique. C'est le cas pour le syndrome Ohdo-Madokoro-Sonoda. Il s'agit d'un syndrome de malformations congénitales multiples incluant une DI.

Un autre mode de transmission non-mendélien connu englobe les anomalies d'empreinte parentale. L'empreinte parentale correspond à une marque épigénétique de l'ADN, déposée durant la spermatogénèse ou l'ovogénèse. Celle-ci provoque une inactivation sélective de quelques gènes dans le génome diploïde issu de

la fécondation. Ces gènes, bien qu'à l'état diploïde, ont une expression monoallélique, dépendante de l'origine parentale des chromosomes. Pour les gènes soumis à l'empreinte (inactivation) paternelle, seul l'allèle maternel est exprimé. À l'inverse, seul l'allèle paternel est exprimé pour les gènes soumis à l'empreinte (inactivation) maternelle. Un gène peut être soumis à empreinte seulement dans un tissu particulier ou à un moment particulier du développement. À ce jour, plus de 100 gènes soumis à empreinte ont été caractérisés, mais on estime à environ 1% la fraction des gènes humains présentant une expression différentielle des 2 allèles. Ces gènes sont le plus souvent regroupés dans des domaines chromatiniques et placés sous le contrôle d'un centre d'inactivation [Franklin and Mansuy, 2011].

Les caractéristiques moléculaires de ce phénomène de marquage épigénétique sont connues et permettent d'expliquer certaines maladies humaines liées à des gènes soumis à empreinte [Chamberlain and Lalande, 2010, Lim and Maher, 2009]. Lorsqu'une anomalie génétique est localisée dans une région chromosomique soumise à empreinte, l'expression de la maladie est fonction de la transmission paternelle ou maternelle de celle-ci.

La perte de l'allèle actif provoque l'apparition de ces maladies par une délétion ou une DUP. En effet, dans les cas de disomie uniparentale, bien que la forme diploïde de l'allèle soit conservée, les gènes soumis à empreinte sont dérégulés dans des types cellulaires et à des stades de développement critiques. Les maladies liées à l'empreinte parentale peuvent aussi être la conséquence d'une mutation du centre d'inactivation, ou de l'expression anormale de l'allèle normalement silencieux (« relaxation d'empreinte »).

Les syndromes de Prader-Willi (ou PWS, MIM :176270) et d'Angelman (ou AS, MIM :105830), décrits dans la table 1.5 ci-dessous, sont tous deux dus à des anomalies génétiques d'une même région soumise à empreinte du chromosome 15 mais rendent compte de deux symptomatologies différentes et reflètent parfaitement ces anomalies d'empreinte. Le PWS est majoritairement causé par une délétion du locus 15q11-q13 d'origine paternelle, tandis qu'une délétion de ce locus d'origine maternelle expliquent la majorité des cas atteints de AS [Kishino et al., 1997, McAllister et al., 2011].

	Syndrome d'Angelman	Syndrome de Prader-Willi
Clinique	DI sévère, absence de langage, microcéphalie, hypoplasie maxillaire, ataxie cérébelleuse parfois accompagnée d'une épilepsie, EEG caractéristique, comportement joyeux, accès de rires inappropriés avec battements des mains, hyperactivité sans agressivité, attention réduite, hyperexcitabilité, troubles du sommeil	Hypotonie néonatale sévère, hyperphagie avec risque d'obésité morbide, dysmorphie faciale, hypogonadisme (retard de puberté, infertilité), difficultés d'apprentissage, troubles du comportement, petite taille
Génétique	75% des cas : délétion du locus 15q11-q13 d'origine maternelle 7% des cas : disomie uniparentale paternelle du chromosome 15 2-4% des cas : défaut d'empreinte parentale 15% des cas : mutation du gène <i>UBE3A</i> (locus 15q11.2)	70% des cas : délétion du locus 15q11-q13 d'origine paternelle 20-30% des cas : disomie uniparentale maternelle du chromosome 15 1-3% des cas : défaut d'empreinte parentale
Références	[Kishino et al., 1997, Mabb et al., 2011]	[McAllister et al., 2011, Cassidy et al., 2012]

**Table 1.5 :** Les syndromes associés au locus 15q11-q13.

## CHAPITRE 2

### L'APPORT DES NOUVELLES TECHNOLOGIES

#### 2.1 La révolution du séquençage à très haut débit ou Next-Generation Sequencing (NGS) : un bref historique

Jusqu'aux années 1930, le génome était un concept abstrait, puisqu'on n'en connaissait pas le support physique. Puis Morgan démontra que les chromosomes étaient le support de l'hérédité mendélienne découverte en 1900. C'est en 1953 que Watson et Crick découvrirent la structure de l'ADN [Watson and Crick, 1953]. Quelques années plus tard, le dogme central de la biologie moléculaire fut énoncé [Crick, 1970].

En 1977, deux méthodes de séquençage de l'ADN étaient élaborées indépendamment : la méthode des di-déoxynucléotides terminateurs de Sanger et al. [1977] s'appuyant sur une synthèse enzymatique sélective et celle de Maxam and Gilbert [1977] basée sur une dégradation chimique de l'ADN et des coupures sélectives en fonction des différentes réactivités des quatre bases A, T, G et C.

Vint alors une longue phase qu'on appellera ensuite la période Sanger de 1977 à 2006, pendant laquelle la méthode Sanger, largement adoptée, est passée du radioactif au fluorescent et de la lecture visuelle des gels d'acrylamide aux séquenceurs automatiques permettant un gain considérable de temps et de productivité.

En 1996, les premiers génomes d'organismes vivants (*Haemophilus*, Levure) étaient publiés [Fleischmann et al., 1995, Jones, 1995]. Le projet de séquençage du génome humain (*The Human Genome Project*) débutait en 1990, avec la formation du consortium nommé « *the International Human Genome Sequencing consortium* ». Une première version de ce génome, incomplète, était publiée en 2001. Puis, dans la publication suivante en 2004, 99% du génome eucaryotique, c'est-à-dire sans les séquences télomériques et centromériques, était couvert [Consortium, 2004].

Très récemment, la révolution génomique a conduit au développement de nouveaux outils de séquençage à très haut débit, appelés de « nouvelle génération » ou « Next-Generation Sequencing » (NGS).

Jusqu'en 2006, les applications du séquençage de type Sanger dans la recherche des causes monogéniques des maladies rares restaient limitées par les frais et le rendement, même après sélection de régions candidates par cartographie par autozygotie. Une nouvelle génération de séquenceurs à haut débit permet désormais de s'affranchir de ces limites et de séquencer plusieurs gigabases d'ADN en quelques jours et à un coût très inférieur. Nous sommes donc passés d'une approche Sanger permettant d'obtenir de longs fragments

---

(jusqu'à 1 000 pb pour les appareils les plus performants) mais peu de lectures et une unique cible testée, à une approche NGS qui génère des milliers de fragments courts (entre 100 et 500 pb suivant les technologies) couvrant un génome intégralement. C'est notamment grâce à la lecture de plusieurs millions de séquences en parallèle que ces nouveaux séquenceurs ont pu révolutionner les analyses en génomique. Globalement, la technologie de ces appareils de séquençage à haut débit est assez proche et fonctionne en plusieurs étapes : la préparation et l'amplification des molécules d'ADN à analyser, leur fixation ou leur répartition sur un support solide, l'incorporation des bases complémentaires des brins à séquencer couplées à un fluorochrome ou autre groupement chemiluminescent, et enfin la lecture (*base calling*) et la détermination des séquences par alignement (*mapping*). Trois technologies sont apparues de façon quasi simultanée : le pyroséquençage, le séquençage avec des terminateurs réversibles et le séquençage par ligation ; mais toutes sont basées sur un système de détection numérique capable de discriminer les échantillons selon leurs positions sur le support de séquençage. Les principales caractéristiques de chacune de ces méthodes sont illustrées en annexe (figure A.2) tandis que les méthodes utilisées pour les travaux présentés dans ce manuscrit sont détaillées dans le chapitre 15 et en annexe (figure A.3, figure A.4 et figure A.5).

Ainsi, en 2001, la publication de la séquence du génome humain était le fruit de 10 ans de travail par 20 institutions. En 2014, le re-séquençage du génome humain peut être réalisé en deux semaines et coûte environ 1 000 euros.

## 2.2 Les méthodes d'enrichissement spécifique

Parallèlement au développement de ces technologies, des méthodes d'enrichissement spécifique, permettant de cibler des régions d'intérêt, ont été mises au point : la méthode de « capture de séquences » par hybridation était la première décrite.

Si les régions régulatrices de l'ADN, les ARN non codants et autres éléments régulateurs suscitent aujourd'hui beaucoup d'intérêt, il n'en reste pas moins que la majorité des mutations impliquées dans les maladies génétiques mendéliennes a été trouvée dans les régions codantes. L'exome, correspondant à l'ensemble des exons codants du génome, représente 1 à 2% de notre ADN (environ 50 Mb) et semble donc particulièrement intéressant à analyser pour identifier de nouveaux gènes impliqués dans des maladies génétiques. La même année, Hodges et al. [2007], Okou et al. [2007] et Albert et al. [2007] décrivent une technique d'enrichissement de l'exome impliquant une capture par hybridation spécifique, une purification des exons ainsi capturés et le séquençage de ceux-ci. Gnirke et al. [2009] publie la première description de la méthode que nous avons utilisée pour la réalisation des travaux présentés dans ce manuscrit, aujourd'hui commercialisée par Agilent (figure A.6).

La technologie RainDance a ensuite été décrite [Tewhey et al., 2009] : elle repose sur un système très contrôlé générant des PCR en émulsion : jusqu'à 20 000 micro-réactions dans un seul tube (figure A.7). Puis, Turner et al. [2009] proposent un protocole optimisé avec l'utilisation des sondes inversées (ou MIPs pour Molecular Inversion Probes), permettant la capture des séquences d'intérêt (figure A.8). Les 3 méthodes d'enrichissements spécifiques (capture par hybridation spécifique, MIPs, ou PCR ciblées), schématisées en annexe (figure A.9) ont par la suite beaucoup évolué permettant d'optimiser la qualité du séquençage tout en réduisant les erreurs, le coût et la durée.

En ce qui concerne les familles dans lesquelles ségrège clairement une pathologie liée au chromosome X, les

---

séquences codantes spécifiques de ce chromosome présentent évidemment un intérêt singulier. Ainsi, des kits de capture spécifiques des séquences codantes du chromosome X ont été élaborés.

Enfin, il est désormais possible de procéder à un séquençage ciblé, réalisé au moyen d'un kit dont les régions capturées ont été dessinées selon les intérêts d'un projet particulier.

L'enchaînement des méthodes d'enrichissement spécifique et de séquençage haut débit rend donc possible le séquençage massif de tout ou partie d'un génome afin d'en identifier les variations (figure A.10).

## 2.3 Démembrement génétique des DI par NGS, mise en place d'une approche innovante

### 2.3.1 Preuve de concept

La révolution technologique du séquençage haut débit laissait entrevoir d'extraordinaires possibilités pour avancer dans l'élucidation des bases moléculaires des formes monogéniques de DI.

En utilisant le syndrome de Freeman-Sheldon (SFS, MIM 193700) comme preuve de concept, une pathologie rare suivant un mode de transmission autosomique dominant due aux mutations du gène *MYH3*, Ng et al. [2009] démontrent pour la première fois que des mutations responsables de maladies monogéniques peuvent être identifiées directement par la méthode de *Whole Exome Sequencing* ou WES sur un nombre modeste de patients non apparentés, sans analyse de liaison ni étude de gènes candidats. Après avoir séquencé l'exome de 4 patients avec SFS, ainsi que 8 individus sains, les auteurs ont effectivement pu analyser les données en ôtant les variants fréquents présents chez les individus sains et ont ainsi retrouvé le gène *MYH3*, qui avait été identifié auparavant par une approche « gènes candidats » [Toydemir et al., 2006]. Dès lors, les auteurs avançaient que l'application de cette stratégie à l'identification de variants responsables de pathologies récessives donnerait probablement des résultats de façon plus immédiate. En effet, chaque individu porte des dizaines de milliers de variants codants dont 5-10% sont absents des bases de données publiques telles que dbSNP, ou sont présents sans précision concernant la fréquence allélique [Sherry et al., 2001]. Cependant, parmi les variants rares non synonymes, on trouve considérablement moins de variants homozygotes ou hétérozygotes composites que de variants hétérozygotes simples réduisant ainsi le nombre de variants candidats lors des études de maladies récessives. La même année, Jones et al. [2009] identifiaient les mutations du gène *PALB2* comme facteur de susceptibilité d'une forme héréditaire de cancer du pancréas par comparaison des données de WES entre tissus sains et tumoraux d'un même sujet, tandis qu'un autre groupe démontrait l'utilité de cet outil en diagnostic moléculaire [Choi et al., 2009].

La première application du WES au démembrement des maladies génétiques a été publiée l'année suivante au travers d'une étude regroupant 4 patients atteints du syndrome de Miller (MIM 263750), une condition autosomique récessive. Le gène *DHODH* est alors le premier gène dont les anomalies furent découvertes par cette approche révolutionnaire [Ng et al., 2010b].

Au cours de l'année 2010, plusieurs groupes publiaient de nouveaux gènes découverts par séquençage de l'exome par l'étude de quelques individus non apparentés [Lalonde et al., 2010, Ng et al., 2010b,a, Hoischen et al., 2010, Gilissen et al., 2010] ou bien de plusieurs sujets atteints d'une même fratrie [Bilgüvar et al., 2010, Bolze et al., 2010, Walsh et al., 2010, Johnson et al., 2010, Musunuru et al., 2010, Krawitz et al., 2010, Wang et al., 2010]. Depuis 2010, l'approche gène candidat est de plus en plus délaissée au profit de

---

l'analyse du WES. Aujourd'hui, l'intérêt d'une approche combinée d'enrichissement par hybridation et de WES pour identifier des mutations rares dans des pathologies humaines a été largement démontré avec plus de 150 nouveaux gènes de syndromes mendéliens décrits en 2013 [Bamshad et al., 2011, Zhang, 2014].

### 2.3.2 Application aux DI

#### i) DILX

Aujourd'hui, le séquençage massif des gènes du chromosome X chez des familles où une DI ségrège clairement selon un mode de transmission lié à l'X a largement contribué à l'amélioration des connaissances de ces formes de DI et est toujours considéré comme une approche de choix [Mondal et al., 2011, Tsurusaki et al., 2011].

#### ii) DIAD

Le WES a notamment permis la recherche active de mutations *de novo*. En effet, ces dernières années ont vu le développement de plusieurs projets très ambitieux visant à séquencer l'exome de dizaines de patients atteints d'une DI et leurs parents sains (trios) afin d'identifier l'ensemble des mutations survenues *de novo* chez ces patients [Visser et al., 2010, de Ligt et al., 2012, Rauch et al., 2012]. Le WES s'est avéré un outil très puissant pour l'identification des gènes impliqués dans les DIAD.

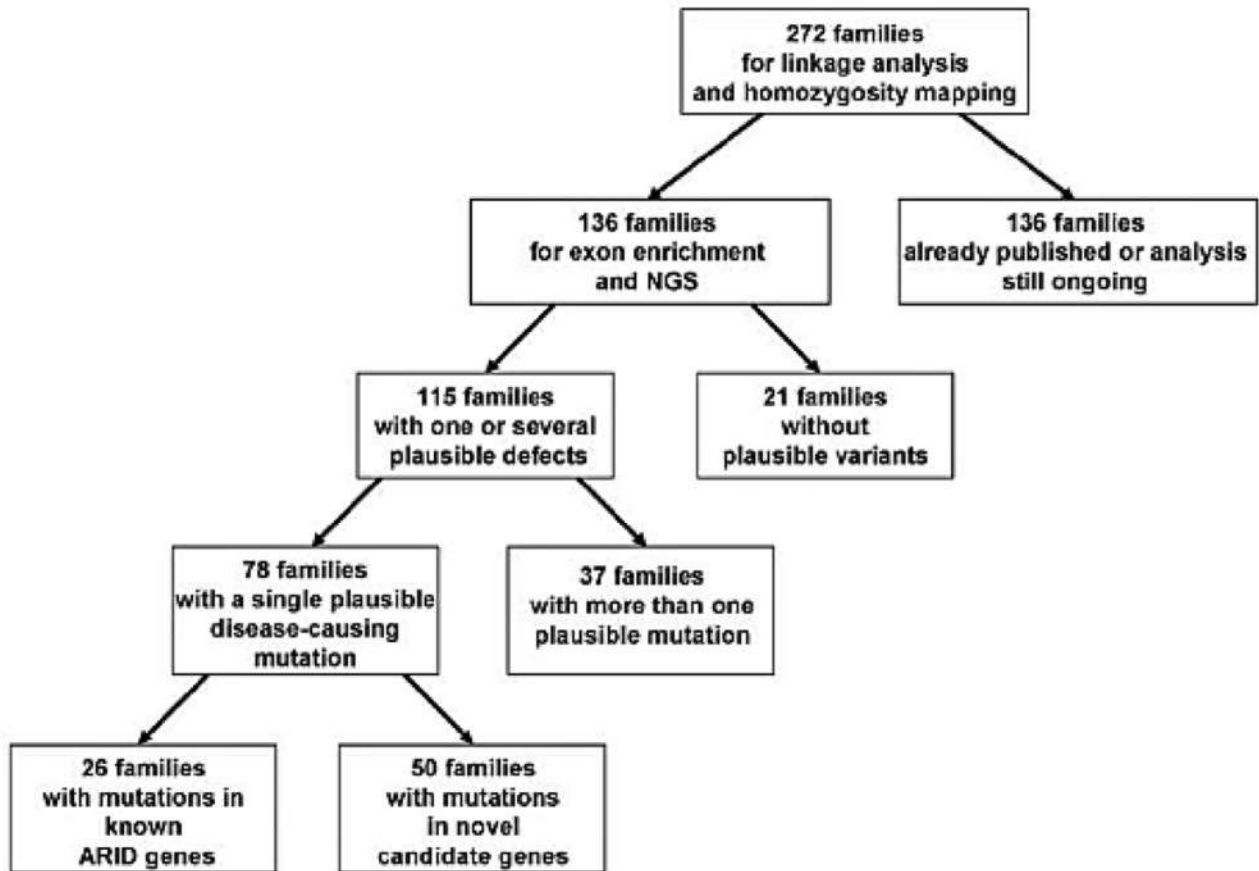
#### iii) DIAR

Enfin, dans le cas de l'étude de familles multiplex et consanguines, plusieurs publications récentes ont démontré l'intérêt de combiner les analyses de cartographie par autozygotie classique et analyses de séquençage d'exome pour identifier rapidement et efficacement des variants localisés dans des régions d'homozygotie ou d'haplo-identité partagées par l'ensemble des enfants atteints. Le premier gène de DIAR identifié par cette double approche a été publié en 2010, *PIGV*, dont les mutations sont responsables du syndrome d'hyperphosphatasie avec déficit intellectuel [Krawitz et al., 2010].

Une variante de cette méthode a été utilisée dans l'étude irano-allemande de Najmabadi et al. [2011] et illustre parfaitement les avantages du séquençage ciblé. Ce travail a débuté par l'identification des régions homozygotes chez 136 familles consanguines atteintes de DIAR au moyen de la cartographie par autozygotie. Dans un second temps, au lieu de séquencer la totalité de l'exome, un kit de capture couvrant l'ensemble des régions candidates a été développé pour séquencer l'ensemble de ces régions chez le probant de chaque famille. L'analyse a permis d'identifier une variation prédite comme potentiellement pathogène dans un gène permettant d'expliquer la pathologie pour 78 des 136 familles étudiées, comme indiqué dans la figure 2.1. Parmi les gènes identifiés, 50 n'avaient pas été impliqués dans la DIAR auparavant et seulement 2 de ces gènes étaient trouvés mutés dans deux familles indépendantes, soit 52 familles révélant une nouvelle cause génétique de DI (22 DI syndromique et 30 DI non-syndromique). À l'inverse, pour 26 familles, les auteurs ont identifié des variants touchant des gènes précédemment associés à une DI principalement syndromique (dont 3 gènes trouvés chez deux familles indépendantes et 20 gènes trouvés chez une seule famille). Chez 37 familles, plusieurs mutations prédites pathogènes ont été identifiées. Dans de tels cas, en l'absence de la description de cas supplémentaires et de toute analyse fonctionnelle, il est difficile d'évaluer si le phénotype



est dû à l'association de ces variations (polygénisme) ou à l'effet d'un seul des deux variants. Enfin, l'analyse des données n'a pas permis l'identification de variants candidats chez 21 familles.



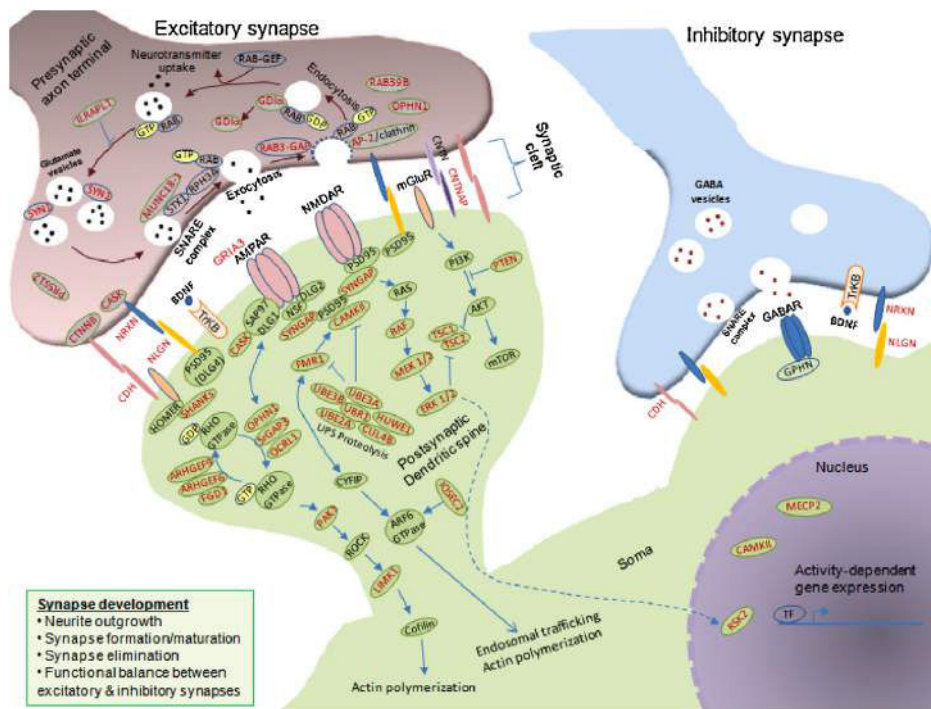
**Figure 2.1** : Résultats obtenus sur l'étude de 136 familles, d'après Najmabadi et al. [2011].

Cependant, lorsque l'hypothèse de transmission génétique n'est pas tranchée, le WES reste l'option la plus intéressante puisque toutes les hypothèses génétiques peuvent alors être testées.

# CHAPITRE 3

## MÉCANISMES PHYSIOPATHOLOGIQUES DES DI, L'ÉTAT DE L'ART

L'étude des effets des mutations des gènes impliqués dans la DI est une étape indispensable pour la compréhension des mécanismes physiopathologiques sous-jacents et le développement de nouvelles approches thérapeutiques. Srivastava and Schwartz [2014] ont récemment publié une revue des mécanismes physiopathologiques impliqués dans la DI. Comme attendu, étant donné l'hétérogénéité des causes génétiques de DI, les fonctions cellulaires associées aux gènes de DI sont multiples. On peut citer par exemple, la modification des protéines, le remodelage de la chromatine et l'assemblage du cytosquelette d'actine. Les auteurs ont mis l'emphase sur l'importance des fonctions cellulaires impliquées dans le contrôle de la structure et des fonctions neuronales, cruciales pour le développement normal des fonctions cognitives (figure 3.1).



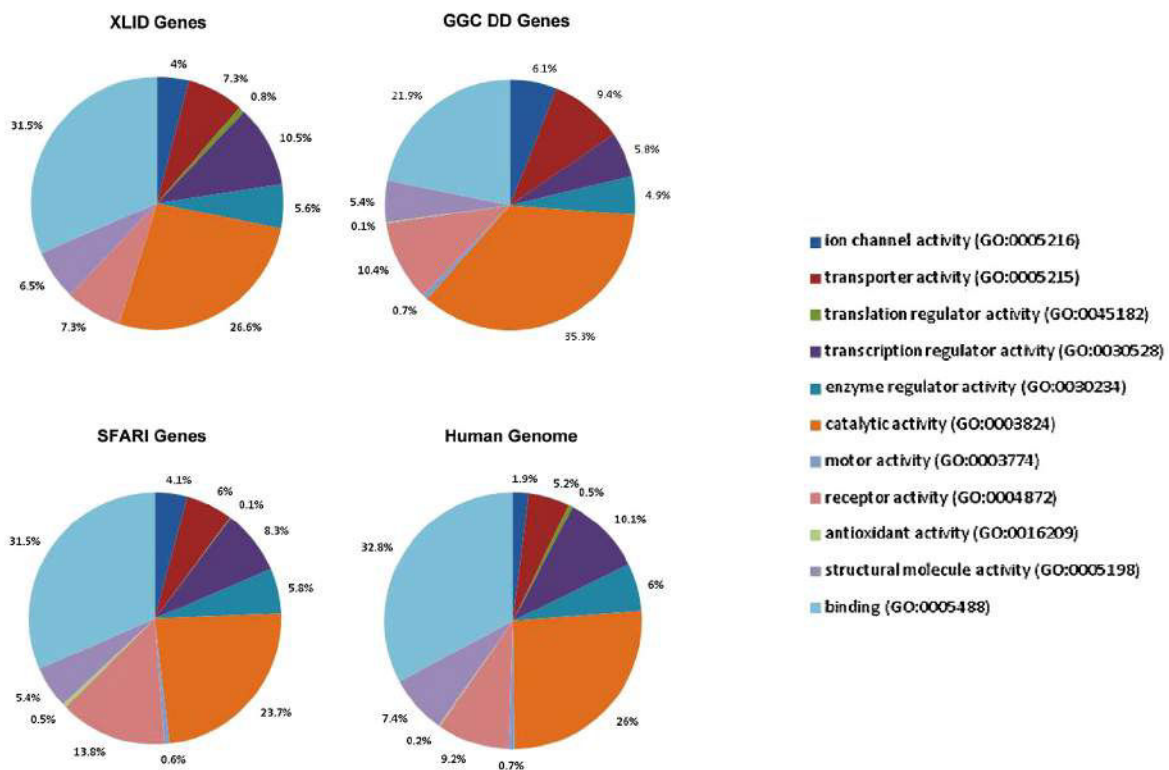
**Figure 3.1 :** Composants synaptiques impliqués dans les conditions neurodéveloppementales, d'après [Srivastava and Schwartz, 2014].

Les produits de gènes impliqués dans la DI et/ou les troubles du spectre autistique sont indiqués en rouge.

En effet, un nombre significatif de protéines liées à la physiopathologie de la DI correspondent à des molécules synaptiques. Elles régulent, entre autre, la transcription, le transport vésiculaire synaptique et le maintien des contacts synaptiques.

Enfin, Srivastava and Schwartz [2014] effectuent une classification de 4 listes de gènes par fonctions moléculaires (figure 3.2) ou processus biologiques (figure 3.3) avec les données de *Gene Ontology* (GO) et l'outil de classification PANTHER (<http://www.pantherdb.org/>) : les 101 gènes de DILX, les 705 gènes associés à un retard de développement (*Greenwood Genetic Center Developmental Delay* ou GGC DD), les 546 gènes de DI et de troubles du spectre autistique (*SFARI* : <https://sfari.org/resources/sfari-gene>) et les 18 331 gènes du génome humain.

De façon inattendue, les auteurs de cette revue observent une distribution similaire des 4 listes dans les diverses catégories de GO et démontrent ainsi que les gènes impliqués dans la DI ne se concentrent pas dans une catégorie de fonctions particulières ; au contraire, leur répartition semble refléter celle du génome humain. Cette observation est sans doute à rapprocher du nombre de gènes exprimés dans le cerveau (cf. p.22) et explique, du moins en partie, l'extrême hétérogénéité génétique de ces affections. Par conséquent, la concentration des recherches de gènes dans une certaine catégorie de GO ne semble pas la méthode la plus adaptée, ni la plus prometteuse.



**Figure 3.2 :** Classification des fonctions moléculaires des gènes des 4 listes, d'après [Srivastava and Schwartz, 2014].

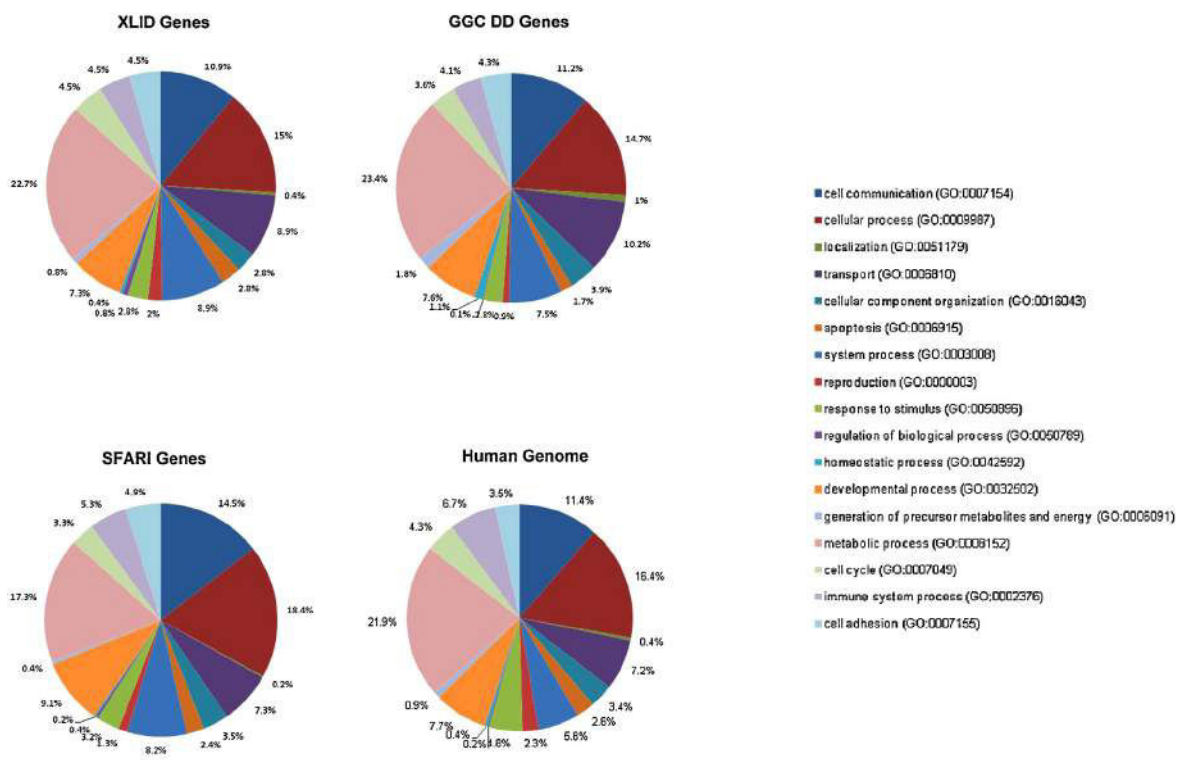


Figure 3.3 : Classification des processus biologiques des gènes des 4 listes, d'après [Srivastava and Schwartz, 2014].

## CHAPITRE 4

### OBJECTIFS DES TRAVAUX

Jusqu'à récemment, toutes les études publiées s'accordaient à montrer que l'étiologie génétique était identifiée dans environ 40% des cas de DI sévère ( $QI < 50$ ) [Rauch et al., 2006]. Depuis 4 ans, avec la révolution technologique apportée par le développement du séquençage à haut débit, l'enrichissement de la connaissance des causes génétiques de la DI s'est considérablement accéléré et il est probable que ce dernier chiffre soit quelque peu sous-estimé. Aussi, si les causes environnementales s'élèvent à plus de 18%, l'étiologie de la DI reste inconnue dans environ 40% des cas, appelées DI cryptogéniques ou idiopathiques. L'ensemble des approches, cartographie, séquençage manuel de gènes candidats par leur(s) fonction(s) et/ou leur localisation et plus particulièrement le séquençage à haut débit, a révélé une remarquable hétérogénéité génétique avec comme fait marquant une fréquence élevée de mutations *de novo* non récidivantes [Hamdan et al., 2011]. Effectivement, début 2013, le nombre de gènes reliés à une DI était évalué à 450 [Ellison and Rosenfeld, 2013], chiffre en constante augmentation. Cette considérable hétérogénéité a pour corollaire une remarquable hétérogénéité physiopathologique [Srivastava and Schwartz, 2014]. L'ensemble de ces études illustre clairement que chacun des gènes ainsi identifié ne rend compte que d'un très faible nombre de cas, ce qui rend complexe la mise en place de leur diagnostic moléculaire. De plus, au-delà de la démonstration de l'extrême hétérogénéité génétique des DI, ces travaux attestent également d'une grande variabilité phénotypique intra- et inter-familiale associées aux mutations de ces gènes. Un même phénotype peut résulter de l'effet des mutations de gènes différents et, inversement, des mutations d'un même gène peut occasionner divers phénotypes. Des gènes connus comme responsables de formes syndromiques de DI peuvent ainsi être également impliqués dans des DI non ou peu spécifiques. Parmi les 40% de DI idiopathiques, on s'attend donc à découvrir de nouveaux gènes de DI mais aussi à trouver des variants pathogènes dans la séquence de gènes déjà caractérisés. Dans ce dernier cas, ceci permettra de mieux caractériser le spectre phénotypique associé aux mutations d'un gène précis.

Mon projet de thèse, dans la continuité des résultats déjà obtenus dans l'équipe du Dr Laurence Colleaux, comportait deux aspects, l'un de génétique formelle, et le second plus fonctionnel. Les objectifs étaient de poursuivre le travail de démembrement génétique des DI, en utilisant de nouveaux outils de séquençage haut-débit chez cinq familles indépendantes dans lesquelles ségrège une forme syndromique de DI, puis d'étudier les conséquences des anomalies moléculaires identifiées afin de mieux comprendre les mécanismes

---

physiopathologiques qui sous-tendent ces affections. Ce projet s'appuie sur une expertise médicale solide, un recrutement exceptionnel de malades et un savoir-faire technologique approprié. Ces travaux sont nécessaires et constituent la première étape dans la compréhension de ces maladies, indispensable avant d'aborder toute approche thérapeutique.

Le premier chapitre de ce manuscrit se concentre sur les formes autosomiques récessives de DI et s'appuie sur le recrutement de plusieurs familles consanguines et multiplex dont les enfants sont atteints d'une DIAR syndromique. Pour chaque famille présentée, la stratégie d'analyse est détaillée puis les résultats sont discutés.

Le second chapitre développe les résultats obtenus par une approche WES chez deux patients non-apparentés, atteints d'une DI syndromique dont les symptômes sont remarquablement proches et justifiant le regroupement de ces deux familles pour une étude simultanée aboutissant à la description d'un nouveau syndrome de DI.

Ces deux chapitres sont suivis d'une dernière partie de discussion et perspectives autour de l'outil de WES, des nouveaux gènes de DI identifiés et des mécanismes physiopathologiques mis en lumière par les analyses fonctionnelles accomplies.

## **Deuxième partie**

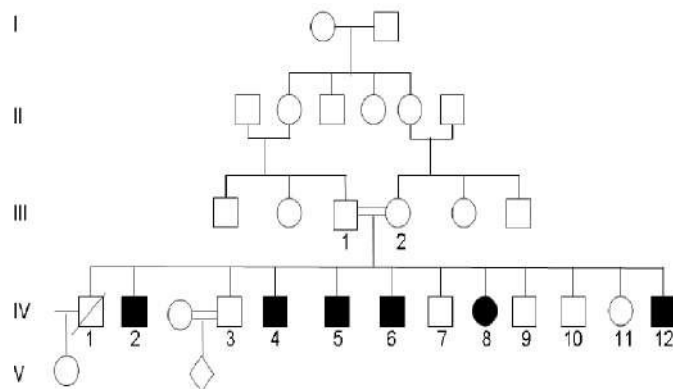
# **Approche Whole Exome Sequencing (WES) pour l'étude de familles consanguines : du monogénique au polygénique**

## CHAPITRE 5

### IDENTIFICATION D'UNE MUTATION DANS UN GÈNE CONNU DE DI -WDR62- ET ILLUSTRATION DE L'INTÉRÊT DU WES

#### 5.1 La famille, les patients

Dans le cadre de l'étude de familles consanguines et multiplex, nous avons eu l'opportunité d'étudier la famille Mo, d'origine algérienne, composée de 6 enfants atteints (une fille et 5 garçons) sur une fratrie de 12 (figure 5.1).



**Figure 5.1 :** Arbre généalogique de la famille Mo.

Elle est suivie par le Dr Abdelkrim Saadi au sein de l'établissement hospitalier de Benaknoun à Alger. Les patients ont une DI syndromique ; ils présentent de manière constante une microcéphalie, une dysmorphie faciale (figure 5.2), un retard psychomoteur et des troubles du comportement.





**Figure 5.2 :** Photographies des patients de la famille Mo.

Les parents de la famille Mo sont indemnes et ont un périmètre crânien (PC) normal (table 5.1).

Mesures à l'âge adulte	Père	Mère
PC (cm)	54	55
T (cm)	168	160
P (kg)	66	70

**Table 5.1 :** Mensurations des parents.

L'aîné (IV.1), né à domicile, ne présente pas d'antécédents pathologiques pré- et post-nataux. Les mensurations à la naissance n'ont pas été prises mais les parents signalent une microcéphalie congénitale. Son développement psychomoteur est retardé ; il a marché à 4 ans, et dit uniquement quelques mots. Aussi, il présente des traits autistiques avec mouvements stéréotypés (jouait avec les lacets pendant le jeune âge) et agressivité. Il a présenté dès l'âge de 6 ans des crises d'épilepsies de l'ordre de 2 à 3 crises par an, jusqu'à l'âge de 20 ans, pour lesquelles il a été traité au phénobarbital. Il ne prend aucun traitement depuis 1986. À l'âge de 45 ans, l'examen objective une microcéphalie avec un PC de 46,5 cm (- 5 DS), une DI avec un langage très réduit, des difficultés à s'habiller, se raser et quelques signes dysmorphiques tels qu'un front raide, un nez proéminent, un palais profond, des yeux enfoncés, une lèvre inférieure éversée et une scoliose. L'examen dermatologique révèle une tache café au lait (TCL) et l'examen neurologique, des réflexes vifs. Le patient IV.4 né à domicile, présentait une microcéphalie congénitale d'après les parents. Les périodes péri- et néonatales se sont déroulées sans incident. Il a marché à 4 ans et ne parle pas. Il a présenté des troubles du comportement : hyperactivité et agressivité avec automutilation. Il frappait sa tête, son visage contre le mur durant l'enfance. Vers l'âge de 18 ans, il a présenté de façon inexplicable un déficit moteur progressif aux 4 membres avec arrêt de la marche vers 30 ans. Il est aujourd'hui grabataire avec troubles de la déglutition.

Le patient IV.5 est né à terme à l'hôpital, après une grossesse normale. Les mensurations à la naissance n'ont pas été prises mais les parents rapportent la notion de microcéphalie. Il a marché à 3 ans, dit quelques mots mais n'a pas de phrases. À l'âge de 45 ans, les mensurations montrent un PC de 48 cm (- 4 DS), une taille de 163 cm et un poids de 65 kg. C'est le moins affecté dans la fratrie ; il est calme, sociable et n'a jamais présenté de crises d'épilepsies. Il peut s'habiller, se raser, exécuter des ordres simples. Il présente un front raide, un nez proéminent, des sourcils épais, un palais profond, un prognathisme de la mâchoire supérieure, un cou large, des mains trapues, un cubitus valgus et des pieds en rotation interne. L'examen neurologique est normal et une TCL est observée à l'examen dermatologique. L'IRM du patient IV.5 est en faveur d'une pachygyrie avec raréfaction des circonvolutions et aspect épaissi de ses dernières en fronto-parieto-temporale de façon bilatérale (figure 5.13).

Pour le patient IV.6, exceptée la microcéphalie signalée par les parents, la grossesse et l'accouchement à l'hôpital se sont déroulées sans incident. Son développement psychomoteur s'est fait avec retard, il a marché à l'âge de 3 ans et pouvait dire quelques mots. Il a aussi présenté des troubles du comportement ; il est craintif, parfois agressif et refuse de se faire examiner. L'examen réalisé à l'âge de 40 ans objective une DI sévère, une microcéphalie (- 4 DS), une petite taille (154 cm), un front raide et court, un nez proéminent, un prognathisme de la mâchoire supérieure, une malposition dentaire, des sourcils épais, une cyphoscoliose et un genu flessum des genoux. Il présente un strabisme divergent de l'œil droit et une TCL.

La patiente IV.8 est âgée de 37 ans. Son histoire prénatale s'est déroulée sans incident, l'accouchement était spontané, sans complications. La microcéphalie congénitale et le petit poids de naissance ont été rapportés par les parents. Un déficit de l'hémicorps gauche a été retrouvé à l'âge de 1 mois. Son développement psychomoteur s'est fait avec retard ; elle a marché à l'âge de 3 ans et ne parle pas. À l'âge de 4 ans elle a présenté des crises épileptiques partielles motrices à gauche, traitées par acide valporique. Récemment, l'examen clinique a montré une microcéphalie (- 4,5 DS), une petite taille (136 cm), un poids de 30 kg et quelques signes dysmorphiques ; front étroit et fuyant, yeux enfoncés, grande bouche ouverte, mauvaise dentition, lèvre inférieure éversée, pieds plats, hallux valgus et hémiparésie spastique gauche à l'examen neurologique.

Le patient IV.12 est âgé de 30 ans. Son histoire prénatale s'est déroulée sans antécédents particuliers, né à terme, à l'hôpital, sans incidents. Les parents rapportent la notion de microcéphalie à la naissance. Il a marché à 2 ans et ne dit aucun mot. Il présente une agitation psychomotrice, une microcéphalie avec un PC à 46 cm (- 5 DS), une taille de 165 cm, un front fuyant, un nez proéminent, des yeux enfoncés, un palais profond, une mauvaise dentition, un rétrognathisme, une TCL et une hyperréflexie.

La table 5.3 ci-dessous résume les caractéristiques cliniques des patients.

Patients	IV.2	IV.4	IV.5	IV.6	IV.8	IV.12
Sexe	M	M	M	M	F	M
Microcéphalie post-natale	+	+	+	+	+	+
Marche indépendante	4 ans	4 ans	3 ans	3 ans	3 ans	2 ans
Âge à la dernière évaluation	44 ans	40 ans	41 ans	40 ans	37 ans	30 ans
DI	sévère	profonde	sévère	sévère	sévère	sévère
Langage	quelques mots	absent	quelques mots	quelques mots	absent	absent

Patients	IV.2	IV.4	IV.5	IV.6	IV.8	IV.12
Épilepsies	crises généralisées	absentes	absentes	absentes	crises partielles motrices	absentes
Examen neurologique	hyperréflexie	quadriparésie spastique d'installation progressive après 18 ans	normal	strabisme	hémiparésie spastique depuis l'âge de un mois	hyperréflexie
Examen dermatologique	TCL	nl	TCL	TCL	nl	TCL
Comportement pendant l'enfance						
Agressivité	+	+	-	+	-	-
Hyperactivité	-	+	-	-	-	+
Stéréotypies	+	-	-	-	-	-
Angoisses	-	-	-	+	-	-
Calme	-	-	-	+	+	-
Sociable	-	-	+	-	-	-
Mesures à l'âge adulte						
PC (cm)	46,5 (-5DS)	47 (-4,5DS)	48 (-4DS)	48 (-4DS)	45 (-5,5DS)	46 (-5DS)
T (cm)	159	nd	163	154	136	165
Pds (kg)	56	nd	65	42	30	46
Dysmorphie						
Front étroit et fuyant	+	+	+	+	+	+
Yeux enfoncés	+	-	-	-	+	+
Nez large et proéminent	+	-	+	+	-	+
Palais ogival	+	-	+	+	-	+
Mauvaise dentition	+	-	-	+	+	+
Lèvre inférieure éversée	+	-	-	-	+	-
Cyphoscoliose dorsale	+	-	-	+	-	-
Prognathisme	-	-	mâchoire supérieure	mâchoire supérieure	-	-
Rétrognathisme	-	-	-	-	-	+
Autres	-	-	cou large, mains et pieds courts et trapus, cubitus valgus, marche avec les pieds en rotation interne	sourcils épais, doigts courts, grosses articulations inter-phalangiennes avec pouces et gros orteils larges, flexum des genoux	grande bouche, hémiparésie, hémiatrophie gauche, rétraction du coude et du poignet, pieds plats, hallux valgus gauche	-
Bilan paraclinique						

Patients	IV.2	IV.4	IV.5	IV.6	IV.8	IV.12
Scanner cérébral (sans injection PC)	nd	nd	nl	nd	nd	nl
IRM	nd	nd	pachygyrie avec raréfaction des circonvolutions et aspect épaissi de ses dernières en fronto- parieto- temporale de façon bilatérale	nd	nd	nd
EEG	nd	nd	Tracé d'interprétation difficile, caractérisé par une activité thêta, plutôt monomorphe, bilatérale et diffuse	nd	Tracé globalement lent, dénué d'anomalies irritatives	Tracé lent pour l'âge, caractérisé par une activité lente type thêta, continue et diffuse
Échographie abdominale	nd	nd	nl	nd	nd	nl
Échographie thyroïdienne	nd	nd	nl	nd	nd	nl
Rx squelette : telethorax, rachis dorsal, lombaires, bassin, mains, et humarus	nl	nd	nl	nd	nl	nl
Marqueurs tumoraux	nd	nd	PSA Totale : 1 (4,5ng/ml), CA19-9 : 1 (0-37 ui/l), CA125 : 12 (0-35 ui/l)	PSA Totale : 1 (0-4,5ng/ml), CA19-6 : 1 (0-37ui/l), CA125 : 16(0-35ui/l)	HCG :1(0- 4ui/l), CA15-3 : 21(0-30ui/l), CA19-9 :1 (0-37ui/l), CA125 : 48 (0-35ui/l)	nd
Dosage, chromato AA et AO	nl	nd	nd	nd	nd	nd
MPS urinaires	nd	nd	-	nd	nd	-
TSH	nd	nd	0,493 (0,27- 4,2mUI/l)	nd	nd	2,18 mUI/l

Patients	IV.2	IV.4	IV.5	IV.6	IV.8	IV.12
Prot /album. sg	nd	nd	nl	nl	nl	nd
Hépatique	nd	nd	nl	nl	nl	nl
Rénal	nd	nd	nl	nl	nl	nl
Cholestérol, TG	nd	nd	nl	nl	nl	nl
Ca, P, Mg	nd	nd	nl	nl	nl	nl

**Figure 5.3 :** Caractéristiques cliniques des patients de la famille Mo.

DI : Déficience Intellectuelle, PC : Périmètre Crânien, T : Taille, Pds : Poids, DS : Déviation Standard, TCL : Tache Café au Lait, TG : Triglycérides, MPS : Mucopolysaccharides, Ca : Calcium, P : Phosphore, Mg : Magnésium, sg : sanguin, nl : normal, nd : non disponible.

## 5.2 Stratégie

L'étude de la famille Mo, d'origine algérienne, a débuté en 2007, avant mon arrivée au sein de l'équipe du Dr Colleaux. L'analyse de cette large famille consanguine a débuté par une approche de cartographie par autozygotie suivie de l'étude de gènes candidats, réalisées l'une et l'autre par Mohammed Zarhrate. À mon arrivée au laboratoire en 2011, une mutation dans un gène candidat avait été découverte. Ma contribution à ce travail commence avec l'étude des effets fonctionnels de cette mutation. Au même moment, nous avons eu l'opportunité de séquencer l'exome d'un unique probant de la famille, nous permettant de confirmer ou de revoir, par une approche de WES, les hypothèses formulées à la suite d'une approche de cartographie par autozygotie et recherche de gènes candidats.

## 5.3 Résultats

### 5.3.1 Cartographie par autozygotie et recherche de gènes candidats

Avant mon arrivée au laboratoire, étant donné qu'une microcéphalie est observée chez les 6 enfants atteints, l'hypothèse d'une liaison génétique à l'un des gènes de microcéphalie connus à cette époque avait été testée par génotypage (*ASPM*, *microcephalin*, *CDK5RAP2*, *CENPJ*).

loci	IV.2	IV.4	IV.5	IV.6	IV.8	IV.12	III.1	III.2
<i>ASPM</i>	152	152/156	156	156	152	nd	152/156	nd
	169	169	169	169	169	nd	169	169
	251	251/253	251/253	251/253	251	nd	251/253	251
<i>microcephalin</i>	168/170	168/170	168/170	170	170	nd	170	168/170
<i>CDK5RAP2</i>	110/116	110/116	110/116	104/112	110/116	nd	112/116	104/110
	203	203	203	203/205	203	nd	203/205	nd
<i>CENPJ</i>	231	237	231/237	231/237	231/237	nd	231/237	231/237
	267	282/284	267/284	267/284	267/282	nd	267/282	267/284

**Table 5.3 :** Génotypage des locis connus de microcéphalie.

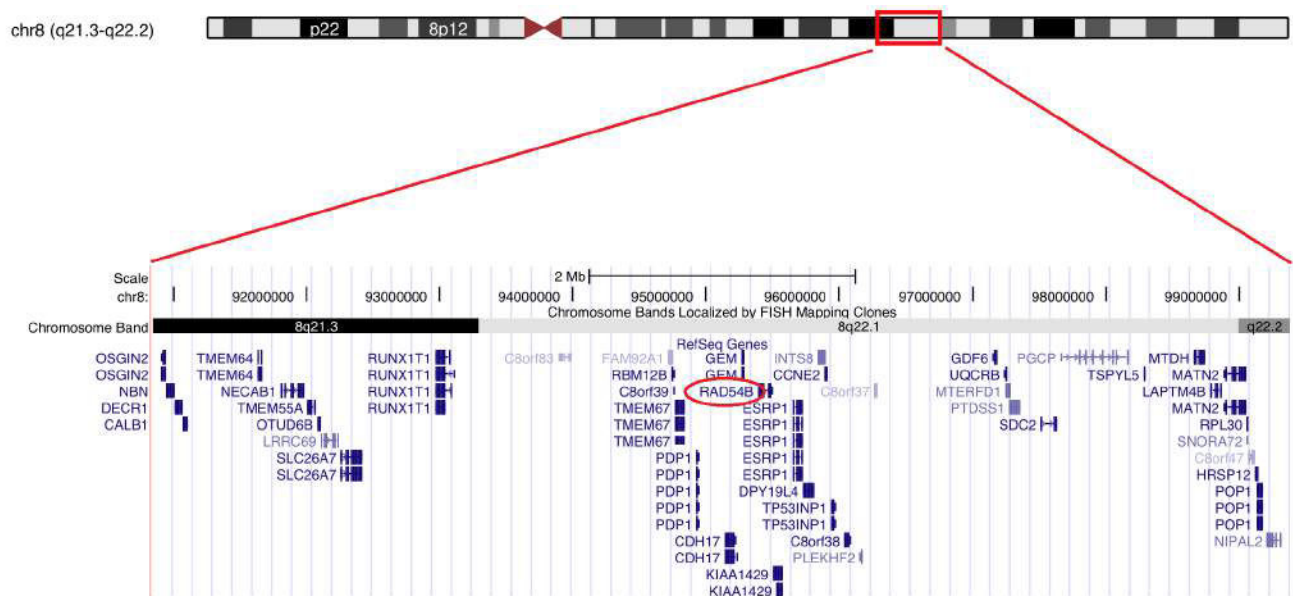
nd : non disponible.

Un tour de génome avait donc été réalisé par génotypage de 10 000 SNPs couvrant la totalité du génome pour 5 des 6 enfants atteints et les deux parents sains (table 16.1). À l'issue de l'analyse des données, 3 régions candidates avaient été retenues avec un lod score supérieur à 2, trouvées homozygotes chez les enfants atteints et hétérozygotes chez les parents sains (table 5.4)

chromosome	position (GRCh37)	position	taille (Mb)	lod score	nombre de gènes
6	109982160-119412952	6q21-q22.31	9,4	> 2,0	53
8	90822759-99282341	8q21.3-q22.2	8,5	2,66	46
19	28424869-42354944	19q11-q13.2	13,9	> 2,0	254

**Table 5.4** : Les régions IBD candidates.

Le génotypage de marqueurs microsatellites localisés dans ces régions candidates pour 5 enfants atteints, 2 enfants sains et les 2 parents sains avait permis d'exclure deux de ces régions. Une région unique d'homozygotie de 8,5 Mb en 8q21.3-q22.2 située entre les marqueurs rs28144 et D8S257 avait été retenue sur le chromosome 8. Le calcul de lod-score donnait une valeur maximale de 2,66. Cette région contenait 42 gènes connus et 4 gènes prédits (figure 5.4).

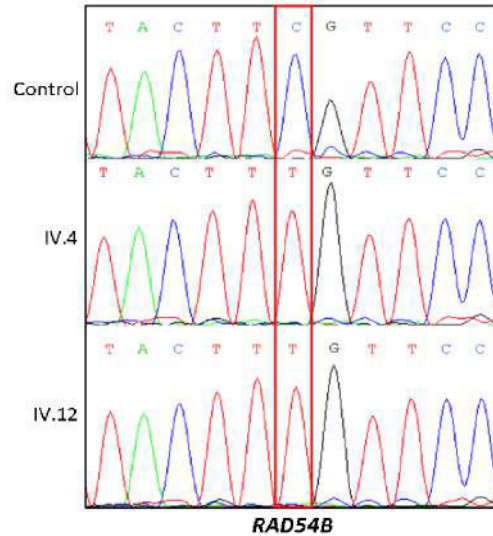


**Figure 5.4** : Carte physique de la région candidate identifiée sur le chromosome 8 (source : UCSC).

Quatre gènes candidats sélectionnés pour leur fonction ou leur profil d'expression avaient d'abord été exclus par séquençage (*RUNX1T1*, *NBN*, *INTS8* et *CCNE2*). De plus, un autre gène, déjà associé à une autre pathologie (*TMEM67* dont les mutations sont responsables d'une forme de syndrome de Joubert, MIM 610688) avait également été exclu. Parmi les autres gènes codants de cette région, *RAD54B*, un gène d'expression ubiquitaire codant une protéine impliquée dans la réparation des lésions de l'ADN et la recombinaison mitotique, semblait être un bon gène candidat. En effet, plusieurs signes cliniques observés dans la famille Mo, comme la microcéphalie congénitale, le retard de développement psychomoteur sévère, le

retard de croissance staturo-pondéral (principalement postnatal) et la dysmorphie faciale, sont régulièrement observés dans d'autres maladies de la réparation [O'Driscoll and Jeggo, 2008].

Le test par séquençage des 9 exons codant de *RAD54B* avait permis l'identification d'une variation dans l'exon 8 : une transition c.1243C>T (figure 5.5).



**Figure 5.5 :** Électrophoregrammes représentant la mutation c.1243C>T du gène *RAD54B* chez deux enfants atteints et un contrôle.

Ce variant co-ségrègeait avec la maladie dans la famille et n'avait pas été retrouvé chez une cohorte contrôle de 602 chromosomes, dont 202 de la même origine ethnique que les patients, écartant ainsi l'hypothèse d'un variant rare. Ce variant entraîne la substitution non-synonyme de l'arginine en position 415 par une cystéine (p.R415C). Cet acide aminé est très conservé chez tous les orthologues de *RAD54B* (figure 5.6), et est localisé 9 acides aminés en amont du domaine hélicase II de la protéine.



**Figure 5.6 :** Alignement des séquences protéiques des orthologues de *RAD54B*.

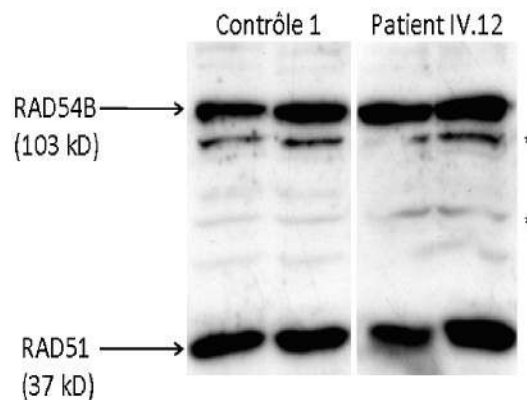
La flèche rouge indique l'arginine 415 altérée chez les patients.

La conservation de l'acide aminé au fil de l'évolution supporte l'idée selon laquelle la mutation affecte la fonction de la protéine. Enfin, divers logiciels de prédiction *in silico* des effets fonctionnels et/ou structuraux de ce variant suggèrent qu'il puisse s'agir d'un variant pathogène (PolyPhen-2 score : 1, SIFT score : 0).

Le gène *RAD54B* est un paralogue du gène *RAD54* des vertébrés et un homologue de la protéine Rdh54/Tid1 de la levure *Saccharomyces cerevisiae* [Hiramoto et al., 1999]. *RAD54B* est une protéine nucléaire. Elle compte 7 domaines hélicases caractéristiques de la famille de protéines SNF2/SWI2 et possède une activité ATPase stimulée par sa liaison à l'ADN double brin. Cette protéine est impliquée dans la réparation des cassures doubles brins (CDB) par recombinaison homologue (figure A.11). Les CDB sont causées par les radiations ionisantes, les pontages inter-brin, le stress oxydatif et les erreurs de la réplication de l'ADN. La réparation des CDB par recombinaison homologue dépend des gènes du groupe *RAD52*, dont *RAD51*, *RAD54* et *RAD54B* font parti. Le profil d'expression du gène *RAD54B* chez l'homme coïncide avec celui de *RAD54* [Tanaka et al., 2000] : l'expression est forte dans la rate et les testicules, des tissus où des événements de recombinaison se déroulent. Le gène s'exprime à un niveau basal dans les autres organes [Hiramoto et al., 1999]. Le variant p.R415C du gène *RAD54B* était donc considéré comme un bon candidat à étudier avec des analyses fonctionnelles afin de mettre en évidence d'éventuels effets de la mutation sur la structure, la stabilité et la fonction de la protéine. Nous avons donc débuté une série d'expériences visant à démontrer l'impact de ce variant sur la fonction de la protéine *RAD54B*.

### 5.3.2 Étude de la stabilité de la protéine *RAD54B* mutée

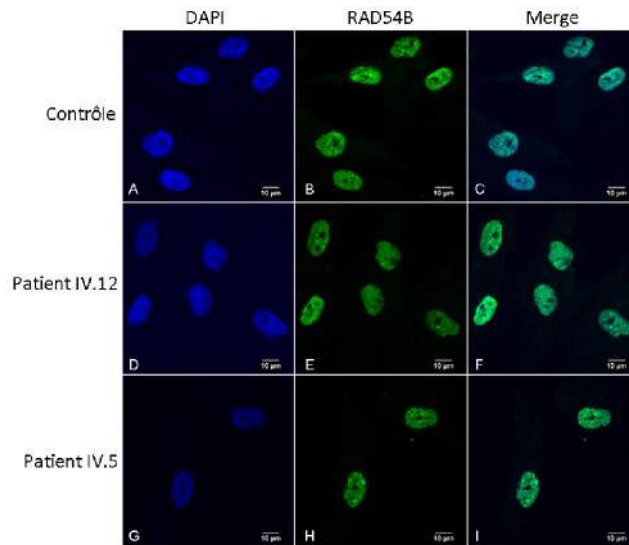
Pour tester l'hypothèse de la pathogénicité du variant p.R415C nous disposons de fibroblastes issus de biopsies cutanées des patients IV.5 et IV.12. Les premiers tests réalisés nous ont permis de montrer que le variant n'affecte ni la stabilité de la protéine (figures 5.7 et 5.9) ni sa localisation cellulaire (figure 5.8).



**Figure 5.7** : Stabilité de la protéine *RAD54B*.

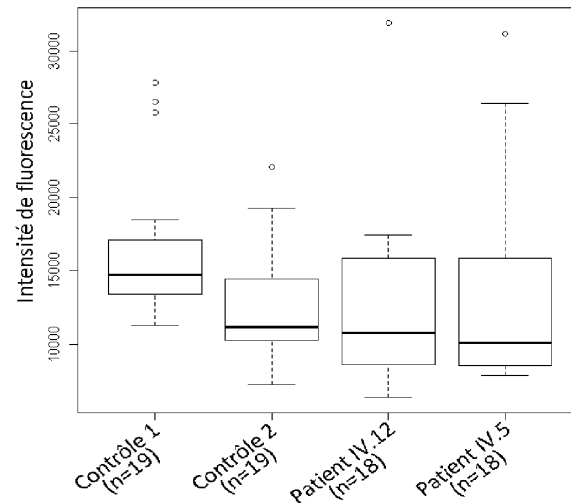
Western blot (WB) : 35  $\mu$ g d'extraits protéiques de patient vs contrôle ont été déposés dans chaque puits. L'intensité du marquage de la  $\beta$ actine étant trop importante, nous avons utilisé la protéine *RAD51* comme témoin de charge. Les deux faibles bandes, marquées d'une astérisque (\*), sont des bandes aspécifiques.





