

L'approche basée sur le génotype déterminé par séquençage haut-débit en première intention et le partage international des données pour identifier de nouveaux gènes et nouveaux syndromes responsables d'anomalies du développement

Sébastien Moutton

▶ To cite this version:

Sébastien Moutton. L'approche basée sur le génotype déterminé par séquençage haut-débit en première intention et le partage international des données pour identifier de nouveaux gènes et nouveaux syndromes responsables d'anomalies du développement. Médecine humaine et pathologie. Université Bourgogne Franche-Comté, 2019. Français. NNT: 2019UBFCI013. tel-02477164

HAL Id: tel-02477164 https://theses.hal.science/tel-02477164

Submitted on 13 Feb 2020 $\,$

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Thèse de doctorat de l'Université Bourgogne Franche-Comté

École doctorale n°554 Environnement Santé Laboratoire Génétique des Anomalies du Développement INSERM U1231

Doctorat de génétique

Par Sébastien MOUTTON

L'approche basée sur le génotype déterminé par séquençage haut-débit en première intention et le partage international des données pour identifier de nouveaux gènes et nouveaux syndromes responsables d'anomalies du développement

Thèse présentée et soutenue à Dijon le 04/11/2019

Président du jury :	Pr Patrick Callier, PU-PH, U1231, Université Bourgogne Franche-Comté
Rapporteurs :	Dr Amélie PITON, Unité de génétique moléculaire, Hôpitaux Universitaire de Strasbourg, Université de Strasbourg
	Dr Christelle GOLZIO, Institut de Génétique et de Biologie Moléculaire et Cellulaire, U1258, Université de Strasbourg
Examinatrice :	Pr Catherine BOILEAU, PU-PH, U1148, Université Paris Diderot
Directrice de thèse :	Pr Laurence FAIVRE, PU-PH, U1231, Université Bourgogne Franche-Comté
Co-directrice de Thèse :	Pr Christel Thauvin-Robinet, PU-PH, U1231, Université Bourgogne Franche-Comté
Membres du jury invités:	Dr Laurence Jego, Ingénieur de recherche, HDR, U1231, Université Bourgogne Franche-Comté (co-encadrante)
	Dr Virginie Carmignac, Ingénieur de recherche, U1231, Université Bourgogne Franche-Comté (co-encadrante)

SOMMAIRE

REMERCIEMENTS
1. INTRODUCTION
1.1. Généralités et historique
1.2. La déficience intellectuelle : définition, classification, physiopathologie
1.2.1. Définition et évaluation11
1.2.2. Causes de la DI14
1.2.2.1. Facteurs environnementaux (10-13%)14
1.2.2.2. Causes génétiques14
1.2.2.3. DI idiopathique
1.2.3. Rôle des gènes dont les variations pathogènes sont responsables de DI
1.2.3.1. DI due à une anomalie de la fonction synaptique18
a. Gènes responsables de DI impliqués au niveau de l'élément pré-synaptique (figures 8 et
9)
al. Génes impliqués dans le trafic vésiculaire
a2. Génes impliqués dans la structure de la synapse
b. Gènes de DI impliqués au niveau de l'élément post-synaptique (figures 8 et 10) 22
b1. Gènes impliqués dans la structure de la synapse
b2. Gènes impliqués dans la réception du signal et codant pour les canaux ioniques 22
b3. Gènes impliqués dans la régulation du taux des protéines post-synaptiques
b4. Gènes impliqués dans le cytosquelette et sa dynamique
b5. Gènes impliqués dans les voies de signalisation intra-cellulaire
1.2.3.2. Régulation épigénétique de la plasticité synaptique
1.2.3.3. Autres voies biologiques et métaboliques impliquées dans la DI
a. Voie de biosynthèse du cholestérol
b. Maladies de surcharge lysosomale
c. Maladies peroxysomales
d. Maladies congénitales de la glycosylation
d1. Anomalie de la voie de la N-glycosylation
d2. Anomalie de la voie de la O-glycosylation
d3. Voie de synthèse du glycosylphosphatidylinositol (GPI) et glycosylation des lipides 36
1.2.3.4. DI syndromique avec malformation cérébrale
a. Holoprosencéphalie : voie SHH
b. Lissencéphalie ou agyrie et pachygyrie

c. Microcéphalie primaire (figure 5A)	39
d. Hétérotopies nodulaires péri-ventriculaires	39
e. Macrocéphalie/mégalencéphalie et polymicrogyrie	40
f. Schizencéphalie, porencéphalie	41
g. Hydrocéphalie	41
h. Anomalies en mosaïque - hémimégalencéphalie et dysplasie corticale focale	42
i. Anomalies de la fosse postérieure	42
i1. Hypoplasie vermienne	42
i2. Malformation de Dandy-Walker	43
i3. Hypoplasie ponto-cérébelleuse	43
j. Anomalies de la myéline et leucodystrophie	44
1.3. Autres pathologies et manifestations cliniques associées à la DI abordées dans le cadre de cette thèse	46
1.3.1. Les troubles du spectre autistique (TSA) : définition et causes	46
1.3.2. Epilepsie : définition et causes	48
1.3.3. Chevauchement des causes génétiques de pathologies neurodéveloppementales et de l'épilepsie	54
1.3.4. Syndrome de Marfan et syndromes marfanoïde	54
1.4. Intérêt des modèles animaux pour étudier les voies physiopathologiques de la DI	57
1.4.1. Modèle murin	57
1.4.1.1. Génome murin et outils utilisés pour la transgenèse	57
1.4.1.2. Application des modèles murins à la DI	59
a. Etude des protéines synaptiques	60
b. Modèle murin de syndrome de l'X fragile	61
c. Modèle murin de syndrome de Rett	61
1.4.2. Modèle drosophile	62
1.4.2.1. Moyens utilisables chez la drosophile pour modéliser la DI	62
1.4.2.2. Exemples de modèles drosophiles de maladies humaines	64
1.4.3. Modèle du poisson zèbre	64
1.4.3.1. Développement de l'espèce et comparaison avec le génome humain	64
1.4.3.2. Moyens utilisables pour manipuler le génome du poisson zèbre	66
1.4.3.3. Applications à la modélisation de la DI	66
1.4.4. Modèle Xenope	67
1.4.4.1. Développement et comparaison avec le génome humain	67
1.4.4.2. Manipulations génétiques et applications	67
1.4.5. Modèles animaux dans l'épilepsie	68

	1.4.6. Avancées thérapeutiques dans le champ de la DI	. 70
	1.4.7. Limites des modèles animaux	. 72
1.	5. Interprétation des données de séquençage haut-débit	. 73
	1.5.1. Principales bases de données utiles à l'interprétation des variants génétiques	. 73
	1.5.2. Vers une harmonisation et une reproductibilité dans l'interprétation des variations génétiques	. 74
	1.5.3. Stratégies classiques d'identification des gènes à l'aide des outils pangénomiques	. 75
	1.5.4. Stratégie pratique de lecture d'exome pour prioritiser les variants en diagnostic et en recherche et rendement des stratégies	. 79
	1.5.4.1. Stratégie dans le cadre du diagnostic	. 79
	1.5.4.2. Stratégie en recherche	. 84
	a. Réanalyse systématique des données en recherche	. 84
	b. Autres stratégies d'identification de nouveaux gènes en recherche	. 85
	b1. Réanalyse en trio	. 86
	b2. Etudes statistiques et d'agrégation de données d'une cohorte pour faire ressortir des récurrences mutationnelles	87
	b3. Comparaison des données d'une cohorte donnée à des bases de données existantes	. 88
	b4. Analyse de variants situés dans des gènes candidats spécifiquement ciblés après interrogation de la base de données d'une cohorte	. 90
	b5. Le recul sur l'analyse basée sur le phénotype	. 90
	1.5.5. Intérêt du partage de données inter-laboratoire	. 91
	1.5.6. De la stratégie phenotype-first vers le genotype-first (figure 38)	. 93
1.	6. Objectifs du travail de thèse	. 95
2. M	IETHODES	. 96
2.	1. Séquençage et interprétation	. 97
2.	2. Stratégie de réanalyse recherche	. 99
2.	3. Stratégie de partage de données1	100
3. RI	ESULTATS	102
3. in	1. Identification de mutations du gène <i>DLG4</i> comme responsables d'un phénotype de déficien tellectuelle - syndrome marfanoïde (article 1)1	ce 102
3. av	2. Implication du gène <i>KCNQ3</i> dans un nouveau phénotype comprenant déficience intellectuel vec épilepsie néonatale potentiellement pharmaco-dépendante (article 2)	lle 106
3. in	3. Etude de l'apport du séquençage d'exome en trio pour l'identification de nouveaux gènes npliqués dans les anomalies du développement (article 3)1	111
3. m	4. Implication de variants hétérozygotes perte de fonction du gène <i>KCNMA1</i> dans un syndrom alformatif avec déficit intellectuel et épilepsie (article 4)	ie 122
2	5 Participation aux autres travaux de l'équine	126

4. DISCUSSION / PERSPECTIVES
4.1. Synthèse du travail
4.2. Perspectives d'évolution des pratiques128
4.2.1. Evolution en diagnostic
4.2.2. Evolution en recherche
4.2.2.1. Séquençage du génome et regroupement de cohortes de patients atteints de syndromes caractérisés sans base moléculaire connue
4.2.2.2. Intérêt du RNA-seq130
4.2.2.3. Autres mécanismes à prendre en considération pour de futures perspectives 131
a. Altération de l'expression génique et/ou en perturbant l'organisation chromatinienne via l'altération des TADs (domaines topologiquement associés)
b. Altération de l'expression génique et/ou en perturbation de l'organisation chromatinienne via l'interruption de gènes ou d'interaction gène(s) - élément(s) régulateur(s) du fait de la mobilité de séquences répétées notamment des rétrotransposons
c. Anomalies épigénétiques altérant la méthylation de l'ADN et/ou les modifications post- traductionnelles des histones132
4.2.2.4. Projets menés dans l'équipe GAD INSERM U1231132
4.2.3. Enjeux de l'interprétation des variations génétiques133
4.3. Conclusion
ANNEXES
REFERENCES
RESUME
SUMMARY

Remerciements

A Laurence et Christel dont la motivation et le dynamisme sont à toute épreuve, je les remercie pour l'encadrement de ce travail et pour tout ce qu'elles m'ont apporté au cours de ma vie professionnelle dijonnaise.

A mes rapporteurs Amélie et Christelle qui ont accepté de prendre cette responsabilité et qui m'ont poussé dans mes retranchements pour améliorer substantiellement ce travail.

A Patrick et Catherine Boileau qui ont accepté de faire partie de mon jury, et Patrick d'en être le président.

A mes collègues de l'équipe médicale de génétique dont les périodes de chevauchement ont été parfois trop courtes (Daphné maître dysmorphologue, Nolwenn Marie et Audrey toutes 3 investies à fond dans le suivi) ou plus prolongées (Arthur qui arrive à jongler entre clinique et labo, Sophie qui arrive à jongler avec les anomalies du développement malgré son orientation onco/cardio, Julian qui arrive à s'investir pour les activités et sorties pour les patients, et Nada qui passait un temps conséquent à se creuser la tête pour les régimes alimentaires dans les maladies métaboliques), mais des interactions toujours sympathiques, efficaces ou non, toujours dans une bonne ambiance, avec cohésion et solidarité.

Aux secrétaires du service clinique pour leur disponibilité, leur gentillesse et leur soutien (Claudine surtout qui avait la plus grosse charge pour me supporter et qui a rempli cette dure mission avec brio, Tiphaine qui va maturer la théorie de la relativité pour ne pas se faire envahir des émotions négatives, Sylvie en qui je reconnais mes troubles attentionnels mais qui sait transmettre sa zenattitude, Sonia x2 (B qui s'est bien adaptée à ses responsabilités croissantes / G qui trouve toujours une façon conviviale de vulgariser les maladies rares et le handicap pour les enfants) + Fatma l'archiviste sur sa période de présence dans le service).

Aux conseillers en génétique pour nos aides et échanges réciproques dans les différents domaines de la génétique et les moments partagés en voiture pour les consultations délocalisées qui les ont rendu plus conviviaux et ont permis de mieux apprendre à se connaître, Geoffrey (pour sa bonne humeur permanente y compris en période de coup dur), Caroline, Amandine, Marion, Elodie pour leur sympathie.

A l'équipe paramédicale avec qui les échanges ont été riches et qui apportent toujours beaucoup, preuve de l'intérêt majeur de la pluridisciplinarité pour permettre d'apporter les compétences de chacun au bénéfice des patients, Lorraine toujours motivée, Annick à fond dans tout ce qu'elle fait et grande superviseuse de prélèvements qui arrive toujours à anticiper et rattraper mes boulettes, Claire et Jenny qui m'ont appris entre autre à manipuler les chiffres et les tests neuropsy, Lou que j'ai trop peu croisé, Christine, Murielle, Isabelle et Marie-Myriam pour leur bonne humeur, Hélène que je n'aurai croisé que très peu de temps.

Aux internes qui sont passés dans le service qu'ils soient issus du DES de génétique Mathilde Aurore et Caroline (au top, merci Caro pour le travail clinique sur les trios !), Aurélien, Basile (bonne continuation dans ta future spé), Benoit, (Julian à cheval avec équipe médicale et qui m'a fait confiance pour faire partie de son jury de thèse de médecine), ou hors spécialité Quentin (vive la neurogénétique !), Ossama, Nawale, Angélique.

Aux membres de l'équipe du laboratoire GAD/B3 et surtout Frédéric (nombreux échanges et travaux communs), Yannis (Grand Maître incontesté de la bio-informatique, des lignes de commande et des outils statistiques), Ange-Line (travaux communs, designeuse en chef de diapositives et grande

organisatrice de loisirs), Christophe (qualités humaines mais aussi en tant que membre du comité de suivi de thèse), Antonio (dynamisme collaborative), Philippine (partenaire de badminton plus que de badmito), Anne-Sophie (avec qui j'aurais bien testé la planche à voile à Arc-sur-Tille), Virginie (en particulier pour l'accueil en pique-nique), Laurence J (trop peu de tennis ensemble), Martin et Thibaud (un peu de foot et d'escalade), Paul (qui tente de jongler entre Besançon et Dijon et qui nous trouve toujours un mécanisme chromosomique sorti du chapeau), Nicolas (qui pourrait aussi être ajouté au DAN mais le temps associé y a probablement été un peu trop court), Cyril, Charlotte, Emilie, Romain, Vincent, Binnaz, Stephan, Justine, Anne-Laure.

A Laurence Colleaux qui avait accepté de faire partie de mon comité de suivi de thèse et qui m'a toujours soutenu depuis que je l'ai rencontré dans mon projet de Master 2 et dont j'ai toujours apprécié les qualités avec sa gentillesse, son efficacité, sa rigueur et son sens de la justice.

Aux membres du laboratoire PBHU notamment les biologistes Patrick, Nathalie, Anne-Laure, Bernard, Anne et toutes les secrétaires et équipe technique que j'ai sollicitées régulièrement.

Aux autres médecins participant à l'activité du service et apportant leur grande compétence et plusvalue Vanessa, Anaïs, Clément.

Aux autres membres de l'équipe qui participent à l'activité générale et dont l'efficacité ne peut être remise en question Elodie, Laurent, Meriem, Céline, Aurore P.

Aux différents professionnels médicaux et paramédicaux avec qui j'ai eu l'occasion d'interagir avec qui j'ai pu échanger, pour contribuer je l'espère à la prise en charge des patients, que ce soit au CHU pédiatres en particulier Frédéric (merci pour l'accueil) Mondher Véronique + Sandra neuropsy et Adélaïde en radio, mais aussi Candace, Raphaelle, Pauline, Clémence, Véronique, Jean-Damien, Delphy, équipe du DAN/CPDPN Thierry, Olivia, Vinciane, Nicolas en passage transitoire, Christelle, Fanny, Fabienne, Régine, Karine, Estelle, Garbielle, Fara et Georges en foetopath, équipe de neurologie, de cardiologie notamment pédiatrique Sylvie et Charlotte, pédiatres et équipes de gynéco-obstétrique mais aussi autres spécialités des hôpitaux périphériques Muriel à Sens, Bertrand, Cyrille, à Auxerre, Sébastien, Benoit, Marianne à Chalon, Elodie Brice Clara à Macon, équipes de prise en charge et rééducation de Bourgogne et parfois un peu au-delà CAMSP CMP(P) IME ERHR avec Alice et autres structures de soin et activité libérale comme Dr Madinier-Chappat.

Aux personnes avec qui j'ai pris plaisir à faire quelques activités de loisirs (foot salle, laser game, bowling) et qui se reconnaitront, à Jérémie mon partenaire de tennis du FDTC qui m'a maintenu en forme.

De façon générale, j'ai été très content de cette expérience dijonnaise, et cela m'a beaucoup apporté, je n'oublierai pas la bonne humeur générale et les relations avec chacun qui priment sur certaines périodes plus difficiles.

A Nada et toute ma famille (qui est beaucoup trop grande pour qu'ils soient tous cités mais ils ne se vexeront pas) pour leur soutien et sans qui tout cela n'aurait pas été possible. Welcome back home.

Life is too short, don't waste it, go surfing, mais de façon générale, profitez bien des bons moments de la vie que l'on pense parfois anodins en évitant les excès. Nous sommes si peu de chose et si éphémères que cela ne vaut pas la peine de se rendre malade pour des choses qui ne méritent pas d'y accorder une si grande importance. Il faut garder espoir dans le futur mais ne pas fermer les yeux sur les problèmes en cours et à venir et ne pas retarder la mise en place de mesures à l'échelle individuelle et collective qui pourraient être bénéfiques pour la planète. Il faut se dire que la vie est belle et continuera avec ou sans l'Homme (peut-être que ce sera ce qui lui arrivera de mieux, rien n'est éternel).