**Figure 5.8** : Localisation cellulaire de la protéine RAD54B.

Immunocytochimie (IC) : le même anticorps a été utilisé. Les acquisitions sont réalisées au microscope confocal LSM700 avec un objectif Plan-Apochromat 63x/1.40 Oil DIC M27 et un logiciel d'acquisition Zen (version 2009). La protéine se retrouve répartie uniformément dans le noyau, à l'exception des zones nucléolaires où elle est absente, chez les patients comme chez les contrôles.



**Figure 5.9** : Intensités de fluorescence du marquage RAD54B.

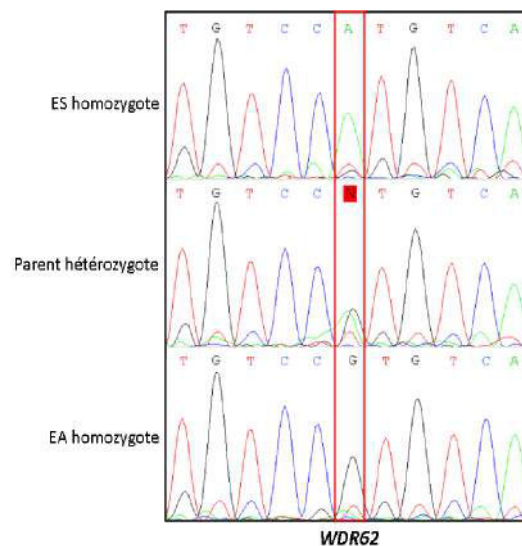
Quantification de fluorescence : les intensités de fluorescence sont mesurées avec imageJ sur les acquisitions illustrées dans la figure 5.8. Les moyennes pour  $n$  noyaux par sujet sont calculées et soumises au test de comparaison des variances (ANOVA) et au test post hoc de Tukey sur le logiciel R. Les niveaux de fluorescence chez les deux patients ne sont pas statistiquement divergents comparés à chacun des contrôles (tableau B.1 en annexe).

### 5.3.3 WES sur un unique probant

Nous nous apprêtons à poursuivre notre travail d'analyse fonctionnelle pour tester les capacités des fibroblastes de patients à réparer les lésions de l'ADN lorsque nous avons eu l'opportunité au printemps 2011 de réaliser un WES pour un des sujets atteints de la famille Mo. Lors de l'analyse des données de WES, nous nous sommes concentrés sur les variants homozygotes, absents des bases de données disponibles (« dbSNP138 » et « the 1 000 Genomes Project »), dont l'effet délétère sur la protéine est le plus probable (non synonymes, affectant des sites d'épissage ou correspondant à des insertions et délétions dans la région codante). Cette analyse nous a permis d'identifier le variant c.1901T>C, p.Met634Thr touchant l'exon 15 du gène *WD-repeat-containing protein 62* (*WDR62*).

L'acide aminé 634, peu conservé, appartient à un domaine « WD40-repeat ». La variation p.M634T est prédite potentiellement pathogène par divers outils de prédictions *in silico* (PolyPhen-2 score : 0.682, SIFT score : 0.04 et « Disease Causing » d'après Mutation Taster). Le gène *WDR62* est localisé sur le chromosome 19 (19q13.12), dans l'une des grandes régions d'homozygotie que nous avons éliminé par l'étude de marqueurs microsatellite en 2007. L'analyse précise de la région révèle que le gène se situe dans une petite région d'homozygotie de 1,5 Mb (entre les marqueurs rs826978 et rs16968761).

Le variant c.1901T>C a été validé par séquençage Sanger et coségrégé avec le phénotype dans la famille (figure 5.10). De plus il n'a pas été retrouvé chez 56 contrôles algériens et 157 contrôles de diverses ethnies, soit 430 chromosomes testés au total.

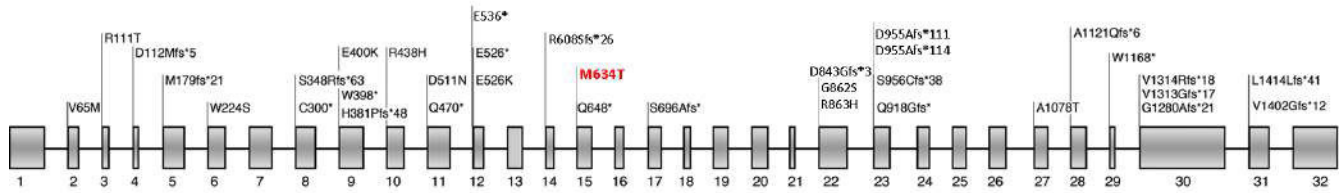


**Figure 5.10** : Électrophoregrammes représentant la mutation c.1901T>C du gène *WDR62* chez la famille Mo.

ES : Enfant Sain, EA : Enfant Atteint.

Bilgüvar et al. [2010] publient en Novembre 2010 les premières mutations de ce gène, identifiées chez des patients avec de sévères malformations cérébrales. Les auteurs proposent que le gène *WDR62* corresponde au locus MCPH2 de microcéphalie primaire autosomique récessive (MIM 613583) dont le spectre phénotypique est plus large et plus sévère qu'une microcéphalie isolée. Depuis 2010, de nombreuses mutations dans ce gène ont été identifiées (figure 5.11), élargissant ainsi le spectre phénotypique associé aux mutations du gène

*WDR62* [Nicholas et al., 2010, Farag et al., 2013, Poulton et al., 2014].



**Figure 5.11** : Mutations *WDR62* identifiées à ce jour, figure modifiée d'après McDonnell et al. [2014].

Le gène *WDR62* (19q13.12) composé de 32 exons code une protéine de 1 523 acides aminés. Les 33 mutations publiées à ce jour sont notées selon le changement peptidique (en noir). La mutation p.Met634Thr, présente chez la famille Mo, correspond à une nouvelle mutation et est notée en rouge.

Le gène *WDR62* code une protéine d'échafaudage (scaffold protein) nucléaire appartenant à la famille des protéines du pôle du fuseau mitotique [Bacino et al., 2012]. Chez l'homme et chez la souris, la protéine *WDR62* est enrichie dans les cellules neuroprogénitrices (Neural Progenitor Cells ou NPCs) et a un rôle dans le développement cortical précoce [Bilgüvar et al., 2010]. L'inactivation du gène *WDR62* engendre une réduction importante de la taille de la tête chez le zebrafish et la souris [Novorol et al., 2013, Xu et al., 2014, Chen et al., 2014]. Chez la souris, Xu et al. [2014] et Chen et al. [2014] constatent une formation anormale du fuseau mitotique dans les NPCs murines, accompagnée d'un arrêt du cycle et une mort cellulaire ; ces anomalies entraînent des défauts dans la neurogenèse et la migration neuronale, responsables du phénotype macroscopique observé. Enfin, Xu et al. [2014] réalisent des expériences de sauvetage phénotypique : les auteurs parviennent à restaurer le phénotype avec le gène humain sauvage dans les NPCs murines mais échouent avec les 5 mutants *WDR62* publiés par Nicholas et al. [2010].

Le variant p.Met634Thr identifié dans la famille Mo apparaît donc comme responsable du phénotype observé chez les patients. Comment évaluer alors la part de responsabilité des deux variants dans le phénotype global des patients ? Avec la revue de la littérature et les résultats de quelques tests fonctionnels, nous avons tenté ci-dessous d'évaluer la vraisemblance des potentiels effets délétères des deux variants.

## 5.4 Évaluation de la contribution de chaque variant

### 5.4.1 Le gène *WDR62*, données de la littérature

La table 5.12 regroupe l'ensemble des cas *WDR62* rapportés dans la littérature jusqu'à novembre 2014 : 34 mutations distinctes (figure 5.11), 105 patients chez 43 familles.

Mutation	Type	N	Ethnicity	Phenotype	Imaging	Reference(s)
p.Val1402Glyfs*12	Frameshift	2	Turkish	severe ID, few words, ambulate at 2y, unable to feed herself, micrognathia and a bulbous nose	MC, diffuse cortical thickening, indistinct gray-white junction, diffuse pachygyria mainly affecting left hemisphere, polymicrogyria, brainstem and cerebellum grossly normal	Bilgüvar et al. [2010]
				na	na	
p.Gly1280Alafs*21	Frameshift	1	Turkish	severe ID, 20m MC, developmental delay	Microdysplasia, hypoplastic CC	
p.Glu526X	Nonsense	1	Turkish	MC, good head control and social smile at 3m, has increased tone on examination	MC, pachygyria and CC hypoplasia	
p.Glu526Lys	Missense	1	Turkish	moderate ID, at 4y seizures, few words, understand simple commands at 6y and ambulate		
p.Gln470X	Nonsense	1	Turkish	severe ID, 3y MC, prognathism, dysconjugated gaze, dysarthria, can ambulate		
				severe ID, MC noted at 1m, celiac, genu varum, arachnodactyly		
p.Trp224Ser	Missense	3	Turkish	severe ID, MC, seizures, dysconjugated gaze, hyperactive reflexes and self-mutilating behaviors, can ambulate. More severely affected than his sister.		
				3y seizures, moderate ID, hyperactivity, dysconjugated gaze, poor verbal skills, can ambulate and understand simple commands	Diffuse polymicrogyria most visible in left posterior frontal lobe associated with a schizencephalic cleft, cortical thickening in bilateral frontal lobes, hypoplasia of CC, atrophy of left side of brainstem	
p.Asp112Metfs*5	Frameshift	3	Mexican	severe ID, 3y MC, hyperactivity, seizures at 2y, micrognathia, ambulate with support, only speak 1-2 word sentences	head CT : MC with Simplified Gyral pattern (MSG)	
				HC on 4th centile at 7y. Walked at 4y, at age of 7 could say a few words. Mild dysmorphic features with low sloping forehead		
				Presented at 12y unable to walk or talk, never sat or crawled or developed speech, GTC seizures at 9m controlled with valproate. Mild dysmorphic features with low sloping forehead	MRI : MSG, enlarged lateral ventricles and relatively smaller right hemisphere	
				At 10m HC -5.4 SD, early motor development normal, walking at 3y, no speech, Mild dysmorphic features with low sloping forehead	MRI : MSG without ventricular enlargement, suggestion of subcortical heterotopias, thin CC, preserved brainstem and cerebellum	
p.Ser348Argfs*63	Splice site	1	Turkish	MC, no seizures	MRI : MC, cerebrum and diminished sulcation, bilateral band heterotopia in post frontal and parietal lobes and thin CC	

Mutation	Type	N	Ethnicity	Phenotype	Imaging	Reference(s)
p.Val1313Glyfs*17	Frameshift	1	Turkish	At birth HC -2.8 SD, dysmorphic features : low sloping forehead, prominent occiput, broad and prominent nasal bridge, wide set eyes and fontanelle. Delayed development at 1y could sit, 18m could move by sliding. Limited verbal expressions HC at 2m -5.3 SD, high palate	Polymicrogyria L>R, MSG on right abnormal CC. cerebellum and brainstem relatively preserved	
p.Ser956Cysfs*38	Frameshift	1	Turkish	MC noted at 6m, walked at 1.5y, few words at 3y seizures at 1.5y	Thin CC, band heterotopia in frontal and parietal lobes	
p.Gly1280Alafs*21	Frameshift	1	Turkish	Walked at 21m, constant drooling, reduced extension of both elbows, bilateral hallux valgus and pes planus and the second toe overlapped the first toe bilaterally	MC, relative sparing of brainstem and cerebellum, thin CC	
p.Val65Met	Missense	3	Saudi Arabia	MC -9.8SD at 8y, unable to sit, no words, spastic quadraparesis, asthma	MSG	
p.Val65Met	Missense	3	Arab	MC -9.2SD at 6.5y, only able to roll, babble and tear paper, spastic quadraparesis, occasional shortness of breath, hyperpigmented skin lesion on the center of his chest extending up over the right shoulder	Diffuse PMG and narrow right tempo-parietal open lip schizencephaly, dysmorphic CC with a thick body and a small genu, brain stem and cerebellum relatively spared	Nicholas et al. [2010]
p.Arg438His	Missense	6	Pakistani	MC btw -4 and -7 SD noted at birth or after few months of life, disproportionate sized face and ears and sloping forehead	Diffuse simplified gyral pattern and cortical indistinct border	
p.Asp511Asn	Missense	3	Pakistani	as compared to the skull, delayed speech acquisition, mild-moderate ID	na	
p.Ala1078Thr	Missense	2	Pakistani		na	
p.Leu1414Leufs*41	Missense	5	Pakistani		na	
p.Val1314Argfs*18	Frameshift	9	Pakistani		na	
p.Val1314Argfs*18	Frameshift	1	Caucasian	Severe ID, MC at -4SD at 10y, ambulatory, incontinent, socially withdrawn, no language	Diffuse simplified gyral pattern most severely affecting frontal cortex and mildly thickened cortex	
p.Met179fs*21	Frameshift	2	Indian	MC, motor milestones mildly delayed, speech grossly delayed, GTC seizures, sloping forehead, inverted nipples, peri-anal pits, low set and prominent ears, impulsive prone for temper tantrums, head banging when frustrated	MC, pachygyria frontal cortex	Bhat et al. [2011]
				MC, walked at 2y, mild ID, delayed milestones, delayed speech acquisition, high-arched palate, sloping forehead, inverted nipples, peri-anal pits, low set and prominent ears	MC and dysplastic cortex involving left front middle region with widened sulcus	

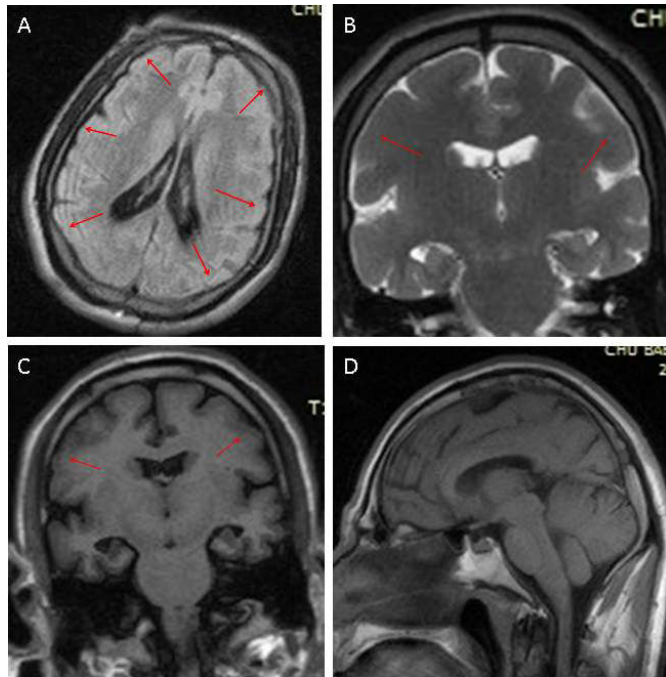
Mutation	Type	N	Ethnicity	Phenotype	Imaging	Reference(s)
p.Cys300X	Nonsense	2	Indian	Mild MR, walked at 4y, speech at 6y, seizures at 9y, prominent ears, sloping forehead Mild MR, walked at 2y, prominent ears, sloping forehead	MC, pachygyria, PMG bilaterally, microencephaly and band heterotopia bilaterally MC, pachygyria and PMG bilateral frontal cortex	Murdock et al. [2011]
mat. p.Ser696Alafs*4 / pat. p.Gln918Glyfs*18	CHM, 2 frameshift	2	Northern european	MC, globally delayed, infantile spasms at 4m, epilepsy, 9y spastic quadraparesis Ambulatory, uses age appropriate language, delayed social development, mild right hemiparesis, mild MC	Bilateral PMG Small left hemisphere, extensive PMG, focus of heterotopia in the right parietal region	Murdock et al. [2011]
p.Gln648X	Nonsense	2	Pakistani	ID, MC at birth, btw -3 and -7SD, abnormal sleeping habits, no facial dysmorphism, no seizures, no autism	Reduced volume of right cerebral hemisphere, prominent CSF space and ill-defined gyri and nuclei pattern	Kousar et al. [2011]
p.Arg438His	Missense	3	Pakistani	ID, MC at birth, btw -3 and -7SD, no facial dysmorphism, no seizures, no autism	na	
p.Val1314Argfs*18	Frameshift	3	Pakistani	ID, MC at birth, btw -3 and -7SD, no facial dysmorphism, no seizures, no autism	na	
p.Asp511Asn	Missense	2	Pakistani	ID, MC at birth, btw -3 and -7SD, no facial dysmorphism, no seizures, no autism	na	
p.Glu400Lys	Missense	2	Mexican	moderate-severe ID, delayed motor development, delayed speech (few words at 8 years), no seizures	CT : MC, pachygyria	Bacino et al. [2012]
p.Arg111Thr	Missense	6	Pakistani	MC -7 to -11SD, muscular dystrophy in one patient, unable to speak clearly, no self-care ability, history of seizure in one patient	na	Sajid Hussain et al. [2013]
p.Trp398X	Nonsense	4	Pakistani	MC -6 to -8SD, alopecia and goiter		
p.Ala1121Glnfs*6	Frameshift	2	Pakistani	MC -10 to -11SD		
p.Trp1168X	Nonsense	4	Pakistani	MC -9 to -11SD, occasional epileptic shocks, no self-care ability skills, self-injurious and aggressive behavior with history of seizures		
p.Arg438His	Missense	4	Pakistani	MC -5 to -14 SD, can not read or do self-care		
p.His381Profs*48	Frameshift	3	Pakistani	2 siblings with MC -6 to -7SD, delay in ambulation, few words, epilepsy	2 siblings with dilated ventricles, CC hypoplasia, schizencephaly	Memon et al. [2013]
p.Arg438His / p.Asp955fsAla*112	CHM, 1 missense and 1 frameshift	2	German	severe ID, MC at -2.3SD at birth, severely delayed language, walked at 2y, atstatic seizures, sloping forehead, convex facial profile, full lips, small chin, low-set and posteriorly rotated ears partly due to reclination of the head, bilateral pes planus and hallux valgus	na small frontal lobes, CC hypoplasia, simplified hippocampal gyration, widened lateral sulci, and cerebellar hypoplasia with an enlarged cisterna magna	Farag et al. [2013]

Mutation	Type	N	Ethnicity	Phenotype	Imaging	Reference(s)
				MC at -2.3SD at birth, ID, left-sided spastic hemiparesis, died of a Wilms tumor at the age of 5y	Head CT at 1m : plump right ventricle, wide arachnoidal spaces, and parieto-occipital hypodensity. EEG at 2y : signal depression on the right hemisphere and epileptic activity on the left hemisphere	
p.Arg608Serfs*26	Frameshift	4	French canadian	2 females : moderate to severe ID, epilepsy, MC -4 to -5SD, sloping forehead, pleasant character, bowel incontinence	MRI revealed only MC	McDonnell et al. [2014]
				2 males : severe ID, epilepsy, MC -3SD, sloping forehead, medically refractory seizures, hypertonia, hyperreflexia/ankle clonus, aggression and self-injury,	na	
p. Glu536X	Nonsense	1	Turkish	Severe ID, seizures, tetraparesis, MC, also carries a change on both <i>TBCD</i> loci ; a maternally inherited duplication including <i>TBCD</i> as well as a paternally inherited missense mutation on the nonduplicated allele	Microdysplasia, pachygyria	Poulton et al. [2014]
p. Asp955Alafs*111	Frameshift	1	nd	MC, able to sit unsupported, cooperative, no speech, non-ambulatory	Simplified gyration with moderately thickened cortex	Müller et al. (manuscript in preparation), from Poulton et al. [2014]
p. Arg863His	Missense	1	Tunisian	MC, moderate developmental delay, insulin dependent diabetes	Symmetric bilateral polymicrogyria	
p. Asp955Alafs*111, p. Gly862Ser	Frameshift, missense	1	na	MC, infantile spasms, astigmatism, severe psychomotor delay, and feeding difficulties	Frontal pachygyria ; unilateral polymicrogyria	
p. Asp955Alafs*114	Frameshift	1	na	MC, no active speech	No gross cortical malformation	Rupp et al. [2014]
p. Glu536X	Nonsense	1	na	MC at birth -2.7SD, no seizures, ID, friendly behavior, hyperactive, walked at 18m, no language at 4y	Pachygyria, CC dysgenesis	
p. Asp843Glyfs*3	Frameshift	3	Pakistani	MC, aggressive behavior, abnormally water mouth	CT-scan : reduced volume of the right cerebral hemisphere and prominent extra axial cerebrospinal spaces with ill-defined gyral and nuclei pattern	Rupp et al. [2014]
				MC, aggressive behavior, abnormally water mouth	na	
				MC, no other abnormalities	na	

**Figure 5.12 :** Patients *WDR62* rapportées dans la littérature, inspiré de Poulton et al. [2014].

MC : Microcephaly, MSG : Microcephaly with Simplified Gyration, PMG : Polymicrogyria, CC : Corpus Callosum, MRI : Magnetic Resonance Imaging, GTC : Generalized Tonic Clonic, CSF : Cerebrospinal Fluid, CHM : Compound Heterozygous Mutation, HC : Head Circumference, CT : Computerized Topography, SD : Standard Deviation, L : Left, R :Right, N : Individuals, pat : paternal, mat : maternal, na : non available, y : years, m : months, btw ; between.

Étant donné les anomalies corticales observées très fréquemment chez les patients rapportés avec des mutations dans le gène *WDR62*, les images IRM étaient nécessaires et représentaient un bon moyen pour estimer la concordance clinique de nos patients avec les patients *WDR62*. Nous avons donc demandé au neurologue suivant ces patients, le Dr Karim Saadi, de réaliser cet examen pour l'un des patients. Les images du patient IV.5 sont présentées ci-dessous (figure 5.13).



**Figure 5.13** : IRM des patients de la famille Mo.

A : axial flair ; B : coronal T2 ; C : coronal IR.

Ces images montrent une pachygyrie diffuse touchant principalement le lobe fronto-parieto-temporal bilatéral, un motif gyral simplifié et un cortex épais tandis que le cervelet et le tronc cérébral sont intacts. Quoique certains cas présentent une microcéphalie isolée [McDonnell et al., 2014], la grande majorité des patients *WDR62* décrits dont les données IRM sont disponibles, ainsi que le patient IV.5 de la famille Mo, présentent des malformations du cortex cérébral avec une anomalie de l'apparence des circonvolutions du cerveau (lissencéphalie). Ces anomalies sont associées à une organisation anormale des couches du cortex en raison d'un défaut spécifique de migration neuronale lors de l'embryogenèse. Elles sont présentes sous forme de pachygyrie et/ou polymicrogyrie, d'une hypoplasie du corps calleux, parfois accompagnées d'un cortex épais et plus rarement d'une schizencéphalie et d'hétérotopies. La table 5.6 rapporte les fréquences associées aux anomalies neurologiques chez les patients *WDR62* publiés et les patients de la famille Mo.



Phénotype	Fréquence chez les patients <i>WDR62</i>	Fréquence chez les patients Mo	Définition	
DI sévère	15/79	6/6		
DI modérée ou légère	44/79	0/6		
Retard de langage	49/97	6/6		
Acquisition de la marche tardive	17/97	6/6		
Épilepsie	23/97	2/6		
Spasticité	4/97	0/6		
Troubles du comportement	11/97	6/6		
Microcéphalie	103/103	6/6		
Gyration simplifiée	15/49	1/1		
Cortex épais	2/48	1/1		
Anomalies du CC	17/48	0/1		
Pachygyrie	14/48	1/1		Circonvolutions cérébrales trop épaisses et trop peu nombreuses
Polymicrogyrie	11/48	0/1		Anomalie locale du cortex cérébral caractérisée par la présence d'un excès de gyrations et de sillons peu profonds
Schizencéphalie	5/48	0/1	Malformation cérébrale caractérisée par une fente linéaire bordée de substance grise traversant tout l'hémisphère, du ventricule latéral à la surface externe du cortex	
Hétérotopies	4/48	0/1	Présence de neurones en dehors du cortex	
Dismorphie faciale	47/97	6/6		
Anomalies des extrémités	4/97	3/6		

**Table 5.6 :** Fréquence des différents signes cliniques observés chez les patients Mo et la totalité des patients *WDR62*.

La revue de l'ensemble des mutations *WDR62* décrites et des caractéristiques cliniques associées montre que la mutation *WDR62* rend compte de la majorité des signes observés chez nos patients (table 5.6). Cet exercice met également en valeur l'extrême variabilité clinique observée chez les patients mutés *WDR62* (table 5.12). En effet, ces mutations semblent être associées à une microcéphalie sévère avec un large spectre de malformations corticales et une atteinte cognitive de degré très variable. D'après Nicholas et al. [2010], cette variabilité pourrait être due au type de mutation, l'effet des mutations non sens étant attendu plus important. Cependant, cette corrélation n'est pas totale comme l'attestent les mutations faux sens

p.Trp224Ser, [Bilgüvar et al., 2010] et p.Val65Met, [Yu et al., 2010] décrites chez des patients présentant un phénotype plus sévère que des patients avec des mutations non sens p.Cys300X, [Bhat et al., 2011] et p.Glu526X, [Bilgüvar et al., 2010] (table 5.12). D'autre part, certaines mutations présentent une variabilité intrafamiliale (par exemple p.Trp224Ser, [Bilgüvar et al., 2010] et p.Arg608Serfs\*26, [McDonnell et al., 2014]). Très récemment, Poulton et al. [2014] ont décrit deux patients porteurs d'une mutation p.Glu536X dans le gène *WDR62* et dont l'un présente un phénotype plus sévère que le second. L'identification, chez ce deuxième patient, d'une variation dans un second gène, tubulin cofactor D (*TBCD*), crucial pour la dynamique des microtubules, explique la sévérité importante de ces symptômes (table 5.12). Les auteurs proposent l'existence de facteurs génétiques modificateurs à l'origine de la variabilité intra-groupe observée chez les patients porteurs de mutations *WDR62*.

## 5.5 Discussion et conclusion préliminaires

En conclusion, l'étude de la famille Mo a permis de mettre en évidence une nouvelle mutation *WDR62*, élargissant encore le nombre de cas répertoriés dans la littérature. Cette mutation est vraisemblablement responsable de la grande majorité des symptômes observés chez les patients de cette famille. Cependant, bien que les tests fonctionnels préliminaires réalisés n'apportent pas d'arguments en faveur d'un effet causal du variant *RAD54B*, nous ne pouvons pas entièrement exclure que celui-ci ne participe pas à un certain degré à la sévérité des symptômes des patients de la famille Mo. Comme suggéré plus tôt par Poulton et al. [2014], le variant *RAD54B* pourrait potentiellement être un facteur génétique modificateur du phénotype induit par la mutation *WDR62*.

Une cohorte de 13 patients avec un tableau clinique similaire à celui de la famille Mo avait été étudiée en 2007 : aucune nouvelle mutation dans le gène *RAD54B* n'avait été identifiée. Il serait intéressant de reprendre cette cohorte aujourd'hui pour tester le gène *WDR62*. Le gène correspondant au locus MCPH2 de microcéphalie primaire autosomique récessive n'était pas caractérisé avant 2010. Par conséquent, ce locus n'avait pas été étudié lors de l'analyse des gènes connus de microcéphalie en première intention en 2007. Avant la révolution NGS, lors de l'étude de grandes familles consanguines, notre attention était ciblée sur les plus grandes régions IBD identifiées par cartographie, anticipant une probabilité supérieure que celles-ci contiennent une mutation causale partagée par l'ensemble des patients. Nos résultats démontrent que cette précédente approche n'était pas sans risque de manquer des variants intéressants dans d'autres régions plus petites. Ce travail illustre l'apport des nouvelles technologies de séquençages à haut débit qui a rendu possible l'identification d'une mutation auparavant ignorée par la méthode de cartographie par autozygotie et, à l'aide des autres cas connus dans la littérature, a permis de mieux caractériser le phénotype des patients.

## CHAPITRE 6

# COMBINAISON DE DEUX MÉTHODES CONDUISANT À LA CARACTÉRISATION D'UN NOUVEAU GÈNE DE DI : TTI2

### 6.1 La famille, les patients

Grâce au soutien de la fondation maladies rares, nous avons eu l'opportunité d'étudier la famille El. D'origine algérienne, cette famille comprend 3 enfants atteints (2 filles et un garçon) sur une fratrie de 12, issus d'une union consanguine. Elle est suivie par le Dr Abdelkrim Saadi au sein de l'établissement hospitalier de Benaknoun à Alger. Les patients présentent une DI syndromique avec notamment une microcéphalie, une dysmorphie faciale, une petite taille et des anomalies du squelette.

### 6.2 Stratégie

Pour l'étude de cette large famille consanguine, débutée en 2011, nous avons opté pour une combinaison des analyses de cartographie par autozygotie et de WES. Un seul enfant atteint a été analysé par WES et les variants candidats ont été filtrés sur la base de leur localisation dans les régions d'IBD.

### 6.3 Résultats : publication n°1, [Langouët et al., 2013]

## Mutation in *TTI2* Reveals a Role for Triple T Complex in Human Brain Development

Maéva Langouët,<sup>1</sup> Abdelkrim Saadi,<sup>2</sup> Guillaume Rieunier,<sup>3</sup> Sébastien Moutton,<sup>1</sup> Karine Siquier-Pernet,<sup>1</sup> Marie Fernet,<sup>4</sup> Patrick Nitschke,<sup>5</sup> Arnold Munnich,<sup>1</sup> Marc-Henri Stern,<sup>3</sup> Malika Chaouch,<sup>2</sup> and Laurence Colleaux<sup>1\*</sup>

<sup>1</sup>INSERM U781, Université Paris Descartes, Sorbonne Paris Cité, Institut IMAGINE, Hôpital Necker-Enfants Malades, Paris, France; <sup>2</sup>Département de Neurologie, Etablissement Hospitalier Spécialisé de Benaknoun, Algiers, Algeria; <sup>3</sup>Centre de Recherche, Institut Curie et INSERM U830, Paris, France; <sup>4</sup>Centre de Recherche, Institut Curie et INSERM U612, Centre Universitaire, Orsay Cedex, France; <sup>5</sup>Plateforme de Bioinformatique, Université Paris Descartes, Sorbonne Paris Cité, Institut IMAGINE, Paris, France

Communicated by Hamish S. Scott

Received 2 April 2013; accepted revised manuscript 5 August 2013.

Published online 16 August 2013 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22399

**ABSTRACT:** Tel2-interacting proteins 1 and 2 (*TTI1* and *TTI2*) physically interact with telomere maintenance 2 (*TEL2*) to form a conserved trimeric complex called the Triple T complex. This complex is a master regulator of phosphoinositide-3-kinase-related protein kinase (PIKKs) abundance and DNA damage response signaling. Using a combination of autozygosity mapping and high-throughput sequencing in a large consanguineous multiplex family, we found that a missense c.1307T>A/p.I436N mutation in *TTI2* causes a human autosomal recessive condition characterized by severe cognitive impairment, microcephaly, behavioral troubles, short stature, skeletal anomalies, and facial dysmorphic features. Immunoblotting experiment showed decreased amount of all Triple T complex components in the patient skin fibroblasts. Consistently, a drastically reduced steady-state level of all PIKKs tested was also observed in the patient cells. Combined with previous observations, these findings emphasises the role of the *TTI2* gene in the etiology of intellectual disability and further support the role of PIKK signaling in brain development and functioning.

Hum Mutat 34:1472–1476, 2013. © 2013 Wiley Periodicals, Inc.

**KEY WORDS:** intellectual disability; *TTI2*; exome sequencing; PIKK

The Triple T complex is an evolutionary-conserved trimeric complex composed of the proteins *TEL2* (MIM #611140), *TTI1* (MIM #614425), and *TTI2* (MIM #614426). *tel2* has been first identified in a screen for yeast mutants with altered telomere length [Lustig and Petes, 1986]. Subsequently, *Tti1p* and its associated protein *Tti2p* were implicated in *tel2*-dependent processes on the basis of their

mutual association in yeast and mammalian cells [Hayashi et al., 2007; Hurov et al., 2010; Kaizuka et al., 2010; Shevchenko et al., 2008; Takai et al., 2007; Takai et al., 2010].

The components of the Triple T complex are mutually dependent on each other, and act as critical regulators of the PIKK abundance and checkpoint signaling. As such, the Triple T complex has been shown to be required for various cellular processes, including DNA damage response, nonsense-mediated decay, and telomerase assembly [Horejsi et al., 2010; Hurov et al., 2010; Izumi et al., 2010; Takai et al., 2007; Takai et al., 2010; Venteicher et al., 2008]. PIKKs are unconventional serine–threonine protein kinases, with homology in the catalytic domain to phosphatidylinositol 3-kinase, which phosphorylates proteins rather than lipids. In mammals, the PIKK family includes the catalytically active members protein kinase, DNA-activated, catalytic polypeptide (PRKDC/DNA-PKcs; MIM #600899), Ataxia Telangiectasia Mutated (ATM; MIM #607585), ATM- and Rad3-Related (ATR; MIM #601215), mechanistic target of rapamycin (MTOR; MIM #601231), and suppressor with morphogenetic effect on genitalia-1 (SMG-1; MIM #607032), as well as the catalytically inactive member transformation/transcription domain-associated protein (TRRAP; MIM #603015).

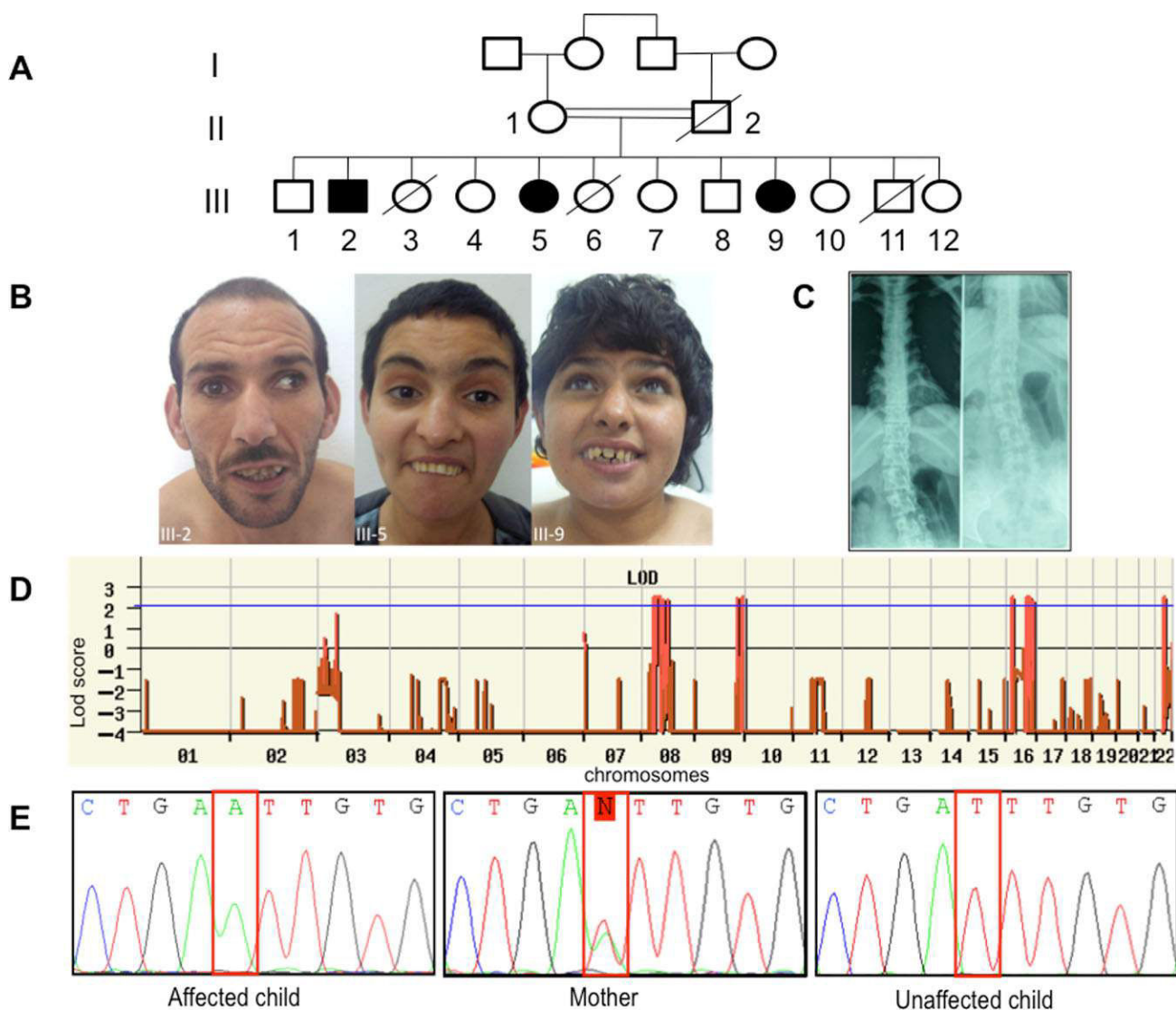
Deficiencies of various signaling pathways regulated by PIKKs cause different diseases. Mutations in ATM, which is crucial for response to DNA double-strand breaks (DSBs) and cellular stresses [Guo et al., 2010], and ATR, which is activated by replication stress and certain DNA repair intermediates, cause ataxia telangiectasia and Seckel syndrome, respectively. DNA-PKcs promotes nonhomologous end joining (NHEJ) and its deficiency causes diminished V(D)J recombination and severe combined immunodeficiency in mice. Finally, the MTOR pathway is implicated in the tumorigenesis and its deregulation is associated with familial cancer syndromes and tuberous sclerosis [Sabatini, 2006; Shaw and Cantley, 2006]. By contrast, the consequences of mutations altering any of the three Triple T subunits have only been studied in yeast or cultured mammalian cells.

We ascertained a sibship of three affected (two female and one male) and six unaffected siblings born to healthy first cousin Algerian parents (Fig. 1A). All patients were born with normal growth parameters after an unremarkable pregnancy and delivery. For all patients, the neonatal period was uncomplicated. At their last visit at the age of 30–36 years, occipitofrontal circumference ranged from –3 to –4 standard deviation (Supp. Table S1). They all had severe cognitive impairment with severe speech delay and behavioral disturbances but no seizures were reported. Short stature and

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

\*Correspondence to: Laurence Colleaux, INSERM U781, Hôpital Necker-Enfants Malades, 149 Rue de Sèvres, 75015 Paris, France. E-mail: laurence.colleaux@inserm.fr

Contract grant sponsors: Centre National de la Recherche Scientifique (CNRS); The Agence Nationale de la Recherche (ANR SAMENTA 2012); The GIS-Maladies RARES; The Fondation pour la Recherche Médicale; Ministère de la Recherche et de l'Enseignement Supérieur; Ligue Nationale Contre le Cancer.



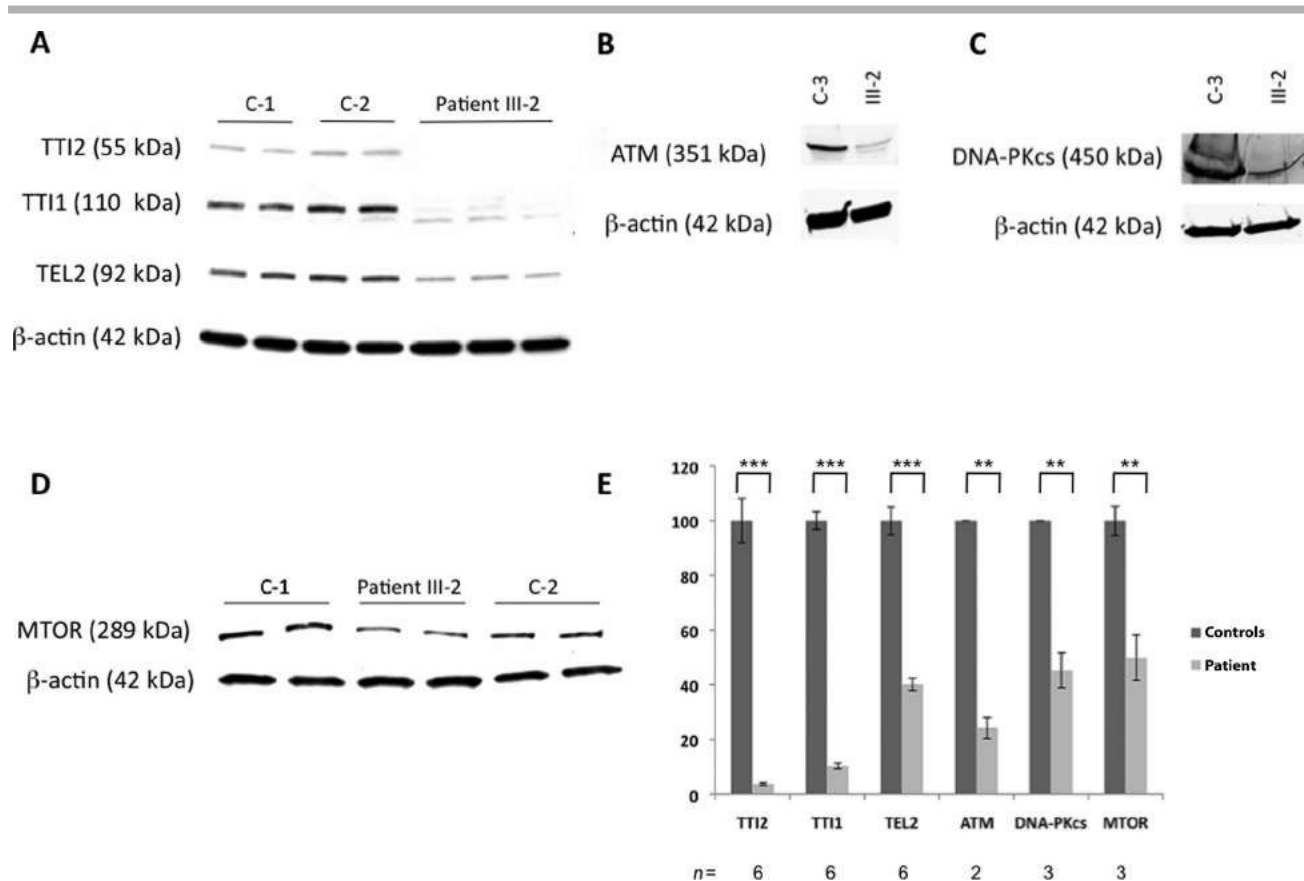
**Figure 1.** A missense mutation in *TT12* underlies a novel form of autosomal recessive syndromic ID. **A:** Pedigree of the family. Shaded symbols indicate the affected individuals. **B:** Facial photographs of patients III-2, III-5, and III-9 showing dysmorphic features consisting of sloping forehead, prominent nose and middle synphris in all patients. Deep-set eyes, large anteverted ears for patient III-2. Dental anomalies for patient III-2 and III-9 and strabismus for patient III-2 and III-5. Legal representative of the patients gave consent for the publication of these photographs. **C:** Spine X-ray of patient III-5 showed dorsal lumbar scoliosis. **D:** Homozygosity mapping results. The y axis represents the LOD scores and the x axis represents the chromosomes. A thick horizontal line indicates the LOD score threshold set at 2.0. Each peak above this line is part of the 14 homozygous regions shared only by the three affected siblings. **E:** Sequence analysis in a patient, the mother and a healthy sib showing the c.1307T>A variant in the *TT12/C8orf41* gene.

vertebral anomalies were noted in all affected sibs (Supp. Table S1 and Fig. 1C). Dysmorphic features including a sloping forehead, deep set eyes, and anteverted large ears were also observed (Fig. 1B). Laboratory analysis indicated normal circulating T, B, and NK cell counts but a mild naïve T (CD4, CD8) lymphocytopenia. Neurological examination, molecular karyotyping, and standard metabolic workup were normal.

Multipoint linkage analysis in all three affected children, the unaffected child III-1, and the mother II-1 revealed 14 homozygous identical-by-descent (IBD) regions shared by all affected sibling (Fig. 1D and Supp. Methods). Collectively, these genomic intervals, which size ranged from 700 kb to 11.5 Mb, encompassed 40.1 Mb, and contained more than 320 known genes (Supp. Table S2).

To identify the disease-causing mutation, we performed whole-exome sequencing on DNA from patient III-9 (Supp. Methods).

Exome sequencing data were first filtered for novel variants located in IBD regions previously defined by autozygosity mapping and focused primarily on nonsynonymous variants, splice acceptor and donor site mutations, and coding insertions/deletions (indels), anticipating that synonymous variants were far less likely to be pathogenic. We regarded variants as novel if they were absent from all publically available data sets, including those of dbSNP132 (<http://www.ncbi.nih.gov/SNP>), the 1000 Genomes Project (<http://browser.1000genomes.org/index.html>), the NHLBI ESP Exome Variant Server (<http://evs.gs.washington.edu/EVS/>), and from in-house exome data containing information for more than 600 samples. Two candidate variants fulfilled these criteria (Supp. Table S3) and were confirmed by Sanger sequencing. Analysis of 130 chromosomes from an ethnic-specific control cohort allowed to discriminate between the two variants and



**Figure 2.** Functional consequences of the *TTI2/C8orf41* mutation **A:** Decreased amount of all Triple T subunits in the patient's fibroblasts. Lysates from controls (C-1 and C-2) and patient (III-2) cultured fibroblasts were subjected to immunoblotting with a polyclonal anti-TTI2 antibody and a monoclonal anti- $\beta$ -actin antibody as a loading control. Each extract was loaded in duplicate (controls) or in triplicate (patient). The same membrane was incubated with a polyclonal anti-TTI1 antibody, a polyclonal anti-TEL2 antibody and a monoclonal anti- $\beta$ -actin antibody as a loading control. TTI1 and TEL2 protein amounts are clearly decreased in the patient compared with controls. **B–D:** Reduced ATM, DNA-PKcs, and MTOR levels in patient fibroblasts. Lysates from control (C-3) and patient (III-2) fibroblasts were subjected to immunoblotting with a polyclonal anti-ATM antibody, a monoclonal anti-DNA-PKcs antibody, a monoclonal anti-MTOR antibody and a monoclonal anti- $\beta$ -actin antibody as a loading control. ATM and ATR protein amounts are clearly decreased in the patient compared with control. **(E)** Densitometry analyses. Means  $\pm$  standard error (SEM) for TTI2/actin, TTI1/actin, TEL2/actin, ATM/actin, DNA-PKcs/actin and MTOR/actin ratios are given and expressed as a percentage of the control values (100%). The number of replicates for each calculation is given (*n*). Significance was calculated using two-way analysis of variance (ANOVA) test with Tukey correction. \*\* *P* value < 0.01, \*\*\* *P* value < 0.001.

after all filtering steps, a single variant mapping within the largest IBD chromosomal region remained (8p12-p11.21, 11.5 Mb). This NM\_025115.1:c.1307T>A, p.Ile436Asn variant was located in exon 6 of *TTI2/C8orf41* (Fig. 1E). Variant data were deposited in a public database (<http://www.lovd.nl/TTI2>). Filtering exome sequencing data according to other genetic mode of inheritance (compound heterozygous mutation) did not provide any candidate variant.

The p.Ile436Asn variant alters a conserved residue (Supp. Fig. S1) and was predicted to be damaging by various in silico tools (PolyPhen-2 score: 0.868, SIFT scores: 0 and disease causing according to Mutation Taster). Eleven patients with a similar phenotype were screened for mutations in *TTI2/C8orf41* but no additional variants could be identified. By contrast, a separate study reported another missense *TTI2* mutation (p.Pro367Leu) cosegregating with intellectual disability (ID) in a large consanguineous Iranian family [Najmabadi et al., 2011]. This variant was also predicted to be deleterious (PolyPhen-2 score: 1, SIFT scores: 0 and disease causing according to Mutation Taster). Collectively, these data suggest that *TTI2* mutations cause a novel form of autosomal recessive ID.

The consequence of the c.1307T>A variant on *TTI2* expression was first assessed by quantitative reverse-transcription PCR (RT-

PCR) of RNAs extracted from cultured skin fibroblasts of patient III-2 but no difference in *TTI2* mRNA expression between patient and control cells was observed (data not shown). Subsequently, the effect of the p.Ile436Asn mutation on TTI2 protein stability was tested by Western blot analysis. Immunoblot analyses of fibroblast lysates showed a markedly reduced amount of the TTI2 protein to almost undetectable level in the patient cells compared with the controls (Fig. 2A and E). Considering that TTI2, as TTI1 and TEL2, is essential for stability and function of the Triple T complex [Hurov et al., 2010; Takai et al., 2007], we investigated the consequence of the mutation on TTT complex stability. We observed a significantly reduced amount of TTI1 and TEL2 proteins in the patient fibroblasts compared with the controls (Fig. 2A). Residual TTI1 and TEL2 levels were less than 10% and 40% of controls, respectively (Fig. 2E). Altogether, our data demonstrate that the p.Ile436Asn mutation caused a dramatic decreased amount of all Triple T complex components.

In mammals, the Triple T complex regulates assembly and stability of PIKK family proteins [Hurov et al., 2010; Takai et al., 2007; 2010]. We thus examined the effect of the p.Ile436Asn *TTI2* mutation on PIKKs stability. We observed that the *TTI2* mutation caused a significant reduction of ATM; DNA-PKcs and MTOR protein



amount in the patients' cells compared with the controls (Fig. 2B–D and Supp. Methods). Residual ATM, DNA-PKcs and MTOR levels were less than 24%, 45%, and 50% of controls, respectively (Fig. 2E). To further confirm these results, siRNA directed against *TTI2* were transfected into HEK 293 cells. Silencing *TTI2* expression had similar effect on TTI1 and TEL2 protein stability to what is observed in the patient cells (Supp. Fig. S2A). Similarly, a significant decrease of ATM and DNA-PKcs protein levels were observed in depleted cells when compared with the control cells (Supp. Fig. S2B). Taken together, these results suggested that the p.Ile436Asn *TTI2* mutation impaired stability of the Triple T complex, leading to drastically reduced steady-state level of all PIKKs tested.

In this study, by combining autozygosity mapping and high-throughput sequencing in a large consanguineous family, we identified a *TTI2* mutation responsible for an autosomal recessive condition characterized by severe cognitive impairment, microcephaly, behavioral troubles, short stature, skeletal anomalies, and facial dysmorphic features. Combined to the previous identification of another *TTI2* mutation cosegregating with ID [Najmabadi et al., 2011], our results suggest Triple T complex impairment as a new pathophysiological mechanism underlying cognitive deficit.

This recently discovered complex plays a key role in the folding/maturation of PIKK proteins. Three members of the PIKK family, namely DNA-PK, ATM, and ATR, play prominent roles in the cellular response to DNA damage through the detection and repair of DSBs [Shiloh, 2003]. MTOR regulates translation, cell growth, autophagy, metabolism, and organization of the actin cytoskeleton [Wullschleger et al., 2006]. SMG-1 facilitates nonsense-mediated mRNA decay [Yamashita et al., 2005]. The sixth known member of the PIKK family, TRRAP, is devoid of kinase activity, and serves as an essential component of several multiprotein chromatin-remodeling complexes [McMahon et al., 2000].

*TTI2* deficient patients share several features with those suffering from defective responses to DNA damage. First, ID and microcephaly are also observed in patients with Nijmegen Breakage, DNA-ligase IV deficiency, Cernunnos-XLF, Seckel, and Bloom syndromes. Second, short stature/growth retardation (Seckel, Fanconi, DNA-ligase IV deficiency, Cernunnos-XLF, Bloom, Nijmegen, and Ataxia telangiectasia syndromes), skeletal abnormalities (Fanconi, Cernunnos-XLF, and Seckel syndromes) are frequently described in DNA repair disorders. Third, facial dysmorphism is a sign of Seckel, Cernunnos-XLF, or Nijmegen Breakage Syndromes. Finally, the mild naïve T (CD4 and CD8) lymphocytopenia found in the *TTI2* patients can also be observed in DNA repair disorders with thymic deficiency.

By contrast, cancer predisposition, a feature commonly observed in DNA repair disorders, is not hitherto reported in our family. Along these lines, it is worth remembering that Seckel syndrome linked to mutations in the ATR gene, as well as Ataxia-telangiectasia-like disorder (A-T-LD, a clinically attenuated version of A-T linked to hypomorphic mutations in the MRE11 gene and causing defective NHEJ DNA-repair pathway), display no evidence of cancer predisposition. In our family, absence of such predisposition could also be a direct consequence of the downregulation of the MTOR pathway, a prototypic survival pathway constitutively activated in many types of cancer, thus preventing from tumorigenesis. Similarly, RNA interference experiments have shown that TRRAP knockdown significantly increased stem cell-like differentiation and suppressed tumor formation [Wurdak et al., 2010]. Although we did not explore the effect of *TTI2* mutation on TRRAP stability, a reduced amount of TRRAP protein in patient cells is also expected to protect against cancer occurrence.

In neurons, mTOR is implicated in the local translation of dendritic mRNAs, a key aspect of dendrite and spine morphogenesis and synaptic plasticity [Campbell and Holt, 2001; Jaworski et al., 2005; Kumar et al., 2005]. Consistently, previous studies revealed a connection between deregulation of the MTOR pathway and several cognitive disorders. Hyperactivation of the MTOR pathway has been observed in autism [Ehninger and Silva, 2011], tuberous sclerosis, Fragile X and Down's syndrome [Troca-Marin et al., 2012] and hypoactivation of MTOR has been linked to Rett syndrome [Ricciardi et al., 2011]. As such, by impairing MTOR signaling in neuronal cells, the *TTI2* mutation could cause altered dendrite morphology and synaptic plasticity leading to the cognitive deficit observed in our patients.

In conclusion, our finding provide the first example of a human disorder related to a defect in Triple T complex, ascribing this new condition to the large family of disorders linked to impaired genome and transcriptome maintenance. These findings provide further support to the hypothesis of a crucial role of PIKK signaling in brain development and functioning.

## Acknowledgments

We are grateful to the patients and their family members for their participation in our study. We thank Sylvain Hanein for technical advices and Janet Hall for helpful comments. We also acknowledge Corinne Antignac for anti-MTOR antibodies.

*Disclosure statement:* The authors declare no conflict of interest

## References

- Campbell DS, Holt CE. 2001. Chemotropic responses of retinal growth cones mediated by rapid local protein synthesis and degradation. *Neuron* 32:1013–1026.
- Ehninger D, Silva AJ. 2011. Rapamycin for treating Tuberous sclerosis and Autism spectrum disorders. *Trends Mol Med* 17:78–87.
- Guo Z, Kozlov S, Lavin MF, Person MD, Paull TT. 2010. ATM activation by oxidative stress. *Science* 330:517–521.
- Hayashi T, Hatanaka M, Nagao K, Nakaseko Y, Kanoh J, Kokubu A, Ebe M, Yanagida M. 2007. Rapamycin sensitivity of the Schizosaccharomyces pombe tor2 mutant and organization of two highly phosphorylated TOR complexes by specific and common subunits. *Genes Cells* 12:1357–1370.
- Horejsi Z, Takai H, Adelman CA, Collis SJ, Flynn H, Maslen S, Skehel JM, de Lange T, Boulton SJ. 2010. CK2 phospho-dependent binding of R2TP complex to TEL2 is essential for mTOR and SMG1 stability. *Mol Cell* 39:839–850.
- Hurov KE, Cotta-Ramusino C, Elledge SJ. 2010. A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. *Genes Dev* 24:1939–1950.
- Izumi N, Yamashita A, Iwamatsu A, Kurata R, Nakamura H, Saari B, Hirano H, Anderson P, Ohno S. 2010. AAA+ proteins RUVBL1 and RUVBL2 coordinate PIKK activity and function in nonsense-mediated mRNA decay. *Sci Signal* 3:ra27.
- Jaworski J, Spangler S, Seeburg DP, Hoogenraad CC, Sheng M. 2005. Control of dendritic arborization by the phosphoinositide-3'-kinase-Akt-mammalian target of rapamycin pathway. *J Neurosci* 25:11300–11312.
- Kaizuka T, Hara T, Oshiro N, Kikkawa U, Yonezawa K, Takehana K, Iemura S, Natsume T, Mizushima N. 2010. Tti1 and Tel2 are critical factors in mammalian target of rapamycin complex assembly. *J Biol Chem* 285:20109–20116.
- Kumar V, Zhang MX, Swank MW, Kunz J, Wu GY. 2005. Regulation of dendritic morphogenesis by Ras-PI3K-Akt-mTOR and Ras-MAPK signaling pathways. *J Neurosci* 25:11288–11299.
- Lustig AJ, Petes TD. 1986. Identification of yeast mutants with altered telomere structure. *Proc Natl Acad Sci USA* 83:1398–1402.
- McMahon SB, Wood MA, Cole MD. 2000. The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc. *Mol Cell Biol* 20:556–562.
- Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P, Zecha A, Mohseni M, et al. 2011. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 478:57–63.
- Ricciardi S, Boggio EM, Grosso S, Lonetti G, Forlani G, Stefanelli G, Calcagno E, Morello N, Landsberger N, Biffo S, Pizzorusso T, Giustetto M, et al. 2011. Reduced AKT/mTOR signaling and protein synthesis dysregulation in a Rett syndrome animal model. *Hum Mol Genet* 20:1182–1196.

- Sabatini DM. 2006. mTOR and cancer: insights into a complex relationship. *Nature reviews. Cancer* 6:729–734.
- Shaw RJ, Cantley LC. 2006. Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature* 441:424–430.
- Shevchenko A, Roguev A, Schafit D, Buchanan L, Habermann B, Sakalar C, Thomas H, Krogan NJ, Stewart AF. 2008. Chromatin central: towards the comparative proteome by accurate mapping of the yeast proteomic environment. *Genome Biol* 9:R167.
- Shiloh Y. 2003. ATM and related protein kinases: safeguarding genome integrity. *Nature reviews. Cancer* 3:155–168.
- Takai H, Wang RC, Takai KK, Yang H, de Lange T. 2007. Tel2 regulates the stability of PI3K-related protein kinases. *Cell* 131:1248–1259.
- Takai H, Xie Y, de Lange T, Pavletich NP. 2010. Tel2 structure and function in the Hsp90-dependent maturation of mTOR and ATR complexes. *Genes Dev* 24:2019–2030.
- Troca-Marin JA, Alves-Sampaio A, Montesinos ML. 2012. Deregulated mTOR-mediated translation in intellectual disability. *Prog Neurobiol* 96: 268–282.
- Venteicher AS, Meng Z, Mason PJ, Veenstra TD, Artandi SE. 2008. Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell* 132:945–957.
- Wullschleger S, Loewith R, Hall MN. 2006. TOR signaling in growth and metabolism. *Cell* 124:471–484.
- Wurdak H, Zhu S, Romero A, Lorger M, Watson J, Chiang CY, Zhang J, Natu VS, Lairson LL, Walker JR, Trussell CM, Harsh GR, et al. 2010. An RNAi screen identifies TRRAP as a regulator of brain tumor-initiating cell differentiation. *Cell Stem Cell* 6: 37–47.
- Yamashita A, Kashima I, Ohno S. 2005. The role of SMG-1 in nonsense-mediated mRNA decay. *Biochim Biophys Acta* 1754:305–315.



**SUPPORTING INFORMATION FOR THE ARTICLE**

**Mutation in *TTI2* reveals a role for Triple T complex in human brain development**

Maéva Langouët, Abdelkrim Saadi, Guillaume Rieunier, Sébastien Moutton, Karine Siquier-Pernet, Marie Fernet, Patrick Nitschke, Arnold Munnich, Marc-Henri Stern, Malika Chaouch, Laurence Colleaux

## **Supp. Materials and Methods**

### **Blood and skin patient samples**

Blood samples were collected for all affected and unaffected siblings and the mother, and skin fibroblasts were cultured from patient III-2. Institutional research ethic approval and written consent were obtained for all participants in the study.

### **Genotyping and linkage analysis**

Single nucleotide polymorphism (SNP) genotyping from individuals II-1, III-1, III-2, III-5, III-9 was carried out on GeneChip® Human Mapping 250K Array (Affymetrix, Santa Clara, CA, USA). Linkage analysis was performed with the Merlin program using a fully penetrant autosomal-recessive inheritance model.

### **Exon enrichment and High-throughput sequencing**

The exome of individual III-9 was captured using the Agilent SureSelect Human All Exon 38 Mb kit (Agilent Technologies, Santa Clara CA, USA) and sequenced on a Genome Analyzer IIx (Illumina, San Diego, CA, USA). 3.8 Gb of sequences were produced allowing a mean sequence coverage of 50 reads per bp position, with more than 83% of targeted bases covered at 15X. Sequence reads were aligned to the human reference genome sequence (assembly GRCh37/hg19) using Burrow-Wheeler Aligner. SNPs/indels were called using Sequence Alignment/Map tools (SAMtools) Pileup. Poorly mapped (less than 3X cover) and low-quality reads (less than 20 quality score) were removed.

### **Cell culture and protein extraction**

Primary fibroblasts were cultured at 37°C under 5% CO<sub>2</sub> in RPMI® + glutamax or OPTI-MEM® + glutamax supplemented with 10% of fetal bovine serum (FBS) and 5% of penicillin/streptomycin (complete medium) (Invitrogen). The adherent cells were washed in phosphate-buffered saline 1X PBS (Invitrogen) and then overlaid with a predetermined volume of the same regular lysis buffer (50 mM Tris pH 8, 170 mM NaCl, 0,5% NP-40, 50 mM NaF and Complete EDTA-free protease inhibitor from Roche Applied Science). Cells were scraped to concurrently dislodge and lyse the cells. After 30 min on ice and several vortex episodes, the lysates were clarified by centrifugation at 5,000 rpm for 1min and stored at -80°C until immunoblotting. The

Bradford protein assay was used to determine the concentration of each sample (B6916-500 Sigma).