1. INTRODUCTION

1.1. Généralités et historique

Les anomalies du développement englobent les syndromes malformatifs et les pathologies neurodéveloppementales comme la déficience intellectuelle (DI) et les troubles du spectre autistique (TSA). Ces manifestations peuvent être isolées ou associées. Ces affections font partie des maladies rares, définies par une prévalence inférieure à 1/2000¹; il est estimé qu'il en existe plus de 7000 dont environ 80% d'entre elles dues à une cause génétique². Collectivement, ces maladies touchent une partie significative de la population puisqu'il est estimé qu'environ 3 à 4 millions de personnes en France et 25 millions en Europe seraient concernées. Au moins 1 malade sur 3 serait sans diagnostic et 97% des maladies ne disposent pas d'un traitement spécifique.



*Figure 1 : Estimations épidémiologiques des maladies rares en France*³.

La majorité des maladies rares touchent des enfants, elles sont responsables de 35% des décès dans la première année de vie et une cause importante d'hospitalisation en pédiatrie et de handicap ^{4,5}. Il est estimé qu'un tiers des enfants nés avec une maladie rares décède avant 5 ans ⁶. Il s'agit donc d'un problème de santé publique. Environ 3% de la population est touchée par une maladie du groupe des anomalies du développement. Les malformations peuvent toucher tous les organes. La déficience intellectuelle (DI) est définie comme «un fonctionnement intellectuel significativement inférieur à la moyenne associé à des limitations des fonctions adaptatives». A ce jour, plus de 3000 maladies faisant

partie du champ des anomalies du développement embryonnaire d'origine génétique sont répertoriées et plusieurs dizaines sont décrites chaque année.

Sur le plan étiologique, les anomalies du développement peuvent être attribuées à des facteurs environnementaux dans 10 à 20 % des cas (diabète, infections, prises de médicaments ou de toxiques au cours de la grossesse) et à des causes génétiques (chromosomiques ou géniques) dans environ 80% des cas ¹. Bien que certains syndromes soient très probablement causés par un défaut de vascularisation au cours de l'embryogenèse, le pourcentage restant non encore élucidé est également probablement d'origine génétique dans la grande majorité des cas mais le diagnostic moléculaire n'est pas posé en raison des limites des techniques utilisées ou disponibles. Néanmoins, l'explosion des découvertes avec les progrès technologiques indique que la part des causes génétiques est bien plus importante. Il est important de poursuivre l'ambition de participer à l'effort fixé par l'Europe, à savoir l'identification d'une large majorité des gènes responsables de maladies avec anomalies du développement à l'horizon 2020.

Les prémices du diagnostic génétique sont survenues à la fin des années 1950 suite à l'identification au microscope d'un chromosome surnuméraire chez les patients atteints de syndrome de Down (trisomie 21) ⁷. Par la suite, de nombreux remaniements de nombre et de structure ont été mis en évidence, en particulier certains sites fragiles propices aux cassures chromosomiques qui étaient particulièrement révélés sous certaines conditions comme le défaut d'acide folique (site FRAXA associé au syndrome de l'X fragile ⁸). L'origine de cette cassure a été identifiée en 1991 par la découverte d'une amplification correspondant au locus du gène *FMR1* ^{9,10}. La particularité du mode de transmission lié à l'X ne concernant qu'un seul chromosome a conduit certaines équipes à se focaliser sur des familles dont l'hérédité était en faveur avec plusieurs garçons atteints de DI reliés par des filles, ce qui a permis des avancées importantes par rapport aux maladies autosomiques ; plus de 100 gènes parmi les quelques 800 gènes de l'X ont été impliqués dans la DI ¹¹. Par la suite, l'augmentation en parallèle des descriptions de syndromes cliniques reconnaissables et les progrès technologiques ont permis d'accroître rapidement les connaissances dans le domaine, notamment en ce qui concerne les gènes autosomiques.

Dans les années 1990-2000, les progrès dans le domaine de la génomique étaient liés au séquençage de gènes par la méthode de Sanger. Cette méthode était coûteuse, nécessitait de disposer de grandes familles touchées par la même maladie et reposait sur l'identification de régions candidates par clonage de points de cassure de réarrangements chromosomiques et/ou cartographie génétique puis le séquençage des gènes candidats situés dans ces régions les uns après les autres selon leur probabilité d'être impliqués (hypothèses sur la base de leurs expressions et fonctions connues). Une fois qu'un gène était identifié comme potentiellement impliqué dans le phénotype d'un patient ou d'une famille, le séquençage de cohortes de patients présentant le même phénotype permettait de confirmer le lien de causalité lorsque d'autres patients étaient porteurs d'altération de ce gène. Par la suite, dans les années

2000, d'autres analyses pangénomiques visant à rechercher des déséquilibres chromosomiques (puce à ADN ou CGH-array) ont permis de cibler des loci responsables d'anomalies du développement et le séquençage de gènes candidats dans des cohortes de patients ayant un phénotype similaire. Une nouvelle génération de séquenceurs à très haut débit a permis d'augmenter largement la génération de données de séquençage à plusieurs gigabases d'ADN pour des temps très inférieurs (quelques jours et jusqu'à quelques heures) et à un coût bien moindre (revue ¹²). A titre de comparaison, en 2001, la publication de la séquence du génome humain était le fruit de 10 ans de travail par 16 grands laboratoires pour un coût de 3 milliards de dollars ¹³. En 2008, le coût du séquençage du génome de James D. Watson était estimé à 1 à 2 millions de dollars ¹⁴; en 2009, le coût était d'environ 50000 dollars ¹⁵ et en 2010 le séquençage pouvait être réalisé en deux semaines pour un coût d'environ 6000 dollars. Actuellement, il est possible de cibler le séquençage à l'exome qui correspond à l'ensemble des séquences codantes et des jonctions intron-exon des gènes soit 1 à 2% du génome. Cela permet, pour un coût moindre (environ 700 dollars), de détecter l'ensemble des variations contenues dans les gènes, c'est-à-dire celles qui sont les plus faciles à interpréter, et de sélectionner celles les plus à même d'entraîner une perturbation dans l'expression ou la fonction du gène. Cette technique peut s'appuyer sur l'analyse comparative des parents du patient testé et permet d'identifier une cause génétique dans des familles où un seul individu est atteint. Cette approche était initialement appliquée au cadre de la recherche des bases moléculaires des maladies dont une origine génétique était suspectée. La preuve de concept a pu être établie dès la fin des années 2000 avec l'identification de nombreux gènes impliqués dans des syndromes reconnus (comme le syndrome de Miller causé par des mutations bialléliques du gène DHODH¹⁶, le syndrome Kabuki dû à des mutations du gène MLL2/KMT2D¹⁷), puis des affections génétiques fortement hétérogènes génétiquement et sans élément distinctif comme la DI.



Figure 2 : Progrès technologiques permettant d'entrevoir la place du séquençage pangénomique en diagnostic

Les modèles cellulaires et animaux ont été régulièrement utilisés pour conforter la pathogénicité de variations génétiques mais certaines recherches chez l'animal ont aussi pu conduire à identifier des

mutations causales avant d'identifier des patients atteints de maladie en lien avec des variations de l'orthologue humain, ce qui a pu permettre de contribuer à apporter des arguments pour incriminer ce gène ^{18,19}.

L'ensemble des différents progrès technologiques a donc permis l'identification de très nombreux gènes impliqués dans ces pathologies mais aussi de décrire des présentations atypiques de syndromes connus. Le transfert progressif en utilisation diagnostique a permis d'augmenter progressivement la proportion de patients pouvant se voir attribuer une cause génétique, donc une augmentation importante du rendement diagnostique, pour apporter un conseil génétique aux familles et parfois proposer des essais de certaines thérapeutiques.

Depuis 2012, plusieurs centres dans le monde ont transféré cet outil à une pratique routinière du diagnostic moléculaire de maladies génétiques mendéliennes. La FHU TRANSLAD a été le premier centre en France à appliquer le séquençage pangénomique d'exome au diagnostic de maladies rares avec anomalies du développement. Cet examen permet d'atteindre un rendement diagnostique sans précédent pour les familles. Toutefois, près de la moitié des examens restent négatifs en l'état actuel des connaissances médicales, et fait l'objet d'initiatives de recherche. Actuellement, avec la baisse des coûts, il est discuté de passer au séquençage de génome entier permettant de s'affranchir de l'étape de capture nécessaire au séquençage d'exome, d'obtenir une meilleure homogénéité de profondeur d'analyse et d'avoir à disposition les données de l'exome mais aussi d'analyser les régions non codantes et les variations de structure.

1.2. La déficience intellectuelle : définition, classification, physiopathologie

1.2.1. Définition et évaluation

Trois organisations ont proposé une définition de la déficience intellectuelle : l'OMS (Organisation Mondiale de la Santé) ²⁰, l'AAIDD (American Association on Intellectuel and Developmental Disabilities) ²¹ et l'APA (American Psychiatric Association) ²². Ces définitions ont en commun trois critères:

- le constat de déficits dans les fonctions intellectuelles comme le raisonnement, la résolution de problèmes, la planification, la pensée abstraite, le jugement, l'apprentissage académique,

l'apprentissage par l'expérience et la compréhension pratique. Ces déficits sont confirmés à la fois par des évaluations cliniques et par des tests d'intelligence personnalisés et normalisés ;

- des limitations significatives du comportement adaptatif en général, c'est-à-dire dans les habiletés conceptuelles, sociales et pratiques apprises qui permettent de fonctionner dans la vie quotidienne ;

- l'apparition de ces déficits intellectuels et de ces limitations adaptatives au cours de la période développementale (précisé avant 18 ans pour l'AAIDD).

En ce qui concerne les 2 premiers critères, plusieurs échelles d'évaluation standardisées ont été développées pour évaluer de façon quantitative les limitations:

- quotient intellectuel (QI) inférieur à la moyenne de la population générale, d'environ 2 écarts types, soit <70, le QI moyen étant fixé par convention à 100 et l'écart type à 15 (figure 3) ^{23,24} : les échelles de Weschler sont les plus utilisées (WPPSI-IV, WISC-IV ²⁵, WISC-V, WAIS-IV), le KABC II, et, plus récemment la NEMI-2, complètent la liste des instruments disponibles ; elles sont utilisées selon l'âge du patient et l'estimation approximative de son niveau de développement pour éviter d'être à un effet seuil (niveau soit trop faible soit trop élevé par rapport à l'intervalle pouvant être étudié). Certaines échelles de développement précoces évaluent un âge de développement et un quotient de développement (QD) plutôt que le QI : il existe 3 principales échelles de développement précoce (Denver-DDST, Bayley III, Brunet-Lézine). Il est aussi possible d'estimer le développement d'un enfant par l'utilisation de questionnaires aux parents, ce qui est plus rapide mais moins précis puisque cela ne repose pas sur l'observation clinique du professionnel : il existe 3 questionnaires parentaux principaux (IFDC-Inventaires français du développement communicatif ; IDE-Inventaire du Développement de l'Enfant ; ASQ-Ages and Stages Questionnaire) ;

- diminution de la performance du comportement adaptatif d'environ 2 écarts types sous la moyenne de la population générale : la Vineland-II (Vineland Adaptive Behavior Scale-Second Edition) est une des rares échelles traduite en français, la France ayant un certain retard dans l'utilisation de ces outils (autres échelles : Adaptive Behavior Scale-School Second Edition ABS-S 2, Adaptive Behavior Assessment System Second Edition ABAS-II, la Scale of Independant Behavior-Revised SIB-R).

En France, la prévalence de la DI légère est estimée entre 10 et 20 pour 1000, ce taux est similaire à ceux retrouvés dans les autres pays européens ou aux Etats-Unis. Concernant la DI sévère, le taux de prévalence en France est plus précis, de 3 à 4 pour 1000, proche d'autres données internationales (entre 2,7 et 4,4 pour 1000). Malgré les nombreuses raisons qui pourraient faire évoluer la prévalence globale des DI dans le temps (meilleurs dépistage prénatal et suivi de grossesse, améliorations des contextes socio-économiques, meilleure survie), la prévalence reste stable dans le temps, en France comme à l'étranger. La survie des personnes avec DI s'améliore avec le temps mais reste inférieure à celle de la population générale ²⁶.



*Figure 3 : Illustration de la distribution du QI dans la population générale*²⁶.

La DI peut s'associer à d'autres troubles neurodéveloppementaux comme les troubles du spectre autistique (TSA), les troubles déficitaires de l'attention avec ou sans hyperactivité (TDA/H), les troubles dys ou multidys (dyslexie, dysorthographie, dysphasie, dyscalculie, dyspraxie) mais chacun d'eux peut aussi représenter un diagnostic différentiel, il est souvent nécessaire de faire passer différentes évaluations par des professionnels paramédicaux spécialisés et de prévoir des synthèses multidisciplinaires pour pouvoir établir le diagnostic le plus précisément possible.

La DI est dite non syndromique lorsqu'elle reste isolée à une anomalie de l'évolution neurodéveloppementale alors qu'elle est syndromique lorsqu'elle s'associe à d'autres manifestations cliniques comme des caractéristiques physiques particulières, des malformations cérébrales ou d'autres organes, des pathologies d'organes ou de systèmes. Certaines associations syndromiques sont parfois caractéristiques et permettent d'évoquer une hypothèse étiologique, qu'elle soit génétique ou non.

1.2.2. Causes de la DI

La DI est une pathologie dont les causes sont très nombreuses, et plusieurs facteurs peuvent parfois se combiner, pouvant augmenter sa sévérité. Selon les études, les différentes causes de la DI incluent (fréquence donnée à titre indicatif, difficile à estimer):

1.2.2.1. Facteurs environnementaux (10-13%)

Ces facteurs peuvent être acquis ou congénitaux comme les complications de la prématurité (4-5%), des complications pré ou périnatales notamment anoxo-ischémiques, un retard de croissance intra-utérin et un petit poids pour l'âge gestationnel ²⁷, des infections principalement anténatales (CMV ²⁸, zikavirus ²⁹) mais aussi péri ou post-natale (encéphalite comme l'herpès ³⁰ et méningites), des intoxications fœtales (alcool ³¹, association de médicaments anti-épileptiques comme le valproate ³², maladie maternelle comme la phénylcétonurie ^{33,34} ou le diabète) ou post-natales (exposition chronique à des métaux lourds comme le plomb et le mercure ³⁵, malnutrition et carences par exemple en iode, origine psycho-sociale).

1.2.2.2. Causes génétiques

Les causes génétiques représentent une part de plus en plus importante des causes de DI grâce aux progrès technologiques et au démembrement progressif des DI non expliquées.

Elles peuvent être liées à :

- des syndromes reconnaissables cliniquement (~ 2%),

- des causes chromosomiques (10-15%), reconnaissables ou non, avec parfois des remaniements récurrents parfois en miroir (délétion-duplication 22q11.2 incluant *TBX1*, délétion ou duplication 17p11.2 incluant *RAI1* responsable du syndrome de Smith-Magenis et de Potocki-Lupski respectivement) ou parfois non récurrent avec des points de cassure variables (syndrome du cri-du-chat, délétion 4p terminale et Wolf-Hirschhorn, délétion 1p36...). L'utilisation croissante de la CGH array en routine a conduit en 2010 à proposer que cet examen soit utilisé en première intention ³⁶, ce qui a été conforté par l'Association des Cytogénéticiens de Langue Française ³⁷.

- des anomalies de l'empreinte parentale : le syndrome de Prader-Willi, le syndrome d'Angelman par atteinte de la copie paternelle ou maternelle respectivement de la région chromosomique 15q11q13 représentent l'exemple le plus caractéristique, les causes pouvant être une délétion, une disomie uniparentale (ou plus rarement une mutation du centre de l'empreinte ou de gènes soumis à empreinte) ^{38,39}.

des causes monogéniques : l'utilisation de la base de données OMIM et des mots-clés en lien avec la DI en 2016 retenait 818 gènes dont les variations pathogènes pouvaient être responsables de ce type d'affection ⁴⁰. Parmi ceux-ci, une surreprésentation de gènes situés sur le chromosome X était notée

(environ 10-15% de gènes de DI en comparaison à une moyenne de 4% du génome), permettant d'expliquer une partie du déséquilibre du sex ratio de 1,3 à 1,4 garçons pour une fille, notamment du fait d'un mode de transmission propre qui permet un repérage plus facile des familles et que ce type d'atteinte a été exploré de plus longue date et de façon plus poussée, allant jusqu'au séquençage Sanger de l'ensemble des gènes dans le cadre d'un consortium international ^{41,42}. Bien que l'on pourrait penser que les formes dites « récessives » liées à l'X sont fréquemment héritées de la mère, la part des mutations *de novo* est estimée à 1/3 à 2/3 des cas selon les gènes ^{26,43–45}.

Dans une population caucasienne ou occidentale, la part des maladies autosomiques récessives dans la DI est estimée à 33% mais les études réalisées dans des pays où la consanguinité est importante rapportent des taux bien plus élevés : l'étude d'une cohorte de 149 patients du Qatar (où la consanguinité est importante et dont le retard du développement et la DI, syndromique ou non, a été l'affection prépondérante) décrit un rendement diagnostique de 60%, d'autant plus important qu'il existe une consanguinité avérée (et 6/89 patients avaient plusieurs maladies) ⁴⁶.

En ce qui concerne les formes autosomiques dominantes de DI, elles concernent environ 50% des cas dont plus de $2/3 \ de \ novo$) $^{26,43-45}$. Les différentes technologies pangénomiques comme la CGH array et le séquençage d'exome et de génome s'appuient donc beaucoup sur cette hypothèse de survenue de *novo* pour des pathologies sporadiques comme la DI 47,48 .

1.2.2.3. DI idiopathique

Elle est estimée à 35-60% des cas selon les études et est probablement en grande partie d'origine génétique.

La figure 4 illustre le lien entre les causes de la DI et les étapes du développement auxquelles elles surviennent :



*Figure 4: représentation schématique de la période neurodéveloppementale à laquelle certains facteurs peuvent être responsables de DI*⁴⁰.

1.2.3. Rôle des gènes dont les variations pathogènes sont responsables de DI

Le développement du cerveau comprend plusieurs étapes de morphogenèse des structures à partir de la différenciation du feuillet embryonnaire épiblastique (phénomène de neurulation), puis prolifération des progéniteurs situés dans les régions périventriculaires, migration radiaire et tangentielle des neurones excitateurs et inhibiteurs respectivement, différenciation, synaptogenèse, apoptose, gliogenèse et myélinisation (figures 4, 5 et 6).

Le néocortex regroupe 2 types cellulaires principaux, les neurones et la glie. Les neurones correspondent aux cellules transmettant les signaux électriques et chimiques et la glie joue un rôle important dans de multiples fonctions en soutien des neurones. Les neurones peuvent être excitateurs ou inhibiteurs et ont des origines différentes. La glie radiaire se différencie à partir des cellules polarisées bordant la région périventriculaire et va permettre de guider la migration des neurones excitateurs vers la surface piale pour former la plaque corticale. Plus les neurones sont produits et migrent tard, plus ils seront situés en périphérie de la surface corticale. Les neurones projettent ensuite leurs axones qui sont guidés grâce à des molécules de signalisation. Les arborescences dendritiques se développent sur lesquelles se forment des épines dendritiques qui seront le siège de la mise en place des synapses.

Chaque étape est coordonnée par des gènes dont les mutations impactant leurs fonctions peuvent être responsables de malformations cérébrales (DI syndromique, parfois létalité) ou des troubles fonctionnels (DI non syndromique). Dans les 2 situations, des troubles neurodéveloppementaux et une épilepsie sont fréquemment associés.



Figure 5 : Illustration des étapes de neurogenèse impliquant la prolifération et la différentiation (A), de migration neuronale et d'arrêt de migration (B), avec représentation schématique du complexe DAG (dystrophin-associated glycoprotein) comprenant des isoformes de dystrophine et le groupement α -dystroglycane permettant l'interaction avec la matrice extra-cellulaire (C)⁴⁹. *MT*, microtubules; α -DG: α -dystroglycan; β -DG: β -dystroglycan.



Figure 6 : Représentation schématique du développement cortical au cours du temps : début au niveau de la zone ventriculaire (VZ) pour les neurones excitateurs et des éminences ganglionnaires pour les interneurones inhibiteurs, prolifération des progéniteurs avant différenciation en neurones et migration vers la plaque corticale pour s'intégrer au réseau⁵⁰.

Plus d'un millier de gènes sont maintenant connus comme impliqués dans la déficience intellectuelle, qu'elle soit syndromique ou non syndromique. Ces gènes codent pour des protéines jouant un rôle dans de nombreuses voies biologiques et métaboliques, impliquées notamment dans la fonction synaptique ou la régulation épigénétique de sa plasticité. Certaines formes de DI syndromique avec malformation cérébrale sont liées à des gènes codant pour des protéines directement impliquées dans le développement embryonnaire cérébral.

1.2.3.1. DI due à une anomalie de la fonction synaptique

Au cours de la phase de développement puis par la suite au cours de la croissance, le cerveau est en maturation permanente avec l'établissement, la destruction, le remodelage des connexions entre les neurones appelées synapse ; ces événements se poursuivent à l'âge adulte et explique les capacités de plasticité cérébrale qui néanmoins se réduisent avec l'âge. Les synapses constituent la zone de neurotransmission dans laquelle un signal électrique est converti en signal chimique. La synapse a une architecture précise avec de nombreuses protéines d'échafaudage permettant de maintenir une proximité physique des éléments pré et post-synaptique. Le versant pré-synaptique comprend les vésicules de sécrétion et protéines impliquées dans la libération des neurotransmetteurs dans la fente lorsqu'un potentiel d'action parvient et dans leur recapture pour le recyclage. La fente synaptique contient notamment des protéines permettant le recyclage des neurotransmetteurs, d'autres protéines les dégradant pour réguler leur action. L'élément post-synaptique (épine dendritique) contient les récepteurs et voies de transduction du signal ^{49,51}. Le glutamate est le neurotransmetteur principal des synapses excitatrices des neurones pyramidaux et le type de stimulation est à l'origine des phénomènes de potentialisation et de dépression à long terme. La potentialisation à long terme (LTP) correspond à une augmentation de l'efficacité de la transmission synaptique et est due à des activations brèves et répétitives alors que la dépression à long terme (LTD) est due à de longues stimulations à faible fréquence. Ces changements liés à la plasticité synaptique peuvent durer de quelques heures à plusieurs années et sont associés à des modifications de taille et de morphologie des synapses, ces 2 phénomènes régulent l'homéostasie et la fonction du réseau neuronal 52. Le GABA est le neurotransmetteur principal des synapses inhibitrices des interneurones.

Une base de données nommée SynGO (Synapse Gene Ontology) a été conçue récemment pour regrouper des informations sur ce qui a trait à la synapse sur le plan biologique et de la recherche. Elle regroupe 179 processus biologiques à partir de 2922 annotations pour 1112 gènes (figures 7 et 8). Cela permet de confirmer une conservation très importante de ces gènes et une intolérance aux variations génétiques par rapport à d'autres groupes de gènes. Il existe ainsi une surreprésentation de variations

associées aux capacités cognitives (troubles neurodéveloppementaux et psychiatriques comme la DI, troubles attentionnels, autisme, trouble bipolaire, schizophrénie)⁵³.



*Figure 7 : Illustration des composants cellulaires et des voies biologiques impliquées dans le fonctionnement synaptique*⁵³.



*Figure 8 : Illustration des protéines, de leur rôle et de leur localisation dans le fonctionnement synaptique*⁵⁴.

a. Gènes responsables de DI impliqués au niveau de l'élément pré-synaptique (figures 8 et 9)

al. Gènes impliqués dans le trafic vésiculaire

Les protéines de la famille RAB appartiennent à la superfamille RAS et sont de petites GTPases impliquées dans le trafic vésiculaire des différents compartiments membranaires, elles jouent donc un rôle dans le bourgeonnement, la formation, la fission, le transport, l'ancrage, et la fusion des vésicules pré-synaptiques, grâce à l'interaction avec d'autres protéines effectrices. Certaines d'entre elles jouent aussi un rôle au niveau post-synaptique ⁵⁵. Elles sont actives ou inactives selon leur liaison au GTP ou au GDP respectivement. Certains gènes codant pour des protéines RAB ont été associés à des maladies génétiques comme *RAB39B* dont les mutations perte de fonction sont associées à une DI non syndromique ⁵⁶ et des faux sens déstabilisant la protéine et entraînant une dégradation prématurée ont été identifiés dans le syndrome de Waisman (DI et syndrome parkinsonien précoce) ⁵⁷.

D'autres gènes impliqués dans ces voies biologiques sont responsables des syndromes de Warburg micro (forme sévère) et de Martsolf (moins sévère) qui ont une hérédité autosomique récessive. Le syndrome Warburg micro comprend une microcéphalie, une microphtalmie, une microcornée, une cataracte congénitale, une atrophie optique, une dysplasie corticale, une hypoplasie du corps calleux, une DI sévère, une spasticité des membres inférieurs et un hypogonadisme. Dans le syndrome de Martsolf, l'hypotonie et la DI sont plutôt légères. Les gènes en cause sont *RAB3GAP1, RAB3GAP2, RAB18, TBC1D20*.

Plusieurs protéines ont un rôle dans la fusion des vésicules présynaptiques (syntaxines, SNAP-25, synaptobrevine 1 / VAMP1, MUNC18 / STXBP1, synaptotagmines) dont les anomalies peuvent entrainer une diminution de la libération des neurotransmetteurs et altérer le fonctionnement synaptique. (figure 9) ^{51,58,59}. Par exemple, les mutations de *STXBP1* responsables d'une perte de fonction ou d'une déstabilisation protéique causent une encéphalopathie épileptique souvent sévère avec syndrome d'Ohtahara ou une DI sévère avec épilepsie ⁶⁰. Les variations faux sens hétérozygotes de *SYT1* (synaptotagmine 1) situées dans 2 régions du domaine C2B et responsables d'une perturbation de l'endocytose causent le syndrome de Baker-Gordon (BAGOS) qui comprend un retard neurodéveloppemental modéré à profond avec hypotonie précoce, anomalies ophtalmologiques, langage absent ou rudimentaire, troubles du comportement, des mouvements hyperkinétiques et des anomalies EEG sans crises convulsives observables ⁶¹.

Des mutations dans d'autres gènes codant pour des protéines ayant un rôle dans les processus présynaptiques, comme *CASK* ⁶², *OPHN1*⁶³, *ILRAPL1* ⁶⁴ ont été rapportées dans la DI non syndromique ou parfois avec malformation cérébrale (comme une hypoplasie ponto-cérébelleuse pour certaines mutations de *CASK* et une hypoplasie vermienne pour *OPHN1*). *PRSS12* représente un exemple de gène codant pour une protéase située au niveau de l'élément présynaptique et relarguée dans la fente synaptique dont le rôle est de réguler le taux de protéines à ce niveau, les variants perte de fonction bialléliques sont responsables de DI modérée à sévère non syndromique ⁶⁵.

a2. Gènes impliqués dans la structure de la synapse

Les gènes codant pour des protéines d'échafaudage situées sur le versant pré-synaptique sont pour la plupart associés à des troubles neurodéveloppementaux comme les neurexines (NRXN1,2,3), les contactines et protéines associées (CNTN6) ⁵⁴. Les gènes *NRXN1* et *CNTNAP2* ont la particularité que les variations perte de fonction hétérozygotes ont été identifiées comme prédisposant aux troubles neurodéveloppementaux ^{66,67} alors que la présence de telles variations à l'état biallélique entraîne un phénotype sévère ressemblant au syndrome de Pitt-Hopkins (DI, troubles autistiques, facies grossier avec l'âge, troubles du rythme respiratoire) ^{68,69}.



*Figure 9 : Illustration du rôle des protéines situées sur le versant pré-synaptique dans le fonctionnement de la synapse*⁵¹.

b. Gènes de DI impliqués au niveau de l'élément post-synaptique (figures 8 et 10)

b1. Gènes impliqués dans la structure de la synapse

On retrouve des gènes codant pour des protéines transmembranaires ayant un rôle dans l'établissement de contacts avec les protéines pré-synaptiques et avec le cytosquelette permettant de délimiter et maintenir un espace entre les membranes cellulaires des 2 neurones, ainsi qu'un rôle dans la transduction du signal, comme les neuroligines (*NLGN3,4*), les protéines contenant un domaine SH3 et de multiples répétitions Ankyrine (SHANK1,2,3), les protéines du groupe HOMER (HOMER1-3, homolog of drosophila), et on trouve des gènes codant pour des protéines de structure sous-membranaire comme les protéines PSD (post-synaptic density) qui ont aussi un rôle de liaison entre les protéines mais également un rôle de transduction du signal, codées par les gènes DLG (discs large homolog of drosophila), au moins 3 parmi les 4 PSD classiques sont associés à la DI dû à des mutations perte de fonction à l'état hémizygote pour *DLG3* (SAP102)⁷⁰, et à l'état hétérozygote pour *DLG2* (PSD-93/chapsyn-110)⁷¹ et *DLG4* (PSD-95)⁷².

b2. Gènes impliqués dans la réception du signal et codant pour les canaux ioniques

Ce groupe de gènes concerne principalement la voie de signalisation glutamatergique excitatrice, et la voie GABAergique inhibitrice.

En ce qui concerne la voie glutamatergique, 2 types de récepteurs sont présents au niveau postsynaptique.

Les récepteurs NMDA (N-methyl-D-aspartate) sont des canaux ioniques composés de di-hétérotétramères ou de tri-hétéro-tétramères dont les sous-unités sont NR1 (codée par GRIN1) qui s'associent parfois à NR3 (NR3A-B codées par GRIN3A-B) pour former un site de liaison à la glycine, et de deux sous-unités NR2 (NR2A-D codées par GRIN2A-D) qui constituent le site de liaison au glutamate. Ils permettent la transmission synaptique lente ⁷³. Des variants tronquants et faux sens perte de fonction des gènes *GRIN1, GRN2A, GRIN2B, GRIN2D* ont été retrouvés chez des patients atteints de DI de sévérité variable, plutôt modérée) avec épilepsie (parfois encéphalopathie épileptique) et parfois des mouvements hyperkinétiques ^{74–77}, ou chez des patients atteints d'épilepsie focale avec atteinte du langage avec ou sans DI ^{78–80}. Certains patients présentent des anomalies de gyration. Des variations gain de fonction ont aussi été identifiées dans *GRIN2B* ⁷⁶. Les mutations impacteraient la transduction du signal et la plasticité synaptique (figure 8).

Les récepteurs AMPA (α-amino-3-hydroxy-5-méthylisoazol-4-propionate) ont un rôle dans la transmission synaptique excitatrice rapide dans le système nerveux central, ils sont composés à partir de 4 sous-unités GluR1-GluR4 (ou GluRA-GluRD ou GRIA1-4) pour s'assembler en complexes hétéro-multimériques ⁸¹. Des variations perte de fonction du gène GRIA3 ont été rapportées comme la cause d'une DI modérée à sévère liée à l'X avec troubles du comportement incluant des traits autistiques, épilepsie, critères physiques particuliers et petite taille ^{82,83} Une délétion chromosomique de novo emportant le 1er exon du gène GRIA2 et une partie du gène GLRB a été identifiée chez un patient atteint de DI, aucune variation de l'autre allèle n'a été identifiée pour chacun des gènes et le gène candidat retenu de par sa fonction était GRIA2⁸⁴. Six patients ont été rapportés comme porteurs de la même mutation faux sens du gène GRIA1, confirmé de novo chez 3 d'entre eux, un des parents indisponible chez les 3 autres et un patient supplémentaire porteur d'un faux sens de novo à une position proche (parmi une cohorte de 17689 patients atteints de troubles neurodéveloppementaux avec pour but d'identifier des événements mutationnels récurrents ou un excès de mutations dans certains gènes); parmi les 4 patients pour lesquels des informations cliniques étaient disponibles, il existait une DI légère à modérée chez 4/4, des traits autistiques chez 3/4, un retard de langage chez 3/4 dont 2 avaient des traits autistiques et un retard moteur chez 2/4, les IRM cérébrales étaient normales ; les études fonctionnelles ont mis en évidence un canal constitutivement ouvert, ce qui représente donc un gain de fonction⁸⁵. Des variations faux sens de novo du gène GRIA4 ont aussi été identifiées comme responsables de DI; la modélisation des mutations était en faveur d'une perturbation de la fonction d'ouverture du canal soit par excès soit par défaut ⁸⁶.

A noter que les récepteurs au glutamate sont divisés en 2 groupes : ionotropiques (NMDA et non-NMDA incluant AMPA et kainate) et métabotropiques (mGluR, divisés en 3 groupes : groupe 1 couplé au métabolisme de l'inositol-phosphate et groupes 2 et 3 couplés à l'activité adénylate cyclase). En ce qui concerne la voie GABAergique, il existe plusieurs types de récepteurs.

Les récepteurs GABA-A sont composés de 5 sous-unités qui s'assemblent pour former un canal Cl-. Chaque sous-unité comprend une région extra-cellulaire variable, 4 domaines trans-membranaires et une région intra-cytoplasmique entre le 3^{ème} et le 4^{ème} domaine trans-membranaire. Les sous-unités sont divisées en classes selon les homologies de séquence avec 70 à 80% d'homologies dans chaque classe et 30 à 50% entre chaque classe ⁸⁷. Les récepteurs métabotropiques GABA-B sont couplés aux protéines G et modulent la transmission synaptique par des voies de signalisation intracellulaire. Ils inhibent la libération présynaptique ou augmentent la perméabilité aux ions K+ ce qui entraîne des potentiels post-synaptiques inhibiteurs ^{88,89}.

Plusieurs sous-unités ont aussi été impliquées dans la DI avec épilepsie voire l'encéphalopathie épileptique et des variations pathogènes du gène *GABRA1* ont été identifiées dans un syndrome proche du Dravet (chapitre 1.3.2) ⁹⁰ mais aussi des variations perte de fonction entraînant une baisse d'expression de la sous-unité à la membrane chez des patients atteints d'épilepsie généralisée dite génétique (chapitre 1.3.2) ⁹¹. De même, des mutations diminuant la fonction de la sous-unité GABBR2 ont pu être identifiées chez certains patients présentant un phénotype similaire au syndrome de Rett sans mutation *MECP2* (chapitre 1.2.3.2) ⁹².

b3. Gènes impliqués dans la régulation du taux des protéines post-synaptiques

Il s'agit d'un point important de régulation pour la plasticité synaptique et les processus de mémorisation. Plusieurs facteurs participent à la régulation de l'expression protéique : modulation au niveau génique par les facteurs de transcription, inhibition de la traduction des ARNm, dégradation protéique notamment par le système ubiquitine-protéasome (UPS).

Plusieurs gènes dont les variations pathogènes sont responsables de DI ont été impliqués dans la voie de régulation de traduction des ARNm et de dégradation protéique ⁹³.

Un des exemples emblématiques est le gène *FMR1* qui code pour la protéine FMRP ; elle est fortement exprimée dans la région somato-dendritique des neurones et se lie aux ARNm, ce qui inhibe la traduction stimulée par les récepteurs neuronaux et cela permet donc de réguler le taux protéique au niveau de nombreuses synapses pour éviter un excès d'activité ⁹⁴. FMRP cible environ 30% des ARNm traduits en protéines qui sont localisées au niveau de la densité post-synaptique, dont les sousunités des récepteurs NMDA (NR1, NR2A, NR2B, NR3A), le récepteur mGluR5, PSD-93, PSD-95, SAPAP1-4, SHANK1-3, HOMER1, SynGAP1 et les neuroligines 1-3 (figure 10) ⁹⁵. La perte d'expression de la protéine FMRP, est responsable du syndrome de l'X Fragile qui associe une DI de sévérité variable, un trouble du spectre autistique, un trouble déficitaire de l'attention, quelques particularités morphologiques faciales et une macro-orchidie post-pubertaire ; il existe un risque un peu augmenté d'épilepsie. La perte d'expression du gène est principalement due à une expansion d'un triplet de nucléotides CGG (estimée en général au-delà de 200) située dans l'intron 1 du gène qui conduit à l'hyperméthylation d'ilots CpG de la région et empêche alors la transcription ⁹⁶.