### **Western blotting experiments and quantification**

30-40µg of protein extracts were resolved by 4-15% SDS-PAGE (SDS-PAGE Mini-PROTEAN® TGX™ #456-1086 Biorad) and transferred to 0.2 µm nitrocellulose membranes (Biorad). Membranes were blocked with 5% nonfat dried milk powder (Invitrogen) diluted in PBS-T (1XPBS with 0.2% Tween-20) and incubated overnight at 4°C with the primary antibodies in 0.5% milk in PBS-T at following dilutions: TTI2 (1/2,500, A303-476A, Bethyl), TTI1 (1/5,000, A303-451A, Bethyl), TELO2 (1/5,000, 15975-1-AP, Proteintech), and β-actin (1/10,000, AM4302, Ambion) used as a control for protein loading. Membranes were then incubated with horseradish peroxidase–conjugated secondary antibodies (donkey anti-rabbit IgG-HRP antibody (sc-2313, Santa Cruz) or donkey anti-mouse IgG-HRP antibody (sc-2314, Santa Cruz) 1/20,000 dilution. Proteins were visualized using ECL-Plus (GE Healthcare). Various exposure times were performed for either autoradiography films and AGFA development (Curix 60 n°1829) or Chemidoc system® (Biorad). The signal intensities were analyzed using relative quantification tool in the ImageLab software (Biorad).

For PIKK immunoblotting experiments, membranes were blocked in Odyssey blocking buffer (LI-COR #927-40000) for 1h at RT and incubated overnight at 4°C with the primary antibodies at the following dilutions: ATM (1/1000, NB110-55475, Novus Biologicals), DNA-PKcs (1/500, MS-423-P0, Thermo Scientific), MTOR (1/1000, 2983, Cell Signaling) and β-actin (1/10,000, A-5316, Sigma). The secondary antibodies coupled to IRDye 800CW (anti-mouse: 926-32212, LI-COR 1/10,000; and anti-rabbit: 926-32213, LI-COR, 1/10,000) were revealed using the Odyssey Quantitative Fluorescence Imaging System (Odyssey®, LI-COR) and band intensities were quantified using Odyssey software (LI-COR).

### **WEB RESOURCES**

The accession number and URLs for data presented herein are as follows:

Ensembl Genome Browser, <http://www.ensembl.org/>

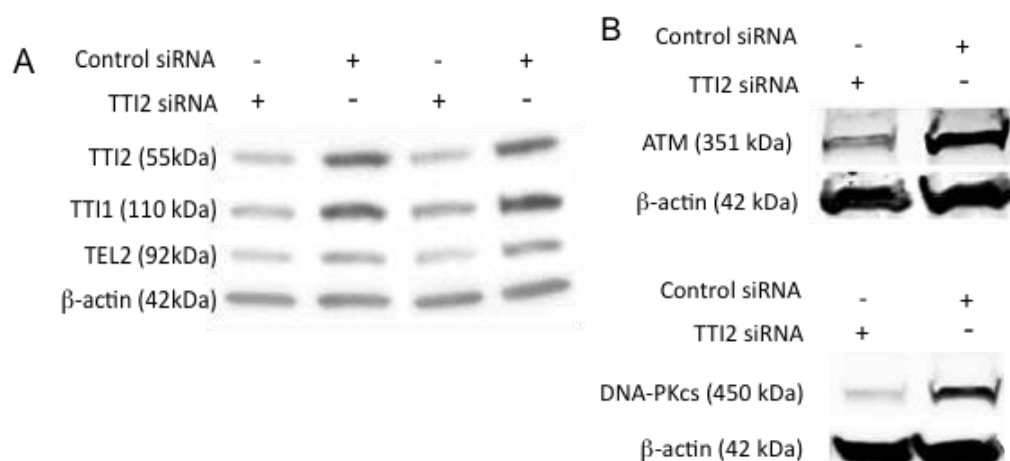
GenBank, <http://www.ncbi.nlm.nih.gov/genbank/>

National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/>

Patient	L	V	V	L	L	K	A	L	L	K	L	N	C	D	V	A	R	D	P	N	L	T	E	P	S
Homo sapiens	L	V	V	L	L	K	A	L	L	K	L	I	C	D	V	A	R	D	P	N	L	T	E	P	S
Bos taurus	L	V	V	L	L	K	A	L	L	K	L	I	C	D	V	A	R	D	S	S	L	T	E	P	C
Canis familiaris	L	V	V	L	L	K	A	L	L	K	L	I	C	D	V	A	R	D	P	N	L	T	E	P	P
Rattus Norvegicus	V	V	V	L	L	K	A	L	L	K	L	I	C	D	I	A	R	D	T	I	P	T	E	T	A
Mus musculus	V	V	V	L	L	K	A	L	L	K	L	I	C	D	I	S	R	D	T	I	P	T	E	T	A
Monodelphis domestica	L	E	V	F	L	K	T	L	L	K	M	V	C	D	V	V	K	D	Q	S	L	T	E	P	P
Gallus gallus	I	P	V	L	L	K	A	L	L	K	M	I	W	D	V	H	T	D	Q	G	S	T	E	P	P
Danio rerio	R	D	S	L	A	Q	S	L	L	R	F	L	V	D	V	S	S	D	S	-	-	S	T	P	E

**Supp. Figure S1.** Multiple Sequence Alignment of TTI2/C8orf41 representative orthologs shows amino acid conservation of the mutated residue across species.



**Supp. Figure S2.** TTI2 silencing in HEK293 cells. HEK293 cells were co-transfected with RNAi oligonucleotides directed against *TTI2* or a negative control RNAi oligonucleotides using Lipofectamine RNAi MAX (Invitrogen) according to the manufacturer's protocols. After 2 days, cells were harvested and knockdown of *TTI2* mRNA expression was confirmed by RT-qPCR. 24 hours later, proteins were extracted and immunoblotted successively with an anti-*TTI2*, anti-*TTI1*, anti-*TEL2* and anti- $\beta$ -actin antibody (**A**) or with anti-*ATM* and anti-*DNA-PKcs* antibody (**B**). *TTI2* silencing results in the depletion of the three TTT complex components as observed in the patients' cells. Similarly, a significant decreased of *ATM* and *DNA-PKcs* protein levels in depleted cells when compared to control cells.

**Supp. Table S1. Summary of the clinical characteristics of children with *TTI2* mutation**

Patient	III-2	III-5	III-9
Gender	M	F	F
born	04/10/1974	12/09/1979	03/05/1987
Prenatal/neonatal period	normal	normal	normal
Severe ID	+	+	+
Walk independently (years)	2.5	3	2.5
Speech	Few words	No words	No words
Stereotypies	+	+	+
Hyperactivity	+	+	+
Agressivity	+	+	+
Sleep disorder	-	-	+
Head circumference	50.5 cm (< - 4 SD)	50 cm (< - 3 SD)	48 cm (- 4 SD)
Height	150 cm (< - 4 SD)	143 cm (< - 3.5 SD)	135 cm (< - 4 SD)
Weight	42kg	48kg	48 kg
Vision	normal	normal	normal
Hearing	normal	normal	normal
Facial dysmorphy	+	+	+
Vertebral anomalies	cyphoscoliosis	scoliosis	cyphoscoliosis
	Dental malposition		Dental malposition
Other features	Hallus valgus	-	

OFC: occipitofrontal circumference, H: height, IgA ,IgM,IgE: immunoglobulin, m: month, NA: not available, nl: normal, y: year

**Supp. Table S2. IBD regions**

chromosome	position (hg19)	position	taille (Mb)	lod score	# gènes
8	31138127-42618286	8p12-p11.21	11.5	2.5327	60
	49536604-50232086	8q11.21	0.7	2.5327	3
	51348758-53047970	8q11.21-q11.23	1.7	2.5327	4
	58390922-60539305	8q12.1	2.1	2.4395	6
	70696034-72166253	8q13.3	1.5	2.4324	7
9	120229555-121661874	9q33.1	1.5	2.5327	2
	133843861-137153374	9q34.12-q34.2	3.5	2.5327	54
	137248166-139367457	9q34.2-q34.3	2.1	2.5327	35
16	18086557-22261697	16p12.3-p12.2	4.2	2.5327	50
	61101396-65382833	16q21	4.3	2.5327	3
	65420673-66861966	16q21-q22.1	1.5	2.5327	15
	68434883-70531366	16q22.1	2.1	2.5327	36
	71242062-72981230	16q22.2-q22.3	1.7	2.5326	23
22	26175564-29465677	22q12.1	3.3	2.5327	23
TOTAL			40.1		321

**Supp. Table S3. Number of variants identified by whole-exome sequencing**

<i>Total variants</i>	<i>128632</i>
Novel or rare homozygous variants (dbSNP130/1000GP/EVS)	4996
Novel or rare homozygous variants in IBD regions	98
Novel and Non-synonymous/indel/SS variants	2
not found in control chromosomes	1
Candidate Gene	<i>TT12</i>



## 6.4 Discussion et perspectives

### 6.4.1 Les principales caractéristiques cliniques observées dans les anomalies de la réparation de l'ADN

Plusieurs études ont démontré que le complexe triple T était nécessaire à la stabilité des protéines PIKKs. Parmi cette famille de 6 membres, 3 (ATM, ATR et DNA-PK) sont impliquées dans les mécanismes de réparation des CDB de l'ADN ; qu'il s'agisse de réparation par recombinaison homologue (Homologous Recombinational ou HR) ou par jonction d'extrémités non homologues (Non-Homologous End-Joining ou NHEJ). C'est la raison pour laquelle nous avons souhaité comparer les signes cliniques observés chez nos patients (MRT39) à ceux observés chez les malades porteurs de mutations affectant ces voies de réparation. Dans la table 6.1 ci-dessous, nous nous sommes donc concentrés sur les pathologies de la réparation dont les gènes responsables sont impliqués dans les mécanismes de réparation des CDB.

L'analyse de cette comparaison montre que la DI, la microcéphalie et le retard de croissance sont des signes retrouvés chez la majorité des malades. Les anomalies du squelette sont retrouvées dans l'anémie de Fanconi, le syndrome de Seckel et le déficit en cernunnos-XLF tandis qu'une dysmorphie faciale est observée dans les syndromes de Nijmegen, de Seckel et le déficit en cernunnos-XLF. En revanche, le tableau clinique des patients *TTI2* diffère significativement de celui de l'ataxie telangiectasie (A-T), due à des mutations du gène *ATM* et de sa variante l'A-T like disorder (A-TLD), causé par des mutations du gène *MRE11*. La DI n'est pas un critère majeur de la description clinique et aucune microcéphalie n'a été rapportée chez les patients souffrant de ces pathologies. Au niveau neurologique, on peut donc discerner deux groupes distincts au sein des pathologies de la réparation de l'ADN ; séparés par une double barre dans la table 6.1 : un ensemble de pathologies neurodéveloppementales (de la MRT30 au syndrome de Bloom) d'une part et d'autre part deux pathologies caractérisée par une neurodégénérescence progressive (à partir de l'âge de 1 à 2 ans chez les patients A-T).

Étant donné que les gènes responsables de l'A-T et l'A-TLD sont aussi impliqués dans les voies de la réparation des CDB (NHEJ et HR), nous pouvons nous demander la raison de cette divergence de phénotype neurologique. Une hypothèse a été proposée pour expliquer ce trait phénotypique [O'Driscoll and Jeggo, 2008, Rulten and Caldecott, 2013]. Lors du développement du cerveau de l'embryon, des vagues de prolifération des neuroprogéniteurs puis de migration et différenciation des cellules neuronales se succèdent, pendant lesquelles sont sollicitées les voies de réparation des CDB HR et NHEJ respectivement. Un dysfonctionnement de ces voies peut conduire à un excès de CDB entraînant une mort cellulaire importante et une microcéphalie. Dans ces cellules, la mort cellulaire serait activée par la protéine ATM et le complexe MRN. Par conséquent, en l'absence de ATM, les neurones portant des CDB non réparées risquent d'échapper à l'apoptose et persister malgré un dysfonctionnement de ceux-ci, entraînant une neuropathie juvénile, dégénérative et progressive.

De plus, un déficit immunitaire et une prédisposition au cancer, deux signes très souvent associés aux anomalies de réparation de l'ADN, ne sont pas rapportés chez nos patients. Ces signes ne sont pas non plus observés chez les patients avec un syndrome de Seckel dont le tableau clinique est le plus proche des patients EI.

Caractéristiques	Patients EI (MRT39)	Syndrome de Seckel	Syndrome de Nijmegen	Déficit en DNA ligase IV	Anémie de Fanconi	Déficit en cernunos-XLF (NHEJ1)	Syndrome de Bloom	Ataxie telangiectasie (A-T)	A-T-like disorder (A-TLD)
MIM	615541	210600	251260	606593	227650	611291	210900	208900	604391
Gène	<i>TTI2</i>	<i>ATR</i>	<i>NBS1/NBN</i>	<i>LIG4</i>	Gènes <i>FANC</i>	<i>XLF/CERNUNNOS</i>	<i>RECQL3/BLM</i>	<i>ATM</i>	<i>MRE11</i>
Substrats et Interaction(s) protéique(s)	complexe TTT et PIKKs	complexe TTT et ATRIP	complexe MRN	XRCC4	FANCD2 mono-ubiquitination, une cible des protéines ATM et ATR	XRCC4, LIG4, DNA-PK	substrat de ATM et ATR	complexe TTT et complexe MRN	ATM et composants du complexe MRN
Type de lésions de l'ADN	DSBs et SSBs	DSBs et SSBs	DSBs	DSBs	liaisons covalentes entre brins d'ADN	DSBs	DSBs	DSBs	DSBs
Voie de la réparation concernée	NHEJ et HR	HR	NHEJ et HR	NHEJ	HR et Réparation des liaisons covalentes	NHEJ	HR	NHEJ	NHEJ et HR
Troubles du comportement	+	+/-	-	-	-	-	-	-	-
<b>Phénotype neurologique</b>	DI	DI	DI	DI	DI : +/-	DI : +/-	DI : +/-	ataxie cérébelleuse progressive et apraxie oculomotrice. 30% des cas avec une DI modérée ou des difficultés d'apprentissage	apraxie oculomotrice
Microcéphalie	+	+	+	+	+/-	+	+	-	-
Retard de croissance	+	+	+	+	+	+	+	+/-	-
Anomalies du squelette	+	+	-	-	+	+/-	-	+	-
Dysmorphie faciale	+	+	+	-	-	+	-	-	-
Immuno-déficience	-	-	+	+	+	+	+	+	+
Prédisposition au cancer	-	-	+	+	+	-	+	+	-

**Figure 6.1** : Caractéristiques cliniques des maladies liées à une anomalie de la réparation.

DSBs : Double-Strand Breaks, SSBs : Single-Strand Breaks, HR : Homologous Recombinational, NHEJ : Non-Homologous End-Joining.

Toutefois, un bilan immunologique a été effectué chez le patient III.2 afin de déceler d'éventuelles altérations. Une légère diminution des pourcentages de lymphocytes T CD4 et T CD8 naïfs et une augmentation des lymphocytes T mémoire, notamment des T CD8 effecteurs mémoire RA+ (TEMRA), ont été observées (table B.2, en annexe). Ce phénotype est observé, plus sévère, dans les pathologies de la réparation avec atteinte thymique, comme le syndrome de Nijmegen, l'A-T et le déficit en cernunnos-XLF.

Enfin, le dosage des marqueurs tumoraux (table B.3, en annexe) s'est révélé normal chez les 3 patients, excepté pour l'alpha-fœtoprotéine (AFP) : plus de 3 fois la normale chez les filles (III.5 et III.9) et plus de 2 fois chez le garçon (III.2). L'augmentation de ce marqueur est caractéristique du carcinome hépato-cellulaire (CHC) qui représente environ 90% des cancers primitifs du foie. Même si les chiffres ne sont pas très élevés, le neurologue en charge de la famille propose de surveiller de façon régulière les patients sur le plan clinique, biologique (notamment par le dosage de l'AFP) et éventuellement radiologique (scanner thoraco abdominal). Le phénotype des patients mutés dans le gène *TTI2* s'inscrit donc clairement dans le spectre des maladies de la réparation de l'ADN.

### 6.4.2 Études des voies de signalisation des PIKKs

Étant donné l'impact observé sur les niveaux protéiques des différentes PIKKs testées, nous avons souhaité étudier certaines des voies de signalisation dans lesquelles ces protéines sont impliquées. Nous avons donc débuté des collaborations avec deux groupes dont les centres d'intérêt rejoignent certaines de ces voies et leur avons envoyé les fibroblastes du patient III.2.

#### i) ATM et la réponse au stress oxydatif

En 2012, nous avons eu l'occasion de collaborer avec le Pr Dominique Stoppa-Lyonnet et l'équipe du Dr Marc-Henri Stern (Institut Curie, Paris) dont le sujet de recherche est la génétique et la biologie des cancers. D'après Barascu et al. [2012], le stress oxydatif entraîne une sur-expression de la lamine B, elle-même responsable d'une altération de l'architecture des noyaux des cellules. La protéine ATM est un capteur important des niveaux de ROS dans les cellules humaines et contrôle la réponse au stress oxydatif. Le stress oxydatif est une caractéristique de l'A-T et une altération de la structure des noyaux est observée dans les cellules de patients avec A-T. Un excès de ROS (Reactive Oxygen Species) pourrait causer les défauts neurologiques observés chez nos patients, nos collaborateurs ont donc souhaité étudier ces paramètres dans les cellules de patients *TTI2*. Les analyses de circularité des noyaux et de niveau de ROS, menées par le Dr Guillaume Rieunier à l'institut Curie, n'indiquent pas de différences significatives entre les fibroblastes de patients et de contrôles (figures A.12, A.13 et A.14, en annexe).

#### ii) Les voies de signalisation de ATM et ATR

Nous avons également établi une collaboration avec le groupe du Pr Mark O'Driscoll (Université de Sussex, Falmer, Brighton, UK) dont les recherches se concentrent sur les maladies de la réparation des lésions de l'ADN. Ce groupe a montré qu'outre la diminution des niveaux protéiques des protéines ATM et DNA-PK, une réduction significative des protéines ATR et SMG1 était également observée (données non présentées).

Dans un second temps, nos collaborateurs souhaitent évaluer les effets de la diminution de ces protéines sur l'activation des voies de signalisation spécifiques dans lesquelles elles sont impliquées. Ils ont étudié la cinétique de la voie de réparation des CDB dépendante de ATM en observant la formation des foci 53BP1 mais aussi le statut phosphorylé vs non-phosphorylé de la protéine CHK1, dépendant de ATR. Les données préliminaires de ces tests n'indiquent pas de défauts majeurs dans ces voies de signalisation. L'utilisation d'inhibiteurs de ATM et ATR a permis de démontrer qu'il existe une activité résiduelle de ces protéines dans les cellules de patients.

### iii) La voie de signalisation mTOR

En parallèle des essais menés dans notre laboratoire, le groupe de M. O'Driscoll a analysé la phosphorylation de la protéine p70-S6Kinase (Thr389) induite par l'EGF (Epidermal Growth Factor) pour évaluer l'activité de mTOR (figure A.15 en annexe). Nos résultats ainsi que les leurs ne montrent pas de défaut significatif (données non présentées).

### 6.4.3 Perspectives et conclusion préliminaires

Les premiers tests réalisés afin d'explorer les voies de signalisation sous-jacentes aux 3 PIKKs analysées dans les cellules de patients n'ont pas permis de mettre en évidence de défauts majeurs dans ces voies. Étant donné qu'il subsiste entre 24 et 50% de protéines, l'activité résiduelle existante est vraisemblablement suffisante pour assurer une signalisation correcte dans les fibroblastes. Cependant, la mutation *TTI2* pourrait avoir un impact plus important dans d'autres types cellulaires pour lesquels la stabilité des PIKKs est primordiale et les voies de signalisation associées à ces protéines sont cruciales. Notamment, la réponse aux dommages de l'ADN, assurant une stabilité génomique et la survie cellulaire, apparaît très importante dans les cellules du cerveau en développement et son dysfonctionnement engendre une atteinte neurologique dans de nombreuses pathologies de la réparation [O'Driscoll and Jeggo, 2008]. Le contrôle de la croissance et de la prolifération cellulaire demeure également une fonction indispensable à la mise en place du système nerveux et au développement en général. La microcéphalie, par exemple, un symptôme largement observé dans le groupe des anomalies de la réparation et chez nos patients, résulte d'une augmentation de la mort cellulaire ou d'un défaut de prolifération des cellules souches neuronales pendant le développement cérébral [O'Driscoll and Jeggo, 2008]. Aussi, une étude d'expression sur des cDNA provenant de divers tissus humain montre que le gène *TTI2* est principalement exprimé dans le thymus, les testicules et le cerveau (figure A.16, en annexe), suggérant l'importance de la stabilité des PIKKs dans ces tissus.

En résumé, le phénotype des patients EI est dû à une anomalie de dosage plutôt qu'à une perte de fonction des PIKKs, puisque les protéines ATM, ATR et DNA-PK sont toujours fonctionnelles, bien que moins concentrées. Cependant la stœchiométrie affaiblie des trois protéines est susceptible d'entraîner une diminution générale de l'efficacité de la réparation des CDB. Nous proposons donc que le phénotype des patients *TTI2* résulte de l'effet cumulatif d'une moindre efficacité des différentes voies associées aux 6 PIKKs, dans des types cellulaires et à des stades de développement où celles-ci sont particulièrement activées.

Dans ce cas de figure précis, l'utilisation des fibroblastes de patients semble limitante. Afin d'évaluer les effets de la mutation p.I436N du gène *TTI2* dans le SNC, d'autres systèmes peuvent apporter de nouveaux éléments de réponses. Il serait intéressant, par exemple, d'évaluer l'effet de l'interférence du gène *TTI2* dans

des cultures neuronales issues de souris et ainsi vérifier notre hypothèse. De la même manière, l'analyse de l'inactivation du gène dans des modèles animaux tels que le zebrafish nous permettrait de vérifier que celle-ci entraîne une atteinte neurologique, comme observé chez les patients.

Enfin, la publication de ce travail en 2013 a conduit très récemment à une nouvelle collaboration avec le Pr Andrew Crosby (University of Exeter Medical School, Exeter, UK). Une famille Iranienne dont les deux enfants atteints présentent un retard psychomoteur, une absence de langage et des retards de développement nous a été confiée. Une unique région IBD de 14 Mb a été identifiée par un tour de génome chez cette famille, dont le degré de consanguinité des parents est faible. Cette région contient le gène *TTI2*. Les cliniciens suivant cette famille ont déclaré que le phénotype s'accorde avec celui des patients EI. Nous avons donc entrepris le séquençage du gène entier chez un des patients de la famille ainsi que chez les parents sains.

## CHAPITRE 7

# LE NGS, UN OUTIL POUR DÉFINIR DE NOUVEAUX MODES DE TRANSMISSION GÉNÉTIQUE

### 7.1 La famille, les patients

La famille Ma, d'origine algérienne, est suivie par le Dr Jeanne Amiel à l'hôpital Necker-Enfants Malades. Les deux enfants issus d'une union consanguine sont atteints d'une DI sévère, un retard psychomoteur, une spasticité et une obésité précoce très sévère qui a débuté à 1 an et demi pour le patient II.1 et à 6 mois pour la patiente II.2.

### 7.2 Stratégie

Afin de déterminer les causes génétiques associées à ce tableau clinique, nous avons analysé par WES chaque membre de la famille. Connaissant l'existence d'une boucle de consanguinité dans cette famille, nous avons fait l'hypothèse d'une transmission autosomique récessive. Bien que ce mode de transmission semble la meilleure hypothèse, nous disposons des données de WES pour l'ensemble de la famille, nous permettant de tester diverses hypothèses lors de l'analyse des résultats.

### 7.3 Résultats : publication n°2 (en préparation)

## **Contiguous mutation syndrome in the era of high-throughput sequencing**

Maéva LANGOUËT (1), Karine SIQUIER-PERNET (1), Sylvia SANQUER (2),  
Christine BOLE-FEYSOT (3), Patrick NITSCHKE (4), Nathalie BODDAERT (1,5),  
Arnold MUNNICH (1), Grazia M S Mancini (6), Robert BAROUKI (2), Jeanne AMIEL  
(7), Laurence COLLEAUX (1)

(1) INSERM UMR 1163, Laboratory of Molecular and pathophysiological bases of  
cognitive disorders, Paris Descartes – Sorbonne Paris Cité University, Imagine  
Institute, Necker-Enfants Malades Hospital, 75015 Paris, France

(2) Metabolic and Proteomic biochemistry service, Necker-Enfants Malades hospital,  
AP-HP, 75015 Paris

(3) Genomic platform, INSERM UMR 1163, Paris Descartes– Sorbonne Paris Cité  
University, Imagine Institute, 75015 Paris, France

(4) Bioinformatic platform, INSERM UMR 1163, Paris Descartes– Sorbonne Paris  
Cité University, Imagine Institute, 75015 Paris, France

(5) Service de radiologie pédiatrique, Hôpital Necker-Enfants Malades, AP-HP,  
75015 Paris, France

(6) Department of Clinical genetics, Erasmus University Medical Center, 3015 CN  
Rotterdam, The Netherlands

(7) Service de Génétique, Hôpital Necker-Enfants Malades, AP-HP, 75015 Paris,  
France

Correspondance to :

Dr. Laurence COLLEAUX

INSERM UMR 1163, Institut IMAGINE, 24 boulevard du Montparnasse, 75015 Paris,  
France

Email: [laurence.colleaux@inserm.fr](mailto:laurence.colleaux@inserm.fr), phone: +33 1 42 75 42 96, fax: +33 1 42 75 42  
21

Prof Jeanne AMIEL

Service de Génétique, Hôpital Necker-Enfants Malades, 149 Rue de Sèvres, 75015  
Paris, France

Email: [jeanne.amiel@inserm.fr](mailto:jeanne.amiel@inserm.fr), phone: + 33 1 44 49 56 48, fax: + 33 1 44 49 51 50

## ABSTRACT

We ascertained two sibs, born to consanguineous parents, with severe encephalopathy, progressive microcephaly and spasticity, thin corpus callosum, and early onset obesity. Whole-exome sequencing identified two homozygous substitutions in two genes 170 kb apart on 7q22.1: a c.1137+1G>T splice mutation in *AP4M1* previously described in a familial case of AP4-deficiency syndrome and the *AZGP1* c.595A>T missense variant. Haplotyping analysis indicated a founder effect of the *AP4M1* mutation, whereas the *AZGP1* mutation arose more recently in the family we describe.

AP4-deficiency syndrome is characterized by severe intellectual disability, progressive microcephaly and spasticity but obesity is not a clinical feature. *AZGP1* encodes an adipokine that stimulate lipolysis in adipocytes and regulates body weight in mice. We propose that the sibs' phenotype results from the combined effects of mutations in both *AP4M1* and *AZGP1* that account for the neurological signs and the morbid obesity of early onset, respectively.

Contiguous gene syndromes are the consequence of loss of two or more adjacent genes and the phenotype reflects a combination of the endophenotypes caused by the haploinsufficiency of each gene. We propose to broaden this concept to disease phenotypes resulting from independent mutations in 2 genetically linked genes causing a contiguous mutation syndrome.

**Key Words:** Whole-exome sequencing, intellectual deficiency, obesity, AP-4 complex, zinc-alpha2-glycoprotein



We investigated two sibs (one female and one male) born to consanguineous Algerian parents (figure 1A) presenting the association of severe intellectual deficiency (ID), psychomotor delay, progressive microcephaly, spasticity, thin corpus callosum, and severe obesity with early onset (table 1 and figure 1B-C).

To identify the disease-causing mutation of this undiagnosed condition, we performed whole-exome sequencing (WES) on peripheral-blood DNA from all family members. We first focused on novel homozygous variants cosegregating with the disease and corresponding to either nonsynonymous (NS) variants, splice acceptor and donor site mutations (SS) or coding insertions/deletions (indels). We regarded variants as novel if they were absent from all publically available data sets, including those of dbSNP138 (<http://www.ncbi.nih.gov/SNP>), the 1000 Genomes Project (<http://browser.1000genomes.org/index.html>), the NHLBI ESP Exome Variant Server (<http://evs.gs.washington.edu/EVS/>), and from in-house exome data containing information for over 600 patient samples. Three candidate variants fulfilled these criteria (table S1) and were confirmed by Sanger sequencing: a splice variant in the *AP4M1* gene (NM\_004722.3: c.1137+1G>T), a missense variation in the *AZGP1* gene (NM\_001185.3: c.595A>T, p.Asn199Tyr) and a missense variation in exon 81 of *HERC2* (NM\_004667.4: c.12445G>A, p.Ala4149Thr) (figure 1D). Filtering WES data for compound heterozygous mutations or heterozygous “de novo” mutations shared by the sibs did not provide any additional candidate variants.

The *AP4M1* splice variant affects the donor splice site of intron 14 (figure 1D) and was previously described in 5 affected sibs from a consanguineous Moroccan family with a highly similar neurological presentation but no obesity was reported [1]. As previously demonstrated by Verkerk and colleagues, this variant leads to skipping of exon 14 in the *AP4M1* transcript extracted from patient’s cultured skin fibroblast (figure 1E). *AP4M1* encodes one of the subunit ( $\mu$ ) of the adaptor protein (AP) complex 4, an evolutionary conserved heterotetrameric complex consisting of two large ( $\alpha$  or  $\beta$ ), a medium ( $\mu$ ), and a small ( $\sigma$ ) adaptin. Mutations affecting all four subunits of AP4 (*AP4M1*, *AP4E1*, *AP4S1*, and *AP4B1*) have been found to cause autosomal recessive AP-4 deficiency syndrome characterized by severe ID, microcephaly, progressive spastic paraplegia, speech delay and growth retardation (table 1 and table S2) [2].

The p.Ala4149Thr substitution in *HERC2* affects an absolutely conserved amino acid residue located within the Regulator of Chromosome Condensation (RCC1) protein

domain and is predicted to be damaging by various *in silico* tools (PolyPhen-2 score: 0.968, SIFT score: 0 and disease causing according to Mutation Taster). The *HERC2* gene is located on 15q13.1 and encodes a large ubiquitin ligase protein. A distinct *HERC2* founder mutation (c.1781C>T, p.Pro594Leu) has been previously associated in the Amish community with a disorder characterized by mild developmental delay, Autism Spectrum Disorder and Angelman-like features (MIM:615516) [3, 4]. Functional studies of the p.Pro594Leu variant demonstrated that it induces protein aggregation and decreased *HERC2* abundance.

The third variant (c.595A>T, p.Asn199Tyr) affects the *AZGP1* gene, encoding the zinc-alpha2-glycoprotein (ZAG). The c.595A>T variant does not affect a highly conserved amino acid residue and its consequences on protein function is unclear (PolyPhen-2 score: 0.946, SIFT score: 0.01 and Polymorphism according to Mutation Taster). *AP4M1* and *AZGP1* are located on chromosome 7q22.1, 170kb apart from each other (figure 1F). To discriminate between a hot spot and a founder mutation, we performed genotyping analysis in all family members as well as in 2 affected sibs from the Moroccan family using microsatellite markers encompassing the *AP4M1* locus. As shown in table S3, both families share a common haplotype supporting the hypothesis of a founder effect of the *AP4M1* mutation. The c.595A>T *AZGP1* mutation was not present in the previously described Moroccan family (table S3). These data strongly support the hypothesis that the *AZGP1* mutation arose more recently in our family and occurs secondarily to the *AP4M1* mutation and on the same haplotype.

Our patients are severely retarded. They have some features of the autistic spectrum but marked and prolonged ocular contacts (table 1). Although two AP-4 patients have been described with a shy and anxious character [5], the majority of patients with AP-4 deficiency are described as shy, amicable and calm (table 1) [1, 2]. Collectively, these data predicted only a minor effect of the p.Ala4149Thr *HERC2* variant identified in our family and demonstrated that the c.1137+1G>T *AP4M1* variant clearly accounts for the neurological features observed in our patients.

However, none of the 33 previously reported AP-4 deficiency cases are reported as obese, while obesity was described as a feature of the 2 sibs' condition as it started very early on in both (before 1 year of age, table 1 and supplemental clinical data). Similarly, no obesity has been reported in *HERC2* mutated patients. Whilst, we cannot exclude that the *HERC2* variant may contribute to the neurological

presentation of our cases, it is unlikely that it accounts for the obesity. By contrast, ZAG is an adipokine secreted by the adipose tissue and playing an important role in the mobilization and utilization of stored lipids [6]. Furthermore, several studies support the role of the ZAG protein in the regulation of body weight in both animal models and humans. Reduced plasmatic ZAG levels have been observed in ob/ob mice [7] and *Azgp1* deficient mice shown increased body weight and decreased adipocytic lipolysis [8]. Moreover, oral administration of human ZAG to ob/ob mice resulted in progressive loss of body weight [9]. Linkage analyses showed linkage disequilibrium mapping of genes influencing human obesity, insulin resistance and Type 2 Diabetes in the 7q22.1 region [10-14] and the rs4215 SNP in *AZGP1* gene is associated with obesity in Chinese populations [15]. Finally, both lower serum concentrations of ZAG and significantly lower ZAG expression in the adipose tissue and liver have been observed in obese subjects [6, 16-18]. Further investigations will be needed to evaluate the ability of the mutant p.Asn199Tyr ZAG protein to induce lipolysis but these results make *AZGP1* an excellent candidate gene for very early onset obesity in humans and may provide interesting clues for novel therapeutic interventions in obese patients.

Altogether, our results suggest that the phenotype observed in our patients results from the additional effects of *AP4M1* and *AZGP1* mutations accounting for the neurological signs on the one hand and the precocious morbid obesity on the other hand. Additional subtle effects of the *HERC2* variant on the neurological presentation, particularly on the communicative skills, cannot be excluded. The current report further demonstrates how consanguinity could increase intrafamilial clustering of multiple hereditary diseases and how WES has considerably empowered our ability to detect such complex events.

The term contiguous gene syndromes (CGS) has been proposed in 1986 to explain the association of unrelated clinical features due to the deletion of multiple genes lying in close proximity to one another on a single chromosome [19]. The phenotypes observed arise as a result of the combination of the endophenotypes from each deleted gene sensitive to haploinsufficiency. While CGS is caused by a single mutational event (i.e a chromosomal deletion), the phenotype we describe is the consequence of two independent mutational hits in two genetically linked genes. The mode of inheritance we describe is also different from a digenic inheritance that refers to the alteration of two interacting genes to cause a phenotype. While the final

demonstration of our hypothesis is awaiting the identification of *AZPG1* mutations in obese patients with normal neurodevelopment, we propose to use the name contiguous mutation syndrome to describe this complex and clinically challenging phenotype caused by independent mutations in genetically linked genes.

## **ACKNOWLEDGEMENTS**

We are grateful to the patients and their family members for their participation in our study. We thank Pr Josué Feingold and Dr Valérie Malan for helpful comments. We also thank M Tisdale for his help in the analysis of the mutant ZAG protein. This study was supported by the Centre National de la Recherche Scientifique (CNRS), the Agence Nationale de la Recherche (ANR SAMENTA 2012), the Fondation pour la Recherche Médicale (DEQ20120323702) and Ministère de la Recherche et de l'Enseignement Supérieur.

*Disclosure Statement:* The authors declare no conflict of interest

## **AUTHOR CONTRIBUTIONS**

LC and JA designed the study. JA and GMM recruited and evaluated the study subjects. NB performed and analysed the human brain imaging. ML and KSP analyzed WES data. ML performed transcript analysis on patient cells. SM contributed to WES data analysis. CBF performed WES. SS and RB measured the ZAG serum level. PN performed the bioinformatic studies. All authors contributed to writing and editing the manuscript.

## **REFERENCES**

1. Verkerk AJ, Schot R, Dumeé B, Schellekens K, Swagemakers S, Bertoli-Avella AM, Lequin MH, Dudink J, Govaert P, van Zwol AL, Hirst J, Wessels MW, Catsman-Berrevoets C, Verheijen FW, de Graaff E, de Coo IF, Kros JM, Willemsen R, Willems PJ, van der Spek PJ, Mancini GM. Mutation in the AP4M1 gene provides a model for neuroaxonal injury in cerebral palsy. *Am J Hum Genet* 2009;85:40-52.
2. Abou Jamra R, Philippe O, Raas-Rothschild A, Eck SH, Graf E, Buchert R, Borck G, Ekici A, Brockschmidt FF, Nothen MM, Munnich A, Strom T, Reis A, Colleaux L. Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am J Hum Genet* 2011;88:788-95.
3. Puffenberger EG, Jinks RN, Wang H, Xin B, Fiorentini C, Sherman EA, Degrazio D, Shaw C, Sougnez C, Cibulskis K, Gabriel S, Kelley RI, Morton DH, Strauss KA. A

- homozygous missense mutation in HERC2 associated with global developmental delay and autism spectrum disorder. *Hum Mutat* 2012;33:1639-46.
4. Harlalka GV, Baple EL, Cross H, Kuhnle S, Cubillos-Rojas M, Matentzoglou K, Patton MA, Wagner K, Coblentz R, Ford DL, Mackay DJG, Chioza BA, Scheffner M, Rosa JL, Crosby AH. Mutation of HERC2 causes developmental delay with Angelman-like features. *J Med Genet* 2013;50:65-73.
  5. Abdollahpour H, Alawi M, Kortum F, Beckstette M, Seemanova E, Komarek V, Rosenberger G, Kutsche K. An AP4B1 frameshift mutation in siblings with intellectual disability and spastic tetraplegia further delineates the AP-4 deficiency syndrome. *Eur J Hum Genet* 2014.
  6. Balaz M, Vician M, Janakova Z, Kurdiova T, Surova M, Imrich R, Majercikova Z, Penesova A, Vlcek M, Kiss A. Subcutaneous adipose tissue zinc-alpha2-glycoprotein is associated with adipose tissue and whole-body insulin sensitivity. *Obesity (Silver Spring)* 2014;22:1821-9.
  7. Mracek T, Gao D, Tzanavari T, Bao Y, Xiao X, Stocker C, Trayhurn P, Bing C. Downregulation of zinc-alpha2-glycoprotein in adipose tissue and liver of obese ob/ob mice and by tumour necrosis factor-alpha in adipocytes. *J Endocrinol* 2010;204:165-72.
  8. Rolli V, Radosavljevic M, Astier V, Macquin C, Castan-Laurell I, Visentin V, Guigne C, Carpenne C, Valet P, Gilfillan S, Bahram S. Lipolysis is altered in MHC class I zinc-alpha(2)-glycoprotein deficient mice. *FEBS Lett* 2007;581:394-400.
  9. Russell ST, Tisdale MJ. Role of beta-adrenergic receptors in the oral activity of zinc-alpha2-glycoprotein (ZAG). *Endocrinology* 2012;153:4696-704.
  10. Sim X, Ong RT, Suo C, Tay WT, Liu J, Ng DP, Boehnke M, Chia KS, Wong TY, Seielstad M, Teo YY, Tai ES. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet* 2011;7:e1001363.
  11. Sell SM, Altungoz O, Prowse AA, Meloni AM, Surti U, Sandberg AA. Molecular analysis of chromosome 7q21.3 in uterine leiomyoma: analysis using markers with linkage to insulin resistance. *Cancer Genet Cytogenet* 1998;100:165-8
  12. Sanghera DK, Blackett PR. Type 2 Diabetes Genetics: Beyond GWAS. *J Diabetes Metab* 2012;3:198 .
  13. Li WD, Li D, Wang S, Zhang S, Zhao H, Price RA. Linkage and linkage disequilibrium mapping of genes influencing human obesity in chromosome region 7q22.1-7q35. *Diabetes* 2003;52:1557-61
  14. Huang J, Sabater-Lleal M, Asselbergs FW, Tregouet D, Shin SY, Ding J, Baumert J, Oudot-Mellakh T, Folkersen L, Johnson AD, Smith NL, Williams SM, Ikram MA, Kleber ME, Becker DM, Truong V, Mychaleckyj JC, Tang W, Yang Q, Sennblad B, Moore JH, Williams FMK, Dehghan A, Silbernagel G, Schrijvers EMC, Smith S, Karakas M, Tofler GH, Silveira A, Navis GJ, Lohman K, Chen MH, Peters A, Goel A, Hopewell JC, Chambers JC, Saleheen D, Lundmark P, Psaty BM, Strawbridge RJ, Boehm BO, Carter AM, Meisinger C, Peden JF, Bis JC, McKnight B, Ohrvik J, Taylor K, Franzosi MG, Seedorf U, Collins R, Franco-Cereceda A, Syvänen AC, Goodall AH, Yanek LR, CM, Müller-Nurasyid M, Folsom AR, Basu S, Matijevic N, van Gilst WH, Kooner JS, Hofman A, Danesh J, Clarke R, Meigs JB, Kathiresan S, Reilly MP, , Klopp N, Harris TB, Winkelmann BR, Grant PJ, Hillege HL, Watkins H, Spector TD, Becker LC, Tracy RP, Marz W, Uitterlinden AG, Eriksson P, Cambien F, , Morange PE, Koenig W, Soranzo N, van der Harst P, Liu Y, O'Donnell CJ, Hamsten A. Genome-wide association study for circulating levels of PAI-1 provides novel insights into its regulation. *Blood* 2012;120:4873-81.

15. Zhu HJ, Dong CX, Pan H, Ping XC, Li NS, Dai YF, Wang LJ, Yang HB, Zhao WG, Gong FY. rs4215 SNP in zinc-alpha2-glycoprotein gene is associated with obesity in Chinese north Han population. *Gene* 2012;500:211-5.
16. Selva DM, Lecube A, Hernandez C, Baena JA, Fort JM, Simo R. Lower zinc-alpha2-glycoprotein production by adipose tissue and liver in obese patients unrelated to insulin resistance. *J Clin Endocrinol Metab* 2009;94:4499-507.
17. Mracek T, Ding Q, Tzanavari T, Kos K, Pinkney J, Wilding J, Trayhurn P, Bing C. The adipokine zinc-alpha2-glycoprotein (ZAG) is downregulated with fat mass expansion in obesity. *Clin Endocrin* 2010;72:334-41.
18. Garrido-Sanchez L, Garcia-Fuentes E, Fernandez-Garcia D, Escote X, Alcaide J, Perez-Martinez P, Vendrell J, Tinahones FJ. Zinc-alpha 2-glycoprotein gene expression in adipose tissue is related with insulin resistance and lipolytic genes in morbidly obese patients. *PLoS One* 2012;7:e33264.
19. Schmickel RD. Contiguous gene syndromes: a component of recognizable syndromes. *J Pediatr* 1986;109:231-41.

## LEGEND TO FIGURES

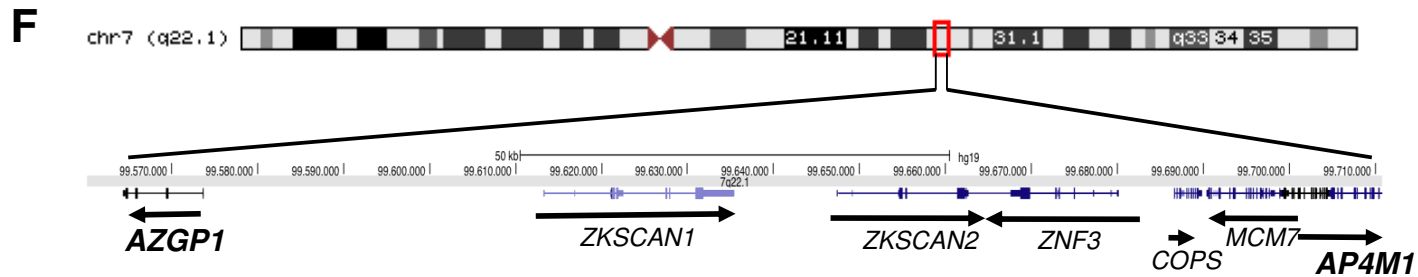
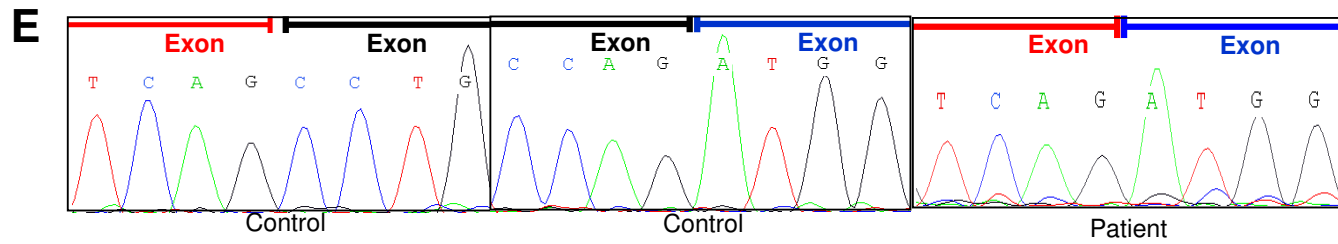
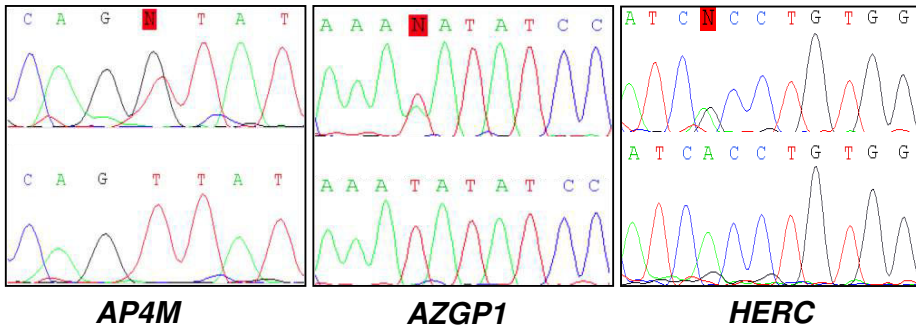
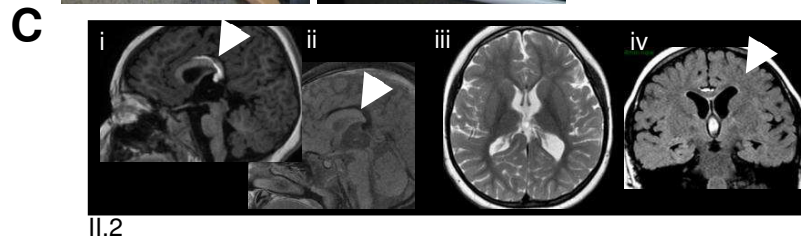
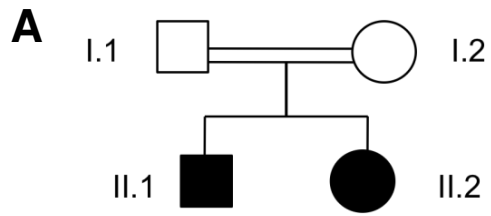
**Figure 1:** The combination of two substitutions in *AP4M1* and *AZGP1* genes underlies the association of two distinct syndromes in both patients.

**A:** Pedigree of the family. Shaded symbols indicate the affected individuals. **B:** Photographs of patients II.1 (left) and II.2 (right). Legal representative of the patients gave consent for the publication of these photographs. **C:** Brain MRI of patient II.2 at 4 years-old (left panel) and a 4 years-old normal control girl (right panel): Sagittal T1 (i) and sagittal T1 fat sat (ii), axial T2 weighted FSE images (iii) and coronal FLAIR weighted FSE images (iv). Sagittal T1 (i) and sagittal T1 fat Sat (ii) of the affected girl show a lipoma of the posterior part of the corpus callosum (white arrows). Axial T2 (iii) show a moderate ventricular dilatation compared to age matched normal control. The Coronal Flair (iv) of the affected girl shows delayed myelination of the white matter (arrow) compared to age matched normal control. **D:** Sequence analysis in a patient (II.1) and an healthy parent (I.1) showing the c.1137+1G>T; the c.595A>T, p.Asn199Tyr and the c.12445G>A, p.Ala4149Thr variants in the *AP4M1*, *AZGP1* and *HERC2* genes respectively. **E:** Exon 14 skipping in *AP4M1* mRNA from patients' fibroblasts compared to controls'. **F:** Drawing showing the 7q22.1 region and the the two mutated genes located 170kb apart.

**Table 1:** Clinical findings in patients with mutations in AP4 subunits

	This report		Previously reported patients with AP4-deficiency
	II.1	II.2	
Sex	M	F	Sex ratio~1
Age at evaluation (years)	7	4	from 2 to 24
ID	Severe	Severe	33/33
Speech	Less than 10 words, echolalia	Never acquired	30/31
Stereotypes	Hand-flapping	Unmotivated laughter, hand-flapping	26/28
Character	Calm	Episodes of heteroagressivity	Shy, amicable and calm for 14/19 unpublished, [2] and [1] Shy and anxious for 2/19 [5]
Neonatal hypotonia	+	+	25/25
Pyramidal syndrome	+	+	33/33
Deambulation	Achieved then lost, wheelchair	Never achieved, wheelchair	31/31
Seizures	a unique episode of prolonged hyperthermic seizure at 10 months of life, treated with valproate until the age of 3 years without new convulsions	-	16/31
MRI examination	Cerebral atrophy	incomplete corpus callosum agenesis and lipoma, delayed myelinisation	19/20 thin corpus callosum ; ventriculomegaly ; thinning and abnormal signal of the periventricular white matter; atrophy of the inferior vermis with cortical atrophy, dilated ventricles, prominent cisterns
Head circumference	-1.5 SD	-2 SD	Microcephaly for 25/32
Stature	+1 SD	+2 SD	Short stature for 12/14
<b>early onset of severe obesity</b>	<b>starting at 18 months of life, weight &gt;+4 SD at 7 years old</b>	<b>starting at 6 months of life, weight &gt;+5 SD at 4 years old</b>	<b>0/33</b>





## Supplementary Material

### Patients

II.1 and II.2 are brother and sister born to first cousin parents originating from Algeria. They were term-born after an uneventful pregnancy and delivery with normal birth parameters. Neonatal hypotonia was reported for both patients. At their last visit at 4 (II.2) and 7 (II.1) years of age, II.1 had a small head circumference (-1.5SD) and II.2 was microcephalic with occipitofrontal circumference on -2 standard deviation (Table 1). They had severe intellectual deficiency with only a few single words. Eye contact was sustained and they both had purposeful handskills. Patient II.2 presented unmotivated laughs and episodes of heteroaggressivity. Patient II.1 could walk with wall support from 15 months of age and then lost the ability to walk at 4 years of age. Patient II.2 never achieved independent walking. Today, both patients are wheelchair-bound. II.1 experienced a unique episode of prolonged hyperthermic at 10 months of life and received Valproate until the age of 3 years with no recurrent episodes. Both presented pyramidal syndrome with hyperreflexia, Babinski sign and spasticity. Excessive weight gain started at 18 months of life for patient II.1 and 6 months of life for patient II.2. At 7 years of age, the weight of patient II.1 is over +4 SD and over +5 SD for patient II.2 at 4 years of age while height remained on the normal range (+1SD and +2SD respectively). Currently, the father has a BMI of 24 and the mother a BMI of 31.5, with a gain of weight in motherhood. No dysmorphic features were observed except a short columella. The neck was short with acanthosis nigricans. Brain MRI showed cerebral atrophy for patient II.1 (figure 1C, left panel) and partial agenesis of the corpus callosum with a lipoma as well as delayed myelinisation for patient II.2 (not shown). Multiple ENT infections motivated amygdalectomy for both patients.

Endocrine and metabolic screening showed no abnormalities except a raised PTH (57 ng/l, N: 10-46) with normal calcemia (2.57 mmol/l, N: 2.2-2.7) in II.2 who was supplemented in Vitamine D. Array-CGH (SpectralChip CC4-V0.3) showed no pathologic CNV.

All human protocols were reviewed and approved by the institutional review board of the Necker Hospital and informed consent was obtained from all subjects involved in this study.

## Whole Exome Sequencing (WES)

Agilent SureSelect libraries were prepared from 3 µg of genomic DNA sheared with a Covaris S2 Ultrasonicator as recommended by the manufacturer. Exome capture was performed with the 50 Mb SureSelect Human All Exon kit (Agilent technologies) using a multiplex approach with molecular barcodes for traceable ID of samples. Sequencing was carried with the SOLiD5500 (Life Technologies) on a pool of barcoded exome libraries. 75+35 paired-end reads were generated and mapped on human genome reference (NCBI build37/hg19 version) using LifeScope (Life Technologies).

Sequences produced allow respectable mean sequence coverage of 39-77 reads per bp position, with more than 74-85% of targeted bases covered at 15X. Sequence reads were aligned to the human reference genome sequence (assembly GRCh37) using Mapreads. SNPs/indels were called using Sequence Alignment/Map tools (SAMtools) Pileup. Poorly mapped (less than 3X cover) and low-quality reads (less than 20 quality score) were removed. An in-house software (PolyWeb) was used to annotate and filter the variants.

## Cell culture and RNA extraction

Primary fibroblasts were cultured at 37°C under 5% CO<sub>2</sub> in RPMI® + glutamax or OPTI-MEM® + glutamax supplemented with 10% of fetal bovine serum (FBS) and 5% of penicillin/streptomycin (complete medium) (Life Technologies). Total RNA was isolated using the RNeasy Mini Kit (Qiagen) according to manufacturer's protocol. RNA concentration was measured by spectrophotometry (Nanodrop, Thermo Scientific).

## Supplementary Tables

**Table S1:** Number of variants identified by whole-exome sequencing

<b>Patients</b>	<b>II.1</b>	<b>II.2</b>
Total variants	48036	52062
Novel or rare variants (dbSNP130/1000GP/EVS/in-house database)	1526	1827
Homozygous for patients and heterozygous for parents	326	370
Coding Non-synonymous/indel/consensus SS variants	65	69
Exclusively novel variants	48	50
Shared by both patients	3	3

**Table S2:** Literature review of patients with mutation in one of the four AP4 genes

References	AP4M1	AP4E1	AP4B1	AP4S1	TOTAL	SEX
[1]	5				5	2F,3M
[20]		2			2	1F,1M
[21] from [22]			2		2	1F,1M
[2]		2	3	3	8	4F,4M
Philippe thesis (2012) <i>Gc2iD doctoral school</i>			4		4	1F,3M
[23]	1	1			2	nd
[24]		2			2	2F
[5]			2		2	1F,1M
[25]	4		2		6	5F,1M
This report	2				2	1F,1M
<b>TOTAL (August 2014)</b>	<b>12</b>	<b>7</b>	<b>13</b>	<b>3</b>	<b>35</b>	<b>18F,15M, 2nd</b>

Nd; no data

**Table S3:** Genotypes of the two families carrying the c.1137+1G>T mutation [1]

markers	genes	NCBI build37 position	This report				[1]		unrelated control
			I.1	I.2	II.1	II.2	87RD35	87RD38	
D7S2409		91056453	219/221	219	219/221	219/221	219	219/223	219
D7S657		92566477	258/260	250/256	256/258	250/258	246	256/258	250/260
<b>D7S479</b>		96093138	121/ <b>123</b>	117/ <b>123</b>	117/ <b>123</b>	<b>123</b>	<b>123</b>	<b>123</b>	121/123
<b>D7S651</b>		98308219	165/ <b>171</b>	159/ <b>171</b>	<b>171</b>	<b>171</b>	<b>171</b>	<b>171</b>	165/167
	<i>AZGP1</i>	99565796	A/T	A/T	T	T	A	A	A
	<i>AP4M1</i>	99704138	G/T	G/T	T	T	T	T	G
<b>D7S2480</b>		100055770	<b>203/215</b>	<b>203/219</b>	<b>203</b>	<b>203</b>	<b>203</b>	<b>203</b>	217/219
D7S515		101703900	152/156	152	152	152	154	154	158
D7S2420		106890015	282/288	284/288	284/288	284/288	280	280	286/288

## Supplementary References

20. Moreno-De-Luca A, Helmers SL, Mao H, Burns TG, Melton AMA, Schmidt KR, Fernhoff PM, Ledbetter DH, Martin CL. Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J Med Genet* 2011;48:141-4.
21. Bauer P, Leshinsky-Silver E, Blumkin L, Schlipf N, Schroder C, Schicks J, Lev Dorit, RO, Lerman-Sagie T, Schols L. Mutation in the AP4B1 gene cause hereditary spastic paraplegia type 47 (SPG47). *Neurogenetics* 2012;13:73-6.
22. Blumkin L, Lerman-Sagie T, Lev D, Yosovich K, Leshinsky-Silver E. A new locus (SPG47) maps to 1p13.2-1p12 in an Arabic family with complicated autosomal recessive hereditary spastic paraplegia and thin corpus callosum. *J Neurol Sci* 2011;305:67-70.
23. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P, Zecha A, Mohseni M, Püttmann L, Vahid LN, Jensen C, Moheb LA, Bienek M, Larti F, Mueller I, Weissmann R, Darvish H, Wrogemann K, Hadavi V, Lipkowitz B, Esmaeeli-Nieh S, Wiczorek D, Kariminejad R, Firouzabadi SG, Cohen M, Fattahi Z, Rost I, Mojahedi F, Hertzberg C, Dehghan A, Rajab A, Banavandi MJS, Hoffer J, Falah M, Musante L, Kalscheuer V, Ullmann R, Kuss AW, Tzschach A, Kahrizi K, Ropers HH. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 2011;478:57-63.
24. Kong XF, Bousfiha A, Rouissi A, Itan Y, Abhyankar A, Bryant V, Okada S, Ailal F, Bustamante J, Casanova JL, Hirst J, Boisson-Dupuis S. A novel homozygous p.R1105X mutation of the AP4E1 gene in twins with hereditary spastic paraplegia and mycobacterial disease. *PLoS One* 2013;8:e58286.
25. Tuysuz B, Bilguvar K, Kocer N, Yalçinkaya C, Caglayan O, Gul E, Sahin S, Comu S, Gunel M. Autosomal recessive spastic tetraplegia caused by AP4M1 and AP4B1 gene mutation: expansion of the facial and neuroimaging features. *Am J Med Genet Part A* 2014;164:1677-85.

## 7.4 Conclusions et perspectives : étude fonctionnelle du variant c.595A>T du gène AZGP1

Les patients de la famille Ma présentent un tableau neurologique très évocateur du syndrome AP-4. Étant donné la constance du tableau clinique des 33 patients décrits avec un tel syndrome, nous n'avons aujourd'hui pas de doute sur la pathogénicité du variant *AP4M1* identifié.

Étant donné la littérature disponible démontrant le rôle de ZAG dans la régulation du métabolisme des tissus adipeux, le gène *AZGP1* nous semblait un candidat prometteur pour expliquer l'obésité observée chez nos patients. Nous avons donc contacté différents groupes afin de constituer une cohorte de patients supplémentaires avec une obésité précoce et morbide afin de séquencer intégralement le gène *AZGP1* chez ces patients. Grâce à la participation du Pr Maïté Tauber et du Pr Jeanne Amiel nous avons sélectionné 12 patients avec une obésité précoce c'est-à-dire avant l'âge de 3 ans. Jusqu'à présent, nous n'avons pas mis en évidence de variants candidats chez ces patients.

Avec la collaboration du laboratoire de biochimie métabolique et protéomique de l'hôpital Necker-Enfants Malades, un dosage de la protéine ZAG par chromatographie liquide couplée à la spectrométrie de masse en tandem a pu être réalisé. Le dosage de la concentration plasmatique de ZAG se fait par LC-MS/MS (Liquid chromatography coupled to tandem mass spectrometry) après un dosage des protéines totales, une dénaturation des protéines, une alkylation des protéines et une digestion par la trypsine [Bondar et al., 2007]. Le tableau 7.1 ci-dessous rassemble les résultats obtenus pour chacun des membres de la famille ainsi que 9 autres sujets. Une très bonne similitude des résultats a été observée entre deux réplicats et l'utilisation d'une gamme de protéine recombinante ou une gamme de peptide.

État	Sujets	Concentration plasmatique de ZAG (mg/L)
sains	S1	50
	S2	52
	S3	88
	S4	71
cancer ORL	C1	110
	C2	74
	C3	50
	C4	78
	C5	93
famille Ma	I.1	99
	I.2	84
	II.1	50
	II.2	63

**Table 7.1** : Dosage des concentrations plasmatiques de ZAG.

Connaissant la variabilité possible entre deux individus sains, nous n'avons pas conclu à une diminution de la concentration plasmatique de ZAG chez les patients. Nous faisons alors l'hypothèse que la mutation n'influe pas la stabilité de la protéine mais peut-être sa fonction. Nous avons alors collaboré avec les groupes du Pr

R. Barouki et du Pr M. Tisdale afin de mettre en place des tests d'efficacité de la lipolyse après injection de la protéine recombinante ZAG sauvage ou mutante. Les premiers tests réalisés ne se sont pas révélés efficaces avec aucune des deux formes de la protéine. La glycosylation de cette protéine est primordiale pour assurer son activité. La production de cette protéine a été réalisée dans des cellules de mammifères. Par conséquent, on s'attend à une glycosylation correcte de la protéine. Nos collaborateurs tentent aujourd'hui de tester le statut glycosylé de la protéine produite. L'activité de la protéine ZAG est dépendante de sa liaison au récepteur  $\beta$ -adrenergic ( $\beta$ -AR) à la surface des adipocytes [Russell and Tisdale, 2012]. Cette capacité d'interaction est fortement altérée par les épisodes de congélations. D'autres approches peuvent être envisagées pour tester l'effet du variant p.Asn199Tyr sur la fonction de la protéine ZAG : un système de surexpression du gène sauvage ou muté dans des lignées cellulaires CHO (Chinese Hamster Ovary) ou l'administration de la protéine recombinante sauvage ou mutée à des souris ob/ob, comme précédemment réalisé par Russell and Tisdale [2012].

## 7.5 Définition du concept de syndrome des mutations contiguës

La notion de syndrome des gènes contigus a été proposée pour rendre compte d'associations cliniques liées à l'haploinsuffisance de plusieurs gènes contenus dans le segment délété. L'analyse de la famille Ma par séquençage haut-débit nous a permis d'élargir ce concept. La définition du SGC est basée sur un unique événement mutationnel ; une délétion affectant le dosage génique de plusieurs gènes. Le phénotype de la famille Ma est dû à deux événements mutationnels distincts affectant deux gènes génétiquement liés. Il ne s'agit pas de digénisme (une pathologie causée par l'association de défauts dans deux gènes distincts) puisque chaque mutation est à l'origine d'un tableau clinique distinct [Schäffer, 2013]. Les patients de la famille Ma cumulent donc deux pathologies monogéniques distinctes à spectre phénotypique non chevauchant. La consanguinité et la liaison génétique des deux gènes sont les facteurs responsables de l'association de ces deux pathologies. Le cas de la famille Ma ressemble au concept du SGC : le tableau clinique des patients est lié aux mutations de plusieurs gènes contenus dans un haplotype délétère. La notion de SGC peut donc être étendue à la transmission simultanée d'événements mutationnels indépendants mais génétiquement liés que nous avons appelé syndrome des mutations contiguës (SMC). Les caractéristiques du digénisme, du SGC et du SMC sont résumées dans la table 7.2.

Caractéristiques	SGC	famille Ma ou SMC	Digénisme
Nombre d'événement mutationnels	1	2	2
Type d'événement mutationnel	CNVs	substitutions	tous types de mutations
Nombre de gènes impliqués	au moins 2	2	2
Liaison génétique	oui	oui	oui ou non
Interaction génétique	non	non	oui
Mode de transmission	dominant	récessif	dominant ou récessif

**Table 7.2** : Comparaison des caractéristiques du SGC, du digénisme et du SMC.

## **Troisième partie**

**Approche WES sur des cas sporadiques  
cliniquement semblables : identification  
d'un nouveau syndrome de DI et  
démonstration du rôle de NONO dans la  
biologie des synapses GABAergiques**



### 8.1 Présentation des familles

La qualité du recrutement du département de Génétique de l'hôpital Necker-Enfants Malades et l'étroite interaction de l'institut *Imagine* avec les activités cliniques du département font de ce laboratoire un lieu privilégié pour développer un programme de recherche basé sur l'analyse exhaustive de patients avec DI. Les patients présentés dans cette partie sont pris en charge par des médecins de l'Hôpital Necker-Enfants Malades ayant une activité de recherche au sein de l'institut *Imagine* ; les docteurs Jeanne Amiel et Marlène Rio dont l'interaction a permis de regrouper ces patients au sein d'une même étude génétique.

Les patients MCCID1 et 2 sont deux garçons, issus de deux familles non apparentées et non consanguines, chacun étant l'unique enfant atteint de la fratrie. Les patients présentent principalement une mégalencéphalie, un corps calleux épais révélé par l'analyse de l'IRM et une DI, abrégé MCCDI. Ces patients souffrent également d'une hypoplasie malaire importante accompagnée de troubles sévères de l'élocution et sont tous deux minces.

Les cliniciennes, exerçant sur le même site, ont pu pratiquer leurs consultations ensemble et sont formelles ; ces deux patients présentent des symptômes identiques, décrivant un même syndrome, non décrit précédemment.

### 8.2 WES sur les trios

Face à l'extrême ressemblance des symptômes, nous avons émis l'hypothèse que les deux patients portent des mutations touchant un même gène. Afin d'identifier la cause génétique permettant d'expliquer ce syndrome, nous avons réalisé le séquençage de l'exome pour les deux trios et analysé les données en privilégiant la recherche de mutations touchant un gène commun aux deux enfants atteints, sous tous les modes possibles de transmission génétique. Les résultats de cette étude ont donné lieu à la soumission de l'article n°3 en Septembre 2014, faisant l'objet du chapitre suivant.

## CHAPITRE 9

RÉSULTATS

### 9.1 Identification des mutations du gène NONO : publication n°3 (soumise)

# **NONO mutations cause syndromic intellectual disability and inhibitory synaptic defects**

Dennis MIRCSOF<sup>1,2\*</sup> and Maéva LANGOUËT<sup>3\*</sup>,  
Marlène RIO<sup>3,4</sup>, Sébastien MOUTTON<sup>3</sup>, Karine SIQUIER-PERNET<sup>3</sup>, Christine BOLE-  
FEYSOT<sup>5</sup>, Nicolas CAGNARD<sup>6</sup>, Patrick NITSCHKE<sup>6</sup>, Ludmila GASPAR<sup>1</sup>, Matej  
ŽNIDARIČ<sup>1</sup>, Olivier ALIBEU<sup>5</sup>, Ann-Kristina FRITZ<sup>7</sup>, David P. WOLFER<sup>7</sup>, Aileen  
SCHRÖTER<sup>8</sup>, Giovanna BOSSHARD<sup>2</sup>, Markus RUDIN<sup>8</sup>, Christina KOESTER<sup>2</sup>,  
Florence CRESTANI<sup>2</sup>, Petra SEEBECK<sup>9</sup>, Nathalie BODDAERT<sup>3,10</sup>, Jean-Marc  
FRITSCHY<sup>2</sup>, Arnold MUNNICH<sup>3</sup>, Jeanne AMIEL<sup>3,4</sup>, Steven A. BROWN<sup>1§\*</sup> and Shiva  
K. TYAGARAJAN<sup>2§\*</sup>, and Laurence COLLEAUX<sup>3§\*</sup>

<sup>1</sup>: Chronobiology and Sleep Research Group, and <sup>2</sup>: Neuromorphology Group,  
Institute of Pharmacology and Toxicology, University of Zurich, 190  
Winterthurerstrasse, 8057 Zurich

<sup>3</sup>: INSERM UMR 1163, Laboratory of Molecular and pathophysiological bases of  
cognitive disorders, Paris Descartes– Sorbonne Paris Cité University, Imagine  
Institute, Necker-Enfants Malades Hospital, 75015 Paris

<sup>4</sup>: Service de Génétique, Hôpital Necker-Enfants Malades, AP-HP, Paris

<sup>5</sup>: Genomic platform, INSERM UMR 1163, Paris Descartes– Sorbonne Paris Cité  
University, Imagine Institute

<sup>6</sup>: Bioinformatic platform, INSERM UMR 1163, Paris Descartes– Sorbonne Paris Cité  
University, Imagine Institute

<sup>7</sup>: Institute of Anatomy, University of Zürich

<sup>8</sup>: Molecular Imaging and Functional Pharmacology Group, University of Zurich

<sup>9</sup>: Center for Integrative Rodent Physiology, University of Zurich

<sup>10</sup>: Service de radiologie pédiatrique, Hôpital Necker-Enfants Malades, AP-HP, Paris

\*These authors contributed equally to this work.

§Correspondance to:

Dr. Laurence COLLEAUX  
INSERM UMR 1163, Institut IMAGINE, 24 boulevard du Montparnasse, 75015 Paris,  
France  
Email: [laurence.colleaux@inserm.fr](mailto:laurence.colleaux@inserm.fr), phone: +33 1 42 75 42 96,/ fax: 33 1 42 75 42  
21

Drs. Shiva TYAGARAJAN and Steven BROWN  
Institute of Pharmacology and Toxicology, University of Zurich, Winterthurerstrasse  
190, 8057 Zurich, Switzerland  
Email: [tyagarajan@pharma.uzh.ch](mailto:tyagarajan@pharma.uzh.ch), [steven.brown@pharma.uzh.ch](mailto:steven.brown@pharma.uzh.ch)

Identifying causes of sporadic intellectual disability remains a considerable medical challenge. Here, we demonstrate that null mutations in the *NONO* gene, a member of the Drosophila Behavior Human Splicing (DBHS) protein family, are a novel cause of X-linked syndromic intellectual disability. Comparing patients to *Nono*-deficient mice revealed related behavioral, craniofacial, and transcriptional anomalies. In brain, these mice also showed deregulation of a large number of synaptic transcripts including the GABA receptor alpha2 subunit, as well as impaired postsynaptic scaffolding of gephyrin, a master organizer of inhibitory synapses. Importantly, alteration of synaptic scaffolding could be rescued by over-expression of *Gabra2* in *NONO*-compromised neurons, suggesting that aspects of this syndrome are potentially treatable. Our data identify *NONO* as a new neurodevelopmental-disease gene and highlight the key role of DBHS proteins in functional organization of GABAergic synapses.

Intellectual disability (ID) is characterized by significant limitations in intellectual functioning and adaptive behavior. Recent developments in next-generation sequencing (NGS) and whole-exome sequencing (WES) have considerably empowered detection of disease variants in ID<sup>1-3</sup>. WES was carried out in parallel in two unrelated male ID patients (MCCID1 and 2), who presented the same slender built-macrocephaly gestalt, facial features, shy behavior, a thick corpus callosum and a smaller cerebellum (Fig. 1a and 1b, Table S1 and Supplementary Note).

Hypothesizing that the same disease gene was shared by the two patients, only the X-linked *NONO* (Non-octamer-containing, POU-domain DNA-binding protein, also known as *p54NRB*) gene emerged as a candidate (Supplementary Table 2).

Capillary sequencing confirmed the *de novo* occurrence of a splice site variant affecting the last base of exon 10 in patient MCCID1 (NM\_001145408.1:c.1131G>A; p.Ala377Ala) (Supplementary Fig.1). The variant identified in patient MCCID2 was a one-base-pair insertion in the last coding exon (NM\_001145408.1:c.1394dup ; p.Asn466Lysfs\*13) inherited from his healthy mother (Supplementary Fig.1).

Immunoblot analyses of cultured skin fibroblast lysates using an anti-*NONO* antibody showed a markedly reduced amount of *NONO* proteins in patient cells compared to controls (Fig 1c, d). These results were confirmed by immuno-cytochemistry using the same antibody (Fig. 1e).

*NONO* belongs to the highly conserved Drosophila Behaviour Human Splicing (DBHS) protein family. This family includes three members in mammals, namely *NONO*, paraspeckle component 1 (PSPC1), and Splicing Factor Proline/Glutamine-Rich (SFPQ, also known as PSF). DBHS proteins are nuclear proteins forming homo- and heterodimers *in vivo*<sup>4,5</sup>, and previous literature documents their

involvement in various aspects of RNA production<sup>6</sup>. Studies *in vitro* suggest that they play a role in transcriptional activation and repression<sup>7-9</sup>, splicing<sup>10,11</sup>, pre-mRNA processing<sup>12</sup> and RNA transport<sup>13,14</sup>. In addition, they are major components of nuclear paraspeckles, which have been recognized as nuclear RNA-holding structures for edited RNAs<sup>15,16</sup> that likely play a role in stress-mediated regulation *via* nuclear retention of transcripts<sup>17-19</sup>. NONO and other DBHS family members also serve as transcriptional cofactors for correct circadian clock function in both flies and mammals, where they regulate the circadian clock *via* interaction with PER proteins<sup>20-23</sup>. However, no study so far has linked impaired function of these proteins to human disease.

Consistent with the established role for the NONO protein within the circadian clock, cultured skin fibroblasts from patients showed reduced amplitude of circadian oscillations (Supplementary Fig. 2). Microarray analysis of cultured skin fibroblast RNAs in patients compared to controls also revealed a marked modification in the global pattern of gene expression between the two groups. Indeed, hierarchical clustering analysis showed that the patient group formed an independent cluster away from the control RNAs, as illustrated by the corresponding heatmap (Fig. 2a). A total of 389 differentially expressed genes were shared by the two patients with 372 transcripts being similarly affected (213 down regulated and 159 up regulated), whereas only 17 transcripts were deregulated in opposite orientation (Fig. 2b, Supplementary Table 3). Notably, increased levels of *PSPC1* (2.22 and 2.78 fold respectively) and *SFPQ* transcripts (1.54 and 1.71 fold respectively) were detected in patients compared to controls. Expression data were confirmed by Western blot analysis showing increased amounts of PSPC1 and SFPQ proteins in patient fibroblasts compared to controls (Fig 1c,d).

To further characterize the physiological role of NONO upon brain development, we analyzed a mouse model in which the *Nono* gene had been disrupted by genetrap (*gt*)<sup>23</sup>. Visual inspection and CT scan revealed that mutant mice displayed a flattened nose mimicking the facial anomalies observed in the patients (Fig. 3a-c, Supplementary Table 4). Similarly, a smaller cerebellum was observed in mutant mice (Fig 3d, S2, Supplementary Table 4) and patients (Fig. 1b, Supplementary Table 1), as well as other structural anomalies (Supplementary Fig. 3, Supplementary Table 1, 4). Behaviorally, mice showed impaired performance in the Morris Water Maze (Fig 3e,f) reflective of the cognitive defects observed in patients, as well as a marked anxiety phenotype documented via prepulse inhibition, open field exploration, and light-dark preference testing (Supplementary Fig. 4). Gene expression analysis in adult dermal fibroblasts from wild type and *Nono*<sup>gt</sup> mice also revealed global transcriptional deregulation in the mutant mice compared to controls in patterns similar to that of human patients and controls. When human and mouse data sets were merged and submitted to a hierarchical clustering with the Spearman correlation similarity measure, the samples were split in two main groups. *Nono*<sup>gt</sup> mice samples segregated with the patients group, whereas wild type mice samples segregated with the human control samples (Fig. 2c and Supplementary Table 5). Taken together, these data demonstrated a regulatory role of NONO conserved through evolution and supported the relevance of the *Nono*<sup>gt</sup> model for further elucidation of the disease mechanism in patients.

Immunofluorescence analysis using anti-Nono antibodies detected strong immunoreactivity in mouse brain, including cortex and hippocampus (Fig. 4a).

Staining was strongest in neuronal nuclei (NeuN-positive) in CA1 and CA3 pyramidal regions and granule cells of the dentate gyrus, but absent from neighboring astrocytes (GFAP-positive) (Supplementary Fig. 5). Gene expression analysis in hippocampi of wildtype and *Nono<sup>gt</sup>* mice identified 882 differentially expressed genes (Fig. 4b), including the two other DBHS family members *Sfpq* and *Pspc1* (Supplementary Fig. 6), as observed in fibroblasts from NONO-deficient patients (Supplementary Table 3). Protein levels of these orthologs were also overexpressed in *Nono<sup>gt</sup>* hippocampi (Supplementary Fig. 7). Interestingly, mRNA levels of *Gabra2*, the GABA<sub>A</sub> receptor alpha2 subunit, were markedly reduced in hippocampi of *Nono<sup>gt</sup>* mice compared to controls, a result confirmed by qRT-PCR and Western blot analyses (Supplementary Fig. 6,9). Finally, pathway analyses suggested that NONO-regulated genes were markedly enriched in Gene Ontology categories related to synaptic functions (Supplementary Table 5). To support this *in silico* prediction, we compared the synaptosomal transcriptomes from *Nono<sup>gt</sup>* and control mice. RNA was extracted from synaptosomal fractions obtained by density gradient ultracentrifugation<sup>24</sup>. The quality of this fractionation was confirmed by verifying enrichment of known synaptically transported RNAs and depletion of known nuclear RNAs relative to whole transcriptome (Fig S7A-C). We found that 30.5% of synaptosomal transcripts, including *Gabra2* (Fig S6D), were deregulated in samples from *Nono<sup>gt</sup>* mice when compared to controls, a significant over-representation compared to whole transcriptome (Table S6, p=0.0007). Hence, our data suggest that NONO might play an important role in the regulation of synaptic RNAs.

*Gabra2* codes for a subunit of the GABA<sub>A</sub> receptors that potentiate the majority of fast synaptic inhibition in the brain. Previous studies have demonstrated that GABA<sub>A</sub>



receptors are anchored postsynaptically by gephyrin, which self-assembles into a scaffold and interacts with the cytoskeleton. Gephyrin binding to GABA<sub>A</sub> receptors is mediated by interaction sites that have been mapped within the intracellular loop of of GABA<sub>A</sub> receptors  $\alpha 1$ ,  $\alpha 2$  and  $\alpha 3$  subunits<sup>25</sup>. We thus tested the consequences of loss of NONO at inhibitory synapses by immunohistological studies of brain sections from wild type and *Nono*<sup>gt</sup> mice, using antibodies directed against GABA<sub>A</sub>R  $\alpha 2$  and gephyrin respectively. Post-synaptic punctate staining for gephyrin and GABA<sub>A</sub>R  $\alpha 2$  was significantly reduced in the CA3 hippocampal region of *Nono*<sup>gt</sup> mice, while staining of the presynaptic vesicular GABA transporter marker (VGAT) was unaffected (Fig. 4c). Quantitative evaluations showed that the number of gephyrin clusters was significantly reduced, but average cluster size was conserved (Fig. 4d). Western blot analyses showed markedly reduced GABA<sub>A</sub>R  $\alpha 2$  levels in total brain lysates and synaptosomes (Supplementary Fig. 9). By contrast, gephyrin levels were similar in wild type and *Nono*<sup>gt</sup> in all compartments analyzed (Supplementary Fig. 9). Taken together, these data suggest that reduced GABA<sub>A</sub>R  $\alpha 2$  levels in *Nono*<sup>gt</sup> mice are likely due to transcriptional deregulation. On the other hand, altered gephyrin clustering is likely a downstream consequence of altered synaptic composition.

In principle, the anomalies observed in *Nono*<sup>gt</sup> mice could arise from either cellular synaptic defects, or broader neurodevelopmental changes. To distinguish between these possibilities and uncover potential cell-autonomous phenotypes, we used high-resolution fluorescent *in situ* hybridization to analyse dissociated hippocampal neurons from wildtype and *Nono*<sup>gt</sup> mice. As in intact brain slices, we observed a significant reduction of *Gabra2* transcripts in *Nono*-deficient neurons but unchanged transcript levels for other postsynaptic markers (collybistin, *Gabra1*) (Supplementary

Fig. 10). At the protein level, these cultures showed the same reduction of gephyrin scaffolding puncta and GABA<sub>A</sub>R α2 levels, demonstrating that primary culture neurons from these mice recapitulated the key cytological features observed in *Nono<sup>gt</sup>* mouse hippocampi. This cellular model was therefore used for functional analysis of NONO in post-synaptic biology.

It has been previously established in cultured primary neurons that over-expression of green fluorescent protein-tagged gephyrin (GFP-gephyrin) has no apparent effect on gephyrin clustering. This tool was used to examine in more detail the effects of NONO upon post-synaptic structures<sup>26</sup>. Inversely to what was observed in *Nono*-deficient mouse neurons, over-expression of Myc-tagged NONO in wildtype primary hippocampal neurons expressing GFP-gephyrin caused a significant increase in density of GFP-gephyrin puncta compared to neurons transfected with GFP-gephyrin alone. By contrast, a reduction in GFP-gephyrin density was observed in cells transfected with a construct over-expressing a mutant form of the NONO protein, myc-NONO RRM (Fig. 4e,f). This construct contains four point mutations in the RNA recognition motif (RRM), and is therefore unable to bind to single-stranded RNAs. Thus, we can conclude that the function of NONO upon synaptic biology is RNA-dependent.

Gephyrin depends on the presence of GABA<sub>A</sub>R to form postsynaptic clusters in GABAergic synapses, and the size and density of gephyrin scaffolding can be correlated to strength and frequency of GABAergic transmission<sup>25,26</sup>. Thus, gephyrin clustering is largely impaired in the mice lacking the GABA<sub>A</sub>R α2 subunit (*Gabra2*-KO)<sup>27</sup>. Therefore, we hypothesized that the altered gephyrin distribution observed in

NONO-deficient neurons might be the consequence of *Gabra2* transcript deregulation, and furthermore that this defect might be rescued by over-expressing GABA<sub>A</sub>R  $\alpha$ 2. Indeed, transfection of plasmids expressing *Gabra2* could rescue the reduced gephyrin cluster density observed in cultured neurons transfected with myc-NONO RRM (Fig. 4g,h), further supporting the view that NONO alters postsynaptic structure by regulating *Gabra2* levels.

In conclusion, this study shows that *NONO*-null mutations cause a novel, clinically recognizable ID syndrome, with slender built-macrocephaly, malar hypoplasia and thick corpus callosum. A mouse model deficient in NONO recapitulates the major features of this syndrome. Studying brains and neurons from these mutant mice suggested that NONO plays an unsuspected role in regulating inhibitory synaptic biology, results also consistent with intellectual and anxiety phenotypes of these mice. Previous studies have identified NONO as a member of a neuronal RNA transport complex<sup>13</sup> and demonstrated increased NONO abundance at synapses in response to synaptic activity<sup>28</sup>, but no study has yet linked dysfunction in NONO or other DBHS proteins to any human disease. Our studies clearly demonstrate that, in the brain, NONO function is primarily mediated by its role in the regulation of RNA expression, and suggest that it may contribute to the local regulation of RNA metabolism underlying activity-dependent regulation of dendritic spine morphology. Finally, because inhibitory synapse structural defects caused by NONO dysfunction could be rescued by increasing the number of GABA- $\alpha$ 2 receptors, this study opens reasoned opportunities for therapeutic trials in this novel intellectual disability syndrome.

## **ACKNOWLEDGMENTS**

We are grateful to the patients and their family members for their participation in our study, and to the Functional Genomics Center Zürich (FGCZ) for transcriptomic services. This program has received a state subsidy managed by the National Research Agency under the "Investments for the Future" program bearing the reference ANR-10-IAHU-01. This study was also supported by the Centre National de la Recherche Scientifique (CNRS), the Fondation pour la Recherche Médicale (DEQ20120323702) and the Ministère de la Recherche et de l'Enseignement Supérieur, as well as by the Swiss National Science Foundation, the Zurich Clinical Research Priority Program "Sleep and Health", the Zurich Fonds zur Förderung des akademischen Nachwuchses, and the Zurich Neurozentrum (ZNZ). DM and SAB and SKT are affiliates of the ZNZ Life Sciences Zurich graduate program, and SAB is a member of the Zürich Center for Interdisciplinary Sleep Research (ZIS).

## **AUTHOR CONTRIBUTIONS**

LC, SAB, and SKT designed the study. JA and MR recruited and evaluated the study subjects. NB performed and analysed the human brain imaging, and DM, PS, AS, and MR performed and analysed the mouse brain imaging. ML analyzed WES data, performed transcriptional analysis and WB analysis on patient cells, and LG performed circadian analyses on patient cells. SM contributed to WES data analysis. CBF performed WES. NC and PN performed the bioinformatic studies. DM executed and analysed all experiments on mouse tissues, and DM and MZ executed and analyzed experiments in rodent cells. DM, CK, FC, AKF, and DPW performed and analysed mouse behavioral experiments. All authors contributed to writing and editing the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

## REFERENCES

1. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* **367**, 1921-9 (2012).
2. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674-82 (2012).
3. Vissers, L.E., de Vries, B.B. & Veltman, J.A. Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis. *J Med Genet* **47**, 289-97 (2010).
4. Fox, A.H., Bond, C.S. & Lamond, A.I. P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol Biol Cell* **16**, 5304-15 (2005).
5. Myojin, R. *et al.* Expression and functional significance of mouse paraspeckle protein 1 on spermatogenesis. *Biol Reprod* **71**, 926-32 (2004).
6. Shav-Tal, Y. & Zipori, D. PSF and p54(nrb)/NonO--multi-functional nuclear proteins. *FEBS Lett* **531**, 109-14 (2002).
7. Amelio, A.L. *et al.* A coactivator trap identifies NONO (p54nrb) as a component of the cAMP-signaling pathway. *Proc Natl Acad Sci U S A* **104**, 20314-9 (2007).
8. Dong, X., Sweet, J., Challis, J.R., Brown, T. & Lye, S.J. Transcriptional activity of androgen receptor is modulated by two RNA splicing factors, PSF and p54nrb. *Mol Cell Biol* **27**, 4863-75 (2007).
9. Park, Y., Lee, J.M., Hwang, M.Y., Son, G.H. & Geum, D. NonO binds to the CpG island of oct4 promoter and functions as a transcriptional activator of oct4 gene expression. *Mol Cells* **35**, 61-9 (2013).
10. Kim, K.K., Kim, Y.C., Adelstein, R.S. & Kawamoto, S. Fox-3 and PSF interact to activate neural cell-specific alternative splicing. *Nucleic Acids Res* **39**, 3064-78 (2011).
11. Patton, J.G., Porro, E.B., Galceran, J., Tempst, P. & Nadal-Ginard, B. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev* **7**, 393-406 (1993).
12. Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M. & Manley, J.L. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. *Genes Dev* **21**, 1779-89 (2007).
13. Izumi, H., McCloskey, A., Shinmyozu, K. & Ohno, M. p54nrb/NonO and PSF promote U snRNA nuclear export by accelerating its export complex assembly. *Nucleic Acids Res* **42**, 3998-4007 (2014).
14. Kanai, Y., Dohmae, N. & Hirokawa, N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron* **43**, 513-25 (2004).
15. Bond, C.S. & Fox, A.H. Paraspeckles: nuclear bodies built on long noncoding RNA. *J Cell Biol* **186**, 637-44 (2009).

16. Nakagawa, S. & Hirose, T. Paraspeckle nuclear bodies--useful uselessness? *Cell Mol Life Sci* **69**, 3027-36 (2012).
17. Chen, L.L., DeCerbo, J.N. & Carmichael, G.G. Alu element-mediated gene silencing. *EMBO J* **27**, 1694-705 (2008).
18. Prasanth, K.V. *et al.* Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249-63 (2005).
19. Zhang, Z. & Carmichael, G.G. The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell* **106**, 465-75 (2001).
20. Brown, S.A. *et al.* PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator. *Science* **308**, 693-6 (2005).
21. Duong, H.A., Robles, M.S., Knutti, D. & Weitz, C.J. A molecular mechanism for circadian clock negative feedback. *Science* **332**, 1436-9 (2011).
22. Kowalska, E. *et al.* NONO couples the circadian clock to the cell cycle. *Proc Natl Acad Sci U S A* **110**, 1592-9 (2013).
23. Kowalska, E. *et al.* Distinct roles of DBHS family members in the circadian transcriptional feedback loop. *Mol Cell Biol* **32**, 4585-94 (2012).
24. Dunkley, P.R., Jarvie, P.E. & Robinson, P.J. A rapid Percoll gradient procedure for preparation of synaptosomes. *Nat Protoc* **3**, 1718-28 (2008).
25. Tyagarajan, S.K. & Fritschy, J.M. Gephyrin: a master regulator of neuronal function? *Nat Rev Neurosci* **15**, 141-56 (2014).
26. Tyagarajan, S.K. *et al.* Regulation of GABAergic synapse formation and plasticity by GSK3beta-dependent phosphorylation of gephyrin. *Proc Natl Acad Sci U S A* **108**, 379-84 (2011).
27. Panzanelli, P. *et al.* Distinct mechanisms regulate GABAA receptor and gephyrin clustering at perisomatic and axo-axonic synapses on CA1 pyramidal cells. *J Physiol* **589**, 4959-80 (2011).
28. Zhang, G., Neubert, T.A. & Jordan, B.A. RNA binding proteins accumulate at the postsynaptic density with synaptic activity. *J Neurosci* **32**, 599-609 (2012).

## FIGURE LEGENDS

### Figure 1: *P54NRB/NONO* mutations and their functional consequences (a)

Photographs and pedigree of both patients. Shaded symbols indicate the affected individuals. (b) Sagittal T1 and axial FSE T2 brain MRI of patient MCCID1 at 9 years-old (left panel) and patient MCCID2 at 8 years old (right panel) showing a thick corpus callosum, a small cerebellum and a Chiari I malformation (arrows). (c) Immunoblots showing a complete absence of the NONO protein and overexpression of PSPC1 and SFPQ proteins in patients' cells compared to controls. (d) Densitometry analyses. Means  $\pm$  standard error (SEM) for the three DBHS proteins relative to the amount of total proteins are given from two independent experiment

and expressed as a percentage of the control values (100%). Significance was calculated using two-way analysis of variance (ANOVA) test with Tukey correction. Here and in subsequent figures, \* P value < 0.05, \*\* P value < 0.01, \*\*\* P value < 0.001 (e) Immunofluorescence showing the complete absence of the NONO protein in patients' cells compared to controls.

**Figure 2. Transcriptome analysis in human and mouse cells (a)** Heatmap cluster analyses indicating similarity in expression profile among probes from the two patients and differences compared to the two controls. High detection signals relative to the mean were colored in red. Low detection signals were colored in green. The cut-off for inclusion in the heatmap was a 1.5-fold alteration of probe expression for both patients. **(b)** Venn diagram showing the number of genes commonly or differentially expressed in the two patients compared to the mean of the two controls. Significant differences are based on a 1.5-fold difference and a P-value of < .05. U= up, D= down. **(c)** Hierarchical clustering analysis of combined mouse and human orthologous genes, resulting in a striking separation highlighting similarity of mouse and human transcriptional dysregulation.

**Figure 3. Functional consequences of NONO deficiency in mice. (a)** Side view of representative *Nono*<sup>gt</sup> mouse (right) compared to wildtype littermate (left). **(b)** CTscan analysis indicating a flattened and distorted nose in *Nono*<sup>gt</sup> mice (right) compared to WT (left). **(c)** Quantification of skull length, width and nose length in *Nono*<sup>gt</sup> mice (black) compared to wildtype littermate (white). *n* = 20 mice per genotype. \*\*\* *P* < 0.001 ; Student's *t*-test. Bars represent mean ± SD. **(d)** MRI scan of representative *Nono*<sup>gt</sup> mouse (right) compared to wildtype littermate (left).

Yellow arrow indicates cerebellum. Also see Supplementary Fig 3 and Supplementary Table 4. **(e-f)** Behavior of *Nono<sup>gt</sup>* mice and WT littermates in Morris water maze, n=16-20 per genotype. Black circles, *Nono<sup>gt</sup>*; open circles, WT. Unless otherwise noted in this and subsequent figure, bars represent means  $\pm$  SEM. **(e)** Gallagher's proximity test scores, i.e. average distance of mice from goal as fraction of total distance. Repeated ANOVA, gene  $p < 0.0237$ , time  $p < 0.001$ , gene x time n.s. **(f)** Wishaw's error, i.e. % path outside an 18cm-wide corridor connecting release point and goal. Gene  $p < 0.0215$ , time  $p < 0.001$ , gene x time n.s.

**Figure 4. Effects of NONO deficiency upon synaptic biology. (a)**

Immunofluorescence labeling of NONO in wildtype (top) and *Nono<sup>gt</sup>* (bottom) mouse coronal brain sections. PL = pyramidal cell layer, DG = dentate gyrus. **(b)**

Scatter plot of hippocampal transcriptome from WT and *Nono<sup>gt</sup>* mice. Red, differentially expressed genes,  $p \leq 0.01$  and log ratio  $\geq 0.5$ . **(c)**

Immunohistochemical staining for inhibitory postsynaptic marker gephyrin (green), GABA type II receptor (red), and the presynaptic marker vGat (blue) in wildtype and *Nono<sup>gt</sup>* mice in the *stratum radiatum* CA3 of the hippocampus. **(d)**

Quantification of gephyrin, GABA type II receptor and vGat density (left) and gephyrin cluster size (right). Means  $\pm$  SD are shown, \*  $P < 0.05$ , \*\*  $P < 0.01$ ; Student's *t*-test (top) and Kolmogorov-Smirnov test (bottom). **(e-h)**

Immunofluorescence analysis of inhibitory synapse morphology *in-vitro*. **(e)**

Primary hippocampal rat neurons expressing control GFP-gephyrin alone, or **(e')** co-expressed with myc-NONO, or **(e'')** co-expressed with RNA binding-deficient myc-NONO-RRM,. Boxed region is magnified beneath. Postsynaptic clustering is



demonstrated by apposition of eGFP-gephyrin clusters (green) to vGat -positive terminals (blue). (f) Quantification of cluster density distribution in 9 neurons from 3 independent experiments., Control x NONO  $^*P < 0.05$ , Control x NONO- RRM  $^*P < 0.05$  using Kolmogorov-Smirnov test. (g) Primary hippocampal rat neurons expressing control GFP-gephyrin alone, or (g') co-expressed with myc-NONO-RRM, or (g'') co-expressed with myc-NONO-RRM and GABRA2. (h) Quantification, showing complete rescue of the impaired gephyrin cluster distribution by GABRA2. Control x NONO RRM  $^*P < 0.05$ , Control x NONO RRM GABRA2  $^*P < 0.05$  using Kolmogorov-Smirnov test.

## ONLINE METHODS

### Patients

All human protocols were reviewed and approved by the institutional review board of the Necker Hospital and informed consent was obtained from all subjects involved in this study.

Case MCCID1 was the second child of healthy, nonconsanguineous parents. Family history is unremarkable. He was born in the 41st week of gestation, after a normal pregnancy and delivery and an Apgar score<sup>29</sup> of 10/10. Parameters at birth were in the normal range (BW 3370g, LW 50.5 cm, HC 34 cm). Developmental milestones were delayed: he was able to walk alone at 3 years of age, and had limited speech with preserved comprehension. He developed absence seizures at the age of 5 years and continued to have seizures with increased frequency. He had strabismus and myopia. At the age of 15 years, he developed severe kyphoscoliosis. On examination at 17.5 years, his height was 1.79 m (+1SD), his weight 50 kg (-1.5SD), and OFC 60.5 cm (+4SD). Hand and feet were narrow with long fingers and toes, overriding toes and bilateral ankylosis of the metacarpophalangeal joint of thumb. He had flat feet with dystrophic nails. He had long and expressionless face, malar hypoplasia, short palpebral fissures, small and open mouth with drooling, high-arched-palate and enamel defect. He had slender build and distal amyotrophy. Speech was limited with simple sentences, severe elocution disability and nasal speech. Extensive metabolic screening was normal. Myotonic dystrophy and fragile X were excluded. Brain MRI demonstrated bilateral megalencephaly, a thick corpus callosum, enlarged white matter, septum pellucidum cyst, and a small cerebellum. High resolution cytogenetic studies (array CGH) identified a de novo 15q13.3 deletion.

Case MCCID2, a male, is the first child of healthy, non consanguineous parents with no medical family history of note. He has 3 healthy younger sisters (Figure 1A). Because hydramnios and short long bones were noted in the 2<sup>nd</sup> trimester of pregnancy, amniocentesis was performed and karyotyping showed normal chromosomes, 46,XY. He was born at 37.6 WG with low birth parameters (BW: 2540 g, BL: 46 cm, OFC: 35.5 cm) and an Apgar score of 10/10. He presented poor sucking, gastrointestinal reflux, stridor, cryptorchidism and hypotonia from birth. He developed convergent strabismus within the first year of age. Motor skills have been delayed with head control achieved at 10 months and walking unaided at 3 years of age. At that time language was limited to a few single words. He was slender built and macrocephalic with weight and height on -2 SD and head circumference on +2 SD. His thorax was long and narrow and he developed kyphoscoliosis and pes planus. He was awkward and slow with weak patella reflexes. He suffered from drooling, persistent deglutition difficulties and severe elocution disability with nasal speech. He could not blow or smack. He had multiple dental caries due to mouth breathing. Chronic otitis media resulted in conductive hearing loss of 40 to 60 dB and has been treated with grommets. Hands and feet were narrow with overriding toes and ankylosis of the metacarpophalangeal joint of both thumbs. Although macrocephalic, his forehead was relatively short and narrow with low frontal and temporal hairline. The face was elongated with upslanting palpebral features, a convergent squint, a thin and high nasal root with deviated nasal septum and large tip with short columella, severe malar hypoplasia, a small open mouth with narrow and high palate, narrow dental arcades and crowding of teeth. Ears were normally placed and folded with hypoplastic ear lobes. He made constant progress, was toilet trained at about 7 years, speaks in sentences and is able to write his name and read

simple words. He has a shy, gentle and cheerful behaviour. When last seen at 15 years of age, puberty was delayed and orthopaedic surgery of the kyphoscoliosis was planned.

Brain MRI performed at 18 months and 8 years showed a thick corpus callosum, asymmetric trigone and lateral ventricles and a Chiari malformation type I (Figure 1B). EEG showed no gross anomaly. Extensive metabolic screening was normal. Skeletal X-rays showed no malformation of the vertebrae. CytoChip™ (BlueGnome, Cambridge) array-CGH did not detect any copy number variant.

### **Whole-exome sequencing**

Agilent SureSelect libraries were prepared from 3 µg of genomic DNA sheared with a Covaris S2 Ultrasonicator as recommended by the manufacturer. Exome capture was performed with the 50 Mb SureSelect Human All Exon kit (Agilent technologies) using a multiplex approach with molecular barcodes for traceable ID of samples.

Sequencing was carried with the SOLiD5500 (Life Technologies) on a pool of barcoded exome libraries. 75+35 paired-end reads were generated and mapped on human genome reference (NCBI build37/hg19 version) using LifeScope (Life Technologies).

Sequences produced allowed a mean sequence coverage of 42-87 reads per bp.

The average coverage was 70X, with more than 75% of targeted bases covered 15X.

Sequence reads were aligned to the human reference genome sequence (assembly GRCh37) using Mapreads. SNPs/indels were called using Genome analysis toolkit and Picard tools. Poorly mapped (less than 3X cover) and low-quality reads (less than 20 quality score) were removed. An in-house software (PolyWeb) was used to annotate and filter the variants.

## **Mice**

Generation of *Nono*<sup>gt</sup> mice was described previously<sup>23</sup>. Mice have been backcrossed greater than 12 generations to C57/Bl6J. All experiments were performed by comparing adult wild type and mutant littermates (2-3 months old). Animal housing and experimental procedures are in agreement with veterinary law of the canton of Zurich.

## **CT and MR Scanning and analysis**

The skulls of 20 mice per genotype were scanned with *in vivo* 3D micro computed tomography (Quantum Fx, Perkin Elmer, Waltham MA). All mice were sacrificed prior to scanning and placed in the micro CT in ventral recumbency with the head centered to the field of view. Scans were taken with an isotropic voxel resolution of 59  $\mu\text{m}$ . The protocol used 90 kV and 100  $\mu\text{A}$  with 50 msec per projection resulting in a total scan time of 3 minutes for 360°. 3D surface reconstructions of all skulls were created using the Quantum Fx viewer and assessed/scored for signs of deformation and morphological anomalies by two independent investigators. Analysis of skull parameters was done as described previously<sup>30</sup>. Micro MR images were acquired as described previously<sup>31</sup> via high-resolution MRI of mouse brain at 9.4T using a cryogenic quadrature transceiver coil. (Inplane image resolution 60x60mm<sup>2</sup>).

## **Mouse Behavioral Experiments**

STATISTICAL MODELS: Data were analyzed using mixed ANOVA models with genotype (KO, WT) as between-subject factor. Within-subject factors were added as needed to explore the dependence of genotype effects on place, time, or stimulus. Significant interactions and, where necessary, significant main effects were further explored by Tukey-Kramer post-hoc tests or by splitting the ANOVA model, as appropriate. One-sample t-tests were used for follow-up comparisons against chance

levels. Variables known to produce strongly skewed distributions and/or frequent outliers were subjected to a log transformation before ANOVA analysis (as indicated, e.g. latency measures). The significance threshold was set at 0.05. The false discovery rate (FDR) control procedure of Hochberg was applied to groups of conceptually related variables within single tests to correct significance thresholds for multiple comparisons.

#### DESCRIPTIONS OF INDIVIDUAL TESTS:

**Prepulse Inhibition:** The session consists of a series of six 40-ms startling pulses of different intensities varying between 100 and 120 dB in order to get an average magnitude of the initial animal's startle response. The animal is then subjected to a succession of 6 discrete trials. Each trial includes a short period of background white noise (control no- stimulus condition), followed by 20-ms pre-pulses (4, 8, 12, 16, and 20 decibels above the background white noise) and 40-ms startling pulses (100 to 120 dB) presented either separately or in combination (pre-pulse + pulse, with a delay of 100 ms between the two stimuli). The session is completed with additional six 40-ms startling pulses and background noise. The time interval between two stimulus presentations can vary between 10 and 20 seconds.

**Elevated plus maze:** The apparatus is a crossbar-shape maze, comprising two symmetrically arranged open arms equipped with 3-mm ledges and two closed arms equipped with 20-cm walls. The intersection of the 4 arms forms a small central zone (5 cm x 5 cm). The maze is elevated 38 cm from the ground. The whole apparatus is made of semi-transparent Plexiglas. The mouse is put on the central platform, the head facing a closed arm. The session lasts 5 min, starting once it enters the four paws in to one arm, whichever close or open. The behavioral parameters are recorded on-line.

**Open-field Test:** The large open-field arena is circular with a diameter of 150 cm, a white plastic floor, and 35 cm high sidewalls made of white polypropylene.

Illumination is by indirect diffuse room light (4 40W bulbs, 12 lux). Each subject is released near the wall and observed for 10 min. The same procedure is repeated the following day, resulting in a total observation time of 20 min. Movements are tracked using Noldus EthoVision. The number of deposited fecal boli is recorded after each session.

**Light-Dark box:** A 20x30 cm lit chamber with transparent Perspex walls (20 cm high) and open top is connected to a 20x15x20 cm polyvinyl-chloride box. The box is dark (ca. 10 lux) and completely enclosed, except for the 7.5x7.5 cm opening connecting it to the lit chamber. The lit chamber is under direct room light (ca. 450 lux). Each subject is released in the middle of the lit compartment and observed for 5 min. Movements are tracked using Noldus EthoVision. Rearings and grooming are recorded using the keyboard event-recorder provided by the video-tracking system.

**Morris Water Maze:** This test was conducted as described previously<sup>32</sup>, in ca. 12-lux light in a 150cm-diameter water pool. Six trials are conducted per day, each separated by 30-60min and lasting 2min, either learning (15x15cm platform 0.5cm above the water surface), or testing (platform 0.5cm below water surface). Three days of acquisition and two days of reversal training are conducted. Movements are tracked using Noldus Ethovision. Mouse performance was evaluated using Gallagher's measure of proximity (the average distance away from the goal during the test) and Whishaw's error (% path outside a 18cm wide straight corridor connecting release point and goal)<sup>33,34</sup>.

## Expression analyses

Total cultured human and mouse skin fibroblast RNAs were isolated using the RNeasy Mini Kit (Qiagen). RNA quality was assessed using RNA Nano LabChips and the 2100 Bioanalyzer (Agilent Technologies) and RNA concentration was measured by spectrophotometry (Nanodrop, Thermo Scientific). Briefly, 100 ng of total RNA was reverse transcribed, and second strand DNA was produced and amplified by *in vitro* transcription in the presence of biotinylated ribonucleotides using the IVT Express kit (Affymetrix). Microarray experiments were performed for 2 controls and 2 patients in duplicate on the Affymetrix Human PrimeView Arrays (a genome wide array with 49293 probe sets), hybridized with fragmented amplified RNA as recommended by the manufacturer. Similarly, mouse RNAs from 2 *Nono<sup>gt</sup>* and 2 WT mice were hybridized on Affymetrix GeneChip® Mouse Genome 430 2.0 Arrays (a genome wide array allowing the analysis of 39,000 transcripts). Fluorescence data were imported into two analysis softwares: Affymetrix® Expression Console™ and R Bioconductor. Gene expression levels were calculated using the RMA algorithm Expression Console and flags were computed using a custom algorithm within R [<http://www.r-project.org/>]. Assuming that a maximum of 80% of genes are expressed we select the 20% lowest values for each microarray to be background expression data measures. We have then computed a threshold at two standard deviations over the mean of the background. All probes which normalized intensity measures were lower than the computed threshold were flagged 0 instead of 1. The list has been created filtering probes flagged as 1 for at least half of the chips. The group comparisons were done using Student's *t* test. To estimate the false discovery rate we filtered the resulting p-values at 5% and used the Benjamini and Hochberg, Bonferroni or without correction. Cluster analysis was



performed by hierarchical clustering using the Spearman correlation similarity measure and average linkage algorithm.

Meta-analysis of human and mouse samples types was performed as previously described<sup>35</sup>.

Mouse hippocampal RNAs were extracted using a GenElute Mammalian Total RNA Miniprep Kit (Sigma, St Louis, MO, USA) according to the manufacturer's instructions. Total RNA was quantified by absorbance spectroscopy and RNA integrity and quality was assessed by 1.0% agarose gel electrophoresis. Total RNA (1 µg) was transcribed to cDNA with SuperScript II (Invitrogen, Carlsbad, CA, USA) using random hexamer primers according to the manufacturer's instructions. For quantitative real-time PCR (qPCR), 20 ng of cDNA was used, and single transcript levels of genes were detected with the HOT FIREPol EvaGreen qPCR Mix (Solis BioDyne, Tartu, Estonia) and an AB7900 thermocycler. Primers used for detection of synaptic transcripts were as follows: *β-actin*, AGTGTGACGTTGACATCCGTA (sense), GCCAGAGCAGTAATCTCCTTCT (antisense); *Gephn*, GGCGACCGAGGGAATGAT (sense), CCACCCAACAAAGAAGGATCTT (antisense); *Gabra1*, GGTTGACCGTGAGAGCTGAA (sense), CTACAACCACTGAACGGGCT (antisense); *Gabra2*, CAGTGGCCCATAACATGACAAT (sense), GGACATTCGGCTTGGACTGT (antisense); *CamKIIa*, CCCCTTTCGCCTACATGTGA (sense), GGCTACAGTGGAGCGGCTTA (antisense). Data were analyzed using the comparative CT method (Schmittgen & Livak, 2008).

For transcriptome analysis using RNA-Seq, the quality of the isolated RNA was determined with a Qubit® (1.0) Fluorometer (Life Technologies, California, USA) and a Bioanalyzer 2100 (Agilent, Waldbronn, Germany). RNAs were then processed using The TruSeq Stranded mRNA Sample Prep Kit (Illumina, Inc, California, USA) according to the manufacturer recommendations. The TruSeq SR Cluster Kit v3-cBot-HS or TruSeq PE Cluster Kit v3-cBot-HS (Illumina, Inc, California, USA) was used for cluster generation using 8 pM of pooled normalized libraries on the cBOT. Sequencing was performed on a Illumina HiSeq 2000 system using the TruSeq SBS Kit v3-HS (Illumina, Inc, California, USA) with paired end 2 X100 reads or single end 1X100 reads.

RNA-seq reads were quality-checked with fastqc which computes various quality metrics for the raw reads. Reads were aligned to the genome and transcriptome with tophat v 1.3.3. Before mapping the low quality ends of the reads were clipped (3 bases from the read start and 10 bases from the read end). Tophat was run with default options. The fragment length parameter was set to 100 bases with a standard deviation of 100 bases. Based on these alignments the distribution of the reads across genomic features was assessed. Isoform expression was quantified with the RSEM algorithm (<http://www.biomedcentral.com/1471-2105/12/323>) with the option for estimation of the read start position distribution turned on.

### **Cell culture and transfection**

Human primary fibroblasts were cultured at 37°C under 5% CO<sub>2</sub> in RPMI® + glutamax or OPTI-MEM® + glutamax supplemented with 10% of fetal bovine serum (FBS) and 5% of penicillin/streptomycin (complete medium) (Life Technologies).

Mouse primary hippocampal neuron cultures were prepared as described previously<sup>37</sup>. Hippocampal cultures were transfected with 0.5 µg of either eGFP–gephyrin or the specific myc–NONO construct according to the protocol described previously<sup>36</sup>. Cells were transfected after 8 DIV and processed for immunofluorescence 7 days later (referred to as 8+7 DIV). In co-transfection experiments the total DNA concentration was maintained at 1.5 µg.

### **Plasmids used for transfection**

The eGFP–gephyrin P1 variant has been described previously<sup>37</sup>. Overexpression of myc–NONO and myc–NONO RRM, were conducted using the plasmids described previously<sup>38</sup>. GABRA2 was created amplifying rat cDNA by primers containing HindIII and XhoI sites and subsequent cloning into the pCR3.1 vector.

### **Western blot**

Human primary fibroblasts protein extracts were prepared on ice in lysis buffer (50 mM Tris pH 8, 170 mM NaCl, 0.5% NP-40 X-100, 50 mM NaF, and Complete EDTA-free protease inhibitor 11697498001, Roche Applied Science). The Bradford protein assay was used to determine the concentration of each sample (B6916-500 Sigma). 20-30µg of protein extracts were separated by 4-20% SDS-PAGE (SDS-PAGE Mini-PROTEAN® TGX Stain Free™ #4568093 Biorad). Stain Free gels were exposed to UV light for 2'30" prior to transfer to 0.2 µm nitrocellulose membranes (Trans-Blot Turbo Transfer Pack #1704158 Biorad). UV-induced fluorescence corresponding to total proteins were first visualized with ChemiDoc MP imaging system, then blocked with 5% nonfat dried milk powder (Invitrogen) diluted in PBS-T (1XPBS with 0.2% Tween-20) for 1-2H and incubated overnight at 4°C in 2% milk in PBS-T with the primary antibodies: NONO, SFPQ, PSPC1, and β-actin used as a control for protein loading. Membranes were then incubated with horseradish peroxidase–conjugated

secondary antibodies. Proteins were visualized using ECL-Plus (GE Healthcare). Various exposure times were performed for either autoradiography films and AGFA development (Curix 60 n°1829) or Chemidoc system (Biorad). Raw signal intensities were first obtained for target proteins as well as total proteins profile, with the same sample as reference, using volume tools in the ImageLab software (Biorad).

Background signal was deducted from each value. Final quantification data were given as a ratio of target signal to total signal.

Mouse brain lysate and sample preparation were performed as described<sup>39</sup>. Each lane was loaded with 50 µg protein and after blotting, the nitrocellulose membrane was directly blocked for 30 min with 1% solution of Western Blocking Reagent (Roche Applied Science, Indianapolis, US).

### **Antibodies**

The following antibodies were used: Human Antibodies: mouse anti-β-actin (AM4302, 1:20,000; Ambion), rabbit anti-NONO (LS-C31127, 1:2,000; LSBio Cliniscience), rabbit anti-SFPQ (A301-320A, Bethyl), mouse anti-PSPC1 (sc-374181, 1:1,000; Santa Cruz), Mouse anti-gephyrin antibody (mAb7a, 1:3000; or 3B11, 1:10000; Synaptic Systems, Gottingen, Germany), rabbit anti-vGAT antibody (1:3000, Synaptic Systems, Gottingen, Germany), guinea pig anti-GABAAR α2 subunit antibody<sup>39</sup>, mouse anti-Myc (1:10000, Roche), rabbit anti-NONO (1:500,<sup>20</sup>), rabbit anti-PSPC1 (1:500,<sup>20</sup>). rabbit anti-SFPQ (1:500,<sup>20</sup>), NeuN (MAB377, Millipore), GFAP (Z0334, DAK=O and MAB360, Millipore), PSD-95 (MA1-045, Affinity Bioreagents), β-actin (MAB1501, Millipore). Secondary antibodies were Donkey anti-rabbit IgG-HRP antibody (1:20,000; sc-2313, Santa Cruz) or donkey anti-mouse IgG-HRP antibody (1:20,000; sc-2314, Santa Cruz).

### **Immunocytochemistry and Imaging on human skin fibroblasts**

Cells were harvested and seeded one day prior to immunocytochemistry experiment with equal cell density for each well. 24H later, cells were rinsed twice in PBS and fixed in IC fixation buffer (FB001 Invitrogen) for 10 minutes at room temperature. Cells were then permeabilized with 0.5% Triton X-100 for 10 minutes at room temperature. After 3 PBS washes, blocking was done with 5% BSA diluted in 0.1% Triton-PBS for 30 minutes at room temperature. Hybridization was performed with the first antibodies used in Western blotting experiments in a moist chamber overnight at 4°C. After 3 x 10 minutes 0.1% Triton-PBS washes, the following detections were carried out with secondary antibodies Alexa488 anti-rabbit IgG (1:400, A-11034, Life Technologies) and Alexa594 anti-mouse IgG (1:400, A-11005, Life Technologies) for 1H at room temperature in a dark and moist chamber. After 3 x 10 minutes 0.1% Triton-PBS washes and a final PBS wash, slides were mounted with ProLong (P36935, Invitrogen). Images were taken with a Zeiss LSM700 microscope fitted with a Plan-Apochromat 40x/1.3 Oil DIC M27 objectif and the Zen 2009 software. Images montage was done using ImageJ.

### **Immunocytochemistry and immunohistochemistry on mouse samples**

Immunocytochemistry was performed as described previously<sup>26</sup>. In short, cells were rinsed in PBS and fixed for 10 minutes in 4% paraformaldehyde at room temperature. Cells were permeabilized with 0.01% Triton X-100 and detection of intracellular proteins were achieved by incubation for 60 minutes at room temperature with primary antibodies diluted in PBS containing 10% normal serum, followed by incubation with secondary antibodies coupled to Cy3 or Cy5 (1:500, Jackson ImmunoResearch) for 30 minutes at room temperature. Finally, coverslips were mounted with fluorescent mounting medium (Dako Cytomation, Carpinteria, CA). The

GABAAR $\alpha$ 2 subunit antibody was incubated in living cultures for 90 minutes in culture medium<sup>40</sup>.

Staining and immunohistochemical analysis of synaptic components was performed as previously described<sup>41</sup>. Briefly, mice were anesthetized with pentobarbital and perfused intracardially with ice-cold, oxygenated ACSF. The brain was extracted and cut in blocks containing the regions of interest for analysis (e.g. hippocampal formation). The Tissue was plunged into ice-cold, freshly prepared fixative (4% PFA in PBS) and postfixed for 90 min, rinsed with phosphate-buffered saline (PBS), cryoprotected overnight in 30% sucrose in PBS and frozen with powdered dry ice and stored at  $-80$  °C. Sections were cut from frozen blocks with a sliding microtome at a thickness of 40  $\mu$ m and were collected free-floating in PBS. They were incubated under continuous agitation in primary solution (Tris buffer (pH 7.4) containing 0.2 Triton X-100, 2% normal serum and the primary antibodies) for 15–48 h at 4 °C, washed in Tris buffer and incubated for 30–60 min at room temperature in secondary antibodies coupled to a fluorochrome.

Immunofluorescence images were captured by laser scanning confocal microscopy, using a 20x, 40x or 64x lens respectively (NA 1.4, 1024 x 1024 pixels, Zeiss LSM 710). Final illustrations were prepared from the maximal intensity projection of stacks of images spaced at 0.5  $\mu$ m. Signals were quantified, using a custom macro created with the ImageJ software. Images were background-subtracted and filtered with a Gaussian filter, but no change in brightness and contrast was applied.

### **High-resolution *in-situ* hybridization**

Dissociated hippocampal neurons were prepared and maintained as previously described<sup>26</sup>. *In situ* hybridization was performed using the QuantiGene (QG) ViewRNA kit from Panomics as previously described with the following modifications.

Cells (DIV 14) were incubated for 2 min at room temperature in PBS and fixed for 15 min using 4% formaldehyde solution (in PBS pH 7.4). After fixation, cells were permeabilized using a detergent solution (Panomics) for 5 min. Cells were washed three times with PBS followed it with in situ hybridization using GABRA2 and pan collybistin (*Arhgef9*) probes designed by Panomics, following the manufacturer's instructions. Briefly, probes were diluted 1:100 in hybridization buffer supplied by Panomics, incubated at 40°C (3hr), washed, hybridized with preamplification oligonucleotides (1:100) at 40°C (40 min), washed, hybridized with amplification oligos (1:100) at 40°C (40 min), washed, and finally hybridized with the label oligos (1:100) at 40° C (40 min). Cover-slips were dried at RT in vertical position. Coverslips were mounted with Dako-DAPI fluorescent mounting Medium (Dako S3023) and left to polymerize overnight at 4°C.