La dégradation protéique par le système d'ubiquitinisation dans le protéasome consiste en l'ajout de molécules d'ubiquitine sur des protéines cible et met en jeu 3 enzymes : la ligase E1 qui active l'ubiquitine, la ligase E2 qui conjugue les monomères d'ubiquitine pour former une chaîne de polyubiquitine et la ligase E3 qui conjugue la chaîne polyubiquitine à sa protéine cible ⁹⁷. La spécificité des protéines à cibler est déterminée par les ligases E3. Ce système régule ainsi de nombreuses protéines impliquées dans la transmission synaptique, incluant PSD-95, GKAP codé par *DLGAP1*, SHANK, CaMKII, les récepteurs NMDA et AMPA ^{93,98}. Plusieurs gènes impliqués dans cette voie biologique ont été associés à une DI (*UBE3A* dans le syndrome d'Angelman, *UBR1* dans le syndrome de Johanson-Blizzard, *HUWE1*, *UBE2A*, *CUL4B*) (figure 10) ⁵¹. Un des exemples emblématiques concerne les mutations perte de fonction du gène *UBE3A* qui sont responsables du syndrome d'Angelman soit par mutation ponctuelle soit par anomalie de l'empreinte parentale (délétion ou disomie uniparentale) lorsque l'anomalie provient du chromosome maternel (région 15q11q13 soumise à empreinte parentale) ^{99,100}.

b4. Gènes impliqués dans le cytosquelette et sa dynamique

Le cytosquelette est une structure dynamique permettant des modifications dans la forme et la taille des cellules grâce aux filaments d'actine, aux filaments intermédiaires et aux microtubules, il a donc un rôle dans la mise en place des épines dendritiques, ce qui constitue un phénomène majeur pour la mise en place des synapses. Par exemple, une des conséquences de la perte d'expression de la protéine FMRP est la diminution d'expression de la protéine MAP1B (microtubules 1B) qui a un rôle dans la stabilisation des microtubules lors de l'allongement des dendrites et des neurites ^{101,102} ce qui conduit à une augmentation du nombre d'épines dendritiques qui restent néanmoins immatures ¹⁰³.

Les protéines Rho GTPases jouent également un rôle dans la dynamique cellulaire au niveau de l'élément post-synaptique en régulant le cytosquelette d'actine ¹⁰⁴. Parmi les différentes protéines de cette famille, Rac1, Cdc42 et RhoA jouent un rôle dans la structure des dendrites et la migration des neurones ^{105,106}. Comme indiqué précédemment, ces protéines sont actives ou inactives selon leur liaison au GTP ou au GDP. Cette liaison est régulée protéines GEF (Guanine Exchange Factors) qui sont des activateurs (GAP pour GTPase Activating Proteins) (ARHGEF6, ARHGEF7, ARHGEF9, FGD1, FRABIN, VAV2), ou des inhibiteurs (GDI pour GDP Dissociation Inhibitors) (OPHN1, MEGAP, OCRL1, ARHGAP23) ¹⁰⁷.

De nombreux gènes impliqués dans ces processus ont été associés à la DI (*ARHGEF6*, *ARHGEF9*, *FGD1*, *OPHN1*, *MEGAP*, *OCRL1*, *PAK3*, *LIMK1*) (figure 10). On retrouve certaines protéines communes à l'élément pré-synaptique comme OPHN1.

Un autre exemple d'implication du cytosquelette dans la DI concerne le gène *DMD* qui code pour la dystrophine, impliquée dans le lien entre les protéines du cytosquelette et celles de la membrane cellulaire qui interagissent avec la matrice extra-cellulaire (figure 5C). Il est exprimé dans plusieurs tissus, en particulier le muscle, le cerveau, l'œil et le cœur. Il existe 8 différentes isoformes qui conduisent à des protéines nommées selon leur poids moléculaire et différant principalement par leur partie N-terminale de la protéine, du fait d'une régulation par des promoteurs tissus-spécifiques (principales isoformes : Dp427, Dp260, Dp140, Dp116 et Dp71) et qui ont donc des spécificités dans leur expression ; en particulier, Dp71 n'est pas exprimé dans le muscle.

Les manifestations cliniques peuvent comprendre une dystrophie musculaire (sévère dans la maladie de Duchenne, moins marquée dans la maladie de Becker), des troubles cognitifs, une cardiomyopathie dilatée. Des études évaluant les capacités cognitives chez les patients porteurs de variations pathogènes de *DMD* ont pu observer 27% de QI<70 (19% répondant à la définition de DI), un trouble des apprentissages chez 44%, un TDA/H chez 32%, un TSA chez 15% et un trouble anxieux chez 27% des patients ^{108,109}.

Les corrélations génotype-phénotype ne sont pas claires entre le type de variations pathogènes retrouvées dans ce gène et le phénotype mais on peut néanmoins en retenir quelques-unes : il a pu être noté que l'atteinte de l'isoforme Dp140 était responsable d'une atteinte cognitive plus marquée ^{110,111}, celle de Dp71 est également responsable de telles manifestations ¹¹². Elles représentent les isoformes les plus courtes ayant une expression importante dans le cerveau fœtal avec une augmentation progressive jusqu'à l'âge adulte. Dp71 est en particulier très exprimé dans l'hippocampe et certaines couches corticales. A noter que certaines mutations en phase ont été associées à une DI sans déficit musculaire : délétion des exons 3 à 9, duplication des exons 13 à 27, délétion partielle de l'exon 46, délétion des exons 46 à 55 ou 56 et 57 chez des garçons mais aussi chez des filles ^{112,113}, délétion de 3 nucléotides de l'exon 67 perturbant l'expression de Dp71 avec élévation légère des CK sans myopathie y compris après 60 ans ¹¹⁴.

Par ailleurs, la maladie d'Alexander est due à des mutations du gène *GFAP* codant pour un des composants principaux des filaments intermédiaires exprimé dans les astrocytes matures (cf. chapitre 1.2.3.4.e) ¹¹⁵.

b5. Gènes impliqués dans les voies de signalisation intra-cellulaire

L'activation des récepteurs des neurotransmetteurs entraîne l'activation de voies de signalisation cellulaire postsynaptique, dont 2 ont été particulièrement étudiées : la voie Ras-MAPK-ERK et la voie PI3K-AKT-mTOR (figure 10)^{49,51}.

En ce qui concerne la voie Ras-MAKP-ERK, les mutations des gènes impliqués entraînant une activation de cette voie sont responsables d'un groupe de pathologies appelé RASopathies (*PTPN11*,

SOS1, RASA2, KRAS, NRAS, RAF1, CBL, SHOC2, BRAF, MEK1, MEK2, HRAS, CRAF, NF1, SPRED1, RASA1) ¹¹⁶. Ce groupe de pathologie inclut le syndrome de Noonan (NS), le syndrome de Noonan avec lentigines multiples, le syndrome cardio-facio-cutané (CFC), le syndrome de Costello, le syndrome de Legius, le syndrome malformation capillaire - malformation artério-veineuse (CM-AVM), la neurofibromatose de type 1. Chaque RASopathie a ses spécificités cliniques mais il existe des caractéristiques communes, dont une tendance à une petite taille, des anomalies cardiovasculaires, des anomalies musculosquelettiques, des lésions cutanées, et un risque tumoral accru. Une DI est retrouvée dans plusieurs syndromes (comme CFC, Costello), alors que le syndrome de Noonan, la neurofibromatose de type 1 et le syndrome de Legius sont plutôt associés à un risque de trouble des apprentissages. La voie Ras-MAPK-ERK contribue à la plasticité post-synaptique par différents mécanismes ¹¹⁷:

- en modifiant le nombre de récepteurs AMPA au niveau de la membrane cellulaire : une augmentation de la signalisation ERK entraîne une augmentation de la quantité de récepteurs AMPA au niveau de la membrane post-synaptique ¹¹⁸.

- en régulant la transcription et la traduction donc la synthèse des protéines par la régulation du facteur de transcription CREB, important pour la mémoire à long terme (figure 10)¹¹⁹

 - en augmentant la phosphorylation de la synapsine 1 dans les interneurones GABAergiques, ce qui réduit la LTP et la mémoire ¹²⁰

- en augmentant l'activation de la synapsine 1 qui se fixe aux vésicules synaptiques pour activer la libération de neurotransmetteurs sous l'effet de la liaison du BDNF (Brain-derived neurotrophic factor) au récepteur présynaptique TRKB (récepteur tyrosine kinase type B), ce qui active la voie ERK au niveau pré-synaptique ¹²¹.

La voie PI3K-mTOR joue également un rôle majeur dans la signalisation intra-cellulaire, en influençant la traduction qui est importante pour les processus de mémorisation ¹²². Les anomalies de la voie PI3K-AKT-mTOR ont été décrites ci-dessus (chapitre 1.2.3.4.e).



*Figure 10 : réseau protéique et voies biologiques post-synaptiques impliquant des gènes dont les variations pathogènes sont responsables de DI. Protéines sur fond violet si le gène correspondant a été impliqué dans la DI ; histones en rouge répressives, en vert activatrices ; (m) : groupements méthylés*⁵¹.

1.2.3.2. Régulation épigénétique de la plasticité synaptique

En dehors de l'information génétique contenue dans la séquence d'ADN en elle-même, la structure chromatinienne joue un rôle majeur pour permettre les processus biologiques comme la réplication de l'ADN, la recombinaison, la transcription, la réparation de l'ADN. La structure élémentaire de cette organisation est le nucléosome qui correspond à environ 150 paires de bases d'ADN enroulé 2 fois autour d'un octamère d'histone. Il est estimé qu'il existe entre 5 et 51 états chromatiniens différents selon les paramètres et les organismes analysés. La structure chromatinienne, déterminée par l'état de méthylation de l'ADN, les modifications post-traductionnelles des histones,

l'interaction à des complexes protéiques, les molécules d'ARN non codantes, la position des nucléosomes, permet de définir le profil d'expression génique d'une cellule donnée ^{123–126}.

De multiples gènes impliqués dans ces mécanismes de régulation épigénétiques sont responsables de DI syndromique ou non. Leur rôle est :

- soit d'apposer des marques épigénétiques (exemples : *DNMT1, DNMT3B, FTO* pour la méthylation de l'ADN ; *CREBBP, CUL4B, EHMT1, EP300, EZH2, HLCS, HUWE1, KAT6B, KMT2A, KMT2D, KMT2C, NSD1, WHSC1, UBE2A* pour la modification des histones),

- soit de les effacer (méthylation/déméthylation de l'ADN, modification post-traductionnelles des histones) (exemples : *HDAC4*, *HDAC8*, *KDM5C*, *KDM6A*, *PHF8*),

- soit de lire ces marques épigénétiques et de remodeler la structure chromatinienne. La famille des DEAD/ H ATPase hélicase incluant la sous-famille SWI/SNF est notamment impliquée dans la régulation de la localisation des nucléosomes (exemples : *ACTB, ARID1A, ARID1B, ATRX, CHD2, CHD7, CHD8, SMARCA2, SMARCA4, SMARCB1, SMARCE1, SRCAP, SS18L1*),

- d'autres gènes sont aussi impliqués dans le remodelage chromatinien (reconnaissance et liaison à des marques épigénétiques appartenant le plus souvent à des complexes activateurs ou inhibiteurs de transcription) (exemples : *ASXL1, BCOR, CHMP1, CTCF, GATAD2B, HCFC1, KANSL1, MBD5, MECP2, PHF6, POGZ, SKI, MED12, MED17, MED23, NIPBL, RAD21, SALL1, SMC1A, SMC3*).

Une liste non exhaustive de gènes dont les variations pathogènes sont responsables de maladie génétique incluant très fréquemment une DI est consultable en annexe 1¹²⁷.

Il peut exister un chevauchement phénotypique des syndromes associés à ces gènes dont les mutations ont des conséquences épigénétiques. Par exemple, des mutations *de novo* des gènes *MBD5*, *KMT2C*, *SMARCB1*, et *NR113* ont été identifiés chez des patients atteints du syndrome de Kleefstra dont le gène principal en cause est *EHMT1*. Ces gènes sont tous impliqués dans la régulation épigénétique : MBD5, KMT2C et NR113 interagissent avec EHMT1 et SMARCB1 interagit avec KMT2C ¹²⁸. Le syndrome de Kleefstra se caractérise par une DI de sévérité variable, un retard de croissance, une brachydactylie, des anomalies génitales chez le garçon, des troubles du comportement et du sommeil ¹²⁹.

A l'inverse, parfois certains gènes peuvent être responsables de différents syndromes en fonction du type ou des conséquences des mutations. Par exemple *EHMT1* a pu également être impliqué dans un phénotype de DI et syndrome marfanoïde (cf. chapitre 1.3.4). De même, selon les types de mutations, le gène *SMARCB1* peut être responsable de syndrome de Coffin-Siris ou de Kleefstra (figure 11).



*Figure 11 : Illustration du réseau épigénétique des gènes impliqués dans la DI. A. Réseau protéique. B. Exemple d'illustration de corrélation entre un groupe de gènes donné un des syndromes reconnaissables cliniquement*¹²⁷.

A noter qu'un nombre significatif de gènes associés à une DI, un excès de croissance et une macrocéphalie ont aussi un rôle dans ces voies de régulation épigénétique : *NSD1* (syndrome de Sotos), *EZH2* (syndrome de Weaver), *DNMT3A* (syndrome de Tatton-Brown-Rahman), *EED* (syndrome de Cohen-Gibson), *CHD8*, *HIST1H1E* (syndrome de Rahman) (les autres gènes impliqués étant des facteurs de transcription *NFIX*, *GPC3*, *BRWD3* ou faisant partie de la voie de signalisation mTOR déjà décrite précédemment au chapitre 1.2.3.4.e *PTEN*, *AKT3*, *PIK3CA*, *MTOR*, *PPP2R5D*) ¹³⁰.

Grâce aux progrès technologiques, des analyses épigénétiques génome entier sont possibles chez les patients atteints de syndromes génétiques et porteurs de variations dans ce type de gènes impliqués dans la régulation épigénétique. Des épi-signatures spécifiques de certains syndromes ont ainsi pu être identifiées, comme pour l'état de méthylation de l'ADN dans les syndromes de Sotos (par mutation *NSD1*)¹³¹, ATRX (α-thalassémie liée à l'X par mutation du gène ATRX/XNP)¹³², de Claes-Jensen par mutation de KDM5C¹³³, ou du spectre Coffin-Siris - Nicolaides-Baraitser¹³⁴.

Des progrès majeurs dans la compréhension de la régulation épigénétique de la plasticité neuronale dans les troubles du développement neurologique ont été permis grâce aux études fonctionnelles dans le syndrome de Rett ¹³⁵ (figure 12) : il est dû à des mutations perte de fonction du gène *MECP2* situé sur le chromosome X, touche principalement des filles, et se caractérise par une régression psychomotrice survenant entre 1 et 3 ans, faisant suite à un développement quasi normal au cours de la 1ère année de vie, une microcéphalie secondaire, une perte de l'utilisation volontaire des mains, des stéréotypies manuelles, une quasi absence de langage, et des traits autistiques. La protéine MeCP2 a un rôle de liaison aux groupements méthyl-CpG spécifiquement dans le cerveau ¹³⁶. On retrouve des répétitions de dinucléotides CpG (5'-CG-3') réparties de façon non uniforme dans le génome, avec un enrichissement au niveau des régions promotrices et/ou du premier exon de 60% des gènes humains, ce qui reflète l'importance de ces répétitions dans la régulation de l'expression génique ¹³⁷. L'hyperméthylation des îlots CpG entraîne une inhibition de l'expression des gènes. Ces régions méthylées sont reconnues par des protéines du groupe des MBP (Methyl Binding Proteins), auquel MeCP2 appartient, dont le rôle est de recruter des complexes de remodelage chromatinien empêchant l'accès des facteurs de transcription à l'ADN ¹³⁸.



Figure 12 : Illustration des différentes fonctions de MeCP2 comme répresseur transcriptionnel par le recrutement de cofacteurs répresseurs (A), comme activateur transcriptionnel (B), comme interagissant avec l'histone H1 pour coordonner l'expression génique (C), comme régulateur des cellules gliales (D), comme impliqué dans la morphologie neuronale (E), et régulation par les micro-ARN qui ont notamment un rôle dans les phénomènes d'addiction (F) ¹³⁹.

Sur le plan morphologique, cela s'accompagne d'anomalies de densité et de maturation dendritique, et sur le plan fonctionnel d'une altération de la plasticité synaptique des neurones excitateurs et d'une inhibition accrue, la baisse d'activité de MeCP2 entraîne la dérégulation des gènes cible dans les neurones matures ^{136,140–142}. Cependant, des études complémentaires ont montré que MeCP2 a également un effet d'activation de la transcription ¹⁴³ et jouent par ailleurs d'autres fonctions nucléaires comme la régulation et la liaison à des ARN via leurs protéines d'interaction ¹³⁹. MeCP2 est aussi exprimé dans les cellules gliales et joue également un rôle dans ces cellules pour la morphogenèse dendritique des neurones, ce qui pourrait aussi participer à la progression de la maladie ^{141,142,144}.

1.2.3.3. Autres voies biologiques et métaboliques impliquées dans la DI

a. Voie de biosynthèse du cholestérol

Elle est essentielle pour l'organogenèse. Un défaut d'une enzyme dans cette voie est responsable d'une DI avec microcéphalie et malformations diverses.