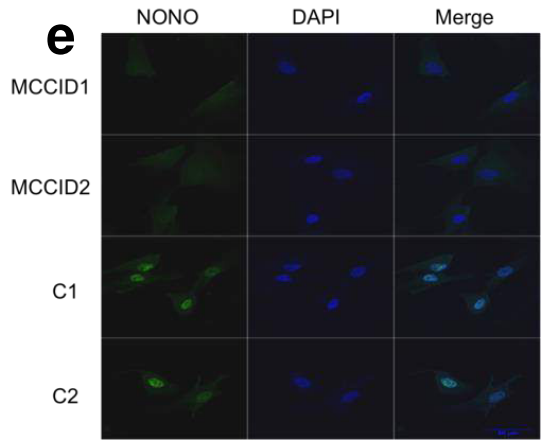
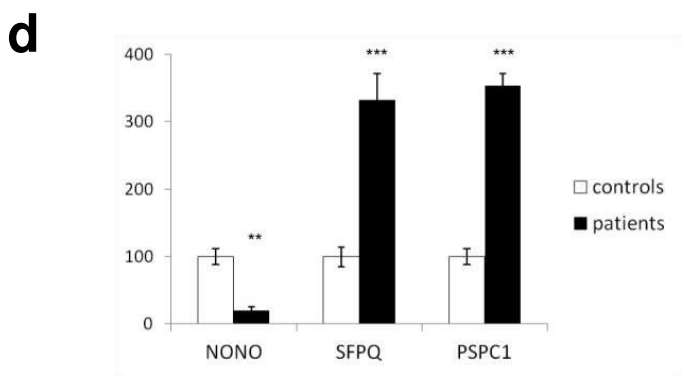
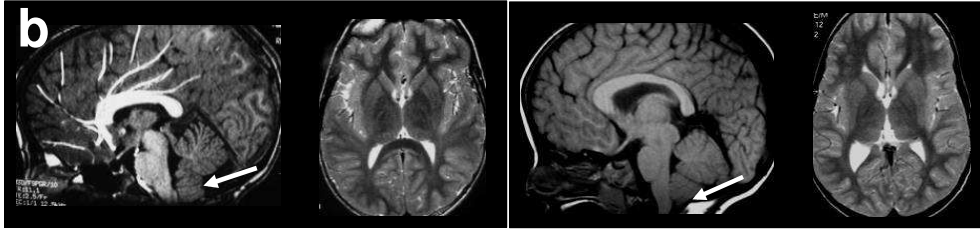
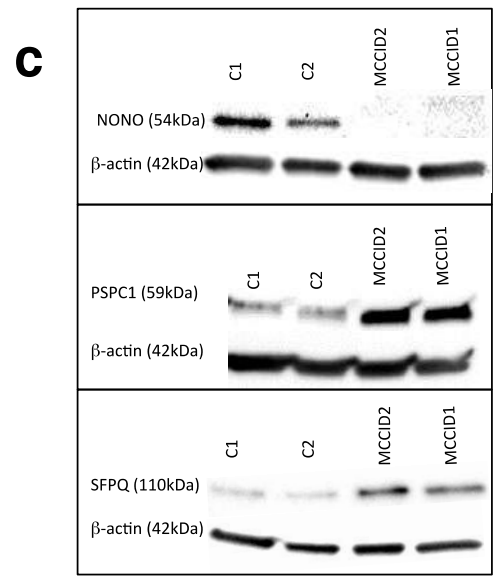
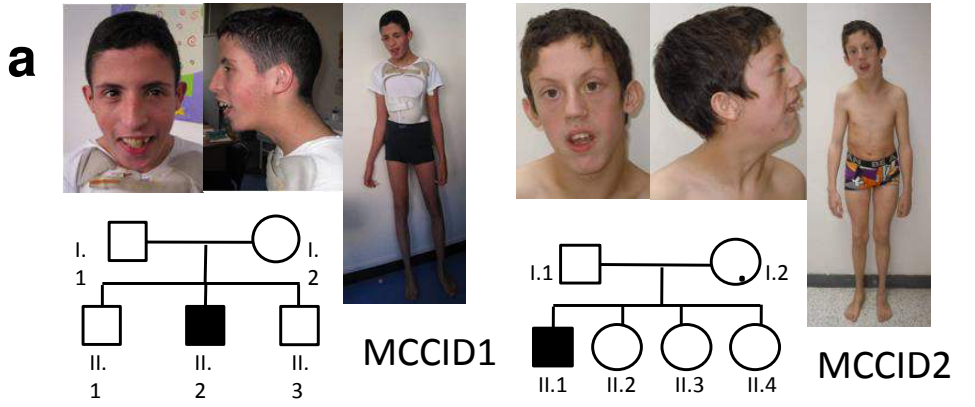
### **Synaptosome preparation**

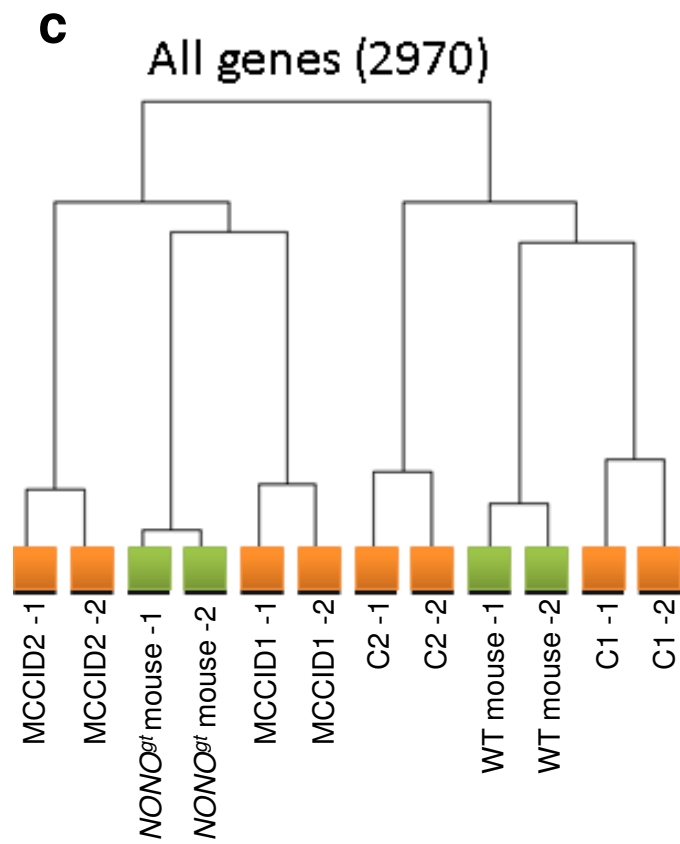
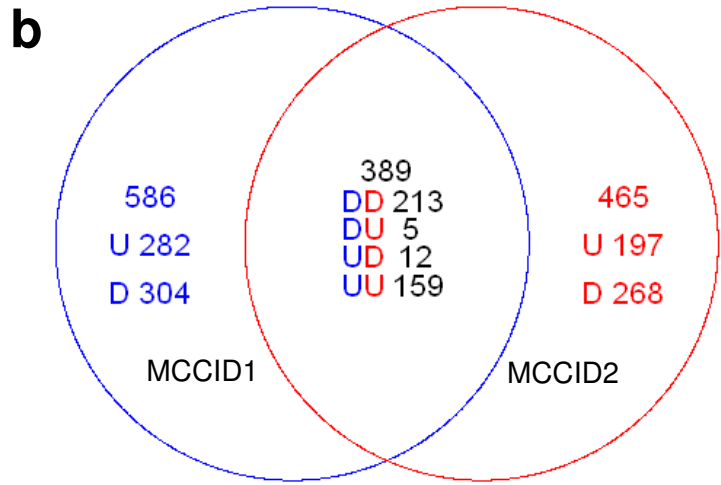
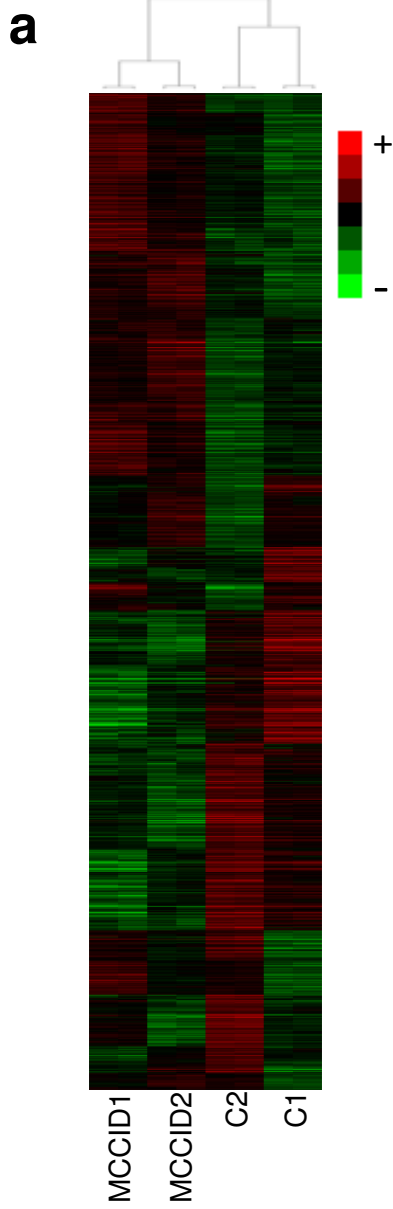
Synaptosomes have been prepared as described previously<sup>24</sup>. In brief, mouse brains were homogenized in 5 ml homogenization buffer (0.32 M sucrose, 1 mM EDTA pH 7.4, 1 mM dithiothreitol, phenylmethanesulfonyl fluoride solution (Sigma, 93482-50ML-F), complete mini-protease inhibitor (Roche Diagnostics) for 10 sec using a polytron. The homogenate was centrifuged at 1,000g for 10 min at 4 °C yielding the nuclear fraction (Nuc) and the supernatant (Sup). The supernatant was centrifuged at 31,000g for 5 min at 4°C using a discontinuous Percoll gradient. The layer between 3% and 10% of Percoll were collected, washed in 30 ml of homogenization buffer and further centrifuged at 22,000 × g for 15 min at 4°C. The pellet was resuspended in in EBC buffer (50 mM Tris-HCl pH 8.0, 120 mM NaCl and 0.5% NP-40) containing complete mini-protease inhibitor (Roche Diagnostics) and phosphatase inhibitor cocktail 1 and 2 (Sigma–Aldrich)) for Western blot analysis or lysis buffer for RNA

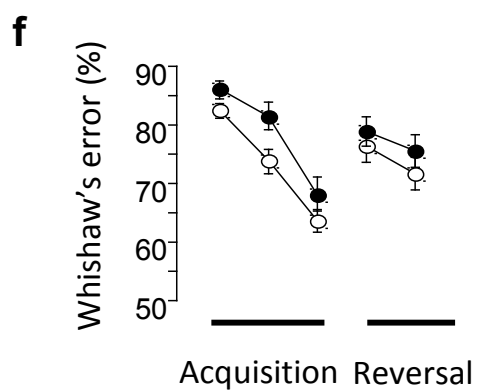
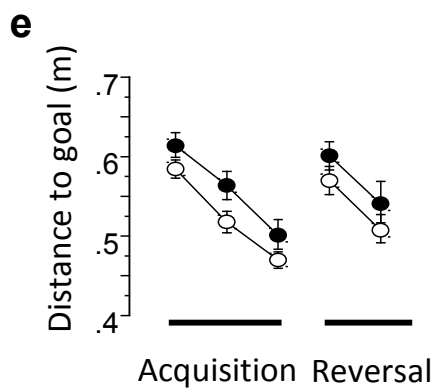
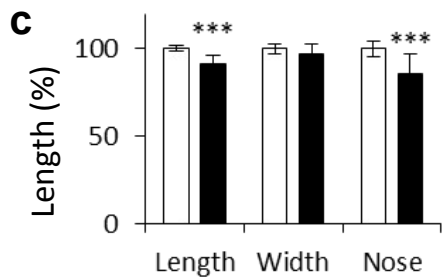
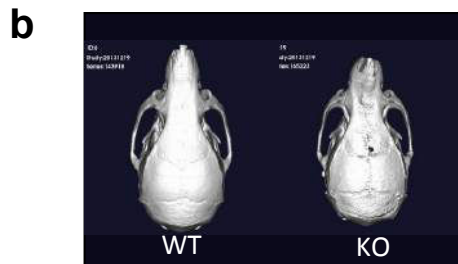
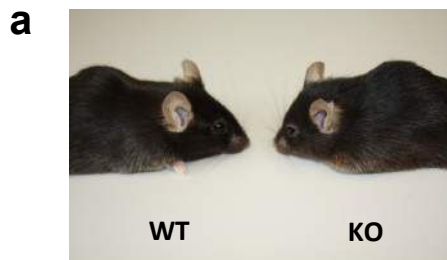
extraction (GenElute Mammalian Total RNA Miniprep Kit, Sigma).

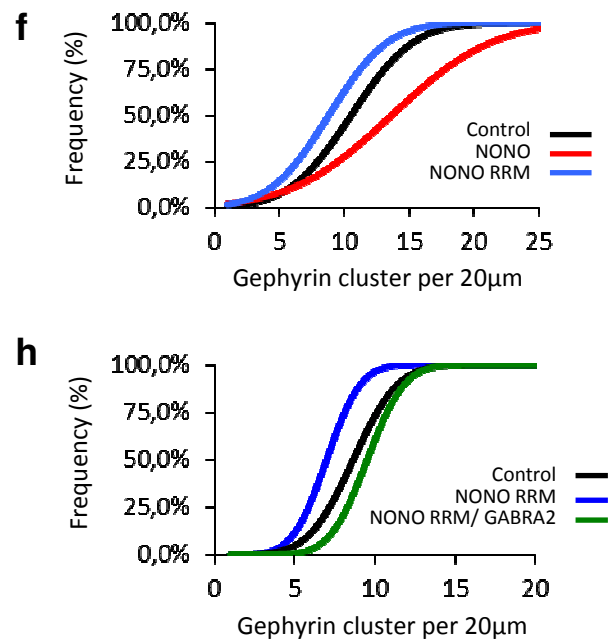
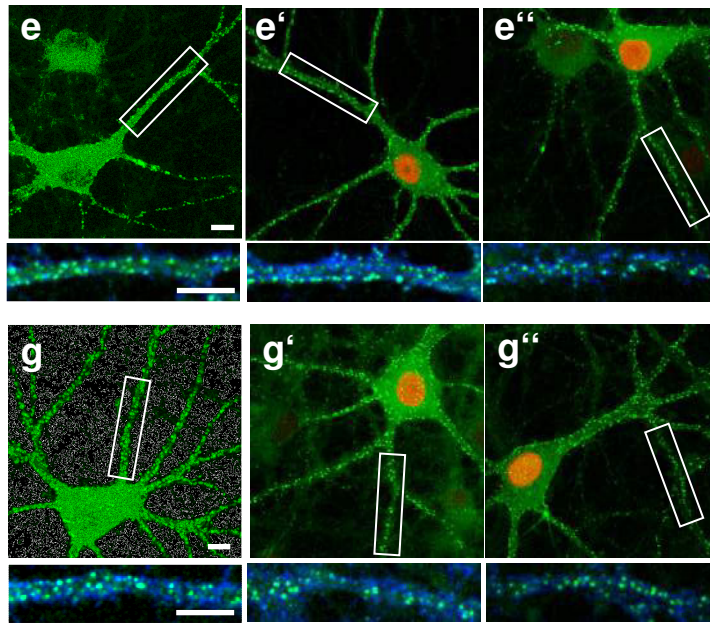
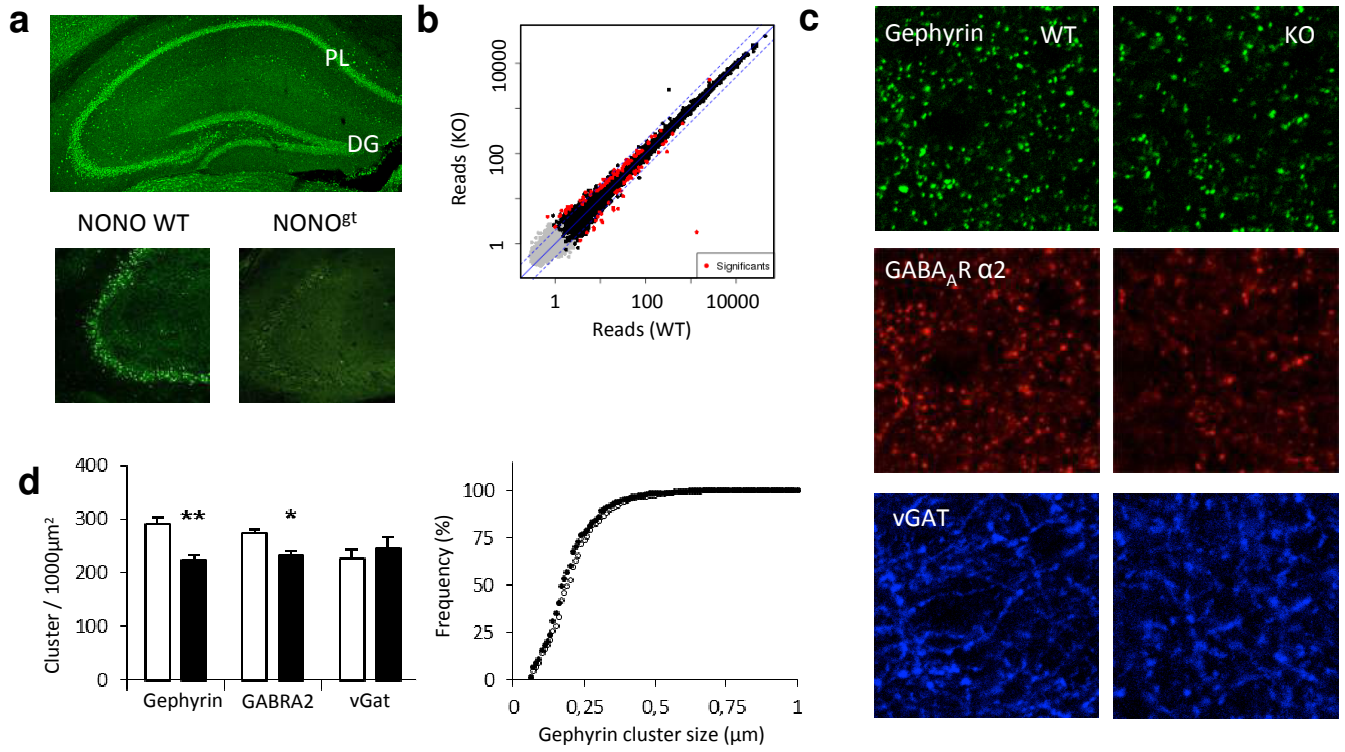
29. Apgar, V. A proposal for a new method of evaluation of the newborn infant. *Curr Res Anesth Analg* 32, 260-7 (1953).
30. Yang, Y.F. et al. A 1.1Mb deletion in distal 13q deletion syndrome region with congenital heart defect and postaxial polydactyly: additional support for a CHD locus at distal 13q34 region. *Gene* 528, 51-4 (2013).
31. Baltes, C., Radzwill, N., Bosshard, S., Marek, D. & Rudin, M. Micro MRI of the mouse brain using a novel 400 MHz cryogenic quadrature RF probe. *NMR Biomed* 22, 834-42 (2009).
32. Mohajeri, M.H. et al. Intact spatial memory in mice with seizure-induced partial loss of hippocampal pyramidal neurons. *Neurobiol Dis* 12, 174-81 (2003).
33. Gallagher, M., Burwell, R. & Burchinal, M. Severity of spatial learning impairment in aging: development of a learning index for performance in the Morris water maze. *Behav Neurosci* 107, 618-26 (1993).
34. Whishaw, I.Q. Cholinergic receptor blockade in the rat impairs locale but not taxon strategies for place navigation in a swimming pool. *Behav Neurosci* 99, 979-1005 (1985).
35. Cros, J. et al. Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. *Immunity* 33, 375-86 (2010).
36. Buerli, T. et al. Efficient transfection of DNA or shRNA vectors into neurons using magnetofection. *Nat Protoc* 2, 3090-101 (2007).
37. Lardi-Studler, B. et al. Vertebrate-specific sequences in the gephyrin E-domain regulate cytosolic aggregation and postsynaptic clustering. *J Cell Sci* 120, 1371-82 (2007).
38. Kuwahara, S. et al. PSPC1, NONO, and SFPQ are expressed in mouse Sertoli cells and may function as coregulators of androgen receptor-mediated transcription. *Biol Reprod* 75, 352-9 (2006).
39. Fritschy, J.M. & Mohler, H. GABAA-receptor heterogeneity in the adult rat brain: differential regional and cellular distribution of seven major subunits. *J Comp Neurol* 359, 154-94 (1995).
40. Brunig, I., Suter, A., Knuesel, I., Luscher, B. & Fritschy, J.M. GABAergic terminals are required for postsynaptic clustering of dystrophin but not of GABA(A) receptors and gephyrin. *J Neurosci* 22, 4805-13 (2002).
41. Notter, T., Panzanelli, P., Pfister, S., Mircsof, D. & Fritschy, J.M. A protocol for concurrent high-quality immunohistochemical and biochemical analyses in adult mouse central nervous system. *Eur J Neurosci* 39, 165-75 (2014).



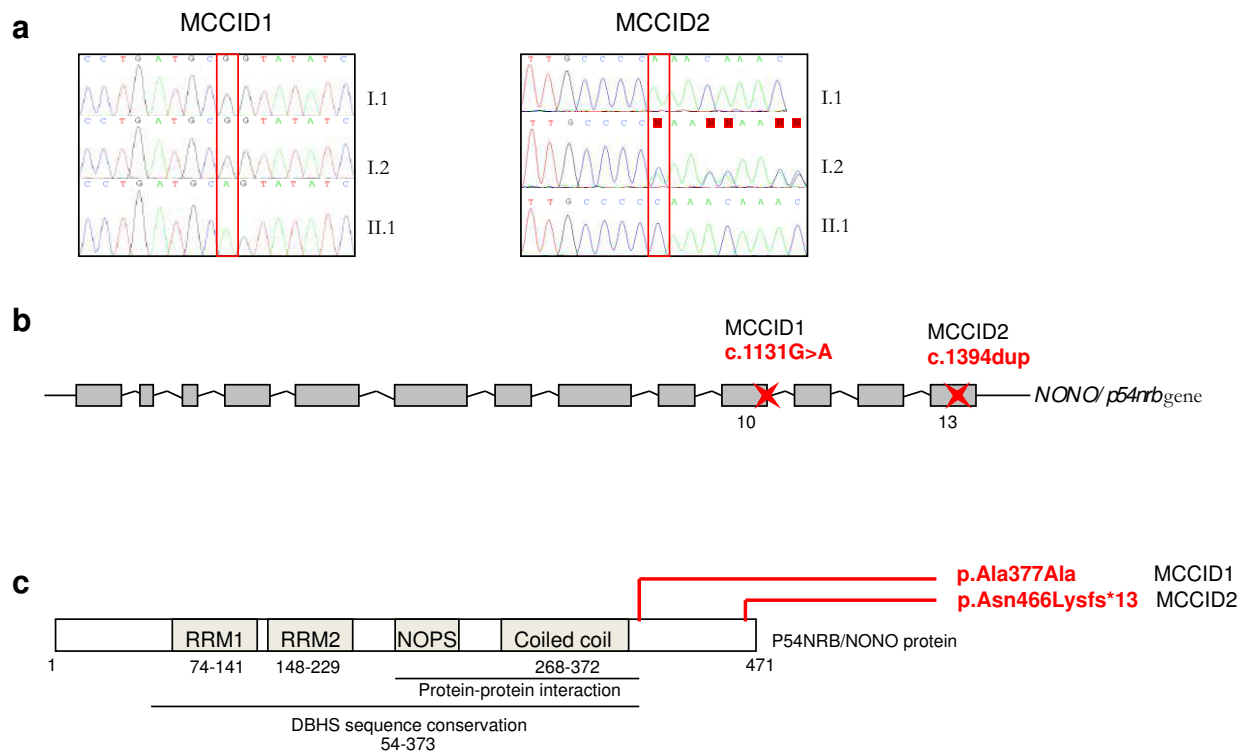




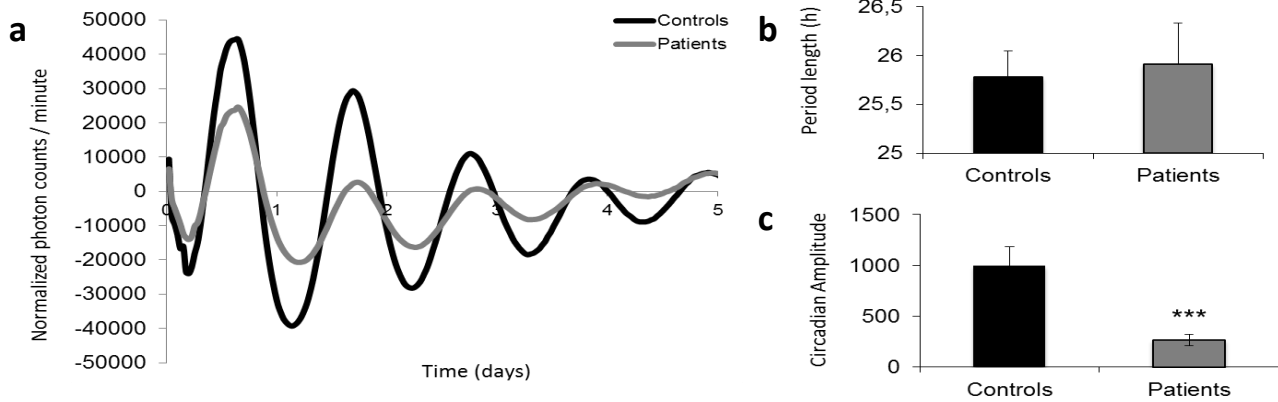




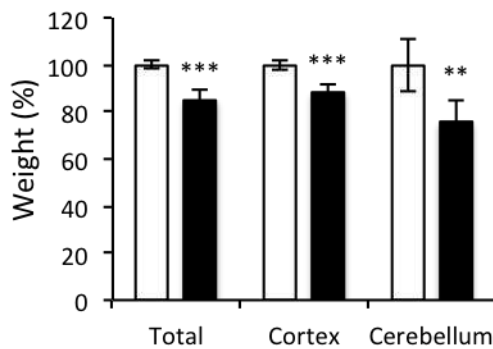
## SUPPLEMENTARY INFORMATION



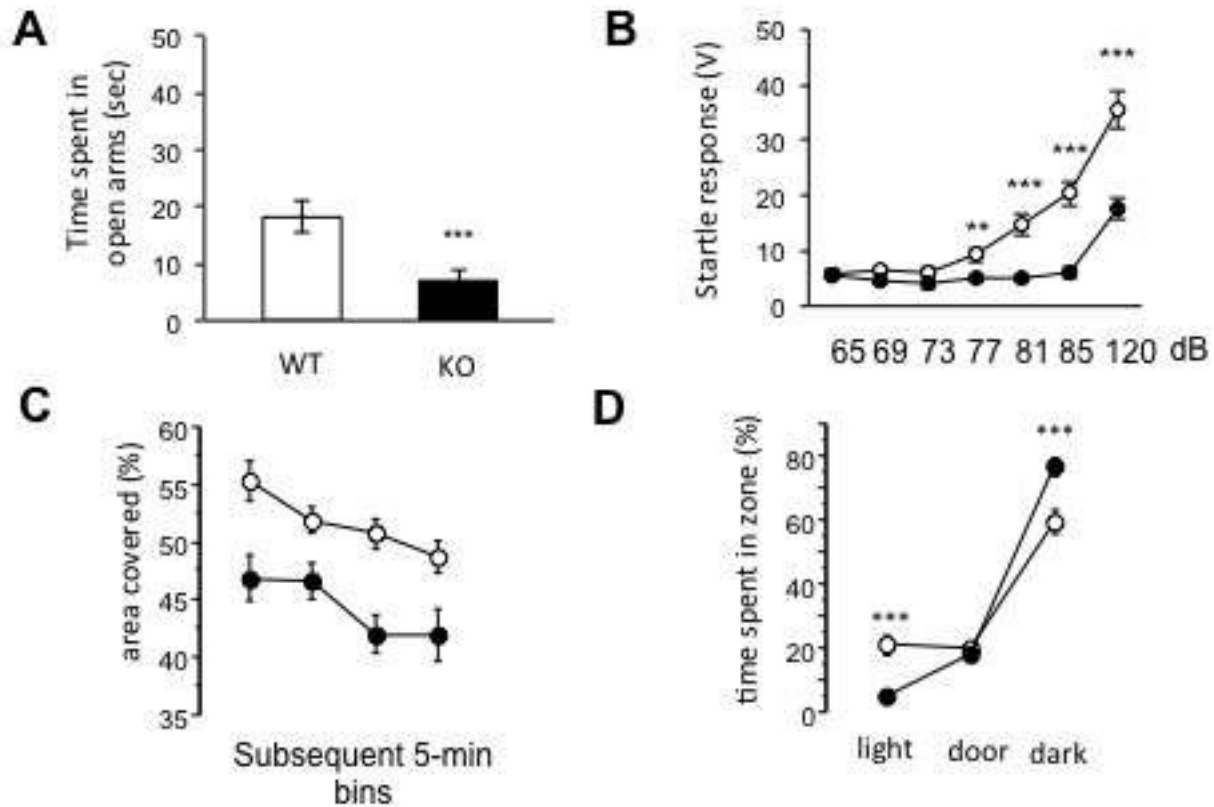
**Supplementary Figure 1. Molecular analyses of patients MCCID1 and MCCID2.** (a) Sanger sequencing chromatograms showing the NONO mutations in the probands and their parents. (b) Schematic representation of the NONO transcript showing exon structure. (c) Schematic overview of the NONO protein showing the different functional domains.



**Supplementary Figure 2. Reduced circadian amplitude in patient fibroblasts.** (a) Graph of average circadian oscillations of bioluminescence from *Bmal1-luciferase* reporter-infected fibroblasts from patients (grey) or siblings (black). Y-axis, background-subtracted bioluminescence in photons per minute; X-axis, time in days. Depicted curve is the average of three independent experiments in technical quadruplicate. (b) Circadian period measured in cells from controls (black bars) and patients (open bars). (c) Circadian amplitude measured in the same cells (arbitrary units). Student t-test,  $p < 0.001$ . Values for both panels are the average of three independent experiments with fibroblasts from both patients and their siblings in technical quadruplicate.

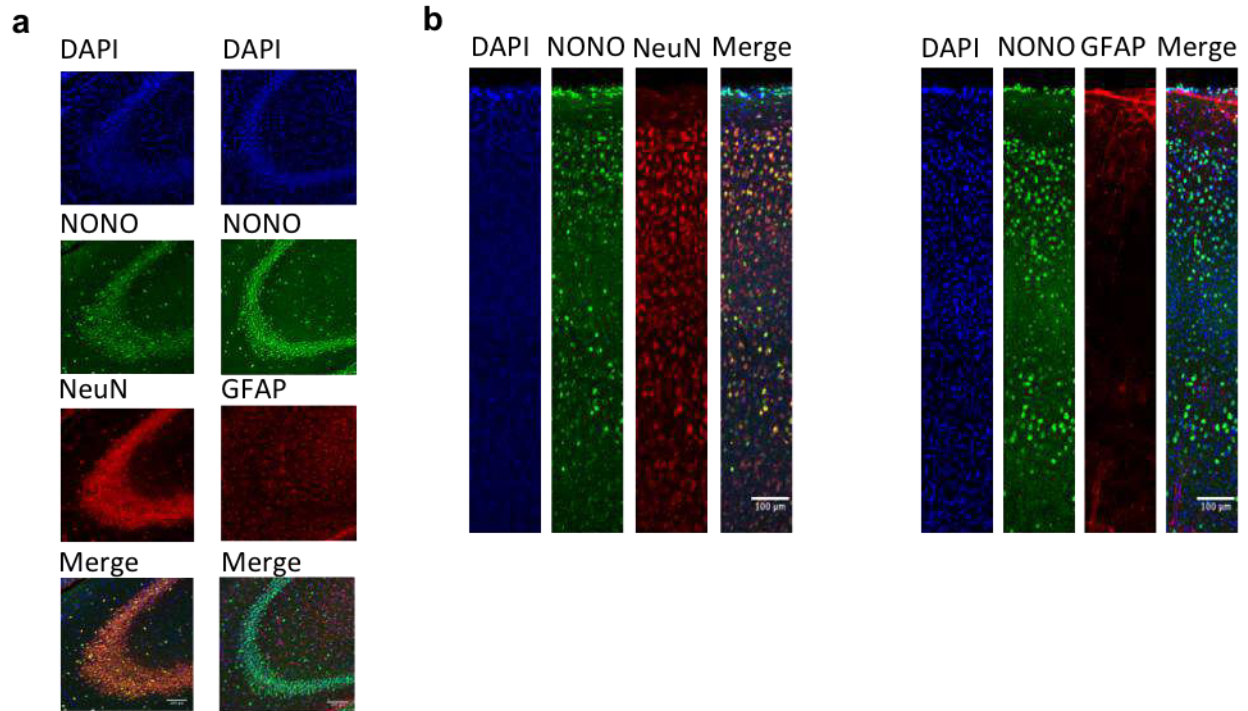
**a****b**

**Supplementary Figure 3. Brain morphology in *Nono*<sup>gt</sup> mice.** (a) Photograph of representative brains from WT (left) and *Nono*<sup>gt</sup> (right) mice. (b) Quantification of weight of whole brain, cortex, and cerebellum. N=5-9 mice per genotype. Student t-test. White bars, WT. Black bars, *Nono*<sup>gt</sup>. In this and subsequent figures, \*\*\* p<0.001, \*\* p<0.01, \* p<0.05. Quantification of all brain parameters from different morphological tests is shown in detail in Supplementary Table 4.

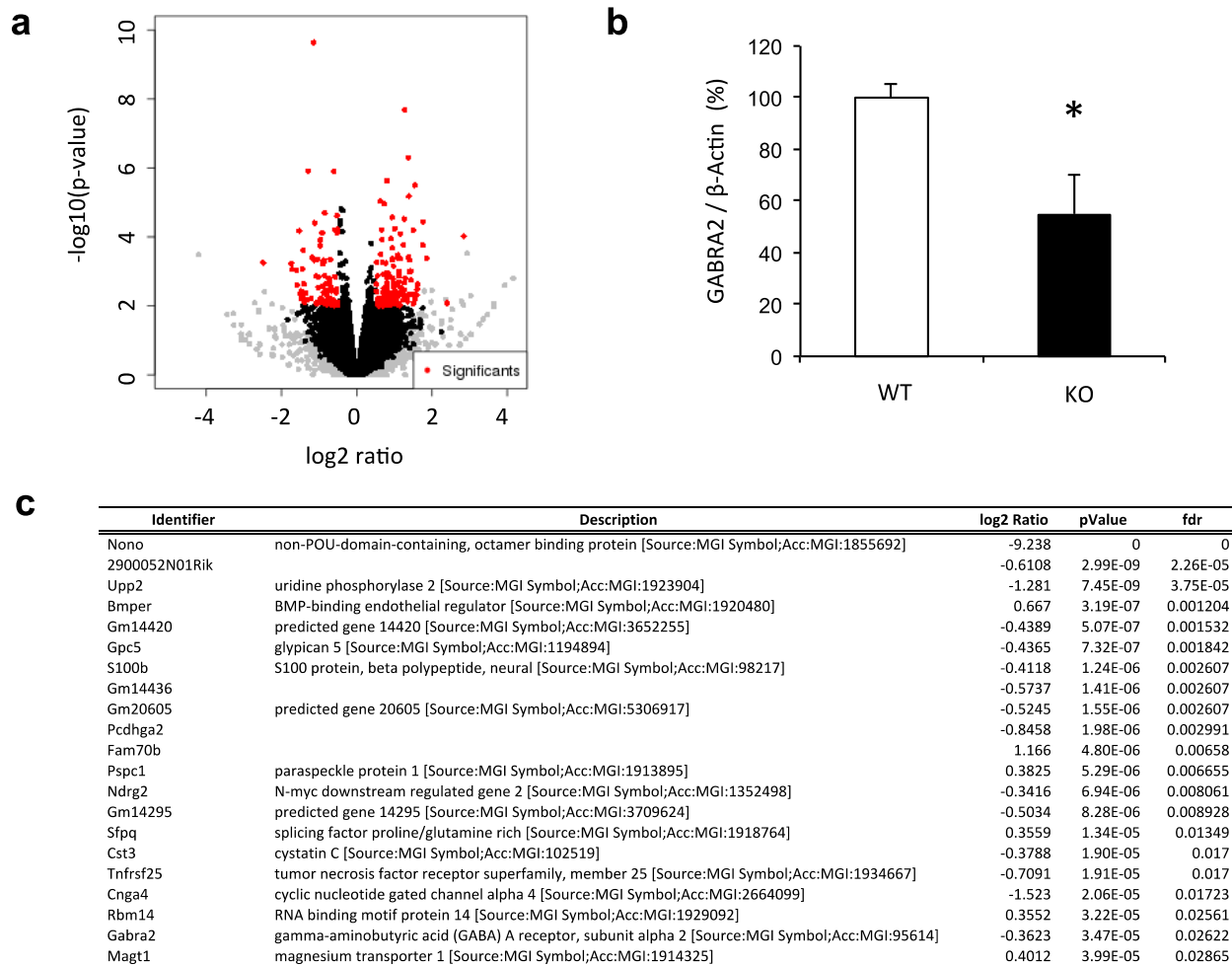


**Supplementary Figure 4. Anxiety-related phenotypes in *Nono*<sup>gt</sup> mice.** (a) Percentage of time spent in open arms of elevated plus-maze. N=13-14 mice per genotype. Student t-test,  $p < 0.001$  (b) Post-conditioning startle response in prepulse inhibition test. Y-axis, whole-body startle response in volts; X-axis, stimulus in decibels. N=13-14 mice per genotype. Student t-test,  $p < 0.01$  or  $0.001$  as indicated. (c) Open-field exploration. Y-axis, percentage of area surface tiles visited. X-axis, subsequent 5-minute intervals after commencement of test. N=18 mice per genotype, repeated ANOVA, gene  $p < .0002$ , time  $p < .0001$ , time x gene n.s. (d) Light-dark transition test. Y-axis, percentage of time spent in zones indicated on X-axis. N=18 mice per genotype, repeated ANOVA, gene  $p < .0001$ , zone  $p < .0001$ , zone x gene  $p < 0.0001$ . In all panels, *Nono*<sup>gt</sup> mice are represented by black bars/circles, compared to WT littermates (open).

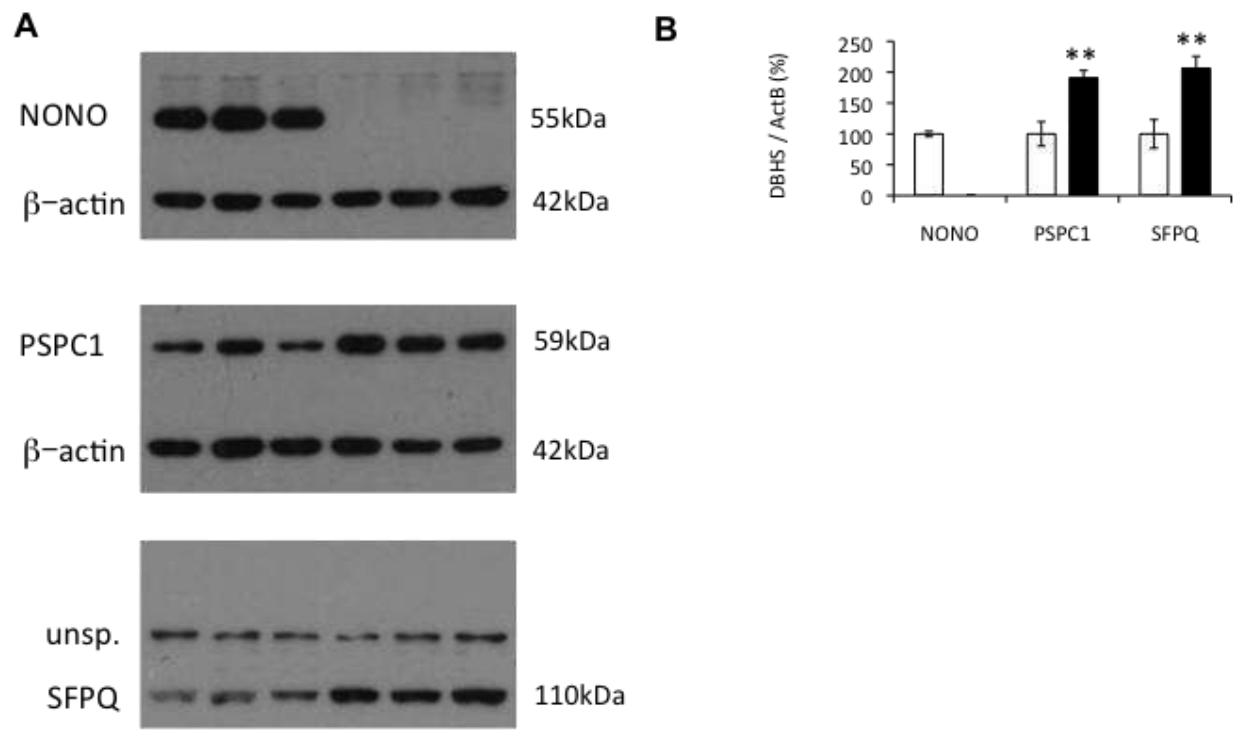




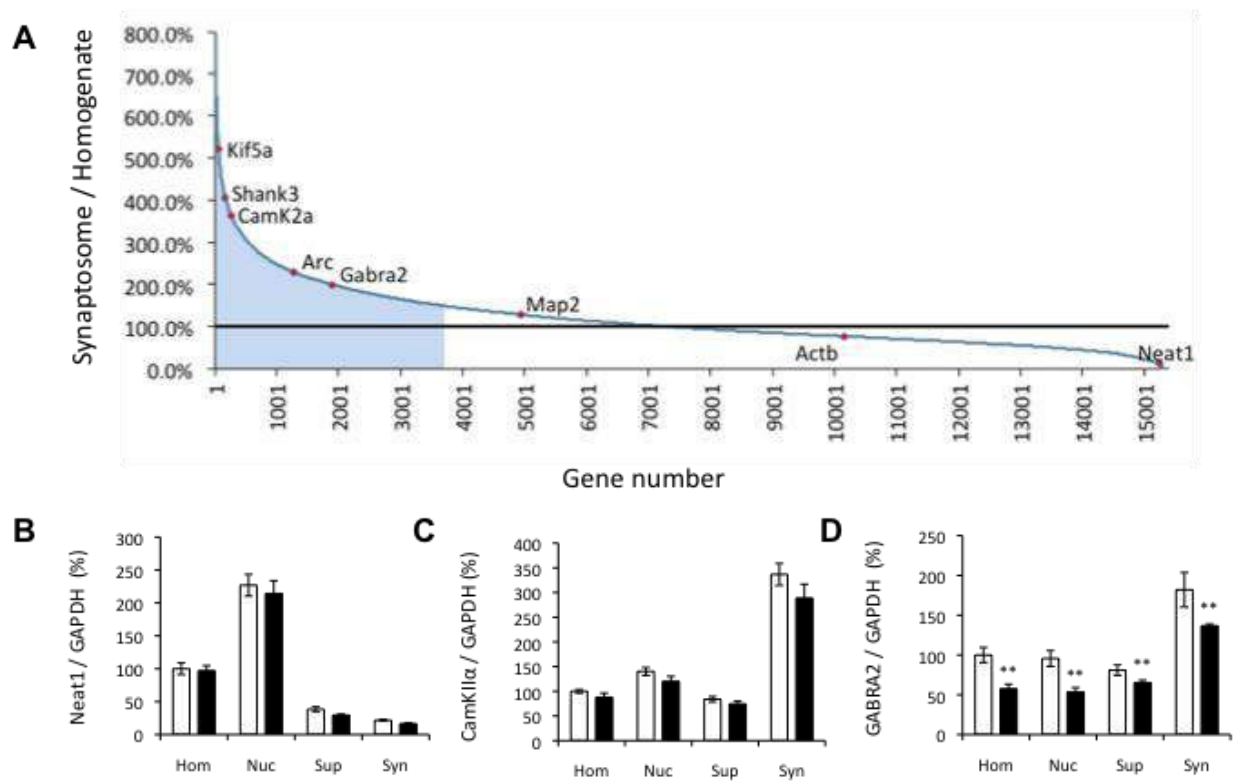
**Supplementary Figure 5. Cell type and layer-specific localization of NONO in mouse brain.** (a) Staining of hippocampal cell nuclei by DAPI (blue), anti-NONO (green), and neuron-specific anti-NeuN or astrocyte-specific anti-GFAP (red, left or right column respectively). (b) Identical staining of cortex.



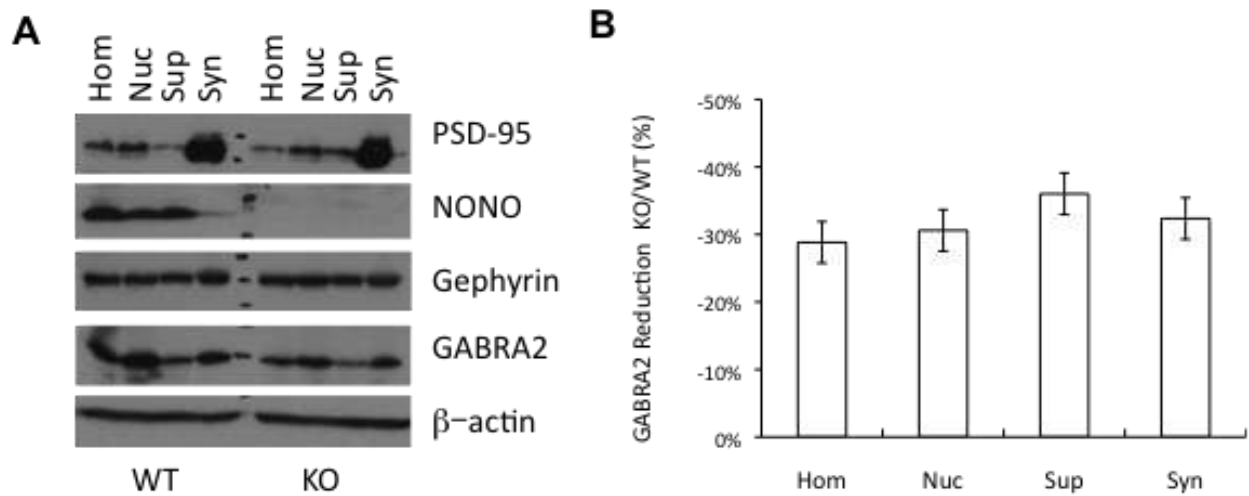
**Supplementary Figure 6. Widespread dysregulation of transcription in *Nono<sup>gt</sup>* mice hippocampi.** (a) Volcano plot of deregulated genes. Red dots,  $p < 0.01$  and  $\log_2 > 0.5$ . (b) Reduced GABRA2 protein levels in hippocampi of *Nono<sup>gt</sup>* mice. White bars, WT. Black bars, *Nono<sup>gt</sup>*. Student t-test,  $N=4$ . (c) List of most severely deregulated transcripts, showing upregulation of sister DBHS family members *Sfpq* and *Pspc1*. The full dataset is available as GEO Accession number XXX.



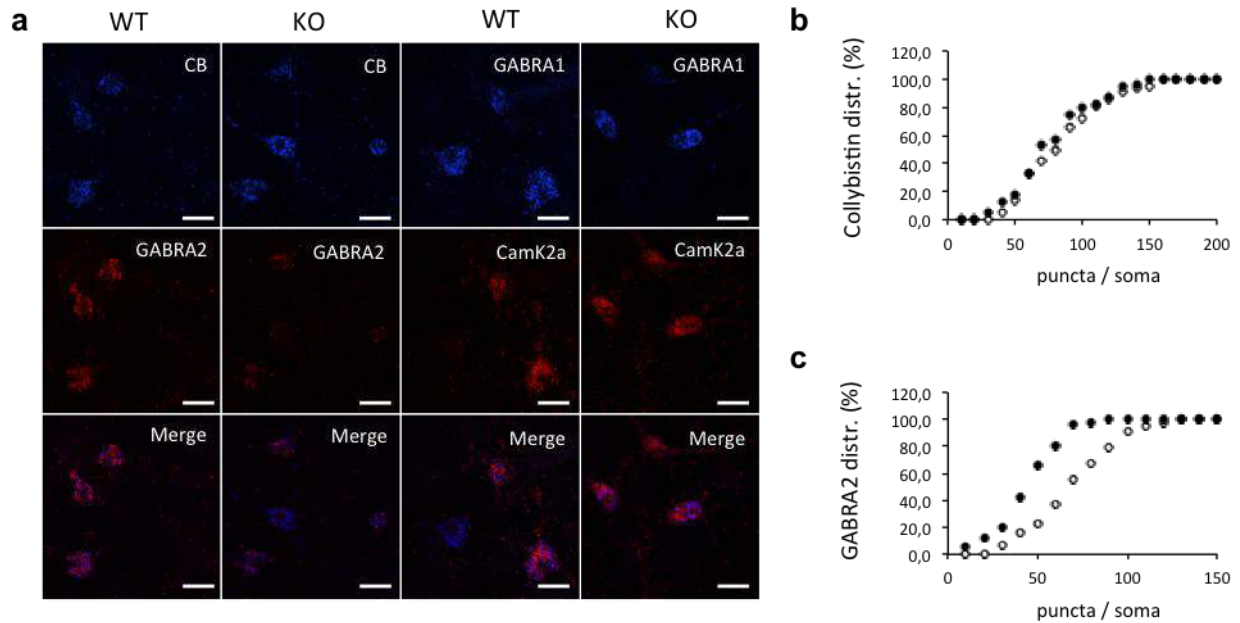
**Supplementary Figure 7. Overexpression of PSPC1 and SFPQ protein in *Nono*<sup>gt</sup> mice hippocampi.** (a) Western blots of hippocampal protein extracts from wildtype and *Nono*<sup>gt</sup> mice hippocampi, probed with anti-NONO, anti-PSPC1, anti-SFPQ, and anti-β-actin. N=3 mice per genotype. (b) Quantification of (A). White bars, WT. Black bars, *Nono*<sup>gt</sup>. Student t-test, p<0.01.



**Supplementary Figure 8. Transcript abundance in different neuronal compartments in *Nono*<sup>gt</sup> mice.** (a) Profile of the mouse forebrain synaptic transcriptome. Y-axis, ratio of transcript reads from synaptosome RNA sequencing compared to total. Selected transcripts previously characterized to be transported to synapses (*Kif5a*, *Shank3*, *CamK2A*, *Arc*, *Gabra2*), present throughout the cell (*Map2*, *Actb*), or retained in the nucleus (*Neat1*) are indicated to demonstrate the quality of synaptosomal transcript enrichment. Blue shading, transcripts more than 1.5x enriched. (b) Quantification of *Neat1* compared to *Gapdh* in whole-cell homogenate, purified nuclei, supernatant, or gradient-purified synaptosomes from WT (white bars) or *Nono*<sup>gt</sup> mice (solid bars). (c) Quantification of *CamkIIα* compared to *Gapdh*. (d) Quantification of *Gabra2* compared to *Gapdh*.



**Supplementary Figure 9. Reduction of synaptic GABRA2 in *Nono<sup>gt</sup>* mice.** (a) Western blot showing fractionation of mouse forebrain into different neuronal compartments by density gradient centrifugation. Immunohistochemistry using antibodies against NONO, gephyrin, and GABA is pictured. As a control,  $\beta$ -actin and dendrite-enriched PSD-95 are also shown. (b) Quantification of the reduction in GABRA2 levels in each compartment.



**Supplementary Figure 10. Specific reduction of *Gabra2* transcript levels in cultured *Nono<sup>gt</sup>* neurons.** (a) Single-transcript-resolution RNA *in-situ* hybridization to detect abundance of *collybistin* (CB, blue), *Gabra1* (blue), *Gabra2* (red), and *CamKII* (red) in WT or *Nono<sup>gt</sup>* neurons. Two transcripts were tested in different colors in a single plate of cells, and depicted in a single column. (b) Quantification of *collybistin* transcript distribution in puncta number per cell for 44-56 cells from 2 experiments. (c) Similar quantification of *Gabra2*. Solid circles, *Nono<sup>gt</sup>* neurons; open circles, WT.

**Supplementary Table 1: Clinical characteristics of patients with *NONO* mutations**

	<i>MCCID1</i>	<i>MCCID2</i>
Antenatal findings	-	hydramnios
Term born	41 WG	WG
Birth parameters (BW, BL and OFC in grammes and cm)	3370 g, 50.5 cm, 34 cm	2540 g, 46 cm, 35.5 cm
Neonatal history	-	poor sucking, GR, hypotonia
Motor skills	walked at 3 y	walked at 3 y
Language	short sentences	short sentences
Nasal speach	++	++
Elocution disability	+ (severe)	+ (severe)
Drooling	+	+
ID	+	+
Epilepsy	+ (5 years)	-
Ophthalmologic findings	convergent strabismus, myopia	convergent strabismus
Behaviour	shy, gentle and cheerful	shy, gentle and cheerful
Sleep disorder	-	-
Age at examination	17 y	15 y
Growth parameters (W, H and OFS in SD)	-1.5SD, + 1SD, +4SD)	-2 SD, -2 SD, +2SD
Puberty	-	delayed
Slender built	+	+
Scoliosis	+	+++
Arachnodactyly	+	-
MCP joint ankylosis of P1	+ (bilateral)	+ (bilateral)
Pes planus	++	++
Facial features		
Long face	+	+
Palpebral fissures	up slanting	up slanting
Malar hypoplasia	++	++
Nose	Thin and high nasal root, deviated nasal septum	Thin and high nasal root, deviated nasal septum
Mouth	open	small and open
Palate	narrow and high arched	narrow and high arched
Teeth	crowding	crowding and dental caries
EEG	Abnormal	No gross anomaly
Brain MRI		
Corpus callosum	Thick (+ cyst of the septum pellucidum)	Thick
ventricles	nl	asymmetric lateral ventricles
Cerebellum	hypoplastic	Chiari malformation type I

Abbreviations are as follows:

WG : weeks of gestation; BW : birth weight; BL : birth length, OFC : occipitofrontal circumference, W : weight, H, height, y : year, SD : standard deviation; ID : intellectual deficiency, nl : normal, GR : gastro-oesophageal reflux, - : absent, + : present

**Supplementary Table 2 : Number of variants identified by whole-exome sequencing**

<i>Patients</i>	<i>MCCID1</i>	<i>MCCID2</i>
Total variants	64010	71478
Novel or rare variants (dbSNP130/1000GP/EVS/in-house database)	1022	1030
Coding Non-synonymous/indel/consensus SS variants	255	248
exclusively novel variants	88	87
Shared gene whose mutations co-segregate in the families	1	1



**Supplementary Table 3: Differentially regulated transcripts in patients and *Nono*-deficient mice.** The full dataset is available from ArrayExpress (accession E-MTAB-2894 and E-MTAB-2895),

DE Genes	Profils (MCCID1, <i>NONOgt</i> mice, MCCID2)	Full name	MCCID1			<i>NONOgt</i> mice			MCCID2		
			p-value	Fold	Delta	p-value	Fold	Delta	p-value	Fold	Delta
ACTA2	UUU	Actin, alpha 2, smooth muscle, aorta	0,00	2,93	28,06	0,00	34,96	220,94	0,02	2,34	40,08
BHLHE40	UUU	basic helix-loop-helix family, member e40	0,01	1,81	116,37	0,03	4,28	1309,55	0,01	1,63	90,20
CNN1	UUU	calponin 1, basic, smooth muscle	0,01	1,62	113,90	0,01	106,91	545,44	0,00	1,53	319,56
COL11A1	UUU	collagen, type XI, alpha 1	0,00	13,89	5906,34	0,00	697,63	4069,19	0,00	9,92	4084,12
COL12A1	UUU	collagen, type XII, alpha 1	0,00	2,06	44,11	0,01	460,70	6067,24	0,00	1,59	24,50
DKK3	UUU	dickkopf 3 homolog ( <i>Xenopus laevis</i> )	0,01	1,52	28,35	0,03	3,36	3000,41	0,01	1,67	235,72
DLG1	UUU	discs, large homolog 1 ( <i>Drosophila</i> )	0,00	1,75	31,84	0,01	1,58	660,25	0,00	1,82	35,24
FST	UUU	follistatin	0,01	1,61	266,91	0,05	10,45	63,58	0,01	1,60	261,20
GPC4	UUU	glypican 4	0,00	4,85	818,41	0,03	3,68	2496,10	0,00	2,07	226,85
LMCD1	UUU	LIM and cysteine-rich domains 1	0,01	1,62	73,52	0,00	83,09	479,42	0,03	2,41	828,09
MEX3B	UUU	mex-3 homolog B ( <i>C. elegans</i> )	0,01	2,84	171,59	0,02	2,18	322,22	0,04	1,56	51,83
NCAM1	UUU	neural cell adhesion molecule 1	0,00	4,20	204,70	0,02	6,66	3332,05	0,02	1,69	43,95
PSPC1	UUU	paraspeckle component 1	0,01	2,22	494,70	0,03	2,57	615,45	0,01	2,78	721,39
RHOJ	UUU	ras homolog family member J	0,00	3,91	108,47	0,05	2,76	631,44	0,00	2,28	47,76
RNF128	UUU	ring finger protein 128, E3	0,01	2,01	33,71	0,01	11,57	102,80	0,01	3,24	74,43
SALL1	UUU	ubiquitin protein ligase sal-like 1 ( <i>Drosophila</i> )	0,02	2,81	139,02	0,00	57,88	280,83	0,04	2,46	71,50
SORBS2	UUU	sorbin and SH3 domain containing 2	0,00	1,79	9,28	0,03	6,09	23,75	0,01	2,19	33,14
SOX11	UUU	SRY (sex determining region Y)-box 11	0,00	1,74	8,54	0,01	30,78	155,19	0,00	4,28	37,92
TSHZ2	UUU	teashirt zinc finger homeobox 2	0,01	2,47	89,36	0,02	2,80	10,19	0,04	1,80	48,72
TWIST2	UUU	twist homolog 2	0,00	2,14	300,21	0,00	4,50	1279,81	0,00	1,98	259,25

		(Drosophila)									
WISP1	UUU	WNT1 inducible signaling pathway protein 1	0,00	2,40	74,30	0,01	2,21	3716,63	0,01	3,38	125,83
ABCA5	DDD	ATP-binding cassette, sub-family A (ABC1), member 5	0,00	0,41	32,25	0,04	0,25	256,57	0,00	0,59	22,15
AHR	DDD	aryl hydrocarbon receptor	0,01	0,40	579,72	0,01	0,57	31,72	0,03	0,56	429,28
ALCAM	DDD	activated leukocyte cell adhesion molecule	0,01	0,41	97,88	0,01	0,25	486,49	0,02	0,39	101,38
ARHGAP18	DDD	Rho GTPase activating protein 18	0,00	0,47	796,59	0,00	0,21	1146,53	0,00	0,43	859,15
CCBE1	DDD	collagen and calcium binding EGF domains 1	0,00	0,25	625,52	0,02	0,18	1406,97	0,00	0,49	420,06
EBF3	DDD	early B-cell factor 3	0,00	0,28	89,36	0,02	0,24	133,74	0,04	0,37	77,64
GJC1	DDD	gap junction protein, gamma 1, 45kDa	0,03	0,58	510,20	0,02	0,38	853,36	0,05	0,60	138,37
GLCC1	DDD	glucocorticoid induced transcript 1	0,03	0,50	85,92	0,01	0,32	177,94	0,02	0,37	107,83
GLRX	DDD	glutaredoxin (thioltransferase)	0,00	0,60	299,02	0,04	0,33	293,86	0,00	0,48	390,86
HAS2	DDD	hyaluronan synthase 2	0,02	0,62	281,41	0,01	0,32	223,83	0,00	0,44	412,14
IGF2BP3	DDD	insulin-like growth factor 2 mRNA binding protein 3	0,00	0,29	304,67	0,05	0,42	362,53	0,00	0,20	344,29
MECOM	DDD	MDS1 and EVI1 complex locus	0,00	0,55	11,32	0,03	0,49	710,76	0,00	0,63	9,14
NAV2	DDD	neuron navigator 2	0,00	0,58	262,16	0,00	0,20	155,82	0,01	0,56	277,64
NEGR1	DDD	neuronal growth regulator 1	0,01	0,56	231,28	0,01	0,04	105,79	0,03	0,58	219,71
PTER	DDD	phosphotriesterase related	0,00	0,44	157,03	0,00	0,62	262,85	0,01	0,34	42,49
RDH10	DDD	retinol dehydrogenase 10 (all-trans)	0,04	0,59	258,51	0,02	0,45	273,84	0,02	0,39	380,29
RGS20	DDD	regulator of G-protein signaling 20	0,00	0,43	65,24	0,03	0,11	325,08	0,00	0,64	40,72
RICTOR	DDD	RPTOR independent companion of MTOR, complex 2	0,00	0,45	41,60	0,01	0,55	452,22	0,01	0,54	34,98
TLE1	DDD	transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	0,00	0,42	66,91	0,02	0,50	265,44	0,01	0,52	56,10
TRIOBP	DDD	TRIO and F-actin binding protein	0,01	0,52	69,16	0,04	0,29	1158,17	0,00	0,46	77,63
WDR20	DDD	WD repeat domain 20	0,00	0,46	99,70	0,02	0,66	60,93	0,00	0,45	102,71

D downregulated; U: upregulated

**Supplementary Table 4. Quantification of brain morphology in *Nono<sup>gt</sup>* mice.**

	WT (%)	Stdev	NONO KO	Stdev	p Value
Brain weight	100.00	1.88	85.06	4.87	p < 0.0001
Cortex	100.00	2.13	88.25	3.69	p < 0.0001
Cerebellum	100.00	11.27	75.73	8.91	p = 0.0004
Skull lenght	100.00	1.77	91.56	4.62	p = 2.49E-7
Skull widht	100.00	2.61	96.85	6.00	p = 3.45E-6
Nose lenght	100.00	3.69	84.97	9.06	p = 6.55E-7
Relative nose length	32.47	0.90	29.86	2.67	p = 0.00019
Skull Width/Length	45.74	1.06	48.10	2.46	p = 0.00032

**Supplementary Table 5. Gene ontology localization of proteins identified as dysregulated in *Nono<sup>gt</sup>* mice.**

#	Localizations	p-value	FDR
1	<a href="#">extracellular space</a>	2.809E-13	1.103E-10
2	<a href="#">extracellular region part</a>	3.250E-13	1.103E-10
3	<a href="#">extracellular matrix</a>	2.233E-11	5.053E-09
4	<a href="#">extracellular region</a>	7.617E-11	1.293E-08
5	<a href="#">plasma membrane</a>	6.585E-10	8.943E-08
6	<a href="#">cell periphery</a>	3.689E-09	4.174E-07
7	<a href="#">cell surface</a>	7.807E-09	7.573E-07
8	<a href="#">cytoplasmic part</a>	2.082E-08	1.767E-06
9	<a href="#">proteinaceous extracellular matrix</a>	4.814E-08	3.632E-06
10	<a href="#">organelle</a>	9.166E-07	6.224E-05
11	<a href="#">extracellular matrix part</a>	1.813E-06	1.119E-04
12	<a href="#">vesicle</a>	3.265E-06	1.848E-04
13	<a href="#">neuron part</a>	4.957E-06	2.589E-04
14	<a href="#">plasma membrane part</a>	6.855E-06	2.981E-04
15	<a href="#">neuronal cell body</a>	7.218E-06	2.981E-04
16	<a href="#">external side of plasma membrane</a>	7.241E-06	2.981E-04
17	<a href="#">basement membrane</a>	7.464E-06	2.981E-04
18	<a href="#">neuron projection</a>	8.693E-06	3.279E-04
19	<a href="#">membrane</a>	1.022E-05	3.653E-04
20	<a href="#">membrane-bounded vesicle</a>	1.111E-05	3.771E-04
21	<a href="#">cell body</a>	1.958E-05	6.330E-04
22	<a href="#">cell projection</a>	2.556E-05	7.890E-04
23	<a href="#">intracellular organelle</a>	2.726E-05	8.047E-04
24	<a href="#">extracellular vesicular exosome</a>	3.309E-05	9.362E-04
25	<a href="#">extracellular membrane-bounded organelle</a>	3.969E-05	1.036E-03
26	<a href="#">extracellular organelle</a>	3.969E-05	1.036E-03
27	<a href="#">membrane-enclosed lumen</a>	5.667E-05	1.425E-03
28	<a href="#">cell projection part</a>	7.825E-05	1.898E-03
29	<a href="#">integral component of plasma membrane</a>	8.650E-05	2.025E-03

30	<a href="#">intrinsic component of plasma membrane</a>	1.612E-04	3.648E-03
31	<a href="#">organelle lumen</a>	1.755E-04	3.844E-03
32	<a href="#">dendrite</a>	2.060E-04	4.372E-03
33	<a href="#">pericentriolar material</a>	2.329E-04	4.792E-03
34	<a href="#">basal lamina</a>	3.408E-04	6.807E-03
35	<a href="#">side of membrane</a>	4.723E-04	9.115E-03
36	<a href="#">protein complex</a>	4.833E-04	9.115E-03
37	<a href="#">haptoglobin-hemoglobin complex</a>	5.051E-04	9.270E-03
38	<a href="#">intracellular organelle part</a>	5.319E-04	9.505E-03
39	<a href="#">membrane raft</a>	5.680E-04	9.730E-03
40	<a href="#">perinuclear region of cytoplasm</a>	5.802E-04	9.730E-03
41	<a href="#">membrane-bounded organelle</a>	5.875E-04	9.730E-03
42	<a href="#">lysosomal lumen</a>	6.163E-04	9.964E-03
43	<a href="#">anchoring junction</a>	6.784E-04	1.037E-02
44	<a href="#">histone acetyltransferase complex</a>	6.835E-04	1.037E-02
45	<a href="#">macromolecular complex</a>	6.872E-04	1.037E-02
46	<a href="#">nucleoplasm part</a>	7.722E-04	1.119E-02
47	<a href="#">organelle part</a>	7.744E-04	1.119E-02
48	<a href="#">vacuolar lumen</a>	9.221E-04	1.304E-02
49	<a href="#">intracellular organelle lumen</a>	9.534E-04	1.321E-02
50	<a href="#">lysosome</a>	9.928E-04	1.322E-02
51	<a href="#">lytic vacuole</a>	9.928E-04	1.322E-02
52	<a href="#">cytosol</a>	1.061E-03	1.385E-02
53	<a href="#">MHC class II protein complex</a>	1.163E-03	1.490E-02
54	<a href="#">perikaryon</a>	1.266E-03	1.591E-02
55	<a href="#">collagen type I trimer</a>	1.421E-03	1.750E-02
56	<a href="#">protein acetyltransferase complex</a>	1.469E-03	1.750E-02
57	<a href="#">acetyltransferase complex</a>	1.469E-03	1.750E-02
58	<a href="#">dendritic spine</a>	1.582E-03	1.852E-02
59	<a href="#">paraspeckles</a>	1.670E-03	1.916E-02
60	<a href="#">neuron spine</a>	1.693E-03	1.916E-02

**Supplementary Table 6. Quantification of synaptic transcripts in *Nono*<sup>gt</sup> mice.**

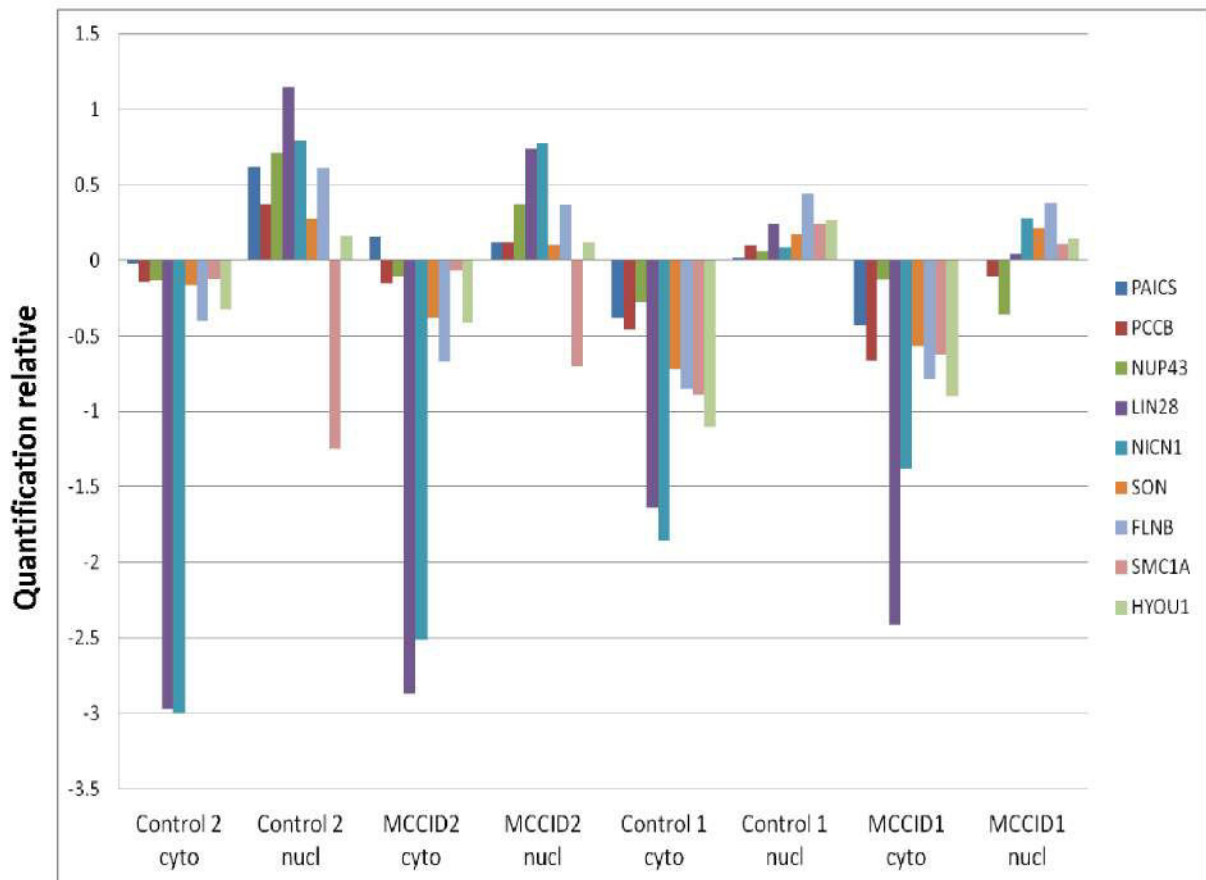
Synaptically enriched genes	3668
Total expressed genes	15393
	23.83%
NONO regulated genes	886
NONO regulated genes with synaptic localization	270
	30.47%

Chi<sup>2</sup> = 11.109, with 1 degree of freedom.