Par exemple, le déficit en 7-DHCR (7-dehydrocholesterol réductase), dernière enzyme de la voie, est responsable du syndrome de Smith-Lemli-Opitz, qui est caractérisé par un retard de croissance pré et post-natal, des malformations cardiaque, des extrémités (polydactylie post-axiale des mains et des pieds, syndactylie II-III des orteils), fente palatine, anomalies génitales chez le garçon. Il s'agit de l'affection la plus fréquente touchant cette voie.

D'autres pathologies plus rares touchant des enzymes en amont ont des phénotypes chevauchant comme la desmostérolose et la lathostérolose qui sont dues à des déficits des enzymes 3-betahydroxysterol-delta-24-reductase et 3-beta-hydroxysteroid-delta(5)-desaturase (lathosterol 5desaturase) respectivement.¹⁴⁵

b. Maladies de surcharge lysosomale

Cette entité regroupe de très nombreuses pathologies dont certaines peuvent s'associer à des troubles cognitifs et fréquemment des anomalies de substance blanche, le plus souvent il existe une dégradation progressive des symptômes notamment des troubles cognitifs avec une DI qui s'installe dans les formes à début précoce.

La maladie de Tay-Sachs représente un exemple de ce groupe de pathologies, il existe un spectre de sévérité variable dû à un déficit en hexosaminidase A. La forme précoce se caractérise par l'apparition progressive de symptômes dont une faiblesse musculaire et une perte des capacités motrices, perte d'attention, augmentation de la réaction de sursaut, une neurodégénérescence pouvant s'accompagner d'épilepsie, une perte visuelle, une spasticité conduisant à un état grabataire et un décès prématuré en général avant 4 ans. Une forme moins sévère est caractérisée par l'apparition plus tardive et plus progressive des symptômes avec évolution vers des troubles psychiatriques. ¹⁴⁶

La leucodystrophie métachromatique par mutation du gène *ARSA* responsable d'un déficit de l'activité arylsulfatase A est responsable d'une dégradation progressive des fonctions neurologiques y compris cognitives à un âge de début variable et s'accompagne d'anomalies de la substance blanche ¹⁴⁷.

D'autres maladies lysosomales ont un profil évolutif similaire comme la maladie de Krabbe (début par une irritabilité importante, une spasticité et un retard neurodéveloppemental avant l'âge d'un an dans sa forme précoce) par déficit en galactocérebrosidase (gène *GALC*)¹⁴⁸, la maladie de

Niemann-Pick A par déficit en sphingomyélinase acide (gène *SMPD1*)¹⁴⁹, la maladie de Canavan par déficit en aspartoacylase ¹⁵⁰, la gangliosidose à GM1 par mutation du gène *GLB1* entraînant un défaut de fonction de la β -galactosidase ¹⁵¹.

Certaines mucopolysaccharidoses (notamment de type I par déficit en Alpha-L-Iduronidase) s'accompagnent d'un retard neurodéveloppemental avec régression ¹⁵². De même, en plus de symptômes principalement squelettiques, les mucolipidoses de type II/III par mutation du gène *GNPTAB* codant pour les sous-unités alpha/beta de la N-acetylglucosamine-1-phosphotransferase peuvent se manifester par quelques difficultés d'apprentissage ¹⁵³.

Les céroïdes-lipofuscinoses représentent un groupe de maladies de surcharge lysosomale caractérisées par un âge de début variable selon le type et le gène en cause avec une détérioration intellectuelle et neurologique progressive, une épilepsie myoclonique, une atteinte rétinienne dans la plupart des cas et un décès précoce. Ces maladies se transmettent selon une hérédité autosomique récessive, les gènes suivants ont été impliqués dans ce groupe de maladie : *PPT1, TPP1, CLN3, CLN5, CLN6, MFSD8, CLN8, CTSD, DNAJC5, CTSF, ATP13A2, GRN, KCTD7*¹⁵⁴.

c. Maladies peroxysomales

Le spectre des maladies peroxysomales est large, allant de troubles neurodéveloppementaux légers à très sévères. La forme la plus sévère correspond au syndrome de Zellweger : les patients ont souvent des signes précoces dont une hypotonie, des troubles de déglutition, il peut exister des traits morphologiques particuliers, notamment une fontanelle large, des anomalies de migration neuronale comme une polymicrogyrie, des anomalies osseuses comme une chondrodysplasie ponctuée, des kystes rénaux et une insuffisance hépatique ; le décès est précoce. Les formes plus légères ne sont pas associées à des malformations mais ont des altérations progressives de certains organes ou systèmes (dystrophie rétinienne, déficit auditif de perception, neurodégénérescence avec ataxie, polyneuropathie, leucodystrophie, dysfonction hépatique, insuffisance surrénalienne, lithiase des voies urinaires de type oxalique, amélogenèse imparfaite) ; il existe souvent une hypotonie et un retard neurodéveloppemental mais les fonctions cognitives peuvent ne pas être touchées.

Il s'agit le plus souvent de maladies autosomiques récessives. Des mutations dans 13 gènes impliqués dans les fonctions peroxysomales (biogenèse et assemblage, voie métabolique incluant import et recyclage) ont été décrites : *PEX1, PEX2 (PXMP3), PEX3, PEX5, PEX6, PEX10, PEX11b, PEX12, PEX13, PEX14, PEX16, PEX19, PEX26*¹⁵⁵.

d. Maladies congénitales de la glycosylation

Plus de 130 types de maladies de glycosylation ont été répertoriées comme associées à un défaut dans une voie de glycosylation, les voies principales étant celles de la N-glycosylation, de la O-glycosylation, des glycosylphosphatidylinositols et des dolichols.

La majorité de ces maladies se manifestent dès l'enfance avec des troubles de déglutition, un retard neurodéveloppemental, une atteinte hépatique, des anomalies neurologiques dont une hypotonie, des hypoglycémies, une entéropathie avec perte protéique, des anomalies ophtalmologiques, des anomalies immunologiques, des atteintes cutanées et des signes squelettiques ; cependant, certaines maladies ont une expression isolée à un organe ou un système comme EXT1, 2, 3 dans la maladie des exostoses multiples. De nombreux types concernent très peu de patients et l'étendue du spectre phénotypique n'est pas bien connue.

d1. Anomalie de la voie de la N-glycosylation

Elle correspond à l'ajout par lien covalent d'une chaîne de sucres à la chaine latérale d'un groupe amide d'une asparagine dans un motif peptidique Asn-X-Ser/Thr. Elle fait intervenir plusieurs enzymes successives au niveau du réticulum endoplasmique avec des modifications dans l'appareil de Golgi.

Les gènes impliqués uniquement dans cette voie sont : *PMM2, MPI, ALG6, ALG3, DPM1, MPDU1, ALG12, ALG8, ALG2, DPAGT1, ALG1 (HMT-1), ALG9, DOLK (DK1), RFT1, DPM3, ALG11, SRD5A3, DDOST, MAGT1, TUSC3, ALG13, PGM1, MGAT2, STT3A, STT3B, SSR4, MOGS (GCS1), SLC35C1, B4GALT1, SLC35A2, GMPPA.*

Certains gènes jouent un rôle dans plusieurs voies métaboliques dont celle de la N-glycosylation : COG7, SLC35A1, COG1, COG2, COG8, COG5, COG4, TMEM165, COG6, DPM2, DHDDS, MAN1B1, PGM3.

En dehors de *MAGT1*, *ALG13* et *SLC35A2* associés à une hérédité liée à l'X, et *GANAB* et *PRKCSH* associés à une hérédité autosomique dominante, tous les autres gènes sont associés à un mode de transmission autosomique récessif.¹⁵⁶

Le diagnostic biochimique consiste à étudier la glycosylation de la transferrine qui contient habituellement une chaîne de 2 ramifications de groupements carbohydrates reliés par une asparagine avec 4 groupes d'acide sialique. Dans l'anomalie de profil de type 1, on observe une diminution de la tetrasialotransferrine et une augmentation des formes asialotransferrine et disialotransferrine, ce qui correspond à un défaut dans les étapes précoces de la voie métabolique. Dans l'anomalie de profil de type 2, il existe une augmentation des formes trisialo- et monosialo-transferrine par défaut de ramification, correspondant à des étapes avancées de cette voie ^{157,158}.

Une des maladies les plus fréquentes est le déficit en phosphomannomutase de type 2 (gène *PMM2*) et dont les signes caractéristiques sont des mamelons ombiliqués, une anomalie de la répartition des graisses avec aspect de lipodystrophie de la région fessière, une hypoplasie cérébelleuse avec parfois malformation de Dandy-Walker, des anomalies de coagulation, des déformations squelettiques, un vieillissement prématuré. Les formes très précoces conduisent généralement au décès dans les
premiers mois à années de vie et celles débutant dans l'enfance sont associées à une DI de degré variable ¹⁵⁹.

d2. Anomalie de la voie de la O-glycosylation

Cette voie comprend plusieurs étapes d'ajout de sucres à des sérines, thréonines, ou hydroxylysines par des glycosyltransférases situées dans l'appareil de Golgi. Il existe 7 différents types de chaînes sucrées rajoutées qui sont classées en fonction du 1^{er} sucre ajouté à l'acide aminé : O-Nacetylglucosaminyl (GlcNAc), O-N-acetylgalactosaminyl (GalNAc), O-galactosyl (Gal), O-xylosyl (Xyl), O-mannosyl (Man), O-glucosyl (Glc), O-fucosyl (Fuc).

Plusieurs gènes sont associés à des maladies humaines. Par exemple le groupe des α -dystroglycanopathies correspond à un défaut dans la O-mannosylation (cf. chapitre 1.2.3.1.b, figure 5C).

Un autre exemple est représenté par les variations pathogènes du gène *OGT* (codant pour O-linked beta-N-acetylglucosamine, O-GlcNAc transferase, gène situé sur le chromosome X) qui ont été incriminées dans un tableau de DI, avec signes associés variables comme une microcéphalie, des signes pyramidaux et des signes morphologiques (hypertélorisme, oreilles basses implantées, nez large, lèvres épaisses, mamelon surnuméraire, rétrognatisme, doigts longs et fins) chez 5 garçons de 3 familles différentes ^{160,161}.

d3. Voie de synthèse du glycosylphosphatidylinositol (GPI) et glycosylation des lipides

L'ancre GPI correspond à une chaîne de glycolipides formée d'une structure principale d'éthanolamine phosphate, permettant la liaison aux protéines, suivie de 3 mannoses, un résidu glucosamine, un résidu phosphatidylinositol (PI), et une queue lipidique saturée permettant l'incorporation des GPI dans la bicouche lipidique des membranes (figure 13). L'assemblage se fait de façon séquentielle grâce à de multiples enzymes successives débutant sur le versant cytosolique du réticulum endoplasmique. La 1^{ère} étape est le transfert de GlcNAc à partir de l'UDP-GlcNAc au phosphatidylinositol (PI) pour former GlcNAc-PI ; cette étape de GPI- GlcNAc transférase est médiée par 6 enzymes (PIG-A, PIG-C, PIG-H, PIG-P, PIG-Q, PIG-Y). La 2^{ème} étape est la dé-N-acétylation en GlcN-PI par PIG-L ce qui intègre la chaîne à l'intérieur du réticulum endoplasmique. Puis le GlcN-PI est acylé par PIG-W pour former GlcN-(acyl)PI et des chaines de sucres dérivées du Mannose sont ajoutées par PIG-M et PIG-V. PIG-N modifie un Mannose avec un groupe ethanolamine phosphate et PIG-B, PIG-O et PIG-F permettent d'obtenir un groupement GPI mature. Cette voie permet à termes d'amarrer des centaines de protéine à la membrane plasmique qui auront de multiples fonctions.

Désormais, la plupart des gènes codant pour ces protéines ont été impliqués dans une anomalie du développement associant DI, hypotonie, épilepsie, malformations (squelettiques notamment des pouces, ophtalmologiques et cardiaques) dû à des mutations perte de fonction avec un mode de



transmission autosomique récessif pour tous les gènes sauf *PIGA* lié à l'X. Un des marqueurs de la maladie est l'hyperphosphatasie (élévation des phosphatases alcalines par défaut de son accrochage à la membrane plasmique qui se retrouve dans le sang circulant) et des tests ont été développés par cytométrie en flux pour détecter des perturbations de l'exposition de protéines à la surface des cellules sanguines notamment immunitaires ^{162–164}.

Figure 13 : Structure du glycosylphosphatidylinositol pour permettre l'amarrage des protéines aux membranes cellulaires.

Les glycosphingolipides et leurs dérivés sialylés nommés gangliosides sont exprimés principalement dans les neurones. Les défauts de leur dégradation sont responsables d'accumulation et sont inclus dans les maladies lysosomales. A l'inverse, le défaut de biosynthèse des gangliosides comme ST3GAL5-CDG et B4GALNT1-CDG est exceptionnel et est responsable de maladies neurodégénératives sévères. Les patients présentent une paraparésie spastique, une DI sévère, une épilepsie et des signes extra-neurologiques comme une dysplasie squelettique, des signes morphologiques et des anomalies pigmentaires ^{165,166}.

1.2.3.4. DI syndromique avec malformation cérébrale

a. Holoprosencéphalie : voie SHH

Il s'agit d'une malformation relativement fréquente estimée à 1/250 des grossesses et 1/8000 des naissances vivantes ^{167,168}, la plupart des cas conduisant à une fausse couche, étant détecté pendant la grossesse ou pouvant conduire à des demandes d'interruption de grossesse pour raison médicale devant le pronostic réservé. Il existe 3 grades de sévérité décroissante : forme alobaire (un seul ventricule et absence de scissure inter-hémisphérique), forme semilobaire (séparation partielle des 2 hémisphères cérébraux avec un seul ventricule), et forme lobaire (ventricules séparés mais séparation

incomplète des régions frontales des hémisphères) ; il existe également une forme mineure où il peut exister une hypoplasie du corps calleux, un hypotélorisme, et une incisive supérieure médiane ¹⁶⁹.

En dehors de la trisomie 13 qui représente la cause la plus emblématique, de nombreuses variations chromosomiques, le plus souvent pour des cas isolés (1q43q44, 3p25.2p22.1, 2p15, 2p11.2, 16p13.11, 3p22, 20q11.2, 3q25.32, 4q12, 5q35.3, 5q35.3, 7p22.1, 7p22.1, 8q23.3q24.11, trisomie 8 en mosaïque, 10p15.3p14, 11q13.4, 12q21.32, 14q23.1, 14q23.1, 15q11.2, 16p13.11, 16p11.2, 18q22.1, 19q13.42q13.43, 22q11.21, Xq25q28), et de nombreux gènes (*SHH*, *ZIC2*, *GLI2*, *SIX3*, *FGF8*, *FGFR1*, *DISP1*, *DLL1*, *TGIF1*, *SUFU*, *STIL*, *GAS1*, *TDGF1*, *CDON*, *FOXH1*, *NODAL*, *BOC*) ont été impliqués avec possibilité de pénétrance incomplète, variabilité d'expression, et parfois une hérédité complexe oligogénique ¹⁷⁰. Tous ces gènes sont impliqués dans la régulation de la voie sonic hedgehog qui définit un gradient morphogénétique de la ligne médiane du diencéphale pour permettre une séparation normale des hémisphères cérébraux et des régions optiques ^{171,172}.

b. Lissencéphalie ou agyrie et pachygyrie

Il s'agit d'une malformation pour laquelle le cerveau a des gyri et sillons absents ou très grossiers, le cortex est souvent anormalement épais avec 4 couches primitives et il existe parfois une hétérotopie sous-corticale en bande, séparée du cortex par une fine zone de substance blanche. Cette dernière malformation peut être isolée et représente un équivalent mineur de lissencéphalie ; les ventricules cérébraux ont souvent une morphologie anormale et le corps calleux est souvent hypoplasique Cela est dû à une anomalie de migration des neurones entre la 9^{ème} et la 13^{ème} semaine de grossesse.

De nombreux gènes sont en cause dont le mode de transmission est variable, soit dans une forme syndromique (syndrome de Miller-Diecker par délétion hétérozygote de la région 17p13.3 incluant les gènes *PAFAH1B1* et *YWHAE*, syndrome de Proud ou XLAG - X-linked lissencephaly and ambiguous genitalia - par mutation du gène *ARX*, *CTU2*...) soit non syndromique (*PAFAH1B1*, *RELN*, *TUBA1A*, *NDE1*, *LAMB1*, *KATNB1*, *CDK5*, *TMTC3*, *MACF1*, *DCX*, *ARX*...). La plupart des gènes ont un rôle dans la migration en jouant en particulier sur le positionnement du noyau dans la cellule (*PAFAH1B1*, *NDE1*) ou sur le cytosquelette (tubulinopathies comme *TUBA1A*)¹⁷³.