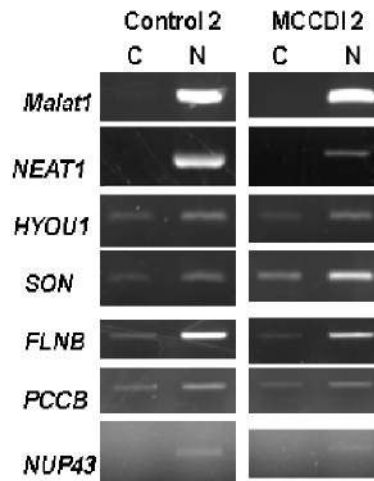
P = 0.0009, 2-tailed

## 9.2 Données supplémentaires : rôle de NONO au sein des paraspeckles

L'une des nombreuses fonctions de NONO décrites dans la littérature est son rôle au sein des paraspeckles : des structures ribo-nucléoprotéiques récemment découvertes, localisées dans l'espace interchromatique et dont le composant principal est le long ARN non codant *Neat1* (*MEN $\epsilon$ / $\beta$* ). Le rôle principal des paraspeckles est d'assurer la rétention nucléaire des ARN hyperédités (Adenosine vers Inosine). Il a été montré que les protéines NONO et PSF ainsi que l'ARN *Neat1* sont nécessaires à l'intégrité des paraspeckles dans les cellules HeLa [Sasaki et al., 2009, Clemson et al., 2009, Chen and Carmichael, 2009, Sunwoo et al., 2009]. Nous avons donc souhaité étudier les conséquences des mutations *NONO* sur la localisation cellulaire des ARN édités. Pour cela, nous avons réalisé des expériences de fractionnement cellulaire visant à séparer les ARN cytoplasmiques des ARN nucléaires (protocole détaillé dans la partie *Matériel et Méthodes*, figure 18.1). La qualité du fractionnement a été vérifiée grâce aux marqueurs nucléaires *Neat1* et *Malat1* ainsi que 4 ARN non-édités mais naturellement enrichis dans le noyau d'après Sasaki et al. [2009] : *SMC1A*, *FLNB*, *HYOU1*, et *SON*. Nous avons ensuite évalué l'efficacité de la rétention nucléaire pour 5 ARNm connus pour être édités et retenus au noyau : *LIN28*, *NUP43*, *NICN1*, *PAICS* et *PCCB* [Chen and Carmichael, 2009, Chen et al., 2008]. Comme l'indiquent les figures 9.1 et 9.2, aucune différence significative de localisation des 5 ARN édités n'est observée dans les fibroblastes de patients par rapport à des cellules issues de sujets sains.



**Figure 9.1 :** Distribution cytoplasmique vs nucléaire des ARN testés : résultats de qPCR.



**Figure 9.2** : Distribution cytoplasmique vs nucléaire des ARN testés : résultats de RT-PCR.

La majorité des rôles de NONO sont partagés avec PSF et les deux protéines forment un hétérodimère [Myojin et al., 2004, Kuwahara et al., 2006, Passon et al., 2012]. Nous proposons donc que la sur-expression des protéines PSF et PSCP1 observée dans les cellules de nos patients et chez les souris *Nono<sup>gt</sup>* compense l'absence de protéine NONO via la formation d'un complexe PSF/PSCP1 en plus grande quantité capable d'assurer seul les principales fonctions des paraspeckles. Une compensation similaire a été observée récemment par Li et al. [2014b] dans des MEFs de souris déficientes pour le gène *Nono*.

Enfin, il est possible que les conséquences de l'absence de NONO sur la formation et la fonction des paraspeckles soit dépendante du type cellulaire analysé. Ainsi, *Neat1*, essentiel à la formation des paraspeckles, n'est pas exprimé dans tous les types cellulaires [Nakagawa and Hirose, 2012]. De plus, la littérature fournit de nombreux exemples des rôles cellule-spécifique des paraspeckles. Ainsi, certains auteurs proposent l'existence d'un mécanisme de libération des transcrits retenus dans ces structures, en réponse au stress [Prasanth et al., 2005, Bond and Fox, 2009, Nakagawa and Hirose, 2012]. À l'inverse, selon d'autres auteurs, la formation des paraspeckles est déclenchée en réponse au stress [Shelkownikova et al., 2014]. Ces auteurs suggèrent qu'une altération de l'assemblage des paraspeckles dans les neurones contribue à la physiopathologie FUSoptahies (groupe de maladies neurodégénératives). Selon un autre modèle, les paraspeckles ont un rôle de séquestration des protéines telles que les DBHS, permettant de contrôler leur action [Nakagawa and Hirose, 2012, Hirose et al., 2014]. Toutefois, toutes ces études mettent en avant une étape supplémentaire et non essentielle dans le contrôle de l'expression des gènes, permettant une régulation rapide, fine et subtile, dépendante des conditions de stress, du type cellulaire et des conditions expérimentales [Nakagawa and Hirose, 2012].

En conclusion, l'absence d'altération de la rétention nucléaire liée aux paraspeckles dans les cellules de patients n'est pas en faveur de l'implication de ce mécanisme dans la physiopathologie de cette affection.



### 10.1 Syndrome MCCID : un syndrome polymalformatif lié à une dérégulation transcriptionnelle globale

En accord avec le rôle de NONO dans la biogenèse des ARN messagers, l'absence de cette protéine cause un large déséquilibre de l'expression des gènes dans les cellules de patients. Les catégories de GO les plus significativement enrichies en gènes différentiellement exprimés (DE) sont listées dans la table B.4, en annexe. Certaines de ces catégories, comme la croissance, la prolifération, le développement et la morphologie, sont compatibles avec plusieurs des signes observés chez les deux patients MCCID tels que l'allure marfanoïde, la mégalencéphalie et l'hypoplasie malaïre, en accord avec des études précédemment publiées [Park et al., 2013].

D'autre part, les résultats *in silico* de notre analyse transcriptomique ainsi que d'autres publications suggèrent un effet oncogène de *NONO* [Liu et al., 2014b, Tsofack et al., 2011, Schiffner et al., 2011, Pavao et al., 2001, Miyamoto et al., 2008, Ishiguro et al., 2003, Kim et al., 2006], soulignant le besoin de mettre en place une surveillance de nos patients pour détecter l'apparition d'éventuelles tumeurs. Enfin, plusieurs gènes DE sont impliqués dans le développement du squelette (*ITGB1BP1*, *PBX1*, *GDF6*). Par ailleurs, Hata et al. [2008] montrent que la protéine NONO stimule la fonction de SOX9 pendant la chondrogenèse chez la souris [Hata et al., 2008]. Ces observations pourraient donc expliquer les anomalies squelettiques observées chez les patients.

En conclusion, la dérégulation transcriptionnelle globale observée chez nos patients nous permet de classer le syndrome MCCID dans le large groupe des pathologies liées à un défaut de régulation de l'expression des gènes.

### 10.2 Contribution des anomalies du rythme circadien au phénotype des patients MCCID

Il a été démontré que NONO, comme les deux autres protéines de la famille DBHS, régule le cycle circadien chez la drosophile et les mammifères via son interaction avec les protéines PER [Brown et al.,

2005, Duong et al., 2011, Kowalska et al., 2012].

De nombreuses études sur des modèles animaux ont démontré l'importance du cycle circadien dans le développement et le fonctionnement du système nerveux. Chez la drosophile, le rythme circadien contrôle les modifications structurelles quotidiennes des neurones et des synapses [Pyza and Gorska-Andrzejak, 2008, Damulewicz and Pyza, 2011, Mehnert et al., 2007]. La formation de la mémoire à long terme est également modulée par l'horloge biologique chez le poisson-zèbre, chez lequel apprentissage et mémorisation se font plus efficacement pendant la journée que la nuit [Rawashdeh et al., 2007]. Ainsi, un traitement à la mélatonine pendant la journée mime la suppression nocturne de formation de la mémoire. Chez la souris, les oscillations circadiennes favorisent le remodelage synaptique induit par l'apprentissage et la formation de la mémoire [Liston et al., 2013]. De plus, Ruby et al. [2008] montrent qu'une altération du système circadien affecte les fonctions cognitives telles que la mémoire et l'apprentissage chez le hamster.

Chez l'homme, les troubles du sommeil et/ou du rythme circadien sont très fréquents dans les désordres cognitifs et de récentes études ont suggéré des mécanismes neurologiques communs entre le sommeil et la cognition permettant d'expliquer l'apparition conjointe de ces symptômes [Foster et al., 2013] : des analyses transcriptomiques évoquent une dérégulation du rythme circadien chez les patients bipolaires [Akula et al., 2014]. Plusieurs publications rapportent une diminution de la concentration de mélatonine chez des patients souffrant de troubles du spectre autistique [Nir et al., 1995, Kulman et al., 2000, Tordjman et al., 2005, Melke et al., 2008]. L'haploinsuffisance du gène *RAI1*, responsable du syndrome de Smith-Magenis (SMS, MIM 182290), est caractérisée par une DI, des anomalies congénitales multiples, une obésité, des anomalies neurocomportementales et des troubles du sommeil. Récemment, Williams et al. [2012] ont montré que *RAI1* contrôle directement l'expression du gène *CLOCK*, un acteur central du cycle circadien. Enfin, le gène *OPHN1*, responsable d'une forme de DILX, intervient également dans la régulation du cycle circadien dans l'hippocampe [Valnegri et al., 2011].

Ces données, combinées à l'observation d'une réduction de l'amplitude des oscillations circadiennes dans les fibroblastes de patients MCCID suggèrent que l'absence de protéine NONO puisse également altérer le développement cognitif des patients via son effet sur le rythme circadien.

Aucun trouble du sommeil n'est décrit chez nos deux patients mais aucune investigation spécifique n'a été menée. C'est la raison pour laquelle nous souhaitons aujourd'hui poursuivre l'exploration de ces patients en ciblant l'analyse du rythme circadien. Cette étude sera conduite en demandant aux parents de remplir un agenda de sommeil pendant 15 jours (voir en annexe : figure A.17) pour éventuellement révéler des particularités comme le décalage de phase.

Si nous obtenons l'accord des familles, une analyse complète du sommeil sera également réalisée en milieu hospitalier pour mesurer :

- l'activité électrique du cerveau par électro-encéphalogramme (EEG),
- l'activité musculaire par un électromyogramme (EMG),
- l'activité cardiaque par un électrocardiogramme (ECG),
- l'activité oculaire par un électrooculogramme (EOG),
- la respiration : le débit d'air circulant par le nez pour étudier la respiration et détecter les apnées du sommeil et l'activité des muscles respiratoires grâce à une sangle autour du thorax et de l'abdomen,
- les positions pendant le sommeil par un enregistrement vidéo,
- le niveau d'oxygène dans le sang grâce à un capteur fixé au bout du doigt.

Enfin, sur le plan biologique, doser la mélatonine (2-3 points dans la journée) permettrait de tester le rythme circadien à l'échelle des patients et ainsi déterminer si une anomalie du cycle circadien contribue au phénotype observé.

### 10.3 Anomalie des synapses GABAérgiques et pathologies neurodéveloppementales : des pistes thérapeutiques

L'exploration du modèle murin *Nono<sup>gt</sup>* a permis de mettre en évidence une dérégulation transcriptionnelle comparable à celle observée dans les fibroblastes de patients. Plus particulièrement, l'expression des transcrits synaptiques dont le transcrite du gène *GABRA2* est affectée dans l'hippocampe des souris *Nono<sup>gt</sup>*. Avec l'analyse du réseau de protéines post-synaptiques, nos données démontrent clairement une désorganisation des synapses inhibitrices dans les neurones de ces souris. Nos travaux démontrent également la réversibilité, *in vitro*, de ce phénotype qui peut être corrigé en induisant la sur-expression du gène *GABRA2*. Le GABA (acide  $\gamma$ -aminobutyrique) est le principal neurotransmetteur inhibiteur du système nerveux central chez les mammifères. L'équilibre dynamique entre excitation et inhibition des signaux synaptiques est crucial pour le bon fonctionnement des circuits neuronaux. Nous proposons donc que le phénotype cognitif observé chez nos patients soit principalement la conséquence d'un déséquilibre entre excitation et inhibition.

De nombreuses maladies neurodéveloppementales telles que les syndromes de l'X fragile, de Rett, de Down, de Tourette, les troubles du spectre autistique, la schizophrénie ou la neurofibromatose sont liées à une dérégulation de la transmission GABAérgique. Depuis plusieurs années, de nombreux laboratoires ont donc testé diverses molécules modulant la transmission GABAérgique sur des modèles cellulaires et animaux de ces maladies [Deidda et al., 2014].

En dehors des sites de liaison du GABA, une variété d'autres sites récepteurs topographiquement distincts sont en effet capables de reconnaître des substances pharmacologiquement actives, en compétition ou non avec le GABA comme les benzodiazépines (BZDs), les barbituriques, les neurostéroïdes, les convulsivants, l'alcool, le muscimol, la bicuculline et la picrotoxine. Ces substances interagissent de manière allostérique avec les sites récepteurs au GABA et modulent la réponse GABA<sub>A</sub>. La plupart des benzodiazépines (BZD) et le muscimol sont des agonistes qui favorisent l'ouverture du canal Cl<sup>-</sup> par le GABA et ont donc un effet inhibiteur. Elles ont des propriétés pharmacologiques communes : anxiolytiques, hypnotiques, anti-convulsivantes, myorelaxantes et peuvent avoir un effet amnésiant. Par conséquent, elles ont potentiellement les mêmes indications et les mêmes effets indésirables. Des agonistes du récepteur GABA<sub>A</sub> -THIP, Benzodiazépine Midazolam et L-838,417- ont également été utilisés respectivement sur les modèles murins du syndrome de l'X fragile (souris Knock-Out (KO) *Fmr1*), de Rett (souris KO *MECP2*) et de l'autisme (agoniste spécifique des sous-unités  $\alpha 2$  et 3 : L-838,417, souris BTBR T+ *Itpr3<sup>tf</sup>* /J) et ont permis l'amélioration du phénotype [Olmos-Serrano et al., 2011, Voituron and Hilaire, 2011, Han et al., 2014]. Certaines de ces molécules font déjà l'objet d'essais cliniques. C'est le cas du traitement à la bumétanide des patients atteints d'un syndrome de l'X fragile ou de troubles du spectre autistique permettant une amélioration des scores autistiques et intellectuels [Lemonnier et al., 2012, Hadjikhani et al., 2013, He et al., 2014]. De plus, le ganaxolone, un neurostéroïde capable de stimuler les récepteurs GABA<sub>A</sub>, est actuellement testé en phase 2 d'un essai clinique regroupant des enfants atteints du syndrome de l'X fragile (NCT01725152, <http://clinicaltrials.gov/>) suite aux résultats

encourageants obtenus sur le modèle murin ([Heulens et al., 2012]).

Des agonistes spécifiques des sous-unités  $\alpha 2$  et 3 sont testés pour leur action anxiolitique/analgésique dans le traitement de la schizophrénie, faisant l'objet d'un autre essai clinique (phase 1) [Lewis et al., 2008]. Enfin, un essai clinique débutera bientôt avec le recrutement des patients avec troubles du spectre autistique, pour la phase 1 de l'agoniste spécifique sous-unités  $\alpha 2$  et 3, AZ7325 [Han et al., 2014].

Compte tenu des observations faites sur le modèle murin *Nono<sup>gt</sup>*, il serait intéressant de tester des molécules agonistes du GABA<sub>A</sub> chez les souris.

En conclusion, notre travail a permis d'identifier un nouveau syndrome cliniquement reconnaissable et lié à des mutations du gène *NONO*. Les travaux menés dans les cellules de patients et les souris *Nono<sup>gt</sup>* ont permis une avancée majeure dans la compréhension de la physiopathologie de cette maladie. Enfin, ils suggèrent que le syndrome MCCDI est une nouvelle maladie neurodéveloppementale liée au dysfonctionnement du système GABAérgique, ouvrant la voie d'un essai thérapeutique pour ces malades.

## **Quatrième partie**

# **Discussion et perspectives : Apport et limites du WES pour le démantèlement génétique de la DI**

# CHAPITRE 11

## BILAN DES TRAVAUX DE THÈSE

La DI, handicap le plus fréquent chez l'enfant et le jeune adulte, reste de cause inconnue dans environ 40% des cas. L'objectif de ce travail était de contribuer à l'identification de nouvelles causes génétiques de DI par l'analyse de 5 familles indépendantes et, ainsi, améliorer la compréhension de la physiopathologie de ces conditions. La table 11.1 résume les résultats obtenus par l'analyse de WES pour ces 5 familles : ces travaux nous ont d'abord mené à l'identification d'une nouvelle mutation dans un gène connu de DI, *WDR62*, chez une famille consanguine et multiplex, la famille Mo. Le WES nous a ensuite permis de définir un nouveau mode d'hérédité ; le syndrome des mutations contiguës, avec l'étude de la famille Ma dont le phénotype est lié à une mutation du gène *AP4M1* et une autre mutation du gène *AZGP1*, un excellent candidat permettant de caractériser une forme très précoce d'obésité morbide. Nous avons également identifié des variants dans deux gènes *RAD54B* et *HERC2* qui pourraient jouer un rôle modificateur du phénotype associé aux mutations des gènes majeurs *WDR62* (Mo) et *AP4M1/AZGP1* (Ma), respectivement. Enfin, l'analyse des trois dernières familles (Ei, *MCCID1* et *MCCID2*) a donné lieu à la caractérisation de deux nouveaux gènes de DI, *TTI2* et *NONO*.

famille	gène(s) majeur(s)			potentiels gènes modificateurs
	nouveau gène de DI	gène connu de DI	nouveau mode d'hérédité	
famille Mo	-	+	-	+
famille Ei	+	-	-	-
famille Ma	-	+	+	+
familles <i>MCCID</i>	+	-	-	+/-

**Table 11.1** : Bilan de l'analyse par WES des 5 familles présentées.

À l'issue de l'identification de ces deux nouveaux gènes de DI, l'analyse fonctionnelle des effets des mutations sur les cellules de patients et celle d'un modèle murin (*Nono<sup>gt</sup>*) a permis de démontrer la pathogénicité de ces mutations. D'autre part, ces études nous ont permis de mettre en lumière les mécanismes physiopathologiques responsables des phénotypes observés dans ces trois familles : le contrôle de l'intégrité du génome et la régulation transcriptionnelle, des processus physiologiques qui avaient déjà été associés à la DI auparavant.

---

Nos travaux apportent donc la première démonstration du rôle des gènes *TTI2* et *NONO* lors de la mise en place du SNC et soulignent encore une fois l'importance de ces voies au cours du développement.

# CHAPITRE 12

## RÉFLEXIONS AUTOUR DE L'OUTIL WES

### 12.1 Évolution et avantages de l'approche WES au cours des 4 dernières années

#### 12.1.1 Études familiales

Lorsque nous avons commencé l'analyse de familles par WES en 2011, nous avons d'abord opté pour une stratégie associant la cartographie par autozygotie et le WES : c'est le cas de la famille EI, qui nous avait alors servi d'étude pilote dans notre laboratoire et de la famille Mo, qui a ainsi marqué la transition de la cartographie vers le WES. Déjà, pour les projets suivants en 2012, le séquençage à haut-débit devenant plus accessible financièrement, nous changions de stratégie et préférons réaliser un WES pour une majorité de sujets par famille, nous permettant d'effectuer les études de ségrégation directement sur les données de séquençage ; c'est le cas de la famille Ma. Aujourd'hui, une avancée possible du WES serait d'intégrer des outils de génotypage aux interfaces d'analyse ; ainsi le travail de génotypage réalisé autour de la famille Ma et de celle portant la même mutation publiée par Verkerk et al. [2009], pourrait être pratiqué directement sur les données de WES.

#### 12.1.2 Études de cohorte de patients

De surcroît, nos résultats de l'analyse des familles MCCID, combinées aux données de la littérature, confirment l'efficacité du WES pour l'étude de cohorte de patients. Bien que la taille de la cohorte, quand elle est importante, soit un facteur bénéfique à la réussite de ce type d'études, nous avons montré qu'elle n'est pas un facteur limitant : en effet, la force de ces études réside principalement dans la constitution de la cohorte. Les critères de constitution de celle-ci sont donc cruciaux : grouper judicieusement des patients au sein d'un même projet de WES semble à l'origine du succès de ces études. On mesure alors l'importance de la qualité des descriptions cliniques et de l'interaction entre les différents acteurs du corps médical.

Pour conclure, lors de l'analyse des premiers projets de WES au sein de notre équipe en 2011, il était nécessaire de vérifier par séquençage Sanger la ségrégation de nombreux variants candidats au sein d'une



famille entière ou de tester le caractère rare des variants dans des cohorte de contrôles sains. Depuis 4 ans, ceci n'est plus nécessaire : le coût d'un WES a diminué de façon importante nous permettant de réaliser un WES pour davantage de sujets d'une même famille ou d'une cohorte de patients. De concert avec ce contexte financier et l'approche ayant rapidement fait ses preuves, le nombre d'études par WES publiées, ou simplement réalisées, a décuplé ; si bien qu'aujourd'hui, il est aisé de comparer les variants identifiés à de nombreuses bases de données -telles que *Exome Variant Server* (EVS), base de donnée interne- et d'en évaluer ainsi le caractère « nouveau ».

## 12.2 Limites du WES

Nous avons vu que le WES est un bon outil pour identifier les bases moléculaires des DI d'origine monogénique. Cependant, nous sommes confronté aux limites de cet outil lorsque l'analyse met en évidence de multiples variants candidats. Comment alors évaluer la part de responsabilité de chaque variant dans le phénotype ? Comme détaillé dans ce manuscrit, les données de la littérature et les études fonctionnelles sont nécessaires pour valider la pathogénicité d'un variant. Ainsi, le nombre de patients répertoriés avec des mutations dans le même gène ont permis de valider les variants identifiés dans les gènes *WDR62* et *AP4M1* pour les familles Mo et Ma, respectivement, tandis que les études fonctionnelles sur les cellules de patients et l'étude d'un modèle murin ont permis de valider ceux identifiés dans les gènes *TTI2* et *NONO* pour les familles El et MCCID, respectivement. Bien que la pathogénicité du gène majeur ait été validée, nous ne pouvons exclure l'effet des variants supplémentaires identifiés, dont la ségrégation dans la famille est compatible avec l'hypothèse d'un effet de ces variants sur le phénotype des patients. Nous avons proposé qu'il s'agisse de variants contenus dans des gènes modificateurs ; ils modulent l'expression clinique et/ou la sévérité de la maladie. Les gènes modificateurs sont d'un intérêt tout particulier puisqu'ils permettent en partie de proposer une explication à la variabilité phénotypique interfamiliale. Cependant, la démonstration de leur implication dans la pathologie n'est que partielle : pour valider leur rôle, l'observation de cas semblables ou des études fonctionnelles sont nécessaires.

À l'inverse du problème lié aux candidats multiples, la seconde difficulté à laquelle nous sommes parfois confronté avec le WES, correspond à l'absence totale de variants candidats à l'issue de l'analyse des données. Cette situation peut parfois être la conséquence d'une complication liée à la technique, comme le design de la capture, la qualité de la couverture ou les erreurs d'alignement. Toutefois, ce cas de figure peut également refléter les limites du WES devant des pathologies dont la transmission n'est pas liée à un variant codant : variants localisés dans les régions UTR, introniques ou intergéniques, mode de transmission non-mendélien, pathologie polygénique ou multifactorielle. Dans de telles situations, le WES ne semble pas suffisant pour déchiffrer les mécanismes moléculaires sous-jacents à ces pathologies.

# CHAPITRE 13

## PERSPECTIVES POUR L'IDENTIFICATION DES GÈNES DE DI ; L'APPROCHE DE DEMAIN

La révolution NGS s'est accompagnée d'une véritable accélération de la description des diverses causes génétiques de DI. Cette avancée est majeure : la génétique des DI est aujourd'hui un domaine en perpétuelle évolution et chaque mois de nouveaux gènes sont caractérisés. Pourtant, les nouveaux gènes identifiés ces dernières années ne rendent compte que d'un faible nombre de patients et de trop nombreuses familles restent encore sans explications.

### 13.1 Amélioration de la démarche diagnostique

#### 13.1.1 Mise en place d'un réseau d'échange d'informations : CARID

Afin de pallier les difficultés de diagnostic liées à l'extrême hétérogénéité des DI, un consortium international composé de 7 groupes, appelé *Consortium for Autosomal Recessive Intellectual Disability* (CARID), a été construit par le Pr A. Reis (Erlangen, Allemagne). À l'instar du consortium « EuroMRX » conçu dans les années 90, ce projet a pour objectif la création d'un cadre favorisant l'échange d'informations cliniques et génétiques, comme les données de WES, concernant les familles étudiées dans les différentes équipes. Ces échanges apporteront certainement de nouveaux éléments de réponses pour toutes les familles pour lesquelles les variants identifiés par l'analyse de WES sont considérés comme de bons candidats sans pour autant être validés.

#### 13.1.2 Vers un test diagnostique unique de la DI

Actuellement, afin d'optimiser la démarche diagnostique des patients et de leur famille, plusieurs projets de recherche translationnelle entreprennent la mise au point d'un test diagnostique de la DI utilisant les outils de séquençage haut-débit. Le service de génétique médicale de l'hôpital Necker-Enfants Malades en étroite collaboration avec les plateformes de génomique et de bio-informatique ont récemment terminé la conception d'un kit diagnostique haut débit de la DI, dont la preuve de concept est en cours. Ce kit de capture permet de limiter le séquençage uniquement aux régions d'intérêt : les séquences codantes de 253 gènes explicitement

impliqués dans la DI (81 gènes de DIAD, 77 gènes de DIAR et 95 gènes de DILX). Parallèlement, 5 autres groupes travaillaient sur un projet similaire et chacun a sélectionné une liste quelque peu différente, incluant de 217 à 481 gènes. Très récemment, un de ces groupes a publié les résultats d'une étude pilote, l'analyse de 217 gènes ciblés pour 106 patients avec un succès du diagnostic pour 25% des patients, validant ainsi cette approche [Redin et al., 2014]. Pour les années à venir, l'objectif de l'ensemble des services de génétique médicale en France est la création d'un kit commun (composé de la même liste de gènes) et évolutif, utilisé en première intention pour le diagnostic de la DI à l'échelle nationale.

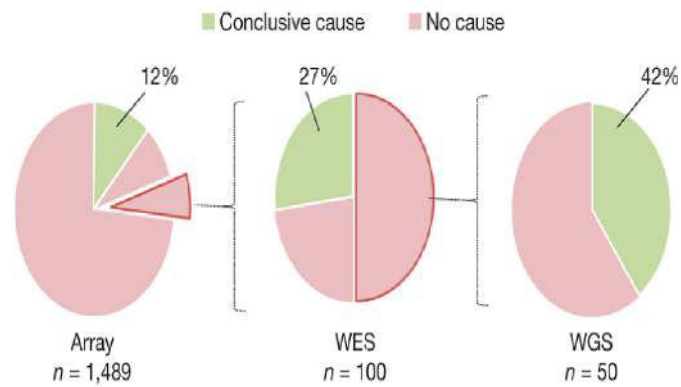
Comme avancé par les résultats de Redin et al. [2014], ce projet promet également de nouvelles améliorations dans la recherche de corrélations génotype/phénotype cliniquement pertinentes pour définir de nouveaux syndromes et aider à la reconnaissance de ces nouvelles entités. Il permettra aussi d'étendre le spectre phénotypique des mutations d'un gène connu. Enfin, pour les familles dont cette approche ne permettrait pas de diagnostic moléculaire, une approche WES sera engagée.

## 13.2 Exploration du génome entier : premiers pas

Nous avons vu qu'une des limites de l'outil de WES est l'absence de candidats satisfaisants. Devant une telle situation, et connaissant l'extrême hétérogénéité génétique de la DI, la communauté scientifique suggérait, et ce à juste titre, que la concentration de nos recherches uniquement dans les régions codantes du génome devenait à présent une faiblesse plutôt qu'une force et aspirait à étudier les autres régions, moins connues, de notre génome afin d'identifier de nouvelles causes génétiques de DI. Plusieurs groupes ont tenté d'utiliser la méthode de séquençage du génome entier ou *Whole Genome Sequencing* (WGS) afin de détecter des variants impliqués dans le spectre des troubles autistiques [Jiang et al., 2013, Michaelson et al., 2012]. Très récemment, Gilissen et al. [2014] ont publié le premier travail d'identification de causes génétiques de DI par WGS. Cette approche a été réalisée pour 50 trios demeurant sans diagnostic moléculaire à l'issue d'une approche gène candidat, d'une recherche de CNVs et de l'analyse des données de WES. La recherche de SNV (Single Nucleotide Variant) lors de l'analyse des données de WGS a permis d'identifier 132 variants parmi lesquels 84 sont situés dans les régions codantes. Cette observation reflète la possibilité que l'absence de diagnostic moléculaire à l'issue du WES soit parfois due à un problème technique et pas seulement au besoin d'explorer les régions non codantes.

Les auteurs identifient également 8 CNV *de novo* incluant des délétions d'un exon unique, de régions intra-exoniques et des duplications interchromosomiques.

Selon l'hypothèse d'un rôle des mutations non codantes dans l'étiologie génétique de la DI, les auteurs sélectionnent ensuite des variants *de novo* non codants, situés dans les régions promotrices, introniques ou les séquences UTR des gènes de DI connus : 43 variants correspondent à ces critères. Les diverses sources ENCODE disponibles [Consortium, 2012] apportant des éléments de discussion sur les régions non codantes identifiées n'ont pas donné d'informations en faveur d'une éventuelle pathogénicité de ces variants, ce qui confirme la difficulté d'interprétation de ces variants. La figure 13.1 résume les fréquences de diagnostic positif obtenues après chacune des méthodologies, mettant en avant le GWS avec le meilleur taux de réussite (42%).



**Figure 13.1** : Bilan des fréquences de diagnostic positif obtenues selon l'approche utilisée, d'après Gilissen et al. [2014].

Une autre étude par une approche de WGS a suivi celle-ci menant à l'identification d'un variant dans une région codante du gène *NOTCH1* [Stittrich et al., 2014]. La méthode WGS commence donc à émerger mais les chercheurs continuent d'explorer principalement les régions codantes représentant 1 à 2% des données générées par cette approche, dont le coût est 3 à 4 fois plus élevé que le WES. La table 13.1 résume les caractéristiques associées à chacune des deux approches :

Approche	WES	WGS
Coût	1	3 ou 4
Couverture du génome	1 à 2%	100%
Profondeur	>100X	30-80X
Quantité d'informations générée	5Go	100Go
Détection des CNVs	-	+

**Table 13.1** : Comparaison des outils de WES vs WGS.

La force de cette nouvelle approche prometteuse réside donc majoritairement dans les multiples possibilités d'explorations qu'elle offre avec une seule et même opération ; la recherche -aujourd'hui devenue courante- de variants dans les régions codantes et la recherche de variants dans les séquences non codantes. Mais, surtout, la méthode de WGS permet l'identification de CNVs de grandes ou petites tailles, ce qui la différencie de l'approche WES.

En conséquence de l'explosion des approches WES, les logiciels de prédiction des effets de variants codants se sont améliorés et les bases de données disponibles se sont enrichies, offrant ainsi des pistes d'interprétation des variants détectés par WES. Au contraire, en dehors des régions codantes, l'interprétation des variants reste complètement inexplorée et tout reste à mettre en place. Aujourd'hui, il semble plus intéressant de continuer à utiliser le WES dans un premier temps puis, pour les familles pour lesquelles aucun variant candidat n'est retenu à l'issue de celui-ci, tenter de réaliser un WGS.

# CHAPITRE 14

## VALIDATION FONCTIONNELLE DES VARIANTS : DISSERTATION AUTOUR DU CHOIX DES MODÈLES

Avec la révolution NGS, le challenge pour les généticiens a basculé : il résidait dans l'identification de variants et repose à présent dans leur interprétation. Aujourd'hui, la validation fonctionnelle des candidats identifiés par WES est devenue essentielle et celle-ci s'annonce indispensable avec l'apparition des études menées par WGS.

Pour les 5 familles présentées dans ce manuscrit, nous avons eu la chance d'obtenir une biopsie de peau pour au moins un enfant atteint de chaque famille. Nous avons donc réalisé l'ensemble de nos études fonctionnelles sur les fibroblastes primaires offrant une véritable force à nos résultats puisqu'ils nous permettent de valider les effets d'un variant directement dans les cellules des patients. Cependant, la pertinence du modèle n'est pas entière puisque le type cellulaire disponible n'est pas représentatif du tissu atteint dans les pathologies neurodéveloppementales. L'étude de la physiopathologie de ces conditions est alors rapidement contrariée en raison du modèle disponible.

Afin de s'affranchir de cette limite, d'autres modèles peuvent fournir de nouvelles perspectives d'études. Par exemple, des travaux fonctionnels dans le domaine des neurosciences ont souvent été conduits à l'aide de lignées neuronales telles que la lignée PC12 provenant du rat ou la lignée humaine SHSY5Y [Contrepas et al., 2009, Park et al., 2014, Modi et al., 2014]. Des tests peuvent également être menés sur les cultures de neurones primaires provenant de la dissection de cerveaux d'embryons de souris [Silva et al., 2006, Hirata et al., 2013]. Les cultures neuronales sont particulièrement utiles puisqu'elles permettent d'étudier les effets de variants dans un type cellulaire approprié. Pourtant, elles présentent aussi des inconvénients : en effet, le caractère transitoire des manipulations génétiques qui accompagne inévitablement l'utilisation de ce modèle peut parfois devenir limitant, par exemple lorsqu'un processus biologique a besoin d'être évalué dans le temps.

Depuis 2006, les progrès de la connaissance des cellules souches ont ouvert de nouvelles perspectives pour la compréhension de la physiopathologie de nombreuses pathologies humaines. Il est désormais possible de générer des cellules pluripotentes humaines par dédifférenciation de cellules somatiques (telles que les fibroblastes primaires de patients), appelées cellules souches pluripotentes induites ou iPSCs (induced Pluripotent Stem Cells) [Takahashi and Yamanaka, 2006, Takahashi et al., 2007, Yu et al., 2007]. Les iPSCs obtenues peuvent ensuite être re-différenciées en divers types cellulaires dont des précurseurs de neurones ou

---

NPCs (Neural Progenitors Cells), offrant de nouvelles perspectives pour l'étude des maladies neurologiques et aujourd'hui largement utilisées [Chamberlain et al., 2010, Akizu et al., 2013, Martins-Taylor et al., 2014].

Plus récemment, le groupe de Lancaster et al. [2013] est parvenu à établir une méthode de culture, à partir de cellules souches pluripotentes (embryonnaires -embryonic stem (ES)- ou iPSCs), aboutissant à la génération de structures proches du cerveau appelées « cerebral organoids », récapitulant ainsi le développement du cerveau humain. À l'aide de ce nouveau modèle quelque peu révolutionnaire, les auteurs ont pu démontrer un défaut dans le processus de différenciation neuronale dans les « cerebral organoids » issus d'un patient avec microcéphalie, apportant de solides éléments pour expliquer l'apparition du phénotype.

Par ailleurs, les modèles animaux fournissent d'excellentes conditions d'études de la physiopathologie des DI. Dans le cas des mutations du gène *NONO*, le modèle murin du groupe du Pr S. Brown (*Nono<sup>gt</sup>*) a permis une avancée majeure de la compréhension du mécanisme physiopathologique responsable du phénotype des patients MCCID. Toutefois, la génération d'un tel modèle est longue et fastidieuse : c'est pourquoi d'autres modèles sont parfois préférés aux modèles murins, comme la drosophile [van Bon et al., 2013] ou l'ovocyte de Xénope (*Xenopus laevis*) [Barcia et al., 2012]. Le poisson-zèbre (*Danio rerio* ou zebrafish) est également souvent manipulé pour modéliser les maladies du SNC [Tarabeux et al., 2010, Gauthier et al., 2010, Boczonadi et al., 2014, Schaffer et al., 2014]. En effet, la conservation importante de l'organisation du cerveau et du génome par rapport à l'homme, le développement embryonnaire externe, synchronisé et rapide et la transparence de l'embryon sont autant d'atouts facilitant l'utilisation de ce modèle.

Enfin, un nouvel outil de modification génétique a émergé depuis 2009. Cette stratégie s'appuie sur les propriétés d'une enzyme hybride capable de reconnaître des séquences particulières de l'ADN et de couper à cet endroit précisément, mais aussi sur les processus physiologiques de réparation de l'ADN. Il s'agit de la jonction des domaines protéiques d'une nucléase aspécifique avec ceux d'une protéine capable de reconnaître et de lier des séquences spécifiques de l'ADN : ZFNs (Zinc Finger Nucleases), TALENs (Transcription Activator-Like Effector NucleaseS), meganucleases (engineered homing endonucleases) et très récemment, la nucléase basée sur le système CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat)-Cas (CRISPR associated) [Segal and Meckler, 2013, Joung and Sander, 2013, Terns and Terns, 2014]. Ces constructions sont facilement façonnables, permettant de modifier génétiquement n'importe quelle position du génome d'une importante variété de types cellulaires et d'organismes, tout en évitant les altérations génétiques non désirées, inconvénients des précédentes approches de modification génétique.

Cette récente technologie de manipulation du génome ouvre de nouvelles perspectives pour l'étude des conditions neurologiques. Elle offre en effet la possibilité de reproduire exactement la même mutation identifiée chez un patient, dans des cultures de cellules ou sur un modèle animal ; une alternative séduisante, particulièrement lorsque l'obtention de cellules de patients est impossible.

La précision de ces outils et le faible taux de mutations non désirées associé [Suzuki et al., 2014] permet de s'affranchir d'un biais important : le fond génétique. En effet, ils offrent la possibilité de travailler sur des lignées isogéniques, un modèle idéal quand on essaie de montrer qu'une différence phénotypique est due uniquement à une mutation ponctuelle, et notamment pour entreprendre des expériences de sauvetage phénotypique.

L'alliance des techniques de reprogrammation du destin cellulaire et de manipulation du génome offre donc des pistes inédites pour l'étude des conséquences de mutations dans des types cellulaires appropriés et promet de futures grandes avancées dans la compréhension des mécanismes physiopathologiques des conditions neu-

---

rodéveloppementales. Enfin, ces outils, permettant de s'affranchir des effets indésirables associés au transfert d'un transgène, annoncent l'apparition de nouvelles stratégies prometteuses dans le domaine de la thérapie génique [Perez et al., 2008, Segal and Meckler, 2013, Li et al., 2014a].

# **Matériels et Méthodes**



# CHAPITRE 15

## CAPTURE ET SÉQUENÇAGE DE L'EXOME

Depuis 2008, une plateforme de Génomique a été créée sur le site de l'hôpital Necker-Enfants Malades, au sein de l'institut *Imagine*. Elle permet aux équipes du site d'avoir accès à des technologies à un coût réduit. Le séquençage de l'exome (WES) a été réalisé pour 5 familles distinctes, appartenant à 4 projets indépendants et présentés dans ce manuscrit. Pour l'ensemble de ces projets, la capture des séquences codantes s'est effectuée en phase liquide, par la technique Agilent SureSelect, (kit Agilent Human All Exon 38 ou 50 Mb) permettant d'obtenir une banque contenant tous les exons du génome à partir de l'ADN (*Acide Desoxyribonucléique*) génomique. Brièvement, l'ADN génomique est fragmenté : des cassures mécaniques sont provoquées par nébulisation (correspondant au passage de la solution à travers un orifice très étroit sous la pression de gaz) ou bien par l'utilisation d'un ultrasonicateur Covaris diffusant des ondes acoustiques. Après réparation des extrémités des ADN fragmentés, des séquences adaptatrices universelles et spécifiques de la technologie NGS utilisée sont liguées aux extrémités des fragments, suivi ou non d'une amplification PCR (*Polymerase Chain Reaction*). Par ailleurs une séquence spécifique (code-barre ou index) est également rajoutée à chaque banque pour l'identification de chaque échantillon au sein d'un pool. À cette étape, on obtient donc une banque d'ADN « génome complet » ou « whole genome ». Cette banque de fragment est alors dénaturée puis hybridée à un panel d'oligonucléotides ARN biotinylés ciblant les séquences à capturer (Agilent SureSelect<sup>TM</sup>). Les hybrides biotinylés formés sont ensuite capturés à l'aide de billes magnétiques couvertes d'avidine (figure A.6), puis amplifiés et séquencés. Le kit Human All Exon utilisé a évolué au fil du temps, ciblant plus de 180 000 exons et couvrant de 38 à 50 Mb du génome humain selon les versions.

Les informations concernant les runs de séquençage de chacun des projets de WES exposés dans ce manuscrit sont résumées dans le tableau 15.1 ci-dessous :

- les sujets séquencés dans chaque famille,
- le type de capture réalisée,
- le type de séquençage pratiqué et l'appareil utilisé,
- la plateforme utilisée,
- la profondeur de lecture ou couverture moyenne, correspondant au nombre de fois qu'une position donnée du génome a été séquencée (minimum et maximum observés sur l'ensemble des individus d'un même projet),
- l'efficacité de la capture estimée (donnée en % de séquences capturées avec un taux de couverture supérieur

à 15X).

Famille(s)	Gène(s)	Sujets	Capture	Type de Séquençage (Appareil)	Plateforme	Profondeur de lecture moyenne (X)	Taux de couverture ou homogénéité (%)
Mo	<i>WDR62</i> , <i>RAD54B</i>	1 EA	nd	Solexa	Genoscope - Centre National de Séquençage (CNS)	nd	nd
EI	<i>TTI2</i>	1 EA	kit Agilent SureSelect Human All Exon 38 Mb	Genome Analyzer Iix (Illumina)	Genoscope - Centre National de Séquençage (CNS)	70X	plus de 83% des pb couvertes à 15X
Ma	<i>AP4M1</i> , <i>AZGP1</i> , <i>HERC2</i>	Père, Mère, 2 EAs	kit Agilent SureSelect Human All Exon 50 Mb	SOLiD5500 (Life Technologies)	<i>Imagine</i>	39-77X	plus de 74-85% des pb couvertes à 15X
MCCID1 et MCCID2	<i>NONO</i>	2 trios	kit Agilent SureSelect Human All Exon 50 Mb	SOLiD5500 (Life Technologies)	<i>Imagine</i>	42-87X	plus de 75-78% des pb couvertes à 15X

**Table 15.1 :** Les différents projets de WES réalisés.

ES : Enfant Sain, EA : Enfant Atteint, nd : non disponible.

La technique de séquençage utilisée pour 2 des 5 familles était celle d'Illumina, anciennement technologie Solexa, sur le séquenceur Genome Analyzer Iix. Les différentes étapes de cette technologie sont illustrées dans la figure A.3. Elle inclut d'abord le processus de *clustering*, qui consiste en l'amplification clonale de l'échantillon à analyser appelé la banque d'exome ici (ou *library*), produisant les *clusters*. Celle-ci ne s'effectue pas en solution mais directement sur support solide de séquençage appelé *flow cell* : elle comporte 8 canaux, chaque canal servant d'enceinte de réaction pour une banque ou plusieurs, si elles portent des code-barres permettant de les différencier (figure A.4). La réaction de séquençage est alors réalisée directement sur le support où l'ADN a été amplifié (en phase solide). Elle se déroule base après base en ajoutant un mélange contenant toutes les bases, chacune étant associée à un fluorophore différent. L'extrémité de ces bases est protégée pour empêcher l'addition de bases supplémentaires à chaque cycle d'incorporation (terminateurs fluorescents réversibles). Une lecture laser permet alors de détecter simultanément toutes les positions incorporées. Le clivage des fluorophores permet ensuite l'incorporation de la base suivante (figure A.5). La lecture est effectuée ainsi cycle après cycle sur chaque *cluster* de la *flow cell*. Les clusters étant identifiés par leur position sur la *flow cell*, cette méthode permet l'acquisition en parallèle de plus de 3 milliards de séquences. Chaque position étant lue l'une après l'autre, les erreurs principales de cette technologie sont des erreurs de substitution d'une base par une autre.

Une autre technique de séquençage a été utilisée pour 3 des 5 familles : le séquençage par ligation

---

avec le SOLID 5500 (figure A.2). La première étape d'amplification clonale de la méthode de séquençage d'Applied Biosystems est identique à celle du pyroséquençage : la banque d'ADN portant des adaptateurs universels spécifiques subit une étape d'amplification clonale par PCR en émulsion, permettant la production de billes couvertes de clones d'ADN dans des millions de réactions indépendantes dans un seul tube. Ces billes couvertes de clones d'ADN sont fixées de façon covalente sur le support de séquençage appelé *flow chip*. La réaction de séquençage s'effectue ensuite par un système assez complexe de séries de cycles de ligation, de détection de fluorescence et de clivage. Cette technique va permettre non seulement la lecture de la séquence mais inclut un système de correction des erreurs d'incorporations. Grâce à cette méthode, il est possible de lire jusqu'à un milliard et demi de séquences en parallèle. Le système incorporé de correction d'erreurs associées à l'utilisation de la ligase réduit le taux d'erreur de séquençage.

# CHAPITRE 16

## GÉNOTYPAGE

Le génotypage ou « tour de génome » a été réalisé pour 2 familles distinctes présentées dans ce manuscrit. Les informations sont résumées dans la table 16.1 ci-dessous pour chacune d'entre elles :

- le nombre d'individus génotypés par famille,
- le type de puce utilisée,
- la plateforme utilisée.

Famille(s)	Projet	Gène(s)	Individus génotypés	Type de puce utilisée	Plateforme
Mo	EXT2007	<i>RAD54B</i> , <i>WDR62</i>	Père, Mère, 5 EAs	Affymetrix GeneChip® Human Mapping 10K 2.0	Extérieur
EI	IMG2010 _00070	<i>TTI2</i>	Mère, 1 ES, 3 EAs	Affymetrix GeneChip® Human Mapping 250K	<i>Imagine</i>

**Table 16.1** : Les différents projets de cartographie par autozygotie réalisés.

ES : Enfant Sain, EA : Enfant Atteint.

Outre les problèmes techniques, la principale difficulté posée par ces approches de séquençage à haut débit est sans aucun doute de répondre au défi informatique que pose le stockage et, surtout, l'analyse de la quantité massive de données générées par ces technologies. C'est la raison pour laquelle nous travaillons en étroite collaboration avec la plateforme de Bioinformatique de l'institut *Imagine*. Les séquences obtenues sont ensuite analysées à l'aide de logiciels implémentant des algorithmes d'assemblage et d'alignement à une séquence de référence tels que BWA (Burrow-Wheeler Alignment Tool) et SAMtool (Sequence Alignment/Map). Tous les « calls » avec une couverture inférieure ou égale à 3 et un score de qualité inférieur à 20 ont été éliminés. L'ensemble des variants (SNPs et insertion/délétion) identifiés est ensuite annoté grâce au « pipeline » d'analyse puis, une interface graphique, développée par la plateforme de bio-informatique s'appuyant sur la base de données Ensembl (release 61) et permettant de présenter les données issues du séquençage de façon simple et conviviale. Il est ainsi possible de :

- visualiser l'ensemble des variants (SNPs ou insertion/délétion) identifiés dans un ADN test par rapport à la séquence de référence du génome humain,
- filtrer ces variants pour éliminer ceux déjà décrits chez des sujets sains (en utilisant les annotations des databases *dbSNP*, *1000 Genomes Project* et une base de données interne composée de l'ensemble des projets exomes réalisés sur la plateforme),
- prédire les conséquences éventuelles de ces variants sur la fonction de la protéine (en utilisant des logiciels de type Polyphen),
- comparer des échantillons entre eux pour identifier des variants communs ou des variants différents mais concernant les mêmes gènes.

L'interface graphique permet également de présenter les données issues du génotypage : l'analyse par le choix du seuil de taille et de lod score et l'obtention des résultats de régions génétiques candidates ont donc été facilitées.

# CHAPITRE 18

## FRACTIONNEMENT CELLULAIRE

Le texte ci-dessous explique la méthode utilisée pour réaliser l'analyse présentée dans la figure 9.1. Primary fibroblast cells were grown two 75 cm<sup>2</sup> dishes. After reaching 80-90% confluence, cells were harvested using trypsin (lifetechnologies) for 5 minutes at 37°C. Cells were centrifuged at 1500 rpm for 5 min and rinsed with ice-cold PBS. Cell pellets were resuspended by gentle pipetting in 400 µL lysis buffer A [10 mM Tris (pH8.0), 140 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.1% Igepal, 2 mM RNase OUT (lifetechnologies)], and incubated on ice for 10 min. 80 µL of the total RNAs were extracted by Trizol reagent (lifetechnologies). The rest of the lysate was centrifuged at 1,000 g for 3 min at 4°C to isolate the cytoplasmic fractionation and pellet the nuclei. Cytoplasmic fractions were further centrifuged at maximum speed and nuclear pellets underwent two additional washes with 320 µL lysis buffer A before extraction with Trizol. After determining the RNA concentration of each sample with the Nanodrop, 500ng of RNA was used to run semi quantitative and quantitative reverse-transcription (RT) PCR. Primers used for mRNA nuclear and cytoplasmic distribution analysis are located in the downstream of IRAlus (Inverted Repeat *Alu* elements) in the 3'-UTRs to detect only full-length mRNAs, as described in Chen and Carmichael [2009], Chen et al. [2008] (figure 18.1).



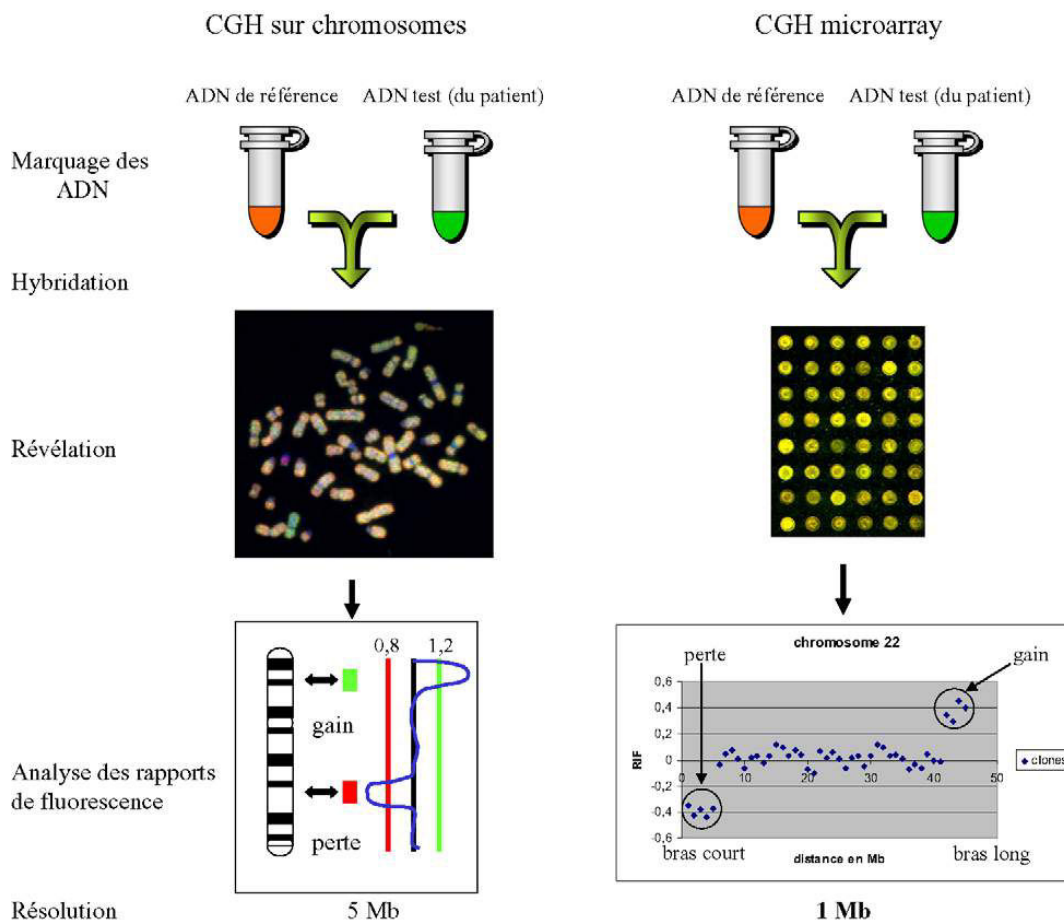
**Figure 18.1 :** Position des amorces permettant de sélectionner les ARN édités, modifié d'après Chen and Carmichael [2009].

Target and housekeeping genes Ct values from Cytoplasmic and Nuclear samples were first normalized to the level of Total sample for each individual. Then these normalized values were treated with the classical comparative Ct method. In controls and patients cells, the majority of full-length *LIN28*, *NUP43*, *NICN1*, *PAICS* and *PCCB* mRNAs are retained in the nucleus as *SMC1A*, *FLNB*, *HYOU1*, and *SON*, mRNAs that do not contain IRAlus in their 3'-UTRs, but are normally enriched in the nucleus. T, total RNA ; C, cytoplasmic RNA ; N, nuclear RNA.

# Annexes

# ANNEXE A

## FIGURES



**Figure A.1 :** Principe de la CGH sur métaphase et sur puce à ADN (CGH array), d'après Sanlaville et al. [2005].

La CGH ou *Comparative Genomic Hybridization* permet une analyse globale du génome avec une résolution supérieure à celle du caryotype. La partie gauche de la figure indique le principe de la CGH sur chromosome : la même quantité d'ADN provenant d'un patient et d'un témoin, chacune marquée par un fluorochrome différent, est co-hybridée sur les chromosomes d'un individu normal. Les signaux générés par les deux fluorochromes sont numérisés et un rapport



de leur intensité respective est établi au niveau des bandes de chacune des paires de chromosomes : il reflète le rapport de la quantité d'ADN du patient par rapport à celle du témoin. La partie de gauche de la figure représente le principe de la CGH sur puces à ADN : des séquences d'ADN, dont on connaît la position sur le génome, sont déposées sur des lames de verres (ou microarray). Le rapport de fluorescence est ici calculé au niveau de chaque dépôt correspondant au fragment d'ADN fixé. Des logiciels permettent un traitement statistique des données et une représentation graphique de celles-ci.