Le syndrome de Baraitser-Winter constitue une forme syndromique de pachygyrie et regroupe un hypertélorisme souvent marqué, un nez large, un ptosis, une saillie de la suture métopique, des sourcils très arqués, un colobome irien et/ou rétinien, un déficit auditif de perception, un aspect proéminent de la ceinture scapulaire, une raideur articulaire évolutive et une pachygyrie à prédominance postérieure (rarement une vraie lissencéphalie ou des hétérotopies neuronales). On peut parfois retrouver une fente labiale et/ou palatine, une duplication des hallux, une malformation cardiaque, du tractus urinaire, une microcéphalie évolutive, et/ou une arthrogrypose. Le degré de DI et d'épilepsie est variable et dépend de la sévérité des malformations cérébrales ¹⁷⁴. Le spectre des α-dystroglycanopathies correspond à une lissencéphalie pavimenteuse par migration excessive d'éléments neurogliaux, au-delà de la limite fixée habituellement par la barrière gliale, dans les espaces méningés, ce qui entraîne un épaississement de la leptoméninge qui adhère fortement au cortex et une obstruction des espaces sous-arachnoïdiens responsable d'une hydrocéphalie constante. Il s'agit d'un spectre malformatif incluant le syndrome de Walker-Warburg (WWS) le plus sévère, le syndrome MEB (muscle eye brain) et un type de dystrophie musculaire congénitale. Ces atteintes sont dues à un défaut de la voie de glycosylation de l'α-dystroglycane (cf. chapitre 1.2.3.3.d2). Les principaux gènes en cause sont *POMT1, POMT2, POMGNT1, FKTN, FKRP, LARGE, ISPD, GTDC2 (POMGNT2), DAG1, TMEM5 (RXYLT1), B3GALNT2, SGK196 (POMK), B3GNT1 (B4GAT1), GMPPB ¹⁷⁵.*

c. Microcéphalie primaire (figure 5A)

Ce signe correspond à un petit périmètre crânien < -3DS congénital avec défaut de croissance post-natale, sans malformation cérébrale associée lorsqu'elle est isolée. Elle s'accompagne de troubles des apprentissages variables pouvant inclure une DI. La cause principale est un défaut de fonction des microtubules et/ou du centrosome, une fonction ciliaire altérée, ou une anomalie dans les voies de réparation de l'ADN qui aboutissent à une diminution de prolifération des progéniteurs entraînant donc une diminution du nombre de neurones et une plus petite taille du cerveau ^{50,176}. Les gènes les plus fréquemment impliqués codent pour des protéines du centrosome et de la région pericentriolaire et plus de 25 sont connus ou suspectés : *MCPH1, CDK5RAP2, CASC5, ASPM, CENPJ, STIL, CEP135, CEP152, ZNF335, PHC1, CDK6, CENPE, SASS6, MFSD2A, ANKLE2, CIT, WDFY3, COPB2, KIF14, NCAPD2, NCAPD3, NCAPH, NUP37, MAP11 ¹⁷⁷.*

d. Hétérotopies nodulaires péri-ventriculaires

Il s'agit d'une anomalie de migration neuronale où certains neurones ne migrent pas normalement vers le cortex entrainant la persistance de certains foyers de substance grise dans la zone périventriculaire.

Le gène majeur impliqué dans ce type de malformation est *FLNA*, situé dans la région Xq28 dont les variations perte de fonction sont en cause. Cette malformation touche principalement des filles, l'atteinte étant létale chez le garçon dans la grande majorité des cas. L'épilepsie est assez fréquente mais la pénétrance est incomplète allant de personnes asymptomatiques à d'autres ayant plus rarement une épilepsie pharmaco-résistantes et une DI. Ce type d'atteinte peut être associé à une dilatation de l'aorte, un syndrome d'Ehlers-Danlos, une thrombopathie ¹⁷⁸. (A noter qu'il existe une corrélation génotype-phénotype puisque les variants faux sens gain de fonction de ce gène sont responsables d'un spectre malformatif de type oto-palato-digital avec dysplasie osseuse, fente palatine postérieure, déficit auditif et plusieurs atteintes malformatives selon la gravité touchant les garçons

plus sévèrement que les filles, avec rarement une DI mais des troubles dans les apprentissages sont possibles ¹⁷⁹).

D'autres causes plus rares ont été décrites : *ARFGEF2* (autosomique récessif, associé à une DI et une microcéphalie), *ERMARD*, *NEDD4L*, *ARF1*, anomalies chromosomiques dans la région 5p ou 5q.

e. Macrocéphalie/mégalencéphalie et polymicrogyrie

La mégalencéphalie correspond à une augmentation du périmètre crânienne >+2/+3DS avec augmentation de taille proportionnelle des structures cérébrales. Elle a été décrite comme associée à plusieurs formes syndromiques avec excès de croissance comme le syndrome de Sotos par mutation du gène *NSD1* mais peut aussi être au 1^{er} plan d'un tableau de DI sévère comme pour les variants perte de fonction des gènes *HERC1* qui code pour une protéine impliquée dans la voie de croissance mTOR ¹⁹ ou *NONO / NRB54* qui joue un rôle dans le métabolisme des ARNm au niveau du noyau cellulaire ¹⁸⁰.

La maladie d'Alexander représente une autre cause de DI avec macrocéphalie. Il existe différents degrés de sévérité de la maladie avec une DI ou une détérioration cognitive, une hydrocéphalie et des anomalies de substance blanche. La maladie se transmet selon une hérédité autosomique dominante et est due à des mutations du gène *GFAP* codant pour la glial fibrillary acidic protein, un des composants principaux des filaments intermédiaires exprimé dans les astrocytes matures du système nerveux central. La protéine a notamment un rôle important dans la morphologie et la mobilité des astrocytes ^{115,181,182}.

La polymicrogyrie consiste en une anomalie de lamination corticale et des circonvolutions cérébrales par la présence d'un excès de petits gyri dans tout ou partie du cortex ; il s'agit d'une des malformations les plus fréquentes (environ 20% des malformations corticales). Elle peut être uni ou bilatérale, symétrique ou non, isolée ou associée à d'autres malformations cérébrales comme des hétérotopies de substance grise, une ventriculomégalie, des anomalies de substance blanche, du corps calleux, du tronc cérébral et/ou du cervelet.

En dehors des causes vasculaires ou infectieuses (CMV) qui sont fréquemment impliquées, certaines causes génétiques sont particulièrement responsables de polymicrogyrie comme des variations de nombre de copie chromosomique (syndrome de délétion 22q11.2 ou 1p36), des gènes de la voie mTOR (*AKT3, CCND2, MTOR, PI4KA, PIK3CA, PIK3R2, PTEN*) en cause notamment dans le syndrome MPPH (Megalencephaly-Postaxial Polydactyly-Polymicrogyria-Hydrocephalus), des tubulinopathies (*TUBA1A, TUBB, TUBB2A, TUBB2B, TUBB3*), le groupe des α-dystroglycanopathies, et de nombreux autres gènes (*ADGRG1 (GPR56) COL3A1, ATP6V0A2, LAMA2, LAMB1, LAMC3, SNAP29, SRD5A3, BICD2, COL18A1, DDX3X, EML1, EOMES (TBR2), EZH2, FIG4, GPSM2*,

GRIN1, GRIN2B, KIFBP (KIAA1279), MAP1B, NDE1, NEDD4L, OCLN, OFD1, PAX6, RAB18, RAB3GAP1, RAB3GAP2, RTTN, TBC1D20, TCTN1, TMEM216, WDR62) incluant certains responsables de maladies métaboliques (hyperglycinémie sans cétose, acidurie fumarique de type 1, acdémie glutarique de type 2, histidinémie, leucinose, maladies mitochondriales, déficit en pyruvate déshydrogénase, maladies peroxysomales) ¹⁸³.

Les mutations des gènes de la voie mTOR sont particulièrement en cause dans le spectre MPPH qui inclut les 2 signes précédents (macrocéphalie - polymicrogyrie) mais peuvent aussi être responsables d'anomalies associées des membres, de pigmentation ou vasculaire comme dans le syndrome MCAP (mégalencéphalie - malformation capillaire) dû à des mutations activatrices en mosaïque du gène *PIK3CA* ¹⁸⁴.

f. Schizencéphalie, porencéphalie

La majorité des causes de schizencéphalie/porencéphalie est clastique par accident vasculaire, mais une des causes génétiques classiques est la collagénopathie de type 4 par mutation des gènes *COL4A1* ou *COL4A2*. La variabilité d'expression est très importante et ce groupe de pathologie peut se présenter par ce type de malformation cérébrale mais aussi une leucoencéphalopathie, des lacunes cérébrales par infarctus, des hémorragies intra-parenchymateuses, des atteintes oculaire (anomalie du segment antérieur de type Axenfeld-Rieger, cataracte congénitale, tortuosité des artères rétiniennes), crampes musculaires, élévation des CK, atteinte rénale, phénomène de Raynaud, arythmie cardiaque, anémie hémolytique. La sévérité de la DI et de l'épilepsie est corrélée à l'atteinte cérébrale ¹⁸⁵.

D'autres causes peuvent être impliquées dans ce type de malformation, qu'elles soient génétiques ou non, comme les coagulopathies (maladie de Willebrand), les thrombopénies et thrombopathies.

g. Hydrocéphalie

Elle correspond à une augmentation du volume du LCR, responsable d'un élargissement des cavités ventriculaires. Elle peut être obstructive, l'exemple le plus caractéristique est la sténose de l'aqueduc du mésencéphale associée aux mutations du gène *L1CAM* dont le phénotype comprend une DI avec altération franche du langage, une spasticité, une agénésie partielle ou complète du corps calleux, des pouces en adduction. Il s'agit d'une pathologie liée à l'X touchant principalement les garçons ¹⁸⁶.

D'autres gènes sont identifiés régulièrement comme associés à des formes d'hydrocéphalie importante sur un mode autosomique récessif (*MDPZ*, *CCDC98*, *EML1*, *WDR81*) ou lié à l'X (*AP1S2*) ¹⁸⁷.

h. Anomalies en mosaïque - hémimégalencéphalie et dysplasie corticale focale

Des mutations somatiques dans la voie PI3K-AKT-mTOR ont été identifiées comme la cause de dysplasie corticale, d'épilepsie avec ou sans DI et d'hémimégalencéphalie (figure 10). Cette malformation cérébrale est caractérisée par un hémisphère cérébral élargi et malformé. La présentation clinique comprend généralement une DI et une épilepsie sévère et pharmacorésistante nécessitant souvent une intervention chirurgicale ^{188,189}.

La dysplasie corticale focale correspond à un versant plus atténué du spectre avec la même voie et les mêmes gènes impliqués, et une présentation clinique plus marquée sur le plan épileptique que sur la DI. Un des exemples caractéristiques est l'implication du gène *DEPDC5* dont des variations hétérozygotes se transmettant selon une hérédité autosomique dominante sont responsables d'épilepsie focale sans malformation cérébrale de pénétrance incomplète ; il a pu être identifié des individus dans certaines familles qui présentaient une dysplasie corticale focale qui correspond en fait à un 2^{ème} événement consistant en une perte d'hétérozygotie en mosaïque dans une région cérébrale. Ce gène code pour une protéine inhibant la voie mTOR et on retrouve comme précédemment une activation excessive de cette voie, cette fois-ci par levée d'inhibition ¹⁹⁰.

Les tubers corticaux représentent également des éléments corticaux focaux de type hamartome pourvoyeurs d'épilepsie, en rapport avec la sclérose tubéreuse de Bourneville (STB), une maladie génétique due à des variations perte de fonction des gènes *TSC1* et *TSC2* dont les protéines forment un complexe inhibant la voie mTOR. Contrairement au gène précédent dont les variations entraînent une expression principalement cérébrale, la STB se manifeste par d'autres symptômes neurologiques (DI et/ou troubles autistiques dans 20 à 30% des cas, nodules sous-épendymaires, astrocytomes à cellules géantes), ophtalmologiques (hamartomes rétiniens), rénaux (angiomyolipomes, kystes, cancers), cutanéo-phanériens (fibromes faciaux, unguéaux, buccaux, tâches cutanées achromiques, zones cutanées en peau de chagrin), pulmonaires (lymphangioléiomyomatose), cardiaques (rhabdomyomes, arythmie) ¹⁹¹.

Une suractivation de la voie mTOR peut être ciblée par la rapamycine (ou sirolimus et molécules apparentées comme everolimus) qui a montré son efficacité dans plusieurs des pathologies précédentes ^{190,192}.

i. Anomalies de la fosse postérieure

i1. Hypoplasie vermienne

Ce signe consiste en une diminution marquée de la taille du vermis sans élargissement des sillons et fissures qui délimitent les régions de cette structure.

Certaines anomalies vermiennes peuvent être relativement isolées en association avec une DI comme pour les mutations du gène *OPHN1*¹⁹³ ou bien s'intégrer dans un syndrome malformatif comme le syndrome de Joubert (dû à des mutations dans de multiples gènes de ciliopathie impliqués dans la fonction du cil primaire et qui associe ataxie, apraxie oculomotrice, troubles respiratoires, troubles des apprentissages variables pouvant inclure une DI, avec le signe classique de la dent molaire par horizontalisation des pédoncules cérébelleux supérieurs ¹⁹⁴) ou le syndrome COACH qui est une variante du syndrome précédent et inclut une fibrose hépatique et parfois des kystes rénaux et un colobome ¹⁹⁵.

i2. Malformation de Dandy-Walker

Il s'agit d'une malformation comprenant une agénésie partielle ou complète du vermis, une dilatation kystique de la fosse postérieure communiquant avec le 4^{ème} ventricule, et une surélévation de la tente du cervelet. Cette malformation peut être isolée et est responsable dans ce cas de troubles neurologiques, notamment une ataxie et une hypotonie, et dans la moitié des cas un retard des acquisitions et une DI.

Dans d'autres situations, elle s'intègre dans une forme syndromique comme des variations chromosomiques, le groupe des ciliopathies (oro-facio-digital/OFD), des α -dystroglycanopathies ou des syndromes plus complexes (PHACES). Le syndrome de Pettigrew en est un exemple, dû à des mutations du gène *AP1S2* qui code pour une protéine faisant partie du complexe adaptateur entre le versant cytoplasmique des vésicules de l'appareil de Golgi et la clathrine, ce qui entraînerait un défaut de développement et de fonction synaptique ¹⁹⁶.

Le syndrome 3C aussi appelé Ritscher-Schinzel est un syndrome malformatif comprenant des anomalies crânio-faciales (front et occiput proéminents, fentes palpébrales en bas et en dehors, dépression de la racine du nez, micrognatisme, oreilles basses implantées), une malformation cardiaque (communication inter-auriculaire ou inter-ventriculaire, sténose aortique), et une malformation cérébelleuse correspondant à un Dandy-Walker ; le retard neurodéveloppemental et les troubles cognitifs sont sévères ¹⁹⁷.

i3. Hypoplasie ponto-cérébelleuse

Cette malformation consiste en une hypoplasie à la fois du tronc cérébral, en particulier du pont, et du cervelet et est responsable d'un retard neurodéveloppemental très sévère avec neurodégénérescence fréquente. De nombreux gènes sont en cause : *VRK1, EXOSC3, EXOSC8, EXOSC9, TSEN54, TSEN2, TSEN34, SEPSECS, PCLO, RARS2, TOE1, CHMP1A, AMPD2, CLP1, TBC1D23, COASY*¹⁹⁸.

Parmi eux, certains codent pour des sous-unités d'endonucléase pour l'épissage d'ARN de transfert, très exprimés dans ces régions du système nerveux (gènes TSEN) ¹⁹⁹. Les protéines codées par les gènes EXOSC font partie d'un complexe impliqué dans la régulation et la dégradation rapide des ARN riches en éléments AU des exosomes par leur activité exonucléase 3'-5' ^{200,201}.

VRK1 a été surtout étudié sur le versant de la prolifération cellulaire et de la tumorigenèse car il existe une boucle de régulation avec p53 qui est altérée dans certaines tumeurs. VRK1 phosphoryle aussi d'autres facteurs de transcription comme c-JUN, ATF-2, et CREB et est important pour la progression de la phase G1 à S du cycle cellulaire. Par ailleurs, VRK1 joue un rôle au niveau de l'enveloppe nucléaire pour la formation des pores et dans l'organisation de la chromatine en phosphorylant l'histone H3 et BAF au cours de la mitose ce qui entraîne la perte d'interaction de cette dernière avec les protéines de l'enveloppe dont l'emerine. Tous ces éléments indiquent un rôle dans le développement et la maintenance du système nerveux central ²⁰².

j. Anomalies de la myéline et leucodystrophie

Un nombre important de pathologies ayant entre autres pour signes des anomalies de substance blanche peut s'accompagner de DI. Ces anomalies peuvent correspondre à :

- une atteinte dysmyélinisante comme pour les mutations ou duplications du gène *PLP1* associées au syndrome de Pelizaeus-Merzbacher incluant des signes neurologiques notamment une spasticité et une ataxie. D'autres gènes sont associés à une atteinte plus syndromique comme le syndrome 4H dû à des mutations des gènes *POLR3A*, *POLR3B*, *POLR1C* qui associe hypodontie, hypomyélinisation et hypogonadisme hypogonadotrope). Les autres gènes pouvant être impliqués sont : *GJC2/GJA12*, *AIMP1*, *HSPD1*, *FAM126A*, *TUBB4A*, *RARS*, *PYCR2*, *VPS11*, *HIKESHI*, *UFM1*, *EPRS*, *TMEM106B*, *AIMP2*, *DEGS1*²⁰³.

- à une atteinte progressive après myélinisation normale : ce champ regroupe un nombre très important de pathologies dont les maladies métaboliques par trouble des fonctions lysosomales et peroxysomales (chapitre 1.2.3.3.b/c) ; un autre type d'atteinte consiste en une leucodystrophie avec perte de la substance blanche (Vanishing White Matter) dû à des mutations bialléliques des gènes *EIF2B1-5* entraînant un trouble des fonctions astrocytaires et qui s'accompagne de signes neurologiques variables comme une ataxie, une spasticité, une altération des fonctions cognitives ayant un début également variable ²⁰⁴.

Au total, les gènes impliqués dans la DI peuvent être responsables d'atteinte cognitive isolée en jouant un rôle dans le fonctionnement du système nerveux central (DI non syndromique) mais ils peuvent aussi avoir un rôle dans la morphogenèse cérébrale et de multiples organes et entraîner une forme syndromique (signes neurologiques, syndrome malformatif associé).

Un article publié en 2016 illustre notamment le rôle des 818 gènes impliqués dans la DI (figure 14).



*Figure 14 : Analyse pour les 818 gènes impliqués dans la DI en 2016 des processus biologiques (a) et des compartiments cellulaires (b) mis en cause grâce au logiciel REViGO (Reduce + Visualize Gene Ontology) et à la base de données DAVID (Database for Annotation, Visualization, and Integrated Discovery)*⁴⁰.

e CenvelopeD e

cell

projection

organelle

organ

cillary membrar

45

ucleoid

synaptic cleft

/esicula

raction

1.3. Autres pathologies et manifestations cliniques associées à la DI abordées dans le cadre de cette thèse

1.3.1. Les troubles du spectre autistique (TSA) : définition et causes

Cette dénomination est apparue dans la 5^{ème} version du DSM (Diagnostic and Statistical Manual of Mental Disorders), édition actualisée en 2013. Les critères diagnostiques sont les suivants ²⁰⁵:

A - Des déficits persistants dans la communication sociale et les interactions sociales dans de multiples contextes :

Déficits de la réciprocité socio-émotionnelle, allant, par exemple, de l'approche sociale anormale et l'incapacité d'échanger dans une conversation ; au partage réduit d'intérêts, d'émotions, ou d'affect ; à l'échec d'engager ou de répondre à des interactions sociales. Déficits dans les comportements de communication non verbaux utilisés pour l'interaction sociale, allant, par exemple, de la communication verbale et non verbale mal intégrée ; à des anomalies dans le contact visuel et le langage du corps ou des déficits dans la compréhension et l'utilisation de gestes : à un manque total d'expressions faciales et de communication non verbale. Déficits dans le développement, le maintien et la compréhension des relations, allant, par exemple, de difficultés à adapter le comportement en fonction de divers contextes sociaux à des difficultés à partager les jeux imaginatifs ou à se faire des amis ; à l'absence d'intérêt pour les pairs.

<u>B - des modes restreints et répétitifs de comportements, d'intérêts ou d'activités :</u>

Mouvements moteurs, utilisation d'objets, ou parole stéréotypés ou répétitifs (par exemple, stéréotypies motrices simples, aligner des jouets ou retourner des objets, écholalie, phrases idiosyncrasiques (hors contexte)). Insistance sur l'adhésion inflexible à des habitudes ou modes ritualisés de comportement verbaux ou non verbaux (par exemple, une détresse extrême en cas de petits changements, difficultés avec les transitions, modes de pensée rigide, rituels de salutation, besoin de prendre le même itinéraire ou de manger la même nourriture tous les jours). Intérêts très restreints et circonscrits qui sont anormaux dans leur intensité ou leur orientation (par exemple, un fort attachement à des objets inhabituels, des intérêts excessivement circonscrits ou poursuivis avec une persévération excessive). Hyper- ou hyporéactivité à des stimuli sensoriels ou niveau d'intérêt inhabituel pour les aspects sensoriels de l'environnement (par exemple, indifférence apparente à la douleur/température, réaction négative à des sons ou des textures spécifiques, sentir ou toucher excessivement des objets, fascination visuelle pour des lumières ou mouvements).

<u>C</u> - Les symptômes doivent être présents dans la période de développement précoce (mais peuvent ne devenir pleinement manifestes qu'après que les exigences sociales dépassent les capacités limitées, ou peuvent être masqués par des stratégies apprises plus tard dans la vie).

<u>D</u> - Les symptômes causent une altération cliniquement significative du fonctionnement actuel dans les domaines sociaux, scolaires ou professionnels, ou d'autres domaines importants.

<u>E</u> - Ces perturbations ne sont pas mieux expliquées par la déficience intellectuelle (trouble de développement intellectuel) ou un retard global de développement. La déficience intellectuelle et le trouble du spectre de l'autisme surviennent fréquemment ensemble ; pour poser les deux diagnostics de trouble du spectre de l'autisme et de déficience intellectuelle, la communication sociale devrait être inférieure à celle prévue pour le niveau de développement général.

Une étude visant à comparer les critères de l'édition précédente DSM-IV-TR avec le DSM-5 a évalué 773 enfants remplissant les critères dont 288 qui présentaient un autre trouble neurodéveloppemental combiné. La concordance pour les troubles autistiques était bonne (0,78) et celle pour les troubles envahissant du développement non spécifié moyenne (0,57 et 0,59). Les enfants remplissant les critères DSM-IV-TR mais pas ceux du DSM-5 (71 enfants) avaient plus souvent des symptômes plus légers ou des symptômes en rapport avec une autre affection. Les enfants qui remplissaient les critères de TED du DSM-IV-TR mais pas du DSM-5 (66 enfants) ou inversement (120 enfants) avaient moins souvent une DI et étaient plus souvent des filles. La sensibilité et la spécificité du DSM-5 étaient donc meilleures (95% et 78% respectivement)²⁰⁶.

La prévalence est estimée actuellement à 1 personne/59 du fait d'une meilleure sensibilisation du dépistage et des modifications des critères diagnostiques, avec un sex ratio homme/femme autour de 3 ²⁰⁷. Les causes de l'autisme sont très variables, avec des facteurs environnementaux et génétiques, et une certaine redondance avec les voies physiopathologiques de la DI mais une architecture génétique plus complexe traduite par le fait que le rendement diagnostique dans les troubles autistiques sans DI est bien plus faible que dans la DI avec ou sans traits autistiques.

Les facteurs génétiques combinent des variations *de novo*, des variations héritées rares et des variations communes. Le rôle des gènes impliqués est très variable, chevauchant avec les gènes impliqués dans la DI (chapitre 1.2.3) et résumé sur les figures 8 et 15 ^{54,208,209} :

- facteurs de transcription et remodelage chromatinien influençant l'expression génique comme *TBR1, CHD8, ADNP, POGZ*.

- modificateurs épigénétiques comme MECP2

- régulation post-transcriptionnelle comme *FMR1* et autres protéines se liant aux ARN comme *CELF4* ou *ELAVL3*.

- architecture de la synapse (neurexines *NRXN1*, neuroligines *NLGN1,3,4*, *SHANK1,2,3*, cadhérines *CDH9,10*, contactines *CNTN4,6*, *CNTNAP2*)

- fonction de la synapse impactant la neurotransmission : canaux ioniques (*SCN2A, CACNA1C, CACNA1E, SCN1A, KCNQ3, KCNQ5, KCND2*), vésicules synaptiques (*SYN1, SYN2*), protéine activant la Ras-GTPase synaptique (*SYNGAP1*). Une exploration électrophysiologique

systématique de 10 lignées de cellules souches induites invalidant pour chacune un gène de susceptibilité à l'autisme (*FMR2, ANOS1, ASTN2, ATRX, CACNA1C, CHD8, DLGAP2, KCNQ2, SCN2A* et *TENM1*) par CRISPR-Cas 9 a montré des anomalies à type de diminution du potentiel post-synaptique des synapses excitatrices pour 5 d'entre eux (*ATRX, AFF2, KCNQ2, SCN2A*, et *ASTN2*) par rapport à leurs contrôles respectifs ²¹⁰

- voies physiologiques diverses comme RAS-MAPK, NF1, PTEN, mTOR, TSC1/TSC2, système ubiquitine protéasome UBE3A

- variations de nombre de copies chromosomiques : parmi les facteurs génétiques pouvant toucher ces voies, nombre d'entre eux sont des variants de nombre de copie chromosomique dont une partie significative constitue des facteurs de prédisposition car ces variants sont retrouvés à une fréquence non négligeable bien que moindre dans la population générale comme les variants 1q21 proximal ou distal del/dup, 2p16.3 (*NRXN1*) del, 3p26 del/dup, 7q11.23 dup, 15q11.2 del/dup, 15q13.3 del/dup, 16p11.2 proximal del/dup, 16p11.2 distal del/dup, 16p13.1 del, 17q12 del, 22q11.2 dup ²¹¹.



Figure 15 : Illustration des différentes voies cellulaires et facteurs génétiques pouvant être en cause dans les troubles du spectre autistique ²⁰⁹.

1.3.2. Epilepsie : définition et causes

La « maladie épileptique » est définie par des crises récurrentes et sans facteur déclenchant majeur transitoire (comme un traumatisme ou une infection) à type de décharges neuronales synchronisées au niveau du cerveau. Elle touche jusqu'à 4% de la population générale dans les pays occidentaux ²¹². Environ 20-30% des cas concernent des pathologies acquises primaires ou séquellaires comme des AVC, des tumeurs, des traumatismes et le reste est fortement suspect d'être d'origine génétique ²¹³. Il s'agit en fait d'un groupe de pathologies très hétérogènes au pronostic très variable. De nombreux syndromes épileptiques ont été définis qui se caractérisent par l'âge de début, le type de crise, le profil évolutif, l'association ou non à d'autres symptômes ou signes (comme des troubles neurodéveloppementaux ou des malformations. La classification de l'épilepsie a été révisée par l'ILAE (International League Against Epilepsy) en 2010 ²¹⁴. Les figures 16 et 17 suivantes résument la classification proposée par catégories et par syndromes électro-cliniques :



Figure 16 : représentation schématique de la classification de l'épilepsie selon l'ILAE en 2010²¹⁵.



Figure 17 : Représentation schématique des syndromes électro-cliniques selon l'ILAE en 2010²¹⁵.

Le changement dans la classification inclut notamment le fait que le groupe des « crises généralisées idiopathiques » (sous-entendu présumées génétiques) devient crises généralisées génétiques, renforçant cette hypothèse étiologique. De nombreuses revues sont disponibles reprenant les causes associées à différents types d'épilepsie ^{216–219}.

Un consortium international Epi4k s'attache à démembrer les causes génétiques de l'épilepsie notamment grâce à l'apport du séquençage pangénomique.

De la même façon que pour la DI et les TSA, les causes génétiques de l'épilepsie sont très hétérogènes :

- les variations de nombre de copies chromosomiques expliquent 5-10% des épilepsies de l'enfant, certaines variations impliquant des gènes connus en pathologies sont souvent associées à des troubles neurodéveloppementaux (sévères dans les encéphalopathies ou plus modérés) et d'autres étant clairement des facteurs de prédisposition aux troubles neurodéveloppementaux et à l'épilepsie déjà décrits dans les chapitres 1.2.3 et 1.3 au sujet de la DI et des TSA (notamment délétions récurrentes des régions 15q13.3, 16p13.11 et 15q11.2 retrouvées pour chacune d'elles chez 0,5-1% des patients, odds ratio = 7,5) ²²⁰⁻²²².

- les variations intragéniques peuvent être somatiques, constitutionnelles, et touchent des voies similaires aux pathologies neurodéveloppementales, résumées sur les figures 18 à 20.



Figure 18 : représentation des syndromes épileptiques par rapport aux périodes de début de l'épilepsie et aux gènes associés ²¹⁷.



Figure 19 : précisions sur la classification des causes génétiques des crises convulsives néonatales ²²³.



Figure 20 : Illustration du rôle au niveau neuronal des différents gènes impliqués dans l'épilepsie²¹⁷.

La meilleure compréhension de ces voies physiopathologiques permet d'adapter l'utilisation des médicaments anti-épileptiques et représente un exemple de médecine personnalisée comme cela est illustré pour certains gènes dans le tableau 1 ci-dessous.

*Tableau 1 : Illustration d'exemples d'adaptation thérapeutique en fonction du canal impliqué dans l'épilepsie*²¹⁹.

Gène	Syndrome épileptique	Possible(s) traitement(s)		
SCN1A	Dravet	Eviction des inhibiteurs de canaux Na ⁺		
SCN8A	Encéphalopathie précoce	Inhibiteurs d'ouverture de canaux Na ⁺ (carbamazépine, phenytoïne)		
KCNQ2	Encéphalopathie précoce	Activateurs d'ouverture de canaux K ⁺ (retigabine) ou inhibiteurs d'ouverture de canaux Na ⁺ (carbamazépine)		
KCNT1	Epilepsie partielle migrante	Activateurs de l'ouverture de canal K ⁺ (quinidine)		
GRIN2A	Encéphalopathie précoce	Antagonistes du récepteur NMDA (N-methyl-d- aspartate) (memantine)		
SCL2A1	Déficit en GLUT-1	Régime cétogène		
TSC1/TSC2	Sclérose tubéreuse de Bourneville	Rapamicine et analogues		
ALDH7A1	Epilepsie sensible à la pyridoxine	Vitamine B6		

Par ailleurs, il est aussi décrit un risque d'épilepsie plus important chez les patients atteints de DI syndromique ou non, estimée à environ 15-20% avec certains syndromes génétiques avec DI particulièrement caractérisés par une épilepsie Des études ont aussi évalué les caractéristiques de l'épilepsie dans des syndromes génétiques bien connus. Une revue reprend les données disponibles au sujet de l'épilepsie pour 4 syndromes bien caractérisés (X fragile, Rett, Angelman, sclérose tubéreuse de Bourneville) (tableau 2)²²⁴.

Tableau 2 : Exemple de caractéristiques de l'épilepsie dans des syndromes génétiques associés à la DI et aux TSA, adapté de Leung et al, 2011 ²²⁴.

Syndrome	Gène	Fréquence	Âge de début	Type de crise si épilepsie	Sévérité
X fragile	FMR1	G : 14-14,5% F : 6-8,3%	4-10 ans	CGTC + CP (pour 23% G, 11% F)	Légère à modérée, 85% pharmacosensibilité
Rett	МЕСР2	60-94%	3-5 ans	CGTC + CP pour 31,5%	34-54% pharmacorésistance
Angelman	Empreinte 15q11q13 / UBE3A	86%	Variable, médiane 2,9 ans	Crises multiples pour 60%	77% pharmacorésistance
STB	TSC1, TSC2	82,6%	Variable, médiane 2,4 ans	Crises multiples pour 53%	56-63% pharmacorésistance

1.3.3. Chevauchement des causes génétiques de pathologies neurodéveloppementales et de l'épilepsie

Tous les éléments indiqués ci-dessus montrent le chevauchement des causes génétiques dans les troubles neurodéveloppementaux dont la DI, les TSA, et l'épilepsie et explique pourquoi la prévalence de l'épilepsie dans ces pathologies neurodéveloppementales est supérieure à celle de la population générale. La figure 21 suivante représente un exemple illustrant le chevauchement des causes génétiques dans ces pathologies.



*Figure 21 : diagramme montrant le chevauchement des troubles neurodéveloppementaux causés par des variations perte de fonction de novo dans des gènes identifiés par séquençage d'exome en trio*²²⁵.

1.3.4. Syndrome de Marfan et syndromes marfanoïde

Ce syndrome est une pathologie du tissu conjonctif responsable de l'atteinte de 3 systèmes : système cardio-vasculaire notamment dilatation de l'aorte, atteinte oculaire (ectopie du cristallin et myopie sévère) et le système squelettique.

Les critères cliniques de Ghent initialement établis en 1996 ont été révisés en 2010²²⁶ :

A. En l'absence d'antécédents familiaux:

- (1) Ao ($Z \ge 2$) et EC = MFS*
- (2) Ao ($Z \ge 2$) et FBN1 = MFS
- (3) Ao ($Z \ge 2$) et Syst (≥ 7 pts) = MFS*
- (4) EC et FBN1 connu pour être associé à Ao = MFS

- EC avec ou sans Syst et avec FBN1 non connu pour être associé à Ao ou sans FBN1 = SEC

- Ao (Z < 2) et Syst (\geq 5 avec au moins un signe squelettique) sans EC = MASS

- PVM ET Ao (Z <2) et Syst (<5) sans EC = SPVM

B. En présence d'antécédents familiaux:

(5) EC et HF = MFS

(6) Syst (\geq 7 pts) et HF = MFS*

(7) Ao (Z≥2 après 20 ans, ≥3 avant 20 ans) + HF de MFS (comme défini par les critères précédents) = MFS*

Ao: dilatation de l'aorte au sinus de Valsalva ; EC : ectopie du cristallin ; FBN1 : variation pathogène du gène ; MASS : myopie, PVM, Ao limite Z<2, stries atrophiques, signes squelettiques ; PVM : prolapsus valve mitrale ; SEC : syndrome d'ectopie cristallinienne ; Syst : score systémique ; Z: Z-score.

* : en l'absence de signes en faveur d'un diagnostic différentiel (syndrome Loeys-Dietz ou Ehlers-Danlos notamment).

<u>Score systémique</u> (critère d'atteinte squelettique si score \geq 7 (sur 20)

- signe du pouce ET du poignet = 3 (un seul des 2 = 1)

- pectus carinatum = 2 (pectus excavatum ou asymétrie thoracique = 1)

- déformation de l'arrière pied = 2 ou pied plat = 1

- pneumothorax = 2

- ectasie durale = 2
- protrusion acétabulaire = 2

- diminution du rapport longueur membres supérieurs/membres inférieurs ET augmentation du rapport envergure/taille en l'absence de scoliose sévère = 1

- scoliose et cyphose thoracolombaire = 1

- limitation d'extension des coudes $< 170^\circ = 1$

- critères morphologiques faciaux (3/5) = 1 (dolichocéphalie, énophtalmie, fentes palpébrales orientées en bas et en dehors, hypoplasie malaire, rétrognatisme)

- stries cutanées = 1

- myopie > 3 dioptries = 1

- prolapsus valve mitrale = 1

FBN1 représente le gène majeur responsable de syndrome de Marfan. La protéine joue un rôle structurel dans le tissu conjonctif mais également un support dans la transmission de la voie de signalisation TGFβ. Une variation causale hétérozygote est identifiée chez plus de 90% des patients

²²⁷. La base de données UMD-FBN1 (Universal Mutation Database) répertorie l'ensemble des variations identifiées pour aider à leur interprétation. Sur le plan des corrélations génotype-phénotype faites sur une série de 1013 patients porteurs de variations pathogènes du gène, les variations faux sens touchant une cystéine étaient plus fréquemment associées à une ectopie du cristallin que les autres faux sens, les variations prédites comme tronquantes avaient des signes squelettiques et cutanés plus marqués, les variations situées dans les exons 24 à 32 étaient responsables d'une atteinte plus sévère et plus complète sur le plan des systèmes touchés y compris après exclusion des formes néonatales ²²⁸.

Dans le cadre de la DI, de nombreux patients ont été décrits comme porteurs de morphotypes empruntant des critères squelettiques du syndrome de Marfan et ces patients ont été regroupés dans le groupe des syndromes marfanoïdes (Habitus Marfanoïde - DI). Il s'agit d'un groupe hétérogène génétiquement avec de multiples causes génétiques.