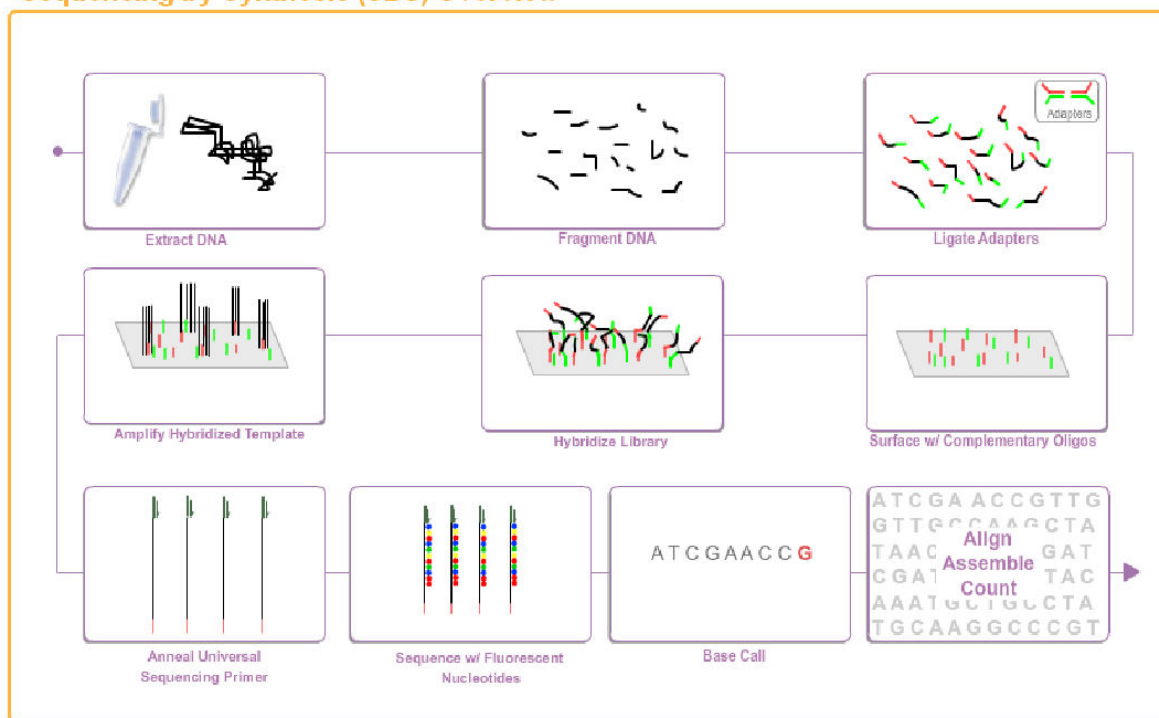
	454 FLX/Roche	Solexa/Illumina	Solid/Life Tech
<b>1-Préparation de banques de NGS</b>	<p>ADN génomique → Fragmentation → Ligation d'adaptateurs</p> <p>But ajout d'adaptateurs : Utiliser toujours la même amorce pour PCR et Séquençage quelque soit l'ADN à séquençer</p>		
<b>2-Amplification clonale</b>  Système miniature d'amplification	<b>PCR en émulsion sur une bille</b>   Amplification clonale dans des microréacteurs	<b>« Bridge » PCR sur un support plan (=Flow Cell)</b>   Colonies d'ADN = polonies	<b>PCR en émulsion sur une bille</b>   Amplification clonale dans des microréacteurs
<b>3-Support pour le séquençage</b>	<b>Billes insérées dans des micropuits</b>  	<b>Flow Cell ayant servi à l'amplification</b>  	<b>Billes fixées à une lame de verre</b>  
<b>4-Séquençage en temps réel :</b> 1 bille ou 1 cluster => 1 Read	<ul style="list-style-type: none"> <li>* Séquençage par synthèse</li> <li>* Pyroséquençage</li> <li>* <b>Longs fragments</b></li> </ul>	<ul style="list-style-type: none"> <li>* Séquençage par synthèse</li> <li>* dye-terminator réversible</li> <li>* <b>petits fragments</b></li> </ul>	<ul style="list-style-type: none"> <li>* Séquençage réversible par ligation</li> <li>* encodage de 2 bases</li> <li>* <b>petits fragments</b></li> </ul>

**Figure A.2 :** Comparaison des principes des 3 principales technologies de NGS.

Pour chaque méthode de séquençage la première étape correspond à la fabrication d'une banque d'ADN génomique : l'ADN à séquençer est d'abord fragmenté. L'ajout d'adaptateurs universels au moment de la fabrication des banques permet d'amplifier et de séquençer des milliers de molécules en parallèle en utilisant un seul type d'amorces. Avec la méthode de pyroséquençage (Roche), après fragmentation, l'ADN subit une étape d'amplification par PCR. Celle-ci est réalisée en émulsion, permettant d'effectuer plusieurs millions de réactions indépendantes dans un seul tube. Le mélange obtenu est ensuite déposé sur une plaque dont chaque emplacement ne peut contenir qu'une seule molécule d'ADN amplifiée. La réaction de séquençage par synthèse est alors initiée base par base. La lecture de chaque base incorporée est révélée à l'aide d'une réaction chemoluminescente et détectée par une caméra CCD (*Charge Coupled Device*, assurant la conversion d'un signal lumineux en un signal électrique). Les erreurs de séquences détectées sont majoritairement des insertions/délétions dues aux répétitions identiques de la même base. Les deux autres méthodes sont détaillées dans le chapitre 15 de ce manuscrit.

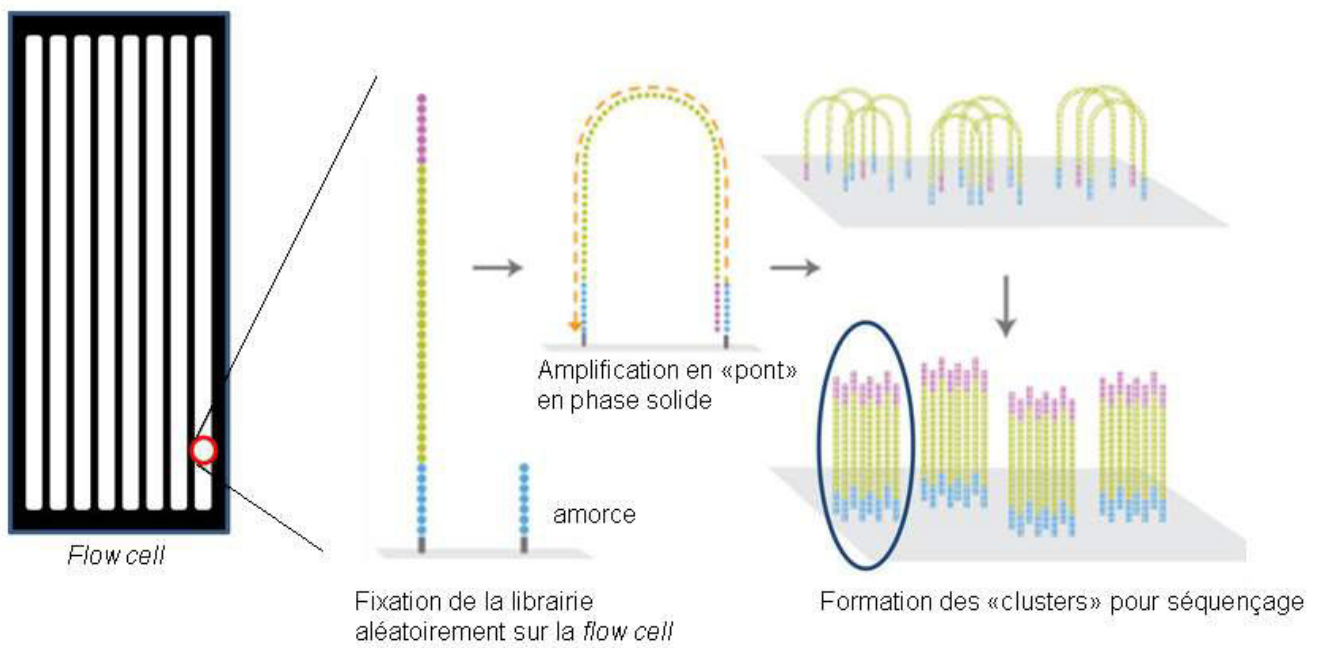
## Illumina Sequencing Technology

### Sequencing by Synthesis (SBS) Overview



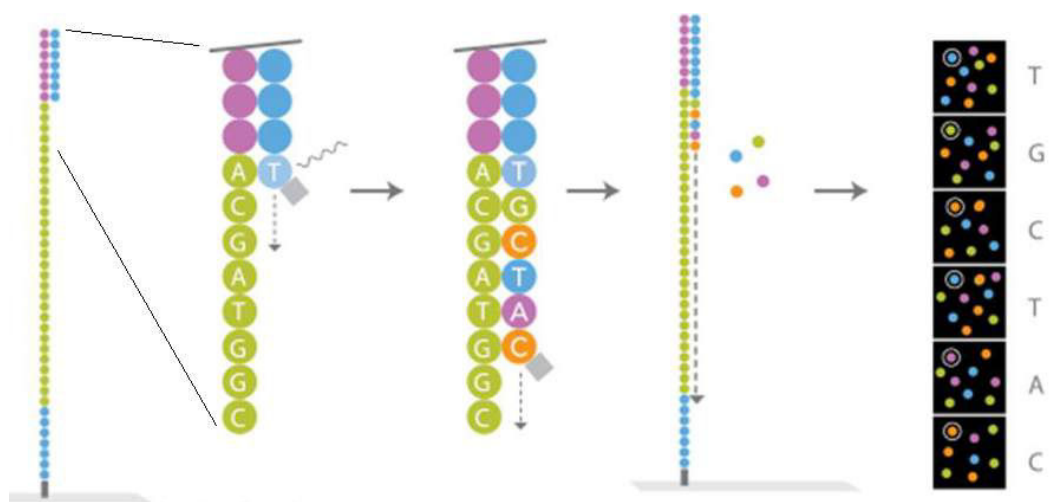
**Figure A.3** : Ensemble des étapes de la technologie Illumina (source : Illumina).

Une description de la technologie Illumina est disponible dans le chapitre 15 de ce manuscrit.



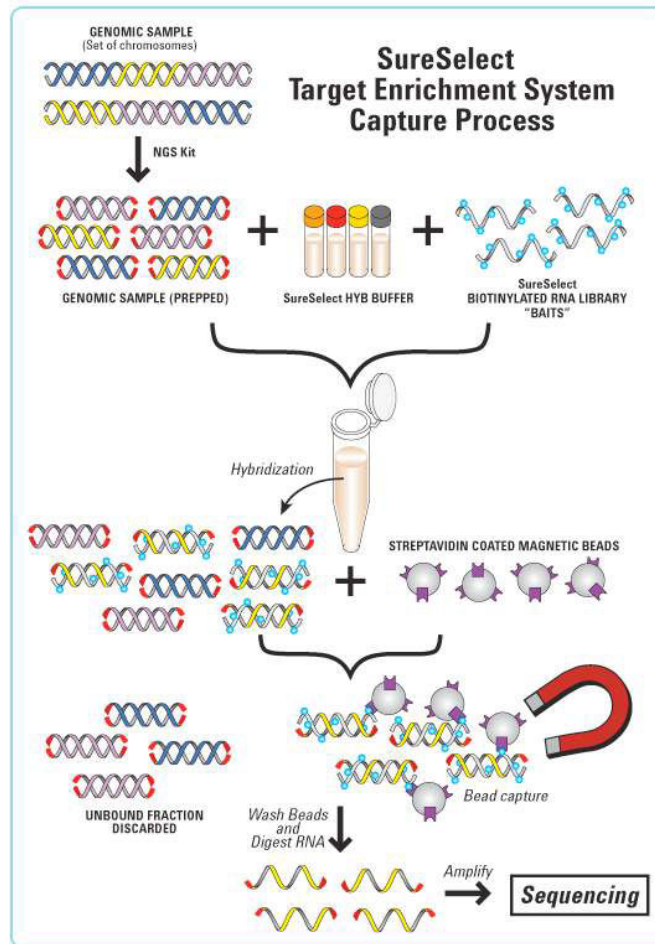
**Figure A.4 :** Étape de fixation et d'amplification de la librairie sur la flow cell (source : Illumina).

Une description de la technologie Illumina est disponible dans le chapitre 15 de ce manuscrit.



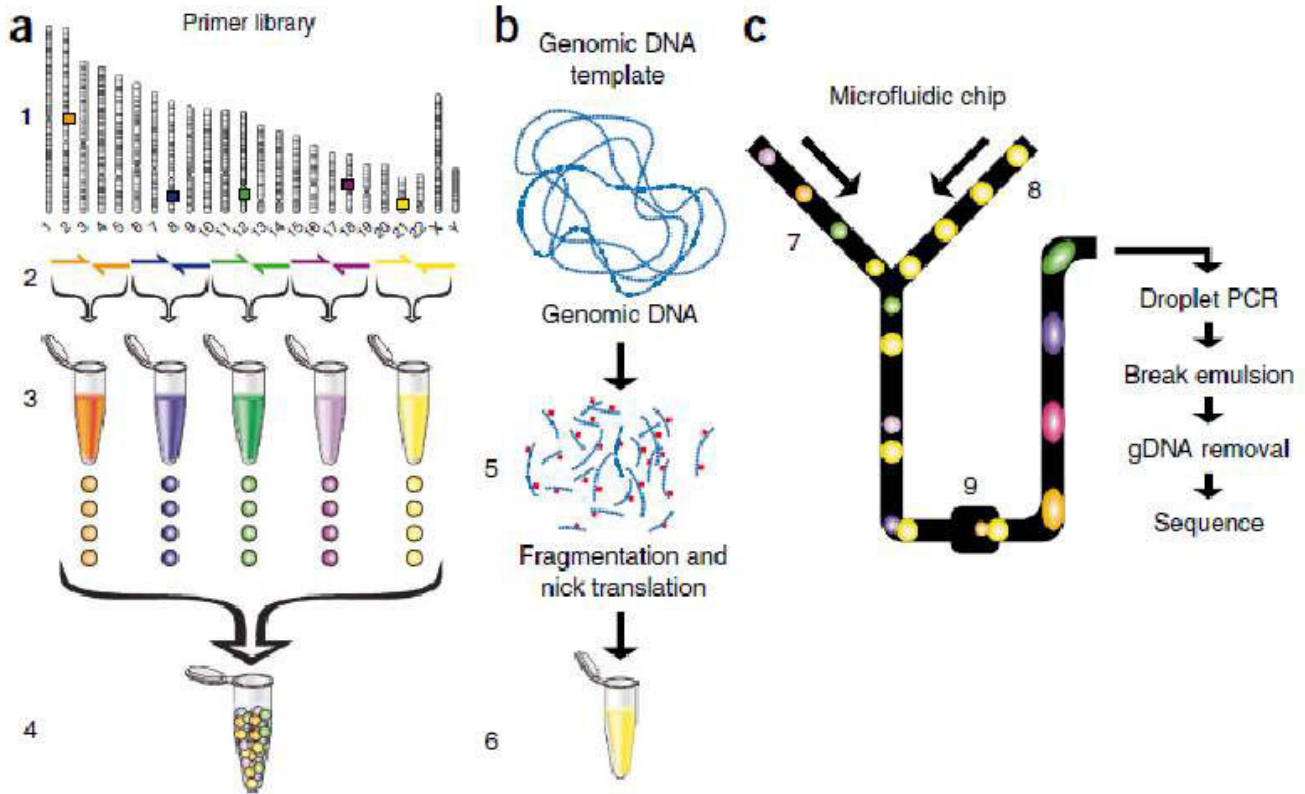
**Figure A.5 :** Réaction de séquençage par la technique Illumina (source : Illumina).

Une description de la technologie Illumina est disponible dans le chapitre 15 de ce manuscrit.



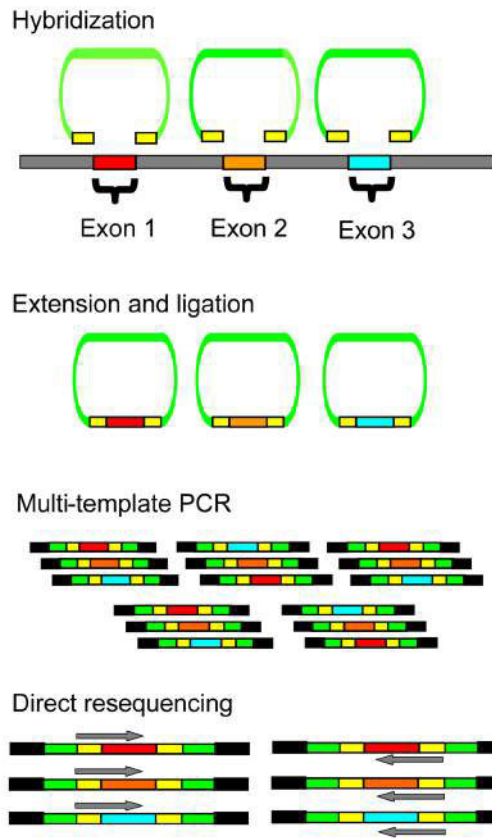
**Figure A.6** : Technique d'enrichissement Agilent SureSelect, modifié de la source : Protocole Agilent SureSelect Target Enrichment System.

Une description de la technique d'enrichissement de l'exome par la méthode Agilent SureSelect est disponible dans le chapitre 15 de ce manuscrit.



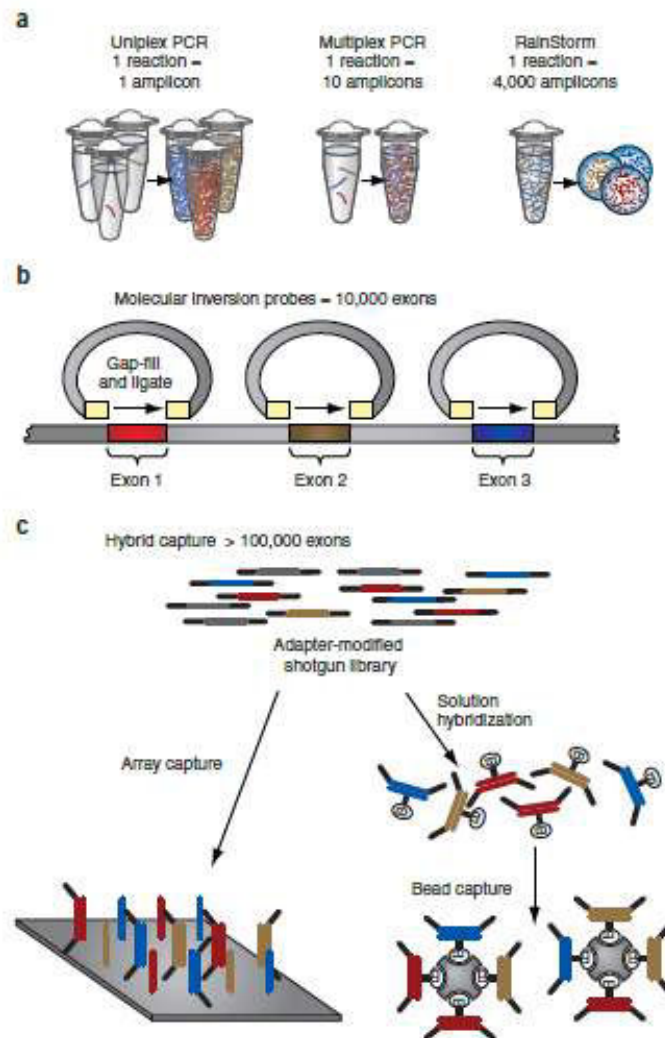
**Figure A.7 :** La technologie RainDance : système de PCR en microgouttelettes, d'après Tewhey et al. [2009].

Le principe de la technologie RainDance repose sur l'utilisation d'une puce microfluidique, permettant la génération de deux types de micro-gouttelettes. Elles contiennent soit des couples d'amorces permettant l'amplification des régions d'intérêt du génome soit le mélange réactionnel nécessaire à la réaction de PCR. Les deux types de micro-gouttelettes sont introduites dans la puce par un micro-canal différent. Les micro-canaux se connectent ensuite entre eux de manière à réaliser le mélange d'un couple d'amorces donné avec les réactifs afin de réaliser une seule réaction de PCR par micro-gouttelette.



**Figure A.8 :** Capture des exons avec les sondes inversées (MIPs), d'après Turner et al. [2009].

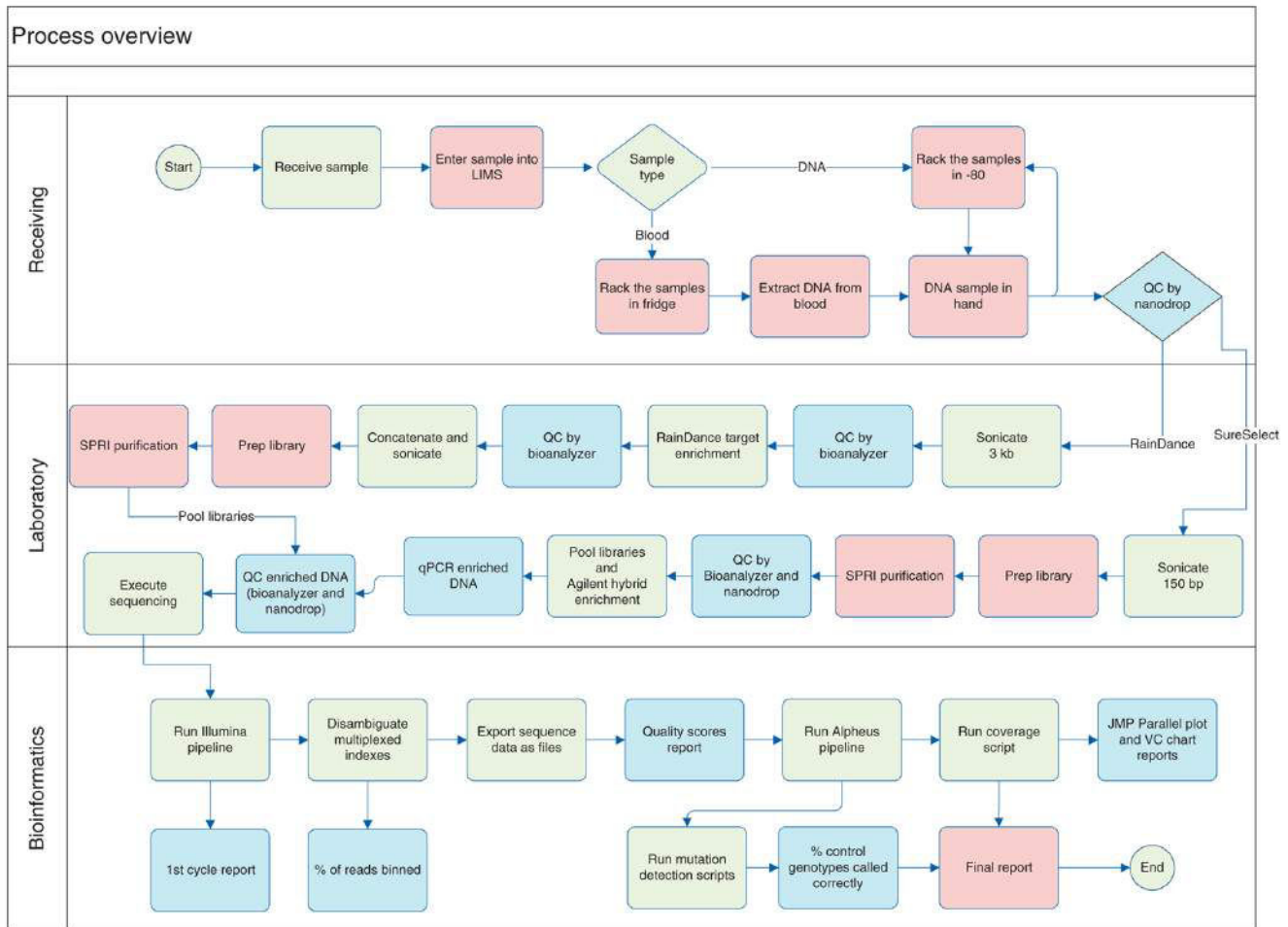
Chaque MIP comprend une région commune (en vert), encadrée de séquences ciblant les régions d'intérêt (les exons), capables de s'hybrider de part et d'autre de celles-ci (en jaune). L'ajout de la polymérase et de la ligase permet la formation d'une forme circulaire incluant la copie de la séquence d'intérêt. Après ajout des adaptateurs (en noir), les amplicons obtenus sont directement séquencés.



**Figure A.9** : Les 3 méthodes d'enrichissement spécifique : PCR ciblées (a), MIPs (b) et capture par hybridation spécifique (c), d'après Mamanova et al. [2010].

Parmi ces diverses méthodes d'enrichissement spécifique, la capture par hybridation spécifique en phase liquide a été majoritairement utilisée par la suite ; pour les résultats présentés dans ce manuscrit et par la communauté scientifique en général.

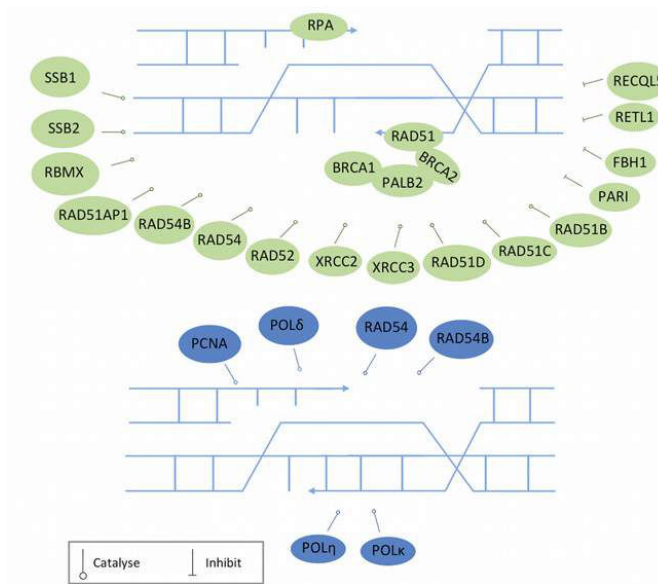




**Figure A.10** : Du génome à la variation, d'après Bell et al. [2011].

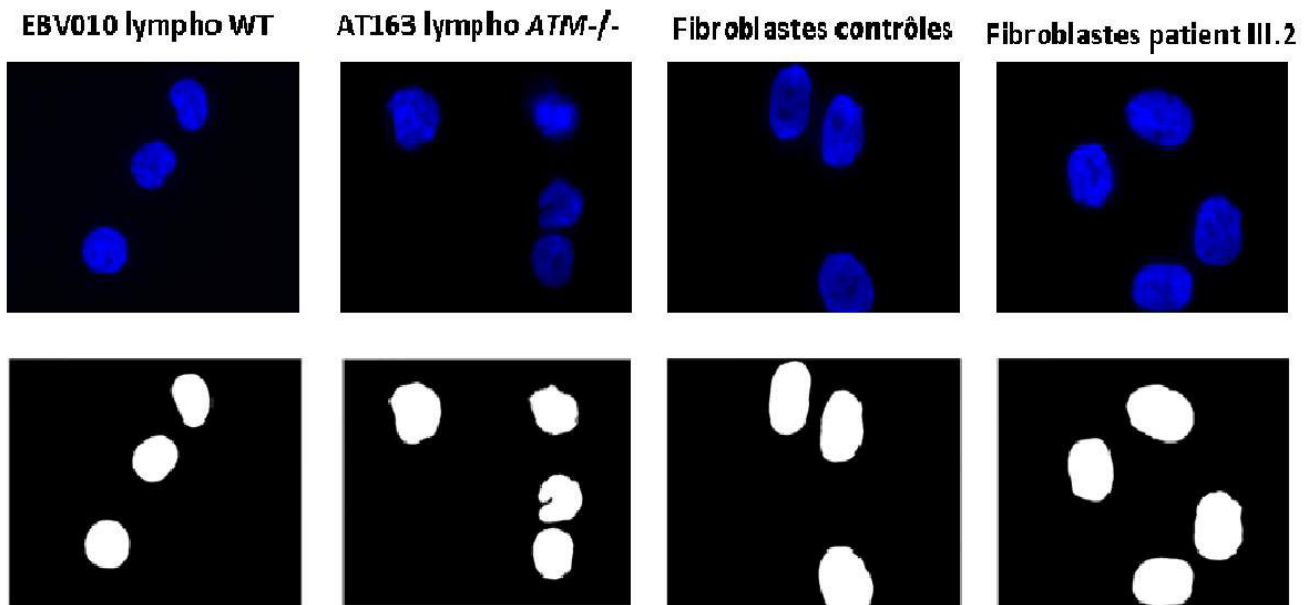
Après l'étape initiale d'examen clinique, les différentes étapes du processus permettant d'identifier des variants candidats est ici détaillé : il commence avec l'extraction de l'ADN des échantillons reçus. À la suite du séquençage et de du traitement bio-informatique, l'étape suivante correspond à l'analyse des données et fait l'objet de ce manuscrit.





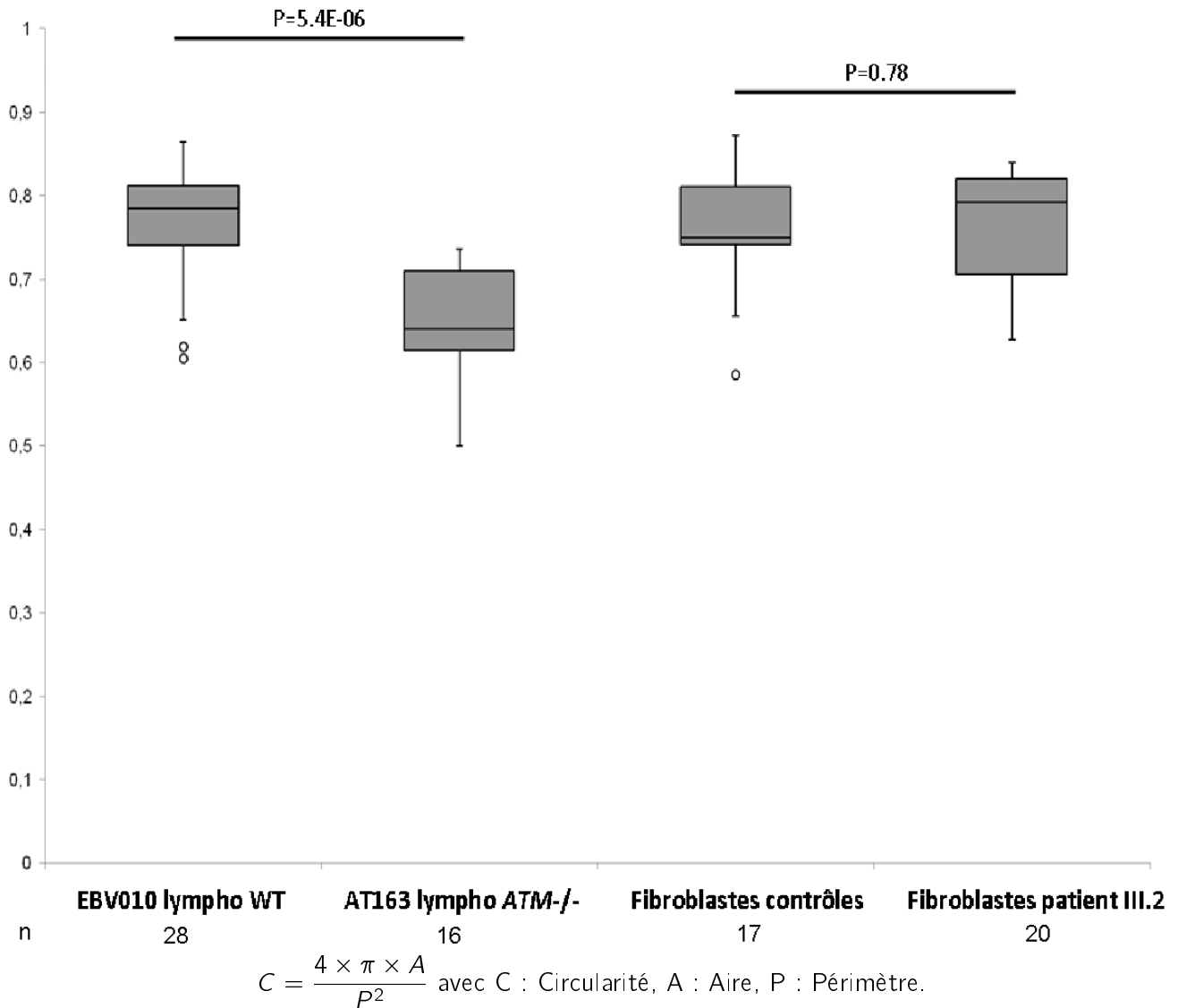
**Figure A.11** : Le mécanisme de réparation des CDB dépendant de la recombinaison homologue d'après Liu et al. [2014a].

Pour que la réparation s'établisse, quelques étapes préliminaires sont indispensables ; les extrémités 5' des CDB sont digérées pour générer des extrémités simple brin de chaque côté de la cassure. Celles-ci sont recouvertes et stabilisées par les protéines RPA. Ensuite, RAD51 se lie à ces extrémités simple brin avec l'intervention de médiateurs de la recombinaison RAD52, RAD55-57, RAD54 et RAD54B. La structure ainsi formée envahit la molécule double brin intacte et commence un dépistage de régions homologues, formant une structure appelée la D-loop. La longueur de la D-loop augmente avec la migration des branches d'ADN. RAD54B est capable, grâce à l'hydrolyse d'ATP, de se déplacer sur l'ADN double-brin (translocation). Son déplacement génère des enroulements de l'ADN, positifs en amont et négatifs en aval. Les super-tours négatifs en aval permettent l'ouverture transitoire de la double hélice d'ADN et facilitent la formation des jonctions entre les régions homologues. Enfin, RAD54B est capable d'interagir directement avec les recombinases RAD51 et DMC1 afin de faciliter le chargement de celles-ci sur l'ADN simple brin. RAD54B stimule également la réaction de formation de la D-loop, un intermédiaire essentiel de la recombinaison homologue [Wesoly et al., 2006, Liu et al., 2014a].



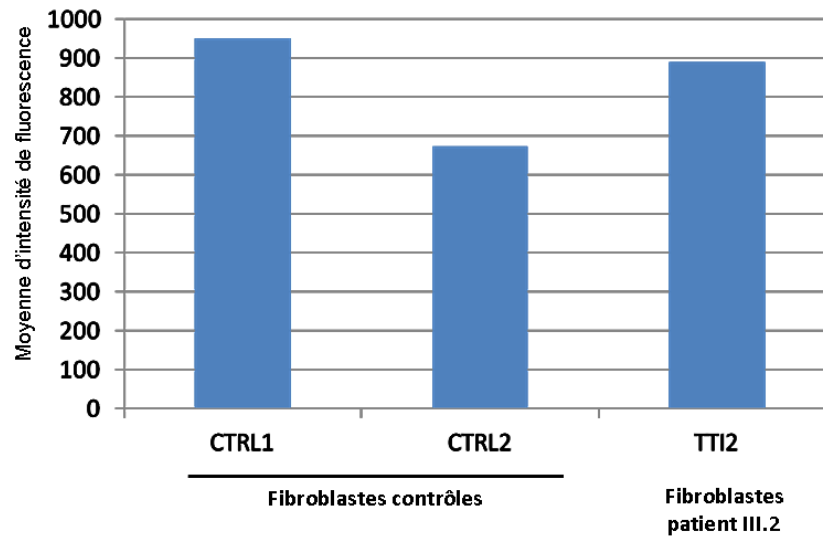
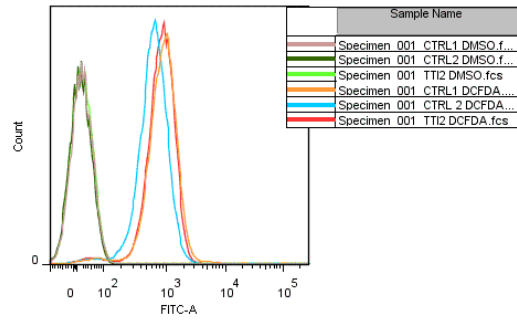
**Figure A.12** : Forme des noyaux des cellules :

marquage au DAPI (première ligne) et contraste (seconde ligne) des fibroblastes du patient III.2, des cellules lymphoïdes AT163 *ATM*<sup>-/-</sup> et des contrôles correspondants (fibroblastes contrôles et cellules lymphoïdes EBV010).



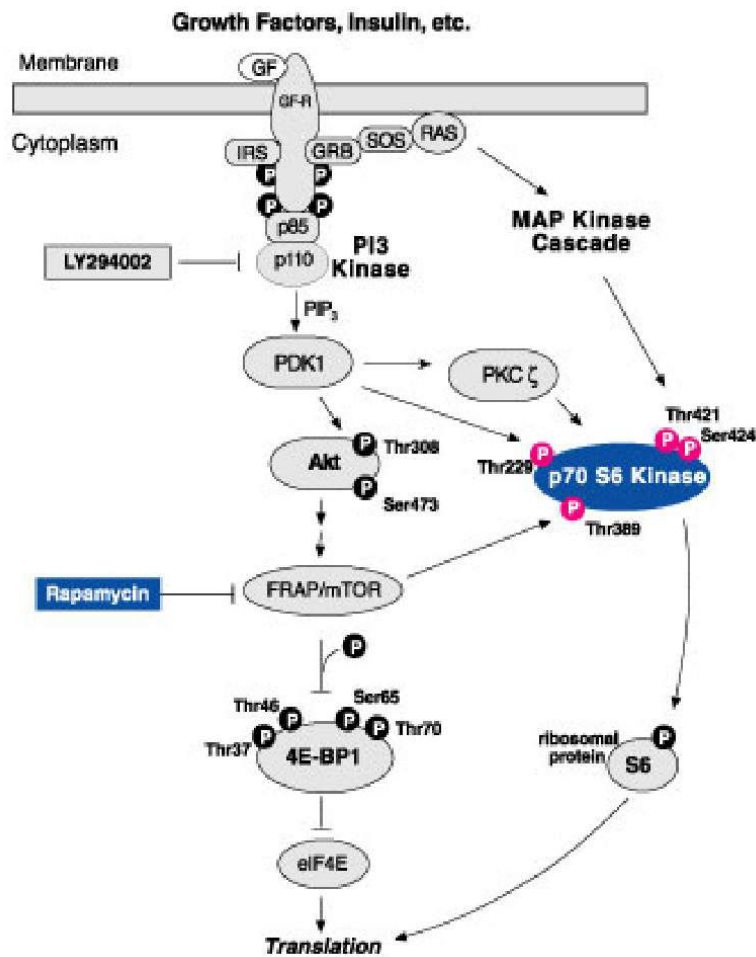
**Figure A.13** : Circularité des noyaux dans les fibroblastes du patient III.2, les cellules lymphoïdes AT163 *ATM*<sup>-/-</sup> et les contrôles correspondants (fibroblastes contrôles et cellules lymphoïdes EBV010).

À l'aide du logiciel ImageJ, la circularité C a été calculée à partir des mesures réalisées sur les images contrastées issues du marquage DAPI présentées dans la figure A.12. Un nombre n de cellules a été utilisé pour déterminer la circularité moyenne de chaque groupe. Un test de student a ensuite été réalisé entre les deux groupes de chaque type cellulaire. Les p-values obtenues sont indiquées sur la figure : elles indiquent une différence entre les cellules lymphoïdes AT163 *ATM*<sup>-/-</sup> et les contrôles lymphoïdes EBV010 mais ne révèlent pas de différences significatives entre les fibroblastes du patient III.2 et les fibroblastes contrôles.



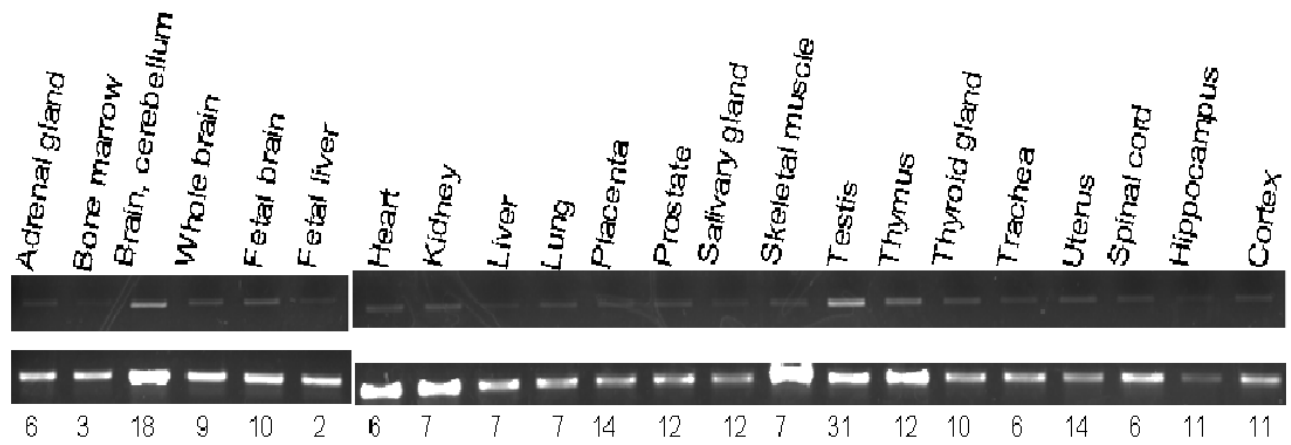
**Figure A.14 :** Niveau de ROS mesuré par DCFDA dans les fibroblastes du patient III.2 et les fibroblastes contrôles.

Afin de déterminer les niveaux de ROS, nous avons mesuré la fluorescence du 2',7'-dichlorofluorescein diacetate (DCFDA) des cellules en conditions de culture classiques. Les premiers résultats de cette analyse n'indiquent pas des niveaux de ROS différents entre les fibroblastes du patient III.2 et les deux contrôles testés.



**Figure A.15** : Schéma de la voie de signalisation mTOR.

En réponse à des stimuli externes à la cellule, tels que les facteurs de croissance, l'insuline etc., la voie mTOR est activée. Une des étapes de cette voie de signalisation conduit à la phosphorylation de la Thréonine (Thr) 339 de la protéine p70 S6 Kinase. L'état de phosphorylation de ce résidu peut être analysé par WB avec l'utilisation d'anticorps spécifiques.



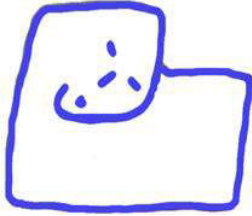
**Figure A.16** : Test de l'expression du gène *TT12* dans divers tissus humain par RT-PCR.

La bande supérieure correspond à l'amplification par RT-PCR du gène *TT12* et la bande inférieure à celle du gène *GAPDH*. Les valeurs (unité arbitraire) correspondent aux résultats de semi-quantification, réalisée par une mesure de pixels avec le logiciel *ImageLab*.

## RESEAU MORPHEE : AGENDA DE VIGILANCE ET DE SOMMEIL

DATE	HEURES	<i>Appréciation par :</i> TB - B - Moy. - M - TM			
Nuit du ... au...	<div style="display: flex; justify-content: space-between; font-size: small; margin-bottom: 5px;"> <span>20</span><span>22</span><span>24</span><span>2</span><span>4</span><span>6</span><span>8</span><span>10</span><span>12</span><span>14</span><span>16</span><span>18</span><span>20</span> </div> <div style="display: flex; justify-content: space-between; font-size: x-small; margin-bottom: 5px;"> <span>21</span><span>23</span><span>1</span><span>3</span><span>5</span><span>7</span><span>9</span><span>11</span><span>13</span><span>15</span><span>17</span><span>19</span> </div>	<b>QUALITE DU SOMMEIL</b>	<b>QUALITE DU REVEIL</b>	<b>FORME DE LA JOURNEE</b>	<b>TRAITEMENT ET REMARQUES PARTICULIERES</b>
<b>EXEMPLE</b>	<div style="display: flex; justify-content: space-between; align-items: center; margin-bottom: 5px;"> <span>↓</span><span style="background: repeating-linear-gradient(45deg, transparent, transparent 2px, #ccc 2px, #ccc 4px); width: 20px; height: 10px;"></span><span>↑</span> </div>	M	TM	Moy	

↓	heure de mise au lit		long réveil	S	Somnolence dans la journée
↑	heure de lever		1/2 sommeil		



Réseau Morphée

Figure A.17 : Agenda de sommeil.

## Réseau de santé consacré à la prise en charge des troubles chroniques du sommeil

### COMMENT REMPLIR L'AGENDA DU SOMMEIL

L'agenda se remplit à deux moments de la journée :

- ✓ le matin au réveil pour décrire le déroulement de la nuit
- ✓ le soir pour reporter ce qui s'est passé dans la journée.

✚ **Le matin** : inscrivez la date (*nuit du 5 au 6 juin par exemple*).

- Notez le moment où vous vous êtes mis au lit par une flèche vers le bas. (*même si c'était pour lire ou pour regarder la télévision*)
- Mettez une flèche vers le haut pour indiquer votre horaire définitif de lever, ou pour préciser un lever en cours de nuit.
- Hachurez la zone qui correspond à votre sommeil. Si vous vous êtes réveillé en pleine nuit et que cet éveil a été gênant, vous interrompez la zone hachurée en regard des horaires du réveil. Le but n'est pas de vous souvenir précisément que l'éveil s'est produit entre 3h17 et 4h12 mais de traduire simplement votre impression avec des horaires qui seront obligatoirement approximatifs.
- Précisez ensuite la qualité de votre nuit, votre forme au réveil, en remplissant les colonnes de droite (*TB = très bon, B = Bon, Moy = moyen, M = mauvais, TM = très mauvais*)
- Précisez si vous avez pris des médicaments, ou s'il s'est passé quelque chose qui a pu interférer avec votre sommeil (*par exemple faire du sport le soir, avoir mal à la tête, avoir de la fièvre...*).

✚ **Le soir** : vous remplissez à nouveau l'agenda

- Pour préciser si vous avez fait une sieste, par une zone hachurée, de la même façon que vous avez noté votre sommeil de nuit. Vous reportez si besoin les moments de somnolence en notant un « S » en regard des horaires correspondants.
- Enfin, vous apprécierez votre forme globale dans la journée.

Vous remplirez cet agenda pendant 3 semaines minimum. S'il vous est possible de le remplir à la fois sur une période de travail et une période de vacances, vous aurez beaucoup plus d'informations.



# ANNEXE B

## TABLES

sujet	contrôle 1	contrôle 2	patient IV.5	patient IV.12
intensité de fluorescence	16516.163	10479.67	10214.147	12735.172
	12000.946	11053.846	7873.695	16277.557
	13790.209	7259.689	8532.948	15868.156
	12728.059	12452.988	8235.793	8592.362
	14378.334	10473.355	10001.808	8572.429
	11634.129	11174.93	8304.264	10222.108
	17092.211	9961.429	9947.047	12808.892
	13666.347	8061.058	8830.468	6340.796
	11298.533	11898.856	12716.068	11349.622
	18466.132	9909.205	9974.518	7241.176
	13171.383	10564.459	10632.542	9059.83
	16400.823	10068.409	8088.789	9118.919
	17044.24	12283.217	26405.407	8688.419
	14751.174	16106.78	25700.577	12545.981
	13757.652	18988.559	15886.306	7197.7
	26516.863	16242.404	15567.369	31928.392
	15779.776	22061.872	31200.538	16778.689
	25815.041	19290.834	20318.268	17398.161
27890.886	12803.042	-	-	
minimum	11298.533	7259.689	7873.695	6340.796
maximum	27890.886	22061.872	31200.538	31928.392
médiane	14751.174	11174.93	10108	10786
moyenne	16458	12691	13802	12374
p-val (vs contrôle 1)	-	-	0.4893268	0.1365181
p-val (vs contrôle 2)	-	-	0.9332697	0.9982291

**Table B.1** : Données d'intensités de fluorescence du marquage RAD54B : mesure du niveau de gris par noyau (unités arbitraires,  $n = 18$  ou  $19$  noyaux par individu).

Populations lymphocytaires	Patient III.2, à l'âge de 38ans	Normes
CD3	81%	64-85%
CD4	48%	34-62%
CD8	30%	14-42%
<b>T CD4 naïfs</b>	<b>14%</b>	<b>30-48%</b>
T CD4 mémoire	63%	
<b>T CD8 naïfs</b>	<b>14%</b>	<b>37-50%</b>
<b>T CD8 central mémoire</b>	<b>4%</b>	<b>6-16%</b>
T CD8 effecteur mémoire	34%	25-37%
<b>T CD8 effecteur mémoire RA+</b>	<b>48%</b>	<b>8-20%</b>
B CD19	6%	6-17%
NK	13%	5-20%
IgG	10.55g/l	6.65-12.8g/l
IgA	1.66g/l	0.7-3.44g/l
IgM	1.09g/l	0.5-2.09g/l
Prolifération lymphocytaire en réponse à la PHA	66cpm/10 <sup>3</sup>	>50cpm/10 <sup>3</sup>
<b>Prolifération lymphocytaire en réponse à la tuberculine</b>	<b>1cpm/10<sup>3</sup></b>	<b>&gt;10cpm/10<sup>3</sup></b>

**Table B.2** : Bilan Immunologique du patient III.2 de la famille EI.

Ig : Immunoglobulines, cpm : coup par minute, g/l : grammes par litre.

Marqueurs	Nom entier	Patient III.2	Patient III.5	Patient III.9	Normes
CA125	Carbohydrate Antigen 125	9,22	10,68	15,26	<35U/ml
CA15-3	Carbohydrate Antigen 15-3	11,01	10 ,38	12,16	<25U/ml
CA19-9	Carbohydrate Antigen 19-9	17,34	4,92	9,21	<39U/ml
<b>AFP</b>	<b>Alpha-FœtoProtéine</b>	<b>13,92</b>	<b>17,83</b>	<b>19,11</b>	<b>&lt;5,8 U/ml</b>
ACE	Antigène Carcino-Embryonnaire	3,94	3,36	3,92	<3,4 ng/l
HCG $\beta$	Hormone Chorionique Gonadotrope $\beta$		3,29	1,83	<4mUI/ml
PSA totale	Antigène Spécifique Prostatique	0,637			<2ug/l

**Table B.3** : Dosage des marqueurs tumoraux chez les 3 enfants atteints de la famille EI.

Catégorie de GO	p-value	Nb de Gènes inclus
Cancer	$8.17^{-17}-2.07^{-03}$	254
Gastrointestinal Disease	$8.17^{-17}-2.02^{-03}$	174
Cellular Growth and Proliferation	$8.11^{-16}-1.81^{-03}$	160
Cellular Development	$2.51^{-09}-1.93^{-03}$	158
Organismal Development	$1.97^{-13}-2.09^{-03}$	151
Tissue Development	$8.71^{-12}-2.09^{-03}$	141
Organismal Injury and Abnormalities	$1.44^{-07}-2.06^{-03}$	135
Cell Death and Survival	$2.87^{-7}-2.09^{-03}$	135
Tissue Morphology	$1.07^{-7}-2.08^{-03}$	118
Cell Morphology	$8.88^{-8}-1.93^{-03}$	115
Embryonic Development	$2.84^{-09}-2.09^{-03}$	110
Organismal Survival	$1.04^{-09}-1.13^{-04}$	109
Cellular Movement	$9.02^{-14}-2.07^{-03}$	106
Organ Development	$2.84^{-09}-2.09^{-03}$	99
Skeletal and Muscular Disorders	$2.36^{-08}-1.57^{-03}$	99

**Table B.4** : Modélisation des fonctions dérégulées chez les patients MCCID.

Les données ont été soumises au logiciel *Ingenuity Pathway Analysis* (IPA) [<http://www.ingenuity.com>] afin de modéliser les relations entre les gènes et les protéines et ainsi définir les voies de signalisation et les processus biologiques significatifs.

- Akizu, N., Cantagrel, V., Schroth, J., Cai, N., Vaux, K., McCloskey, D., Naviaux, R., Van Vleet, J., Fenstermaker, A., Silhavy, J., Scheliga, J., Toyama, K., Morisaki, H., Sonmez, F. M., Celep, F., Oraby, A., Zaki, M., Al-Baradie, R., Faqeih, E. A., Saleh, M. A., Spencer, E., Rosti, R., Scott, E., Nickerson, E., Gabriel, S., Morisaki, T., Holmes, E., and Gleeson, J. (2013). Ampd2 regulates gtp synthesis and is mutated in a potentially treatable neurodegenerative brainstem disorder. *Cell*, 154(3) :505 – 517.
- Akula, N., Barb, J., Jiang, X., Wendland, J., Choi, K., Sen, S., Hou, L., Chen, D., Laje, G., Johnson, K., Lipska, B., Kleinman, J., Corrada-Bravo, H., Detera-Wadleigh, S., Munson, P., and McMahon, F. (2014). Rna-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and gtpase binding in bipolar disorder. *Mol Psychiatry*.
- Albert, T. J., Molla, M. N., Muzny, D. M., Nazareth, L., Wheeler, D., Song, X., Richmond, T. A., Middle, C. M., Rodesch, M. J., Packard, C. J., Weinstock, G. M., and Gibbs, R. A. (2007). Direct selection of human genomic loci by microarray hybridization. *Nat Methods*, 11 :903–905.
- Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., Plouin, P., Carter, N. P., Lyonnet, S., Munnich, A., and Colleaux, L. (2007). Mutations in tcf4, encoding a class i basic helix-loop-helix transcription factor, are responsible for pitt-hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *The American Journal of Human Genetics*, 80(5) :988 – 993.
- Anderson, V., Godfrey, C., Rosenfeld, J. V., and Catroppa, C. (2012). Predictors of cognitive function and recovery 10 years after traumatic brain injury in young children. *Pediatrics*, 129(2) :e254–e261.
- Bacino, C. A., Arriola, L. A., Wiszniewska, J., and Bonnen, P. E. (2012). Wdr62 missense mutation in a consanguineous family with primary microcephaly. *American Journal of Medical Genetics Part A*, 158A(3) :622–625.
- Bahi-Buisson, N., Chelly, J., and des Portes, V. (2006). [update on the genetics of x-linked mental retardation]. *Rev Neurol (Paris)*, 162 :952–963.
- Baldini, A. (2005). Dissecting contiguous gene defects : {TBX1}. *Current Opinion in Genetics & Development*, 15(3) :279 – 284. Genetics of disease.
- Bamshad, M. J., Ng, S. B., Bigam, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., and Shendure, J. (2011). Exome sequencing as a tool for mendelian disease gene discovery. *Nature Reviews Genetics*, 12 :745–755.

- Barascu, A., Le Chalony, C., Pennarun, G., Genet, D., Imam, N., Lopez, B., and Bertrand, P. (2012). Oxidative stress induces an atm-independent senescence pathway through p38 mapk-mediated lamin b1 accumulation. *EMBO J.*, 31 :1080–1094.
- Barcia, G., Fleming, M. R., Deligniere, A., Gazula, V. R., Brown, M. R., Langouët, M., Chen, H., Kronengold, J., Abhyankar, A., Cilio, R., Nitschke, P., Kaminska, A., Boddaert, N., Casanova, J. L., Desguerre, I., Munnich, A., Dulac, O., Kaczmarek, L. K., Colleaux, L., and Nabbout, R. (2012). De novo gain-of-function *kcnt1* channel mutations cause malignant migrating partial seizures of infancy. *Nat Genet*, 44 :1255–1259.
- Basel-Vanagaite, L. (2007). Genetics of autosomal recessive non-syndromic mental retardation : recent advances. *Clinical Genetics*, 72(3) :167–174.
- Basel-Vanagaite, L., Attia, R., Yahav, M., Ferland, R. J., Anteki, L., Walsh, C. A., Olender, T., Straussberg, R., Magal, N., Taub, E., Drasinover, V., Alkelai, A., Bercovich, D., Rechavi, G., Simon, A. J., and Shohat, M. (2006). The *cc2d1a*, a member of a new gene family with c2 domains, is involved in autosomal recessive non-syndromic mental retardation. *Journal of Medical Genetics*, 43(3) :203–210.
- Bashi, J. (1977). Effects of inbreeding on cognitive performance. *Nature*, 266 :440–442.
- Bath, S. C., Steer, C. D., Golding, J., Emmett, P., and Rayman, M. P. (2013). Effect of inadequate iodine status in {UK} pregnant women on cognitive outcomes in their children : results from the avon longitudinal study of parents and children (alspac). *The Lancet*, 382(9889) :331 – 337.
- Behrooz, H. G., Tohidi, M., Mehrabi, Y., Behrooz, E. G., Tehranidoost, M., and Azizi, F. (2011). Subclinical hypothyroidism in pregnancy : Intellectual development of offspring. *Thyroid*, 21 :1143–1147.
- Bell, C. J., Dinwiddie, D. L., Miller, N. A., Hateley, S. L., Ganusova, E. E., Mudge, J., Langley, R. J., Zhang, L., Lee, C. C., Schilkey, F. D., Sheth, V., Woodward, J. E., Peckham, H. E., Schroth, G. P., Kim, R. W., and Kingsmore, S. F. (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Science Translational Medicine*, 3(65) :65ra4.
- Benoist, G., Jacquemard, F., Leruez-Ville, M., and Ville, Y. (2008). Cytomegalovirus (cmv) congenital infection. *Gynecol Obstet Fertil.*, 36(3) :248 – 260.
- Berkel, S., Marshall, C. R., 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, S., and Rappold, G. A. (2010). Mutations in the *shank2* synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nature Genetics*, 42 :489–491.
- Bhat, V., Girimaji, S., Mohan, G., Arvinda, H., Singhmar, P., Duvvari, M., and Kumar, A. (2011). Mutations in *wdr62*, encoding a centrosomal and nuclear protein, in indian primary microcephaly families with cortical malformations. *Clinical Genetics*, 80(6) :532–540.
- Bilgüvar, K., Oztürk, A. K., Louvi, A., Kwan, K. Y., Choi, M., Tatli, B., Yalnizoglu, D., Tüysüz, B., Caglayan, A. O., Gökben, S., Kaymakçalan, H., Barak, T., Bakircioglu, M., Yasuno, K., Ho, W., Sanders, S., Zhu, Y., Yilmaz, S., Dinçer, A., Johnson, M. H., Bronen, R. A., Koçer, N., Per, H., Mane, S., Pamir, M. N., Yalçinkaya, C., Kumandas, S., Topçu, M., Ozmen, M., Sestan, N., Lifton, R. P., State, M. W., and Günel, M. (2010). Whole-exome sequencing identifies recessive *wdr62* mutations in severe brain malformations. *Nature*, 467 :207–210.
- Birk, E., Har-Zahav, A., Manzini, C. M., Pasmanik-Chor, M., Kornreich, L., Walsh, C. A., Noben-Trauth, K., Albin, A., Simon, A. J., Colleaux, L., Morad, Y., Rainshtein, L., Tischfield, D. J., Wang, P., Magal, N., Maya, I., Shoshani, N.,

- Rechavi, G., Gothelf, D., Maydan, G., Shohat, M., and Basel-Vanagaite, L. (2010). Sobp is mutated in syndromic and nonsyndromic intellectual disability and is highly expressed in the brain limbic system. *American Journal of Human Genetics*, 87 :694–700.
- Bjursell, C., Wahlström, J., Berg, K., Stibler, H., Kristiansson, B., Matthijs, G., and Martinsson, T. (1998). Detailed mapping of the phosphomannomutase 2 (pmm2) gene and mutation detection enable improved analysis for scandinavian cdg type i families. *Eur J Hum Genet*, 6 :603–611.
- Boczonadi, V., Müller, J. S., Pyle, A., Munkley, J., Dor, T., Quartararo, J., Ferrero, I., Karcagi, V., Giunta, M., Polvikoski, T., Birchall, D., Princzinger, A., Cinnamon, Y., Lützkendorf, S., Piko, H., Reza, M., Florez, L., Santibanez-Koref, M., Griffin, H., Schuelke, M., Elpeleg, O., Kalaydjieva, L., Lochmüller, H., Elliott, D. J., Chinnery, P. F., Edvardson, S., and Horvath, R. (2014). Exosc8 mutations alter mrna metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. *Nat Commun.*, 5.
- Bolze, A., Byun, M., McDonald, D., Morgan, N. V., Abhyankar, A., Premkumar, L., Puel, A., Bacon, C. M., Rieux-Laucat, F., Pang, K., Britland, A., Abel, L., Cant, A., Maher, E. R., Riedl, S. J., Hambleton, S., and Casanova, J.-L. (2010). Whole-exome-sequencing-based discovery of human fadd deficiency. *The American Journal of Human Genetics*, 87(6) :873 – 881.
- Bond, C. S. and Fox, A. H. (2009). Paraspeckles : nuclear bodies built on long noncoding rna. *The Journal of Cell Biology*, 186(5) :637–644.
- Bondar, O. P., Barnidge, D. R., Klee, E. W., Davis, B. J., and Klee, G. G. (2007). Lc-ms/ms quantification of zn-alpha2 glycoprotein : A potential serum biomarker for prostate cancer. *Clinical Chemistry*, 53(4) :673–678.
- Brown, S. A., Ripperger, J., Kadener, S., Fleury-Olela, F., Vilbois, F., Rosbash, M., and Schibler, U. (2005). Period1-associated proteins modulate the negative limb of the mammalian circadian oscillator. *Science*, 308(5722) :693–696.
- Cans, C., Wilhelm, L., Baille, M. F., du Mazaubrun, C., Grandjean, H., and Rumeau-Rouquette, C. (1999). Aetiological findings and associated factors in children with severe mental retardation. *Developmental Medicine & Child Neurology*, 41(4) :233–239.
- Cassidy, S. B., Schwartz, S., Miller, J. L., and Driscoll, D. J. (2012). Prader-willi syndrome. *Genetics in Medicine*, 14 :10–26.
- Chamberlain, S. J., Chen, P.-F., Ng, K. Y., Bourgois-Rocha, F., Lemtiri-Chlieh, F., Levine, E. S., and Lalande, M. (2010). Induced pluripotent stem cell models of the genomic imprinting disorders angelman and prader-willi syndromes. *Proceedings of the National Academy of Sciences*, 107(41) :17668–17673.
- Chamberlain, S. J. and Lalande, M. (2010). Neurodevelopmental disorders involving genomic imprinting at human chromosome 15q11–q13. *Neurobiology of Disease*, 39(1) :13 – 20. Epigenetics and Neuropsychiatric Disease.
- Chapman, D. A., Scott, K. G., and Mason, C. A. (2002). Early risk factors for mental retardation : role of maternal age and maternal education. *American Journal on Mental Retardation*, 107 :46–59.
- Chen, J. F., Zhang, Y., Wilde, J., Hansen, K. C., Lai, F., and Niswander, L. (2014). Microcephaly disease gene wdr62 regulates mitotic progression of embryonic neural stem cells and brain size. *Nat Commun.*, 5 :3885.
- Chen, L.-L. and Carmichael, G. G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells : Functional role of a nuclear noncoding rna. *Molecular Cell*, 35(4) :467 – 478.