L'étude de 100 patients ayant une atteinte du spectre HM-DI ont été explorés par CGH array et séquençage des gènes décrits comme associés à ce type de phénotype (*MED12, ZDHHC9, UPF3B, FBN1, TGFBR1* et *TGFBR2*). Parmi ces patients, 80 remplissaient les critères squelettiques et 20 présentaient des critères évocateurs sans que le score systémique soit au moins égal à 7. Le rendement de la CGH array dans le 1^{er} groupe de 80 patients était de 12 variations chromosomiques, 1 variation *FBN1* et 1 variation de signification inconnue de *MED12* soit un rendement potentiel de 17%. Le rendement dans le 2^{ème} groupe de 20 patients était de 4 variations *FBN1*, 4 variations chromosomiques (et 2 patients avaient à la fois une variation *FBN1* et une variation chromosomique) soit un rendement de 29%. Aucune variation chromosomique n'était récurrente et certaines concernaient des facteurs génétiques prédisposant aux troubles neurodéveloppementaux. Cette étude indiquait également la possibilité qu'une variation soit responsable de l'ensemble du phénotype mais aussi que parfois, le phénotype clinique puisse être dissocié avec la contribution d'un type de variation expliquant l'aspect marfanoïde et une autre variation en cause dans la DI ²²⁹.

Une étude en cours de finalisation vise à évaluer l'apport du séquençage d'exome pour le diagnostic des patients présentant un phénotype HM-DI. Une cohorte de 64 patients a pu être recrutée grâce à un PHRC national avec l'inclusion de quelques patients étrangers. L'exome a été interprété en solo pour 31 patients et en trio pour 33 patients. Le rendement des trios était de 22/33 avec un gène supplémentaire fortement candidat (70%) et celui des solos était de 13/31 (42%) avec un patient parmi les précédents également porteur d'une variation de signification indéterminée. Cette stratégie a donc permis d'obtenir un diagnostic moléculaire dans 56% des cas, 8 gènes étant mis en cause chez au moins 2 patients (*EHMT1, ARID1B, NFIX, NSD1, ZEB2, DLG4, ATP1A1* et *NUP205*). De même que pour la cohorte précédente, un des patients était porteur d'une variation pathogène de *FBN2* mais l'origine de la DI n'était pas identifiée ; un des patients porteurs d'une variation pathogène de *ZEB2* avait aussi une variation *de novo SYNGAP1* dont les scores de pathogénicité étaient faibles (Chevarin et al, in preparation).

Ces différentes études montrent l'hétérogénéité génétique associée à ce phénotype HM-DI.

1.4. Intérêt des modèles animaux pour étudier les voies physiopathologiques de la DI

Les modèles cellulaires et animaux permettent d'apporter de nombreux éléments pour la compréhension des voies biologiques et physiopathologiques. Parmi les stratégies envisageables, l'utilisation d'un modèle animal permet d'accéder à de multiples informations incluant un phénotypage clinique (malformatif/anatomique, comportemental), et l'étude de voies biologiques cellulaires ou moléculaires. Différentes méthodes sont utilisées pour générer des animaux transgéniques qui sont choisies selon le mécanisme mutationnel ou le type de variants à explorer. Parmi les critères importants orientant vers le choix d'un modèle, on doit retenir la faisabilité de la construction en conformité avec le mécanisme sous-tendant la maladie, la validité clinique où des caractéristiques de la maladie humaine sont retrouvées chez le modèle et la validité prédictive permettant de prédire l'apparition des symptômes ou l'identification de nouveaux traitements ^{230,231}, mais aussi la possibilité de traçabilité génétique pour confirmer le résultat des manipulations faites, l'utilisation d'outils pour visualiser et manipuler les neurones in vivo, et la possibilité de transférer les découvertes vers les patients sur la base de la conservation des voies biologiques. Au final, cela permet de contribuer à la recherche translationnelle pour aider à l'interprétation et la classification des variations génétiques et le développement de molécules thérapeutiques ou de thérapie génique pour tenter d'améliorer le phénotype voire rétablir un phénotype normal.

1.4.1. Modèle murin

1.4.1.1. Génome murin et outils utilisés pour la transgenèse

La première version de la séquence du génome murin a été publiée en 2002 et indiquait qu'il était possible d'aligner environ 50% du génome humain sur celui de la souris ²³². La comparaison des séquences sensibles à la DNase I retrouve un degré de similitudes relativement important ^{233,234} ce qui indique une conservation entre ces 2 mammifères, mais des divergences sont aussi notées (chapitre 1.4.7).

Au cours des dernières décennies, de nombreuses techniques de manipulation génétique ont pu être utilisées pour appréhender la fonction des gènes responsables de maladies génétiques et en particulier de la DI *in vivo*.

- Au début des années 1980, la technique de micro-injection de vecteurs d'ADN exogène à des stades embryonnaires précoces ou dans des pronuclei post-fécondation permettait l'intégration dans le génome murin de séquences cibles. L'injection de cellules souches embryonnaires modifiées à un stade embryonnaire très précoce conduisait à la création de chimères. Dans les 2 cas, le phénotype résultant était étudié et des analyses fonctionnelles pouvaient être menées.

- Dans les années 1990-2000, des technologies de mutagenèse dirigée ont ensuite été développées, améliorant le ciblage des régions d'intérêt.

La technologie du système bactériophage Cre/loxP a été très utilisée dans les modèles murins et consiste à croiser d'une part des souris exprimant la Cre recombinase (spontanément ou sous le contrôle d'activateurs exogènes comme le tamoxifène ou la doxycycline, ou endogènes par l'ajout de séquences régulatrices spécifiques d'un organe ou d'un type cellulaire choisi) avec d'autre part des souris contenant des sites loxP placés à un site ciblé du génome (en pratique au niveau d'un gène d'intérêt) et qui seront reconnus par la recombinase (figure 22) ; l'activation du système permet d'induire des délétions (knock-out), des inversions, des translocations ou l'introduction de séquences (knock-in) permettant ensuite d'étudier leurs conséquences (figure 23) ²³⁵.



*Figure 22 : Procédure permettant d'insérer des sites loxP au niveau de régions cibles du génome qui sera reconnu par la Cre recombinase*²³⁵.



*Figure 23 : Résultats obtenus après activité Cre recombinase selon l'orientation et la localisation des sites loxP insérés*²³⁵.

- La technologie CRISPR-Cas9 a été une révolution dans son potentiel d'applications à partir du début des années 2010 pour permettre de générer de façon encore plus simple des modèles animaux (figure 24) :



Figure 24 : Illustration de la technologie CRISPR-Cas9. a. Système CRISPR RNA (crRNA) contenant des séquences successives visant à cibler de l'ADN par complémentarité associées aux répétitions CRISPR, et séparé du tracrRNA (trans-activating CRISPR RNA); l'interaction de ces 2 ARN par la séquence CRISPR permet l'adressage à la Cas9 (CRISPR-associated protein 9). b. Design d'une séquence hybride combinant le crRNA (ARN guide) et le tracrRNA permettant d'améliorer la spécificité de la séquence à cibler par la Cas9²³⁶.

Le complexe CRISPR-Cas9 cible une région d'ADN spécifique et génère une cassure double brin ; sa réparation se fait de façon variable par réparation homologue ou non homologue pouvant générer des délétions, duplications, inversions variables. Ce système, facile à utiliser, permet une augmentation du « débit » générant des mutants et une analyse ciblée du locus ciblé permet d'étudier le mécanisme de réparation mis en œuvre et l'effet mutationnel ²³⁷ puis il convient d'étudier le phénotype associé.

1.4.1.2. Application des modèles murins à la DI

Toutes ces technologies ont permis de créer de très nombreux modèles animaux pour comprendre la fonction des gènes et les conséquences associées à différents types de mutations. De nombreux tests ont été développés chez la souris pour étudier le profil comportemental et cognitif de façon reproductible :

- le test en champ ouvert (open field test) permet d'évaluer le niveau d'anxiété reflété par le comportement d'exploration d'une souris placée dans un environnement circulaire fermé, le temps passé en périphérie étant considéré comme proportionnel au niveau d'anxiété ²³⁸.

l'inhibition pré-stimulation (pre-pulse inhibition) évalue les capacités d'apprentissage et de mémorisation en faisant précéder une stimulation surprenante et brutale par une faible stimulation, l'effet de surprise est censé s'atténuer avec le temps, indiquant que la souris a pu apprendre le lien entre le 1^{er} et le 2nd événement ²³⁹.

- le labyrinthe de Morris (Morris maze test) est un des tests les plus utilisés pour étudier les capacités d'apprentissage de la reconnaissance spatiale : il consiste à évaluer dans une piscine la rapidité de localisation d'une plateforme cachée où la souris a pied et va donc avoir tendance à se placer pour limiter son temps de nage où des repères visuels ont été placés ; au fur et à mesure de la répétition de l'épreuve, la souris est censée retrouver la plateforme de plus en plus rapidement, la plasticité peut s'étudier en reproduisant plusieurs séries d'épreuves après avoir changé la plateforme ou les repères de place.

- plusieurs tests ont été développés pour évaluer la sociabilité chez la souris pouvant reproduire les traits autistiques (isolés ou pouvant être combinés dans la DI).

a. Etude des protéines synaptiques

Par exemple, des modèles murins invalidant les neurexines et neuroligines au niveau des éléments pré et post-synaptiques ont été créés pour étudier les conséquences de la perte de fonction sur la régulation synaptique. Chez l'Homme, des études d'association ont pu montrer que la délétion hétérozygote du gène *NRXN1* dans la région 2p16.3 représentait un facteur de risque aux troubles neurodéveloppementaux. Il existe 2 isoformes protéiques : NRXN1 α (forme longue) et NRXN1 β (forme courte). Les souris KO (knock-out) pour l'isoforme *Nrxn1\alpha* avaient une diminution de l'intensité d'activation des synapses excitatrices, une diminution de l'apprentissage du test de prepulse inhibition, une augmentation du comportement de toilettage et une amélioration des capacités d'apprentissage moteur ; cette isoforme est probablement responsable de la régulation de la balance excitation-inhibition des synapses excitatrices et la diminution quantitative de protéine est associée à un comportement de type autistique ²⁴⁰.

L'étude d'un modèle murin de surexpression de $NRXN1\beta$ dans les neurones glutamatergiques grâce à un promoteur spécifique a montré une augmentation du comportement de toilettage et une diminution des interactions sociales ce qui est évocateur de symptômes autistiques, ces symptômes étaient réversibles après diminution d'expression de l'isoforme ²⁴⁰.

b. Modèle murin de syndrome de l'X fragile

Un modèle de souris KO *Fmr1* a retrouvé des éléments cliniques et comportementaux proches de la maladie humaine (syndrome de l'X fragile) avec une hyperactivité, une anxiété augmentée, un déficit cognitif, et un trouble de discrimination entre une souris connue et une souris inconnue ²⁴¹. Afin de préciser les conséquences de la perte d'expression du gène dans des cellules spécifiques, des modèles Cre/loxP ont été utilisés pour inactiver le gène dans différents types cellulaires comme les cellules de Purkinje du cervelet (entraînant chez les souris un déficit dans le conditionnement au clignement des yeux médié par le cervelet ²⁴²), ou dans les neurones corticaux et hippocampiques dans les 1^{ères} semaines de vie responsable de modifications des dendrites (à type d'allongement des épines) et de l'expression de marqueurs synaptiques spécifiques à certaines régions cérébrales et une activation de la voie AKT-mTOR (à noter qu'il n'était pas noté de changement comportemental dans ce modèle) ²⁴³.

Chez les mammifères, le gène *FMR1* fait partie d'une petite famille de protéine qui inclut également *FXR1* et *FXR2*. Les 3 membres interagissent avec des ARN et contiennent des motifs habituellement retrouvés dans d'autres protéines ayant des rôles similaires (comme les hnRNP - Heterogeneous nuclear ribonucleoproteins) dont 2 domaines KH et un domaine RGG. Ce groupe de protéine FMR1/FXR contient aussi un domaine d'interaction à l'ARNr 60S et un domaine qui permet l'homo et l'hétérodimérisation avec chacun des membres de la famille protéique.

Afin d'étudier le rôle de gènes homologues et des interactions possibles avec FMRP, plusieurs modèles de souris KO *Fxr1* et *Fxr2* ont été générés. Concernant les souris KO *Fxr1*, un des modèles retrouvait des souris ayant une amyotrophie et une espérance de vie fortement réduites faisant évoquer un rôle similaire de FMRP au niveau des cellules musculaires ²⁴⁴. Concernant les souris KO *Fxr2*, il n'a pas été identifié de différence structurelle par rapport aux contrôles mais des troubles du comportement avec hyperactivité majeure, diminution des capacités de pre-pulse inhibition et de localisation de la plateforme dans la piscine de Morris (water maze test) ²⁴⁵, ce qui indique un rôle au niveau du système nerveux central. Les souris double KO *Fmr1* et *Fxr2* avaient des altérations cognitives encore plus importantes, une perte des rythmes circadiens, une anxiété majeure, des troubles d'apprentissage et de mémorisation ^{246,247}. Ces données sont donc en faveur de relation fonctionnelles entre FMRP et FXR2. Ces exemples montrent comment il est possible d'étudier des voies de signalisation et des interactions protéiques *in vivo*.

c. Modèle murin de syndrome de Rett

Le syndrome de Rett est dû à des variants pathogènes de *MECP2* (Methyl-CpG-binding protein 2) et correspond à une régression des acquisitions dans les 1^{ers} mois de vie avec déficit intellectuel, microcéphalie et TSA incluant des stéréotypies (chapitre 1.2.3.2). La protéine MECP2 est localisée au niveau du noyau cellulaire et se fixe à l'ADN méthylé ²⁴⁸ pour fonctionner comme répresseur

transcriptionnel. Le gène est fortement exprimé dans le cerveau post-natal et a un rôle majeur dans le développement du système nerveux central chez le mammifère. Un modèle murin produit par invalidation cerveau-spécifique du gène grâce au système Cre-loxP a permis de retrouver des éléments cliniques ressemblant au syndrome avec une régression apparaissant vers 5 semaines de vie et décès entre 6 et 12 semaines lorsque l'invalidation est programmée au stade embryonnaire. L'étude morphologique cérébrale a montré une diminution de la taille des neurones sans élément en faveur de neurodégénérescence chez les mutants. Lorsque l'invalidation du gène est programmée en période post-natale, on retrouve des effets similaires mais décalés dans le temps ²⁴⁹. Ces éléments montrent l'intérêt du modèle animal pour explorer le rôle d'un gène dans le déficit intellectuel, et dans cet exemple le rôle de MECP2, pour reproduire le phénotype de la maladie humaine pour servir de modèle au développement de projets thérapeutiques.

1.4.2. Modèle drosophile

Environ ³/₄ des gènes de DI sont conservés chez la drosophile ^{225,250}. Malgré la faible conservation de l'anatomie du système nerveux central entre la drosophile et l'Homme, les voies biologiques sont fortement conservées au niveau moléculaire, cellulaire et synaptique ²⁵¹.

1.4.2.1. Moyens utilisables chez la drosophile pour modéliser la DI

Plusieurs types de modèles peuvent être utilisés pour explorer la DI chez la drosophile (figure 25).



Figure 25 : Illustration des différents moyens permettant de modéliser la DI chez la drosophile²⁵².

A1. L'étude de la jonction neuromusculaire de larve représente un bon modèle pour explorer les anomalies de structure et/ou de transmission synaptique puisque, comme indiqué précédemment, un nombre significatif de gènes de DI concernent une dysfonction de cette structure. L'intérêt est que les synapses sont de grande taille et facilement accessibles pour des tests électrophysiologiques.

A2. La quantification des arborisations dendritiques, en particulier des neurones sensitifs dont certains modèles peuvent reproduire les anomalies d'architecture à type de réduction de la densité qui ont été retrouvées sur des analyses histologiques de cerveaux de patients ayant une DI par exemple causée par des mutations de *CAMK2A*, *SHANK3* ou *IL1RAPL1* ^{253,254}. Une équipe a pu montrer le rôle du gène *minibrain* (orthologue de *DYRK1A*, membre de la famille des kinases régulées par la phosphorylation de tyrosine et qui participe à de nombreux processus cellulaires, dont l'excès d'expression contribue à la DI dans la trisomie 21 et l'haploinsuffisance est responsable de DI modérée à sévère avec microcéphalie) dans l'atteinte de la morphologie dendritique et de la physiologie neuronale par anomalie de la phosphorylation de la β -tubuline et inhibition de sa polymérisation ²⁵⁵.

B/E. L'étude de l'électrophysiologie neuronale chez la drosophile offre un moyen plus facile de manipulation des neurones et des synapses permettant d'évaluer la fonction de neurones uniques à leur réseau et l'étude de la fonction de canaux par patch-clamp.

C. Etude de l'anatomie cérébrale : 2 structures sont particulièrement intéressantes par rapport à la DI et aux TSA chez la drosophile. Le « mushroom body » fait référence à la structure analogue à l'hippocampe chez les mammifères et est impliqué dans les processus d'apprentissage et de mémoire ²⁵⁶ ; il a aussi été proposé comme l'équivalent du cervelet du fait d'une architecture similaire et au cortex par sa similitude du profil d'expression génique ^{257,258}. Le complexe central correspond à une structure riche en arborescences qui intègre de multiples signaux neurosensoriels et qui joue un rôle dans les comportements sociaux et les apprentissages.

D. Tests comportementaux, d'apprentissage et de mémoire :

- test de stimulation conditionnée : par exemple récompense ou choc électrique précédés par une odeur spécifique, ce qui permet d'entraîner par la suite une préférence vers un groupe d'odeur alors qu'il n'y a plus d'événement qui suit l'exposition à l'odeur ;

- étude de la réduction du temps de parade (facilement évaluable par son comportement stéréotypé) d'un male après rejet d'une femelle encore immature ;

- étude des interactions sociales : un des tests consiste à étudier les interactions de 2 drosophiles situées dans 2 espaces séparés par une grille/des mailles, ce qui permet d'appréhender un TSA associé à la DI ;

- étude du phénomène d'habituation, permettant d'évaluer la plasticité synaptique : un test caractéristique consiste à éteindre la lumière de façon répétée ce qui