- Chen, L. L., DeCerbo, J. N., and Carmichael, G. G. (2008). Alu element-mediated gene silencing. *EMBO J.*, 27 :1694–705.
- Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakaloglu, A., Özen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., and Lifton, R. P. (2009). Genetic diagnosis by whole exome capture and massively parallel dna sequencing. *Proceedings of the National Academy of Sciences*, 106(45) :19096–19101.
- Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., and Lawrence, J. B. (2009). An architectural role for a nuclear noncoding rna : Neat1 rna is essential for the structure of paraspeckles. *Molecular Cell*, 33(6) :717 – 726.
- Clifford, A., Lang, L., and Chen, R. (2012). Effects of maternal cigarette smoking during pregnancy on cognitive parameters of children and young adults : A literature review. *Neurotoxicology and Teratology*, 34(6) :560 – 570.
- collective INSERM, E. (2004). *Déficiences et handicaps d'origine périnatale : Dépistage et prise en charge*.
- Consortium, E. P. (2012). An integrated encyclopedia of dna elements in the human genome. *Nature*, 489 :57–74.
- Consortium, I. H. G. S. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431 :931–945.
- Contrepas, A., Walker, J., Koulakoff, A., Franek, K. J., Qadri, F., Giaume, C., Corvol, P., Schwartz, C. E., and Nguyen, G. (2009). A role of the (pro)renin receptor in neuronal cell differentiation. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 297(2) :R250–R257.
- Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227 :561–563.
- Daily, D. K., Ardinger, H. H., and Holmes, G. E. (2000). Identification and evaluation of mental retardation. *Am Fam Physician*, 61 :1059–1067.
- Damulewicz, M. and Pyza, E. (2011). The clock input to the first optic neuropil of *Drosophila melanogaster* expressing neuronal circadian plasticity. *PLoS ONE*, 6(6) :e21258.
- Daoud, H., Gruchy, N., Constans, J.-M., Moussaoui, E., Saumureau, S., Bayou, N., Amy, M., Védrine, S., Vu, P., Rötig, A., Laumonnier, F., Vourc'h, P., Andres, C., Leporrier, N., and Briault, S. (2009). Haploinsufficiency of the *gpd2* gene in a patient with nonsyndromic mental retardation. *Human Genetics*, 124(6) :649–658.
- de Brouwer, A. P., Yntema, H. G., Kleefstra, T., Lugtenberg, D., Oudakker, A. R., de Vries, B. B., van Bokhoven, H., Van Esch, H., Frints, S. G., Froyen, G., Fryns, J.-P., Raynaud, M., Moizard, M.-P., Ronce, N., Bensalem, A., Moraine, C., Poirier, K., Castelnau, L., Saillour, Y., Bienvenu, T., Beldjord, C., des Portes, V., Chelly, J., Turner, G., Fullston, T., Gecz, J., Kuss, A. W., Tzschach, A., Jensen, L. R., Lenzner, S., Kalscheuer, V. M., Ropers, H.-H., and Hamel, B. C. (2007). Mutation frequencies of x-linked mental retardation genes in families from the euromrx consortium. *Human Mutation*, 28(2) :207–208.
- de Ligt, J., Willemsen, M. H., van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., Vulto-van Silfhout, A. T., Koolen, D. A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B. B., Brunner, H. G., Veltman, J. A., and Vissers, L. E. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *New England Journal of Medicine*, 367(20) :1921–1929. PMID : 23033978.
- de Lonlay, P. (2003). [how does one localize a disease gene without a genetic linkage study?]. *médecine/sciences*, 19 :527–528.

- Deidda, G., Fernandez Bozarth, I., and Cancedda, L. (2014). Modulation of gabaergic transmission in development and neurodevelopmental disorders : Investigating physiology and pathology to gain therapeutic perspectives. *Frontiers in Cellular Neuroscience*, 8(119).
- DiLella, A. G., Marvit, J., Lidsky, A. S., Güttler, F., and Woo, S. L. (1986). Tight linkage between a splicing mutation and a specific dna haplotype in phenylketonuria. *Nature*, 322 :799–803.
- Duong, H. A., Robles, M. S., Knutti, D., and Weitz, C. J. (2011). A molecular mechanism for circadian clock negative feedback. *Science*, 332(6036) :1436–1439.
- Ellison, J. W. and Rosenfeld, J A Shaffer, L. G. (2013). Genetic basis of intellectual disability. *Annu Rev Med*, 64 :441–450.
- Farag, H., Froehler, S., Oexle, K., Ravindran, E., Schindler, D., Staab, T., Huebner, A., Kraemer, N., Chen, W., and Kaindl, A. (2013). Abnormal centrosome and spindle morphology in a patient with autosomal recessive primary microcephaly type 2 due to compound heterozygous wdr62 gene mutation. *Orphanet Journal of Rare Diseases*, 8(1) :178.
- Fischbach, B. V., Trout, K. L., Lewis, J., Luis, C. A., and Sika, M. (2005). Wagr syndrome : A clinical review of 54 cases. *Pediatrics*, 116(4) :984–988.
- Fitzky, B. U., Witsch-Baumgartner, M., Erdel, M., Lee, J. N., Paik, Y.-K., Glossmann, H., Utermann, G., and Moebius, F. F. (1998). Mutations in the delta7-sterol reductase gene in patients with the smith-lemli-opitz syndrome. *Proceedings of the National Academy of Sciences*, 95(14) :8181–8186.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., and Merrick, J. M. e. a. (1995). Whole-genome random sequencing and assembly of haemophilus influenzae rd. *Science*, 269 :496–512.
- Foster, R. G., Peirson, S. N., Wulff, K., Winnebeck, E., Vetter, C., and Roenneberg, T. (2013). Chapter eleven - sleep and circadian rhythm disruption in social jetlag and mental illness. In Gillette, M. U., editor, *Chronobiology : Biological Timing in Health and Disease*, volume 119 of *Progress in Molecular Biology and Translational Science*, pages 325 – 346. Academic Press.
- Franklin, T. B. and Mansuy, I. M. (2011). The involvement of epigenetic defects in mental retardation. *Neurobiology of Learning and Memory*, 96(1) :61 – 67. Epigenetics and memory.
- Garshasbi, M., Hadavi, V., Habibi, H., Kahrizi, K., Kariminejad, R., Behjati, F., Tzschach, A., Najmabadi, H., Ropers, H. H., and Kuss, A. W. (2008). A defect in the tusc3 gene is associated with autosomal recessive mental retardation. *American Journal of Human Genetics*, 82 :1158–1164.
- Gauthier, J., Champagne, N., Lafrenière, R. G., Xiong, L., Spiegelman, D., Brustein, E., Lapointe, M., Peng, H., Côté, M., Noreau, A., Hamdan, F. F., Addington, A. M., Rapoport, J. L., DeLisi, L. E., Krebs, M.-O., Joober, R., Fathalli, F., Mouaffak, F., Haghghi, A. P., Néri, C., Dubé, M.-P., Samuels, M. E., Marineau, C., Stone, E. A., Awadalla, P., Barker, P. A., Carbonetto, S., Drapeau, P., Rouleau, G. A., and the S2D Team (2010). De novo mutations in the gene encoding the synaptic scaffolding protein shank3 in patients ascertained for schizophrenia. *Proceedings of the National Academy of Sciences*, 107(17) :7863–7868.
- Gilissen, C., Arts, H. H., Hoischen, A., Spruijt, L., Mans, D. A., Arts, P., van Lier, B., Steehouwer, M., van Reeuwijk, J., Kant, S. G., Roepman, R., Knoers, N. V. A. M., Veltman, J. A., and Brunner, H. G. (2010). Exome sequencing



- identifies wdr35 variants involved in sensenbrenner syndrome. *The American Journal of Human Genetics*, 87(3) :418 – 423.
- Gilissen, C., Hehir-Kwa, J., Thung, D., van, de Vorst, M., van Bon, B., Willemsen, M., Kwint, M., Janssen, I., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H., de Vries, B., Kleefstra, T., Brunner, H., Vissers, L., and Veltman, J. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature*, 511 :344–347.
- Gnrke, A., Melnikov, A., Maguire, J., Rogov, P., Le Proust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., and Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol.*, 27 :182–189.
- Gu, W., Zhang, F., and Lupski, J. (2008). Mechanisms for human genomic rearrangements. *PathoGenetics*, 1(1) :4.
- Hadjikhani, N., Zürcher, N., Rogier, O., Ruest, T., Hippolyte, L., Ben-Ari, Y., and Lemonnier, E. (2013). Improving emotional face perception in autism with diuretic bumetanide : A proof-of-concept behavioral and functional brain imaging pilot study. *Autism*.
- Hamdan, F., Gauthier, J., Araki, Y., Lin, D.-T., Yoshizawa, Y., Higashi, K., Park, A.-R., Spiegelman, D., Dobrzeniecka, S., Piton, A., Tomitori, H., Daoud, H., Massicotte, C., Henrion, E., Diallo, O., Shekarabi, M., Marineau, C., Shevell, M., Maranda, B., Mitchell, G., Nadeau, A., D'Anjou, G., Vanasse, M., Srouf, M., Lafrenière, R., Drapeau, P., Lacaille, J., Kim, E., Lee, J.-R., Igarashi, K., Hukanir, R., Rouleau, G., and Michaud, J. (2011). Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *The American Journal of Human Genetics*, 88(3) :306 – 316.
- Hamdan, F. F., Gauthier, J., Spiegelman, D., Noreau, A., Yang, Y., Pellerin, S., Dobrzeniecka, S., Côté, M., Perreault-Linck, E., Carmant, L., D'Anjou, G., Fombonne, E., Addington, A. M., Rapoport, J. L., Delisi, L. E., Krebs, M.-O., Mouaffak, F., Joober, R., Mottron, L., Drapeau, P., Marineau, C., Lafrenière, R. G., Lacaille, J. C., Rouleau, G. A., and Michaud, J. L. (2009). Mutations in syngap1 in autosomal nonsyndromic mental retardation. *New England Journal of Medicine*, 360(6) :599–605. PMID : 19196676.
- Han, S., Tai, C., Jones, C., Scheuer, T., and Catterall, W. (2014). Enhancement of inhibitory neurotransmission by gabaa receptors having alpha2,3-subunits ameliorates behavioral deficits in a mouse model of autism. *Neuron*, 81(6) :1282 – 1289.
- Hashimoto, S., Boissel, S., Zarhrate, M., Rio, M., Munnich, A., Egly, J.-M., and Colleaux, L. (2011). Med23 mutation links intellectual disability to dysregulation of immediate early gene expression. *Science*, 333(6046) :1161–1163.
- Hata, K., Nishimura, R., Muramatsu, S., Matsuda, A., Matsubara, T., Amano, K., Ikeda, F., Harley, V. R., and Yoneda, T. (2008). Paraspeckle protein p54nrb links sox9-mediated transcription with rna processing during chondrogenesis in mice. *The Journal of Clinical Investigation*, 118(9) :3098–3108.
- Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., van de Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L., Abajian, C., Beckmann, C. F., Bernard, A., Bertagnolli, D., Boe, A. F., Cartagena, P. M., Chakravarty, M. M., Chapin, M., Chong, J., Dalley, R. A., Daly, B. D., Dang, C., Datta, S., Dee, N., Dolbeare, T. A., Faber, V., Feng, D., Fowler, D. R., Goldy, J., Gregor, B. W., Haradon, Z., Haynor, D. R., Hohmann, J. G., Horvath, S., Howard, R. E., Jeromin, A., Jochim, J. M., Kinnunen, M., Lau, C., Lazarz, E. T., Lee, C., Lemon, T. A., Li, L., Li, Y., Morris, J. A., Overly, C. C., Parker, P. D., Parry, S. E., Reding, M., Royall, J. J., Schulkin, J., Sequeira, P. A., Slaughterbeck, C. R., Smith, S. C., Sodt, A. J., Sunkin, S. M., Swanson, B. E.,

- Vawter, M. P., Williams, D., Wohnoutka, P., Zielke, H. R., Geschwind, D. H., Hof, P. R., Smith, S. M., Koch, C., Grant, S. G., and Jones, A. R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489 :391–399.
- He, Q., Nomura, T., Xu, J., and Contractor, A. (2014). The developmental switch in gaba polarity is delayed in fragile x mice. *J Neurosci.*, 34 :446–450.
- Heikura, U., Taanila, A., Hartikainen, A.-L., Olsen, P., Linna, S.-L., Wendt, L. v., and Järvelin, M.-R. (2008). Variations in prenatal sociodemographic factors associated with intellectual disability : A study of the 20-year interval between two birth cohorts in northern finland. *American Journal of Epidemiology*, 167(2) :169–177.
- Heulens, I., D’Hulst, C., Dam, D. V., Deyn, P. P. D., and Kooy, R. F. (2012). Pharmacological treatment of fragile x syndrome with {GABAergic} drugs in a knockout mouse model. *Behavioural Brain Research*, 229(1) :244 – 249.
- Higgins, J. J., Pucilowska, J., Lombardi, R. Q., and P, R. J. (2004). A mutation in a novel atp-dependent lon protease gene in a kindred with mild mental retardation. *Neurology*, 63 :1927–1931.
- Hiramoto, T., Nakanishi, T., Sumiyoshi, T., Fukuda, T., Matsuura, S., Tauchi, H., Komatsu, K., Shibasaki, Y., Inui, H., Watatani, M., Yasutomi, M., Sumii, K., Kajiyama, G., Kamada, N., Miyagawa, K., and Kamiya, K. (1999). Mutations of a novel human rad54 homologue, rad54b, in primary cancer. *Oncogene*, 18 :3422–3426.
- Hirata, H., Nanda, I., van Riesen, A., McMichael, G., Hu, H., Hambrock, M., Papon, M.-A., Fischer, U., Marouillat, S., Ding, C., Alirol, S., Bienek, M., Preisler-Adams, S., Grimme, A., Seelow, D., Webster, R., Haan, E., MacLennan, A., Stenzel, W., Yap, T., Gardner, A., Nguyen, L. S., Shaw, M., Lebrun, N., Haas, S., Kress, W., Haaf, T., Schellenberger, E., Chelly, J., Viot, G., Shaffer, L., Rosenfeld, J., Kramer, N., Falk, R., El-Khechen, D., Escobar, L., Hennekam, R., Wieacker, P., Hübner, C., Ropers, H.-H., Gecz, J., Schuelke, M., Laumonnier, F., and Kalscheuer, V. (2013). Zc4h2 mutations are associated with arthrogryposis multiplex congenita and intellectual disability through impairment of central and peripheral synaptic plasticity. *The American Journal of Human Genetics*, 92(5) :681 – 695.
- Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., Yokoi, T., Nakagawa, S., Bénard, M., Fox, A. H., and Pierron, G. (2014). Neat1 long noncoding rna regulates transcription via protein sequestration within subnuclear bodies. *Molecular Biology of the Cell*, 25(1) :169–183.
- Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., Middle, C. M., Rodesch, M. J., Albert, T. J., Hannon, G. J., and McCombie, W. R. (2007). Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, 39 :1522–1527.
- Hoischen, A., van Bon, B. W., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., de Vries, P., de Reuver, R., Wieskamp, N., Mortier, G., Devriendt, K., Amorim, M. Z., Revencu, N., Kidd, A., Barbosa, M., Turner, A., Smith, J., Oley, C., Henderson, A., Hayes, I. M., Thompson, E. M., Brunner, H. G., de Vries, B. B., and Veltman, J. A. (2010). De novo mutations of setbp1 cause schinzel-giedion syndrome. *Nature Genetics*, 42 :483–485.
- Hoyer, J., Ekici, A., Ende, S., Popp, B., Zweier, C., Wiesener, A., Wohleber, E., Dufke, A., Rossier, E., Petsch, C., Zweier, M., Göhring, I., Zink, A., Rappold, G., Schröck, E., Wiczorek, D., Riess, O., Engels, H., Rauch, A., and Reis, A. (2012). Haploinsufficiency of arid1b, a member of the swi/snf-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *The American Journal of Human Genetics*, 90(3) :565 – 572.
- Irons, M., Elias, E. R., Salen, G., Tint, G. S., and K, B. A. (1993). Defective cholesterol biosynthesis in smith-lemli-opitz syndrome. *Lancet*, 341 :1414.

- Ishiguro, H., Uemura, H., Fujinami, K., Ikeda, N., Ohta, S., and Kubota, Y. (2003). 55 kda nuclear matrix protein (nmt55) mrna is expressed in human prostate cancer tissue and is associated with the androgen receptor. *International Journal of Cancer*, 105(1) :26–32.
- Jiang, Y.-H., Yuen, R., Jin, X., Wang, M., Chen, N., Wu, X., Ju, J., Mei, J., Shi, Y., He, M., Wang, G., Liang, J., Wang, Z., Cao, D., Carter, M., Chrysler, C., Drmic, I., Howe, J., Lau, L., Marshall, C., Merico, D., Nalpathamkalam, T., Thiruvahindrapuram, B., Thompson, A., Uddin, M., Walker, S., Luo, J., Anagnostou, E., Zwaigenbaum, L., Ring, R., Wang, J., Lajonchere, C., Wang, J., Shih, A., Szatmari, P., Yang, H., Dawson, G., Li, Y., and Scherer, S. (2013). Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *The American Journal of Human Genetics*, 93(2) :249 – 263.
- Johnson, J. O., Mandrioli, J., Benatar, M., Abramzon, Y., Deerlin, V. M. V., Trojanowski, J. Q., Gibbs, J. R., Brunetti, M., Gronka, S., Wu, J., Ding, J., McCluskey, L., Martinez-Lage, M., Falcone, D., Hernandez, D. G., Arepalli, S., Chong, S., Schymick, J. C., Rothstein, J., Landi, F., Wang, Y.-D., Calvo, A., Mora, G., Sabatelli, M., Monsurro, M. R., Battistini, S., Salvi, F., Spataro, R., Sola, P., Borghero, G., Galassi, G., Scholz, S. W., Taylor, J. P., Restagno, G., Chio, A., and Traynor, B. J. (2010). Exome sequencing reveals vcp mutations as a cause of familial als. *Neuron*, 68(5) :857 – 864.
- Jones, S., Hruban, R. H., Kamiyama, M., Borges, M., Zhang, X., Parsons, D. W., Lin, J. C., Palmisano, E., Brune, K., Jaffee, E. M., Jacobuzio-Donahue, C. A., Maitra, A., Parmigiani, G., Kern, S. E., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Eshleman, J. R., Goggins, M., and Klein, A. P. (2009). Exomic sequencing identifies palb2 as a pancreatic cancer susceptibility gene. *Science*, 324 :217.
- Jones, S. J. (1995). An update and lessons from whole-genome sequencing projects. *Current Opinion in Genetics & Development*, 5(3) :349 – 353.
- Joung, J. K. and Sander, J. D. (2013). Talens : a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol.*, 14 :49–55.
- Kim, N., Kim, P., Nam, S., Shin, S., and Lee, S. (2006). Chimerdb—a knowledgebase for fusion sequences. *Nucleic Acids Research*, 34(suppl 1) :D21–D24.
- Kishino, T., Lalonde, M., and Wagstaff, J. (1997). Ube3a/e6-ap mutations cause angelman syndrome. *Nature Genetics*, 15 :70–73.
- Kousar, R., Hassan, M., Khan, B., Basit, S., Mahmood, S., Mir, A., Ahmad, W., and Ansar, M. (2011). Mutations in wdr62 gene in pakistani families with autosomal recessive primary microcephaly. *BMC Neurology*, 11(1) :119.
- Kowalska, E., Ripperger, J. A., Muheim, C., Maier, B., Kurihara, Y., Fox, A. H., Kramer, A., and Brown, S. A. (2012). Distinct roles of dbhs family members in the circadian transcriptional feedback loop. *Molecular and Cellular Biology*, 32(22) :4585–4594.
- Krawitz, P. M., Schweiger, M. R., Rödelberger, C., Marcelis, C., Kölsch, U., Meisel, C., Stephani, F., Kinoshita, T., Murakami, Y., Bauer, S., Isau, M., Fischer, A., Dahl, A., Kerick, M., Hecht, J., Köhler, S., Jäger, M., Grünhagen, J., de Condor, B. J., Doelken, S., Brunner, H. G., Meinecke, P., Passarge, E., Thompson, M. D., Cole, D. E., Horn, D., Roscioli, T., Mundlos, S., and Robinson, P. N. (2010). Identity-by-descent filtering of exome sequence data identifies pigv mutations in hyperphosphatasia mental retardation syndrome. *Nature Genetics*, 42 :827–829.
- Kulman, G., Lissoni, P., Rovelli, F., Roselli, M. G., Brivio, F., and Sequeri, P. (2000). Evidence of pineal endocrine hypofunction in autistic children. *Neuro Endocrinol Lett.*, 21 :31–34.

- Kuwahara, S., Ikei, A., Taguchi, Y., Tabuchi, Y., Fujimoto, N., Obinata, M., Uesugi, S., and Kurihara, Y. (2006). Pspc1, nono, and sfpq are expressed in mouse sertoli cells and may function as coregulators of androgen receptor-mediated transcription. *Biology of Reproduction*, 75(3) :352–359.
- Lalonde, E., Albrecht, S., Ha, K. C., Jacob, K., Bolduc, N., Polychronakos, C., Dechelotte, P., Majewski, J., and Jabado, N. (2010). Unexpected allelic heterogeneity and spectrum of mutations in fowler syndrome revealed by next-generation exome sequencing. *Human Mutation*, 31(8) :918–923.
- Lancaster, M. A., Renner, M., Martin, C. A., Wenzel, D., Bicknell, L. S., Hurles, M. E., Homfray, T., Penninger, J. M., Jackson, A. P., and Knoblich, J. (2013). Cerebral organoids model human brain development and microcephaly. *Nature*, 501 :373–379.
- Lander, E. S. and Botstein, D. (1987). Homozygosity mapping : a way to map human recessive traits with the dna of inbred children. *Science*, 236 :1567–1570.
- Langouët, M., Saadi, A., Rieunier, G., Moutton, S., Siquier-Pernet, K., Fernet, M., Nitschke, P., Munnich, A., Stern, M.-H., Chaouch, M., and Colleaux, L. (2013). Mutation in *tti2* reveals a role for triple t complex in human brain development. *Human Mutation*, 34(11) :1472–1476.
- Lejeune, J., Turpin, R., and Gautier, M. (1959). Mongolism ; a chromosomal disease (trisomy). *Bulletin de l'Académie nationale de médecine*, 143 :256–265.
- Lemonnier, E., Degrez, C., Phelep, M., Tyzio, R., Josse, F., Grandgeorge, M., Hadjikhani, N., and Ben-Ari, Y. (2012). A randomised controlled trial of bumetanide in the treatment of autism in children. *Transl Psychiatry*, 2.
- Lewis, D., Cho, R., Carter, C., Eklund, K., Forster, S., Kelly, M., and Montrose, D. (2008). Subunit-selective modulation of gaba type a receptor neurotransmission and cognition in schizophrenia. *Am J Psychiatry*, 165 :1585–1593.
- Li, M., Suzuki, K., Kim, N. Y., Liu, G.-H., and Izpisua Belmonte, J. C. (2014a). A cut above the rest : Targeted genome editing technologies in human pluripotent stem cells. *Journal of Biological Chemistry*, 289(8) :4594–4599.
- Li, S., Li, Z., Shu, F. J., Xiong, H., Phillips, A. C., and Dynan, W. S. (2014b). Double-strand break repair deficiency in nono knockout murine embryonic fibroblasts and compensation by spontaneous upregulation of the pspc1 paralog. *Nucleic Acids Res.*, 6.
- Lim, D. H. and Maher, E. R. (2009). Human imprinting syndromes. *Epigenomics*, 1 :347–369.
- Liston, C., Cichon, J. M., Jeanneteau, F., Jia, Z., Chao, M. V., and Gan, W. B. (2013). Circadian glucocorticoid oscillations promote learning-dependent synapse formation and maintenance. *Nature Neuroscience*, 16 :698–705.
- Liu, C., Srihari, S., Cao, K.-A. L., Chenevix-Trench, G., Simpson, P. T., Ragan, M. A., and Khanna, K. K. (2014a). A fine-scale dissection of the dna double-strand break repair machinery and its implications for breast cancer therapy. *Nucleic Acids Research*, 42(10) :6106–6127.
- Liu, P. Y., Erriquez, D., Marshall, G. M., Tee, A. E., Polly, P., Wong, M., Liu, B., Bell, J. L., Zhang, X. D., Milazzo, G., Cheung, B. B., Fox, A., Swarbrick, A., Hüttelmaier, S., Kavallaris, M., Perini, G., Mattick, J. S., Dinger, M. E., and Liu, T. (2014b). Effects of a novel long noncoding rna, *Incusmycn*, on n-myc expression and neuroblastoma progression. *Journal of the National Cancer Institute*, 106(7).
- Mabb, A. M., Judson, M. C., Zylka, M. J., and Philpot, B. D. (2011). Angelman syndrome : insights into genomic imprinting and neurodevelopmental phenotypes. *Trends in Neurosciences*, 34(6) :293 – 303.

- Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., and Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nat Methods.*, 7 :111–118.
- Mandel, J. L. and Heitz, D. (1992). Molecular genetics of the fragile-x syndrome : a novel type of unstable mutation. *Curr Opin Genet Dev*, 2 :422–430.
- Martins-Taylor, K., Hsiao, J. S., Chen, P.-F., Glatt-Deeley, H., De Smith, A. J., Blakemore, A. I., Lalande, M., and Chamberlain, S. J. (2014). Imprinted expression of ube3a in non-neuronal cells from a prader-willi syndrome patient with an atypical deletion. *Human Molecular Genetics*, 23(9) :2364–2373.
- Martinsson, T., Bjursell, C., Stibler, H., Kristiansson, B., Skovby, F., Jaeken, J., Blennow, G., Strömme, P., Hanefeld, F., and Wahlström, J. (1994). Linkage of a locus for carbohydrate-deficient glycoprotein syndrome type i (cdg1) to chromosome 16p, and linkage disequilibrium to microsatellite marker d16s406. *Human Molecular Genetics*, 3 :2037–2042.
- Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing dna. *Proc Natl Acad Sci U S A*, 74 :560–564.
- McAllister, C. J., Whittington, J. E., and Holland, A. J. (2011). Development of the eating behaviour in prader-willi syndrome : advances in our understanding. *International journal of obesity*, 35 :188–197.
- McDermott, S., Wu, J., Cai, B., Lawson, A., and Aelion, C. M. (2011). Probability of intellectual disability is associated with soil concentrations of arsenic and lead. *Chemosphere*, 84(1) :31 – 38.
- McDonell, L., Warman Chardon, J., Schwartzentruber, J., Foster, D., Beaulieu, C., Consortium, F. C., Majewski, J., Bulman, D., and Boycott, K. (2014). The utility of exome sequencing for genetic diagnosis in a familial microcephaly epilepsy syndrome. *BMC Neurology*, 14(1) :22.
- Mehnert, K. I., Beramendi, A., Elghazali, F., Negro, P., Kyriacou, C. P., and Cantera, R. (2007). Circadian changes in drosophila motor terminals. *Developmental Neurobiology*, 67(4) :415–421.
- Melke, J., Goubran, Botros, H., Chaste, P., Betancur, C., Nygren, G., Anckarsäter, H., Rastam, M., Stahlberg, O., Gillberg, I., Delorme, R., Chabane, N., Mouren-Simeoni, M., Fauchereau, F., Durand, C., Chevalier, F., Drouot, X., Collet, C., Launay, J., Leboyer, M., Gillberg, C., and Bourgeron, T. (2008). Abnormal melatonin synthesis in autism spectrum disorders. *Mol Psychiatry*, 13 :90–98.
- Memon, M., Raza, S., Basit, S., Kousar, R., Ahmad, W., and Ansar, M. (2013). A novel wdr62 mutation causes primary microcephaly in a pakistani family. *Molecular Biology Reports*, 40(1) :591–595.
- Mendola, P., Selevan, S. G., Gutter, S., and Rice, D. (2002). Environmental factors associated with a spectrum of neurodevelopmental deficits. *Mental retardation and developmental disabilities research reviews*, 8 :188–197.
- Michaelson, J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., Wu, W., Corominas, R., Peoples, A., Koren, A., Gore, A., Kang, S., Lin, G., Estabillio, J., Gadomski, T., Singh, B., Zhang, K., Akshoomoff, N., Corsello, C., McCarroll, S., lakoucheva, L., Li, Y., Wang, J., and Sebat, J. (2012). Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell*, 151(7) :1431 – 1442.
- Min, M. O., Singer, L. T., Kirchner, H. L., Minnes, S., Short, E., Hussain, Z., and Nelson, S. (2009). Cognitive development and low-level lead exposure in poly-drug exposed children. *Neurotoxicology and Teratology*, 31(4) :225 – 231.

- Mir, A., Kaufman, L., Noor, A., Motazacker, M. M., Jamil, T., Azam, M., Kahrizi, K., Rafiq, M. A., Weksberg, R., Nasr, T., Naeem, F., Tzschach, A., Kuss, A. W., Ishak, G. E., Doherty, D., Ropers, H. H., Barkovich, A. J., Najmabadi, H., Ayub, M., and Vincent, J. B. (2009). Identification of mutations in *trappc9*, which encodes the *nik*- and *ikk*-beta-binding protein, in nonsyndromic autosomal-recessive mental retardation. *American Journal of Human Genetics*, 85 :909–915.
- Miyamoto, K., Sakurai, H., and Sugiura, T. (2008). Proteomic identification of a *psf/p54nrb* heterodimer as *rnf43* oncoprotein-interacting proteins. *PROTEOMICS*, 8(14) :2907–2910.
- Mochida, G., Mahajnah, M., Hill, A., Basel-Vanagaite, L., Gleason, D., Hill, R., Bodell, A., Crosier, M., Straussberg, R., and Walsh, C. (2009). A truncating mutation of *trappc9* is associated with autosomal-recessive intellectual disability and postnatal microcephaly. *American Journal of Human Genetics*, 85 :897–902.
- Modi, K. K., Jana, A., Ghosh, S., Watson, R., and Pahan, K. (2014). A physically-modified saline suppresses neuronal apoptosis, attenuates tau phosphorylation and protects memory in an animal model of alzheimer's disease. *PLoS One*.
- Molinari, F., Foulquier, F., Tarpey, P. S., Morelle, W., Boissel, S., Teague, J., Edkins, S., Futreal, P., Stratton, M. R., Turner, G., Matthijs, G., Gecz, J., Munnich, A., and Colleaux, L. (2008). Oligosaccharyltransferase-subunit mutations in nonsyndromic mental retardation. *American Journal of Human Genetics*, 82 :1150–1157.
- Molinari, F., Rio, M., Meskenaite, V., Encha-Razavi, F., Augé, J., Bacq, D., Briault, S., Vekemans, M., Munnich, A., Attié-Bitach, T., Sonderegger, P., and Colleaux, L. (2002). Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation. *Science*, 298(5599) :1779–1781.
- Mondal, K., Shetty, A. C., Patel, V., Cutler, D. J., and Zwick, M. E. (2011). Targeted sequencing of the human x chromosome exome. *Genomics*, 98(4) :260 – 265. New Genomic Technologies and Applications.
- Morton, N. E. (1955). Sequential tests for the detection of linkage. *Am J Hum Genet*, 7 :277–318.
- Morton, N. E. (1978). Effects of inbreeding on iq and mental retardation. *Proc Natl Acad Sci U S A*, 75 :3906–3908.
- Motazacker, M. M., Rost, B. R., Hucho, T., Garshasbi, M., Kahrizi, K., Ullmann, R., Abedini, S. S., Nieh, S. E., Amini, S. H., Goswami, C., Tzschach, A., Jensen, L. R., Schmitz, D., Ropers, H. H., Najmabadi, H., and Kuss, A. W. (2007). A defect in the ionotropic glutamate receptor 6 gene (*grik2*) is associated with autosomal recessive mental retardation. *American Journal of Human Genetics*, 81 :792–798.
- Murdock, D. R., Clark, G. D., Bainbridge, M. N., Newsham, I., Wu, Y.-Q., Muzny, D. M., Cheung, S. W., Gibbs, R. A., and Ramocki, M. B. (2011). Whole-exome sequencing identifies compound heterozygous mutations in *wdr62* in siblings with recurrent polymicrogyria. *American Journal of Medical Genetics Part A*, 155(9) :2071–2077.
- Musunuru, K., Pirruccello, J. P., Do, R., Peloso, G. M., Guiducci, C., Sougnez, C., Garimella, K. V., Fisher, S., Abreu, J., Barry, A. J., Fennell, T., Banks, E., Ambrogio, L., Cibulskis, K., Kernytsky, A., Gonzalez, E., Rudzicz, N., Engert, J. C., DePristo, M. A., Daly, M. J., Cohen, J. C., Hobbs, H. H., Altshuler, D., Schonfeld, G., Gabriel, S. B., Yue, P., and Kathiresan, S. (2010). Exome sequencing, *angptl3* mutations, and familial combined hypolipidemia. *New England Journal of Medicine*, 363(23) :2220–2227. PMID : 20942659.
- Mwaniki, M. K., Atieno, M., Lawn, J. E., and Newton, C. R. (2012). Long-term neurodevelopmental outcomes after intrauterine and neonatal insults : a systematic review. *The Lancet*, 379(9814) :445 – 452.

- Myojin, R., Kuwahara, S., Yasaki, T., Matsunaga, T., Sakurai, T., Kimura, M., Uesugi, S., and Kurihara, Y. (2004). Expression and functional significance of mouse paraspeckle protein 1 on spermatogenesis. *Biology of Reproduction*, 71(3) :926–932.
- Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S. S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P., Zecha, A., Mohseni, M., Püttmann, L., Vahid, L. N., Jensen, C., Moheb, L. A., Bienek, M., Larti, F., Mueller, I., Weissmann, R., Darvish, H., Wrogemann, K., Hadavi, V., Lipkowitz, B., Esmaeeli-Nieh, S., Wiczorek, D., Kariminejad, R., Firouzabadi, S. G., Cohen, M., Fattahi, Z., Rost, I., Mojahedi, F., Hertzberg, C., Dehghan, A., Rajab, A., Banavandi, M. J. S., Hoffer, J., Falah, M., Musante, L., Kalscheuer, V., Ullmann, R., Kuss, A. W., Tzschach, A., Kahrizi, K., and Ropers, H. H. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*, 478 :57–63.
- Nakagawa, S. and Hirose, T. (2012). Paraspeckle nuclear bodies : useful uselessness? *Cellular and Molecular Life Sciences*, 69(18) :3027–3036.
- Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., Lee, C., Turner, E. H., Smith, J. D., Rieder, M. J., Yoshiura, K., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D. A., Bamshad, M. J., and Shendure, J. (2010a). Exome sequencing identifies mll2 mutations as a cause of kabuki syndrome. *Nature Genetics*, 42 :790–793.
- Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., J, S., and Bamshad, M. J. (2010b). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*, 42 :30–35.
- Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A., and Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461 :272–276.
- Nicholas, A. K., Khurshid, M., Désir, J., Carvalho, O. P., Cox, J. J., Thornton, G., Kausar, R., Ansar, M., Ahmad, W., Verloes, A., Passemard, S., Misson, J. P., Lindsay, S., Gergely, F., Dobyns, W. B., Roberts, E., Abramowicz, M., and Woods, C. G. (2010). Wdr62 is associated with the spindle pole and is mutated in human microcephaly. *Nature G*, 42 :1010–1014.
- Nir, I., Meir, D., Zilber, N., Knobler, H., Hadjez, J., and Lerner, Y. (1995). Brief report : circadian melatonin, thyroid-stimulating hormone, prolactin, and cortisol levels in serum of young adults with autism. *J Autism Dev Disord.*, 25 :641–654.
- Novorol, C., Burkhardt, J., Wood, K. J., Iqbal, A., Roque, C., Coutts, N., Almeida, A. D., He, J., Wilkinson, C. J., and Harris, W. A. (2013). Microcephaly models in the developing zebrafish retinal neuroepithelium point to an underlying defect in metaphase progression. *Open Biology*, 3(10).
- O’Driscoll, M. and Jeggo, P. A. (2008). The role of the dna damage response pathways in brain development and microcephaly : Insight from human disorders. *DNA Repair*, 7(7) :1039 – 1050. {DNA} Damage Responses and Neurological Disease.
- Okou, D. T., Steinberg, K. M., Middle, C., J, C. D., J, A. T., and Zwick, M. E. (2007). Microarray-based genomic selection for high-throughput resequencing. *Nat Methods.*, 11 :907–909.
- Olmos-Serrano, J. L., Corbin, J. G., and P, B. M. (2011). The gaba(a) receptor agonist thip ameliorates specific behavioral deficits in the mouse model of fragile x syndrome. *Dev Neurosci*, 33 :395–403.

- Pak, C., Garshasbi, M., Kahrizi, K., Gross, C., Apponi, L. H., Noto, J. J., Kelly, S. M., Leung, S. W., Tzschach, A., Behjati, F., Abedini, S. S., Mohseni, M., Jensen, L. R., Hu, H., Huang, B., Stahley, S. N., Liu, G., Williams, K. R., Burdick, S., Feng, Y., Sanyal, S., Bassell, G. J., Ropers, H.-H., Najmabadi, H., Corbett, A. H., Moberg, K. H., and Kuss, A. W. (2011). Mutation of the conserved polyadenosine rna binding protein, *zc3h14/dnab2*, impairs neural function in drosophila and humans. *Proceedings of the National Academy of Sciences*, 108(30) :12390–12395.
- Park, K. H., Park, H. J., Shin, K. S., and Lee, M. K. (2014). Multiple treatments with l-3,4-dihydroxyphenylalanine modulate dopamine biosynthesis and neurotoxicity through the protein kinase a-transient extracellular signal-regulated kinase and exchange protein activation by cyclic amp-sustained extracellular signal-regulated kinase signaling pathways. *Journal of Neuroscience Research*.
- Park, Y., Lee, J.-M., Hwang, M.-Y., Son, G.-h., and Geum, D. (2013). Nono binds to the cpg island of oct4 promoter and functions as a transcriptional activator of oct4 gene expression. *Molecules and Cells*, 35(1) :61–69.
- Passon, D. M., Lee, M., Rackham, O., Stanley, W. A., Sadowska, A., Filipovska, A., Fox, A. H., and Bond, C. S. (2012). Structure of the heterodimer of human nono and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. *Proceedings of the National Academy of Sciences*, 109(13) :4846–4850.
- Pavao, M., Huang, Y.-H., Hafer, L., Moreland, R., and Traish, A. (2001). Immunodetection of nmt55/p54nrb isoforms in human breast cancer. *BMC Cancer*, 1(1) :15.
- Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V., Lee, Y. L., Guschin, D. Y., Rupniewski, I., Waite, A. J., Carpenito, C., Carroll, R. G., Orange, J. S., Urnov, F. D., Rebar, E. J., Ando, D., Gregory, P. D., Riley, J. L., Holmes, M. C., and June, C. H. (2008). Establishment of hiv-1 resistance in cd4+ t cells by genome editing using zinc-finger nucleases. *Nat Biotechnol.*, 26 :808–816.
- Philippe, O., Rio, M., Carioux, A., Plaza, J., Guigue, P., Molinari, F., Boddaert, N., Bole-Feyssot, C., Nitschke, P., Smahi, A., Munnich, A., and Colleaux, L. (2009). Combination of linkage mapping and microarray-expression analysis identifies nf-kappab signaling defect as a cause of autosomal-recessive mental retardation. *American Journal of Human Genetics*, 85 :903–908.
- Poulton, C. J., Schot, R., Seufert, K., Lequin, M. H., Accogli, A., Annunzio, G. D., Villard, L., Philip, N., de Coo, R., Catsman-Berrevoets, C., Grasshoff, U., Kattentidt-Mouravieva, A., Calf, H., de Vreugt-Gronloh, E., van Unen, L., Verheijen, F. W., Galjart, N., Morris-Rosendahl, D. J., and Mancini, G. M. S. (2014). Severe presentation of *wdr62* mutation : Is there a role for modifying genetic factors? *American Journal of Medical Genetics Part A*, 164(9) :2161–2171.
- Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., Zhang, M. Q., and Spector, D. L. (2005). Regulating gene expression through rna nuclear retention. *Cell*, 123(2) :249 – 263.
- Pyza, E. and Gorska-Andrzejak, J. (2008). External and internal inputs affecting plasticity of dendrites and axons of the fly's neurons. *Acta Neurobiol Exp (Wars)*., 68 :322–333.
- Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., Zenker, M., Hüffmeier, U., Thiel, C., Rüschen-dorf, F., Nürnberg, P., Reis, A., and Trautmann, U. (2006). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *American Journal of Medical Genetics Part A*, 140A(19) :2063–2074.
- Rauch, A., Wiczorek, D., Graf, E., Wieland, T., Endeke, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Donato, N. D., Dufke, A., Cremer, K., Hempel, M., Horn, D., Hoyer, J., Joset, P., RÄ¶pke, A., Moog, U., Riess, A.,



- Thiel, C. T., Tzschach, A., Wiesener, A., Wohlleber, E., Zweier, C., Ekici, A. B., Zink, A. M., Rump, A., Meisinger, C., Grallert, H., Sticht, H., Schenck, A., Engels, H., Rappold, G., Schröck, E., Wieacker, P., Riess, O., Meitinger, T., Reis, A., and Strom, T. M. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability : an exome sequencing study. *The Lancet*, 380(9854) :1674 – 1682.
- Rawashdeh, O., de Borsetti, N. H., Roman, G., and Cahill, G. M. (2007). Melatonin suppresses nighttime memory formation in zebrafish. *Science*, 318(5853) :1144–1146.
- Raymond, F. L. and Tarpey, P. (2006). The genetics of mental retardation. *Human Molecular Genetics*, 15(suppl 2) :R110–R116.
- Raymond, F. L., Whibley, A., Stratton, M. R., and Gecz, J. (2009). Lessons learnt from large-scale exon re-sequencing of the x chromosome. *Human Molecular Genetics*, 18(R1) :R60–R64.
- Redin, C., Gérard, B., Lauer, J., Herenger, Y., Muller, J., Quartier, A., Masurel-Paulet, A., Willems, M., Lesca, G., El-Chehadeh, S., Le, Gras, S., Vicaire, S., Philipps, M., Dumas, M., Geoffroy, V., Feger, C., Haumesser, N., Alembik, Y., Barth, M., Bonneau, D., Colin, E., Dollfus, H., Doray, B., Delrue, M., Drouin-Garraud, V., Flori, E., Fradin, M., Francannet, C., Goldenberg, A., Lumbroso, S., Mathieu-Dramard, M., Martin-Coignard, D., Lacombe, D., Morin, G., Polge, A., Sukno, S., Thauvin-Robinet, C., Thevenon, J., Doco-Fenzy, M., Genevieve, D., Sarda, P., Ederly, P., Isidor, B., Jost, B., Olivier-Faivre, L., Mandel, J., and Piton, A. (2014). Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *Journal of Medical Genetics*.
- Regan, R and Willatt, L. (2010). Mental retardation : Definition, classification and etiology. *Monographs in Human Genetics*, 18 :16–30.
- Richardson, S. A. (1989). Issues in the definition of mental retardation and the representativeness of studies. *Research in Developmental Disabilities*, 10(3) :285 – 294.
- Ropers, H.-H. (2006). X-linked mental retardation : many genes for a complex disorder. *Current Opinion in Genetics & Development*, 16(3) :260 – 269. Genetics of disease Andrea Ballabio, David Nelson and Steve Rozen.
- Ropers, H. H. (2007). New perspectives for the elucidation of genetic disorders. *American Journal of Human Genetics*, 81 :199–207.
- Ropers, H. H. (2008). Genetics of intellectual disability. *Curr Opin Genet Dev.*, 18 :241–250.
- Ropers, H. H. (2010). Genetics of early onset cognitive impairment. *Annual Review of Genomics and Human Genetics*, 11(1) :161–187. PMID : 20822471.
- Ruby, N. F., Hwang, C. E., Wessells, C., Fernandez, F., Zhang, P., Sapolsky, R., and Heller, H. C. (2008). Hippocampal-dependent learning requires a functional circadian system. *Proceedings of the National Academy of Sciences*, 105(40) :15593–15598.
- Rulten, S. L. and Caldecott, K. W. (2013). Dna strand break repair and neurodegeneration. *DNA Repair*, 12(8) :558 – 567. Genome maintenance in the Nervous system.
- Rupp, V., Rauf, S., Naveed, I., Windpassinger, C., and Mir, A. (2014). A novel single base pair duplication in wdr62 causes primary microcephaly. *BMC Medical Genetics*, 15 :107.
- Russell, S. T. and Tisdale, M. J. (2012). Role of beta-adrenergic receptors in the anti-obesity and anti-diabetic effects of zinc-alpha2-glycoprotein (zag). *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1821(4) :590 – 599.

- Sajid Hussain, M., Marriam Bakhtiar, S., Farooq, M., Anjum, I., Janzen, E., Reza Toliat, M., Eiberg, H., Kjaer, K., Tommerup, N., Noegel, A., Nürnberg, P., Baig, S., and Hansen, L. (2013). Genetic heterogeneity in pakistani microcephaly families. *Clinical Genetics*, 83(5) :446–451.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). Dna sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74 :5463–5467.
- Sanlaville, D., Lapierre, J., Coquin, A., Turleau, C., Vermeesch, J., Colleaux, L., Borck, G., Vekemans, M., Aurias, A., and Romana, S. (2005). [microarray cgh : principle and use for constitutional disorders]. *Archives de Pédiatrie*, 12(10) :1515 – 1520.
- Sasaki, Y. T. F., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009). Menepsilon/beta noncoding rnas are essential for structural integrity of nuclear paraspeckles. *Proceedings of the National Academy of Sciences*, 106(8) :2525–2530.
- Schaffer, A., Eggens, V., Caglayan, A., Reuter, M., Scott, E., Coufal, N., Silhavy, J., Xue, Y., Kayserili, H., Yasuno, K., Rosti, R. O., Abdellateef, M., Caglar, C., Kasher, P., Cazemier, J., Weterman, M., Cantagrel, V., Cai, N., Zweier, C., Altunoglu, U., Satkin, N., Aktar, F., Tuysuz, B., Yalcinkaya, C., Caksen, H., Bilguvar, K., Fu, X.-D., Trotta, C. R., Gabriel, S., Reis, A., Gunel, M., Baas, F., and Gleeson, J. (2014). Clp1 founder mutation links trna splicing and maturation to cerebellar development and neurodegeneration. *Cell*, 157(3) :651 – 663.
- Schäffer, A. A. (2013). Digenic inheritance in medical genetics. *Journal of Medical Genetics*, 50(10) :641–652.
- Schiffner, S., Zimara, N., Schmid, R., and Bosserhoff, A.-K. (2011). p54nrb is a new regulator of progression of malignant melanoma. *Carcinogenesis*, 32(8) :1176–1182.
- Schmickel, R. D. (1986). Contiguous gene syndromes : a component of recognizable syndromes. *J Pediatr*, 109 :231–241.
- Segal, D. J. and Meckler, J. F. (2013). Genome engineering at the dawn of the golden age. *Annual Review of Genomics and Human Genetics*, 14(1) :135–158. PMID : 23701416.
- Shelkovnikova, T. A., Robinson, H. K., Troakes, C., Ninkina, N., and Buchman, V. L. (2014). Compromised paraspeckle formation as a pathogenic factor in fusopathies. *Human Molecular Genetics*, 23(9) :2298–2312.
- Shen, E. H., Overly, C. C., and Jones, A. R. (2012). The allen human brain atlas : Comprehensive gene expression mapping of the human brain. *Trends in Neurosciences*, 35(12) :711 – 714.
- Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001). dbsnp : the ncbi database of genetic variation. *Nucleic Acids Research*, 29(1) :308–311.
- Shoubridge, C., Tarpey, P. S., Abidi, F., Ramsden, S. L., Rujirabanjerd, S., Murphy, J. A., Boyle, J., Shaw, M., Gardner, A., Proos, A., Puusepp, H., Raymond, F. L., Schwartz, C. E., Stevenson, R. E., Turner, G., Field, M., Walikonis, R. S., Harvey, R. J., Hackett, A., Futreal, P. A., Stratton, M. R., and Gécz, J. (2010). Mutations in the guanine nucleotide exchange factor gene iqsec2 cause nonsyndromic intellectual disability. *Nature Genetics*, 42 :486–488.
- Silva, R., Falcao, A., Fernandes, A., Gordo, A., Brito, M., and Brites, D. (2006). Dissociated primary nerve cell cultures as models for assessment of neurotoxicity. *Toxicology Letters*, 163(1) :1 – 9.
- Srivastava, A. K. and Schwartz, C. E. (2014). Intellectual disability and autism spectrum disorders : Causal genes and molecular mechanisms. *Neuroscience and Biobehavioral Reviews*, 1927(0) :14.

- Stankiewicz, P. and Lupski, J. R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends in Genetics*, 18(2) :74 – 82.
- Stankiewicz, P. and Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annu Rev Med*, 61 :437–455.
- Stittrich, A., Lehman, A., Bodian, D., Ashworth, J., Zong, Z., Li, H., Lam, P., Khromykh, A., Iyer, R., Vockley, J., Baveja, R., Silva, E., Dixon, J., Leon, E., Solomon, B., Glusman, G., Niederhuber, J., Roach, J., and Patel, M. (2014). Mutations in notch1 cause adams-oliver syndrome. *American Journal of Human Genetics*, 95 :275–284.
- Sunwoo, H., Dinger, M. E., Wilusz, J. E., Amaral, P. P., Mattick, J. S., and Spector, D. L. (2009). Men epsilon/beta nuclear-retained non-coding rnas are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Research*, 19(3) :347–359.
- Suzuki, K., Yu, C., Qu, J., Li, M., Yao, X., Yuan, T., Goebel, A., Tang, S., Ren, R., Aizawa, E., Zhang, F., Xu, X., Soligalla, R., Chen, F., Kim, J., Kim, N., Liao, H.-K., Benner, C., Esteban, C., Jin, Y., Liu, G.-H., Li, Y., and Belmonte, J. I. (2014). Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell Stem Cell*, 15(1) :31 – 36.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5) :861 – 872.
- Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4) :663 – 676.
- Takahashi, N., Hashida, H., Zhao, N., Misumi, Y., and Sakaki, Y. (1995). High-density cDNA filter analysis of the expression profiles of the genes preferentially expressed in human brain. *Gene*, 164 :219–227.
- Tanaka, K., Hiramoto, T., Fukuda, T., and Miyagawa, K. (2000). A novel human rad54 homologue, rad54b, associates with rad51. *Journal of Biological Chemistry*, 275(34) :26316–26321.
- Tarabeux, J., Champagne, N., Brustein, E., Hamdan, F. F., Gauthier, J., Lapointe, M., Maios, C., Piton, A., Spiegelman, D., Henrion, E., Millet, B., Rapoport, J. L., DeLisi, L. E., Joobert, R., Fathalli, F., Fombonne, E., Mottron, L., Forget-Dubois, N., Boivin, M., Michaud, J. L., Lafrenière, R. G., Drapeau, P., Krebs, M.-O., and Rouleau, G. A. (2010). De novo truncating mutation in kinesin 17 associated with schizophrenia. *Biological Psychiatry*, 68(7) :649 – 656. New Evidence of Working Memory Dysfunction in Schizophrenia.
- Tarpey, P. S., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C., O'Meara, S., Latimer, C., Dicks, E., Menzies, A., Stephens, P., Blow, M., Greenman, C., Xue, Y., Tyler-Smith, C., Thompson, D., Gray, K., Andrews, J., Barthorpe, S., Buck, G., Cole, J., Dunmore, R., Jones, D., Maddison, M., Mironenko, T., Turner, R., Turrell, K., Varian, J., West, S., Widaa, S., Wray, P., Teague, J., Butler, A., Jenkinson, A., Jia, M., Richardson, D., Shepherd, R., Wooster, R., Tejada, M. I., Martinez, F., Carvill, G., Goliath, R., de Brouwer, A. P., van Bokhoven, H., Van Esch, H., Chelly, J., Raynaud, M., Ropers, H. H., Abidi, F. E., Srivastava, A. K., Cox, J., Luo, Y., Mallya, U., Moon, J., Parnau, J., Mohammed, S., Tolmie, J. L., Shoubridge, C., Corbett, M., Gardner, A., Haan, E., Rujirabanjerd, S., Shaw, M., Vandeleur, L., Fullston, T., Easton, D. F., Boyle, J., Partington, M., Hackett, A., Field, M., Skinner, C., Stevenson, R. E., Bobrow, M., Turner, G., Schwartz, C. E., Gecz, J., Raymond, F. L., Futreal, P. A., and Stratton, M. R. (2009). A systematic, large-scale resequencing screen of x-chromosome coding exons in mental retardation. *Nature Genetics*, 41 :535–543.

- Terns, R. M. and Terns, M. P. (2014). Crispr-based technologies : prokaryotic defense weapons repurposed. *Trends Genet.*, 30 :111–118.
- Tewhey, R., Warner, J. B., Nakano, M., Libby, B., Medkova, M., David, P. H., Kotsopoulos, S. K., Samuels, M. L., Hutchison, J. B., Larson, J. W., Topol, E. J., Weiner, M. P., Harismendy, O., Olson, J., Link, D. R., and Frazer, K. A. (2009). Microdroplet-based pcr enrichment for large-scale targeted sequencing. *Nat Biotechnol.*, 27 :1025–1031.
- Tordjman, S., Anderson, G. M., Pichard, N., Charbuy, H., and Touitou, Y. (2005). Nocturnal excretion of 6-sulphatoxymelatonin in children and adolescents with autistic disorder. *Biological Psychiatry*, 57(2) :134 – 138.
- Toydemir, R., Rutherford, A., Whitby, F., Jorde, L., Carey, J., and Bamshad, M. (2006). Mutations in embryonic myosin heavy chain (myh3) cause freeman-sheldon syndrome and sheldon-hall syndrome. *Nature Genetics*, 38 :561–565.
- Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J. H. J. (1992). Ercc6, a member of a subfamily of putative helicases, is involved in cockayne’s syndrome and preferential repair of active genes. *Cell*, 71(6) :939 – 953.
- Tsofack, S., Garand, C., Sereduk, C., Chow, D., Aziz, M., Guay, D., Yin, H., and Lebel, M. (2011). Nono and raly proteins are required for yb-1 oxaliplatin induced resistance in colon adenocarcinoma cell lines. *Molecular Cancer*, 10(1) :145.
- Tsurusaki, Y., Okamoto, N., Suzuki, Y., Doi, H., Saito, H., Miyake, N., and Matsumoto, N. (2011). Exome sequencing of two patients in a family with atypical x-linked leukodystrophy. *Clinical Genetics*, 80(2) :161–166.
- Turner, E. H., Lee, C., Ng, S. B., Nickerson, D. A., and Shendure, J. (2009). Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods*, 5 :315–316.
- Valnegri, P., Khelifaoui, M., Dorseuil, O., Bassani, S., Lagneaux, C., Gianfelice, A., Benfante, R., Chelly, J., Billuart, P., Sala, C., and Passafaro, M. (2011). A circadian clock in hippocampus is regulated by interaction between oligophrenin-1 and rev-erb alpha. *Nature Neuroscience*, 14 :1293–1301.
- van Bon, B. W., Oortveld, M. A., Nijtmans, L. G., Fenckova, M., Nijhof, B., Besseling, J., Vos, M., Kramer, J. M., de Leeuw, N., Castells-Nobau, A., Asztalos, L., Viragh, E., Ruiten, M., Hofmann, F., Eshuis, L., Collavin, L., Huynen, M. A., Asztalos, Z., Verstrecken, P., Rodenburg, R. J., Smeitink, J. A., de Vries, B. B., and Schenck, A. (2013). Cep99 is required for mitochondrial metabolism and neuronal function in man and fly. *Human Molecular Genetics*, 22(15) :3138–3151.
- Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y.-H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommen, G.-J. B., Blonden, L. A., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A., and Warren, S. T. (1991). Identification of a gene (fmr-1) containing a {CGG} repeat coincident with a breakpoint cluster region exhibiting length variation in fragile x syndrome. *Cell*, 65(5) :905 – 914.
- Verkerk, A. J., Schot, R., Dumeé, B., Schellekens, K., Swagemakers, S., Bertoli-Avella, A. M., Lequin, M. H., Dudink, J., Govaert, P., van Zwol, A., Hirst, J., Wessels, M. W., Catsman-Berrepoets, C., Verheijen, F. W., de Graaff, E., de Coo, I. F., Kros, J. M., Willemsen, R., Willems, P. J., van der Spek, P. J., and Mancini, G. M. (2009). Mutation in the ap4m1 gene provides a model for neuroaxonal injury in cerebral palsy. *The American Journal of Human Genetics*, 85(1) :40 – 52.

- Vissers, L. E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., van Bon, B. W., Hoischen, A., de Vries, B. B., Brunner, H. G., and Veltman, J. A. (2010). A de novo paradigm for mental retardation. *Nature Genetics*, 42 :1109–1112.
- Vissers, L. E., Van Ravenswaaji, C. M., Admiraal, R., Hurst, J. A., De Vries, B. B., Janssen, I. M., Van der Vliet, W. A., Huys, E. H., De Jong, P. J., Hamel, B. C., Schoenmakers, E. F., Brunner, H. G., Veltman, J. A., and Van Kessel, A. G. (2004). Mutations in a new member of the chromodomain gene family cause charge syndrome. *Nature Genetics*, 36 :955–957.
- Voituron, N. and Hilaire, G. (2011). The benzodiazepine midazolam mitigates the breathing defects of mecp2-deficient mice. *Respiratory Physiology & Neurobiology*, 177(1) :56 – 60.
- Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M. K., Thornton, A. M., Roeb, W., Rayyan, A. A., Loulus, S., Avraham, K. B., King, M.-C., and Kanaan, M. (2010). Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein gpsm2 as the cause of nonsyndromic hearing loss dfnb82. *The American Journal of Human Genetics*, 87(1) :90 – 94.
- Wang, J. L., Yang, X., Xia, K., Hu, Z. M., Weng, L., Jin, X., Jiang, H., Zhang, P., Shen, L., Feng Guo, J., li, N., Li, Y. R., Lei, L. F., Zhou, J., Du, J., Zhou, Y. F., Pan, Q., Wang, J., Wang, J., Li, R. Q., and Tang, B. S. (2010). Tgm6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain*, 133(12) :3510–3518.
- Wassif, C. A., Maslen, C., Kachilele-Linjewile, S., Lin, D., Linck, L. M., Connor, W. E., Steiner, R. D., and Porter, F. D. (1998). Mutations in the human sterol delta7-reductase gene at 11q12-13 cause smith-lemlie-opitz syndrome. *The American Journal of Human Genetics*, 63(1) :55 – 62.
- Waterham, H. R., Wijburg, F. A., Hennekam, R. C. M., Vreken, P., Poll-The, B. T., Dorland, L., Duran, M., Jira, P. E., Smeitink, J. A. M., Wevers, R. A., and Wanders, R. J. A. (1998). Smith-lemlie-opitz syndrome is caused by mutations in the 7-dehydrocholesterol reductase gene. *The American Journal of Human Genetics*, 63(2) :329 – 338.
- Watson, J. D. and Crick, F. H. (1953). The structure of dna. *Cold Spring Harb Symp Quant Biol*, 18 :123–131.
- Wesoly, J., Agarwal, S., Sigurdsson, S., Bussen, W., Van Komen, S., Qin, J., van Steeg, H., van Benthem, J., Wassenaar, E., Baarends, W. M., Ghazvini, M., Tafel, A. A., Heath, H., Galjart, N., Essers, J., Grootegoed, J. A., Arnheim, N., Bezzubova, O., Buerstedde, J.-M., Sung, P., and Kanaar, R. (2006). Differential contributions of mammalian rad54 paralogs to recombination, dna damage repair, and meiosis. *Molecular and Cellular Biology*, 26(3) :976–989.
- Williams, S., Zies, D., Mullegama, S., Grotewiel, M. S., and Elsea, S. (2012). Smith-magenis syndrome results in disruption of clock gene transcription and reveals an integral role for rai1 in the maintenance of circadian rhythmicity. *The American Journal of Human Genetics*, 90(6) :941 – 949.
- Xu, D., Zhang, F., Wang, Y., Sun, Y., and Xu, Z. (2014). Microcephaly-associated protein wdr62 regulates neurogenesis through jnk1 in the developing neocortex. *Cell Reports*, 6(1) :104 – 116.
- Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science*, 318(5858) :1917–1920.
- Yu, T. W., Mochida, G. H., Tischfield, D. J., Sgaier, S. K., Flores-Sarnat, L., Sergi, C. M., Topçu, M., McDonald, M. T., Barry, B. J., Felie, J. M., Sunu, C., Dobyns, W. B., Folkerth, R. D., Barkovich, A. J., and Walsh, C. A.

(2010). Mutations in *wdr62*, encoding a centrosome-associated protein, cause microcephaly with simplified gyri and abnormal cortical architecture. *Nature Genetics*, 42 :1015–1020.

Zhang, X. (2014). Exome sequencing greatly expedites the progressive research of mendelian diseases. *Front Med*, 8 :42–57.

Zweier, C., Peippo, M. M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., Reardon, W., Saraiva, J., Cabral, A., Göhring, I., Devriendt, K., de Ravel, T., Bijlsma, E. K., Hennekam, R. C. M., Orrico, A., Cohen, M., Dreweke, A., Reis, A., Nürnberg, P., and Rauch, A. (2007). Haploinsufficiency of *tcf4* causes syndromal mental retardation with intermittent hyperventilation (pitt-hopkins syndrome). *The American Journal of Human Genetics*, 80(5) :994 – 1001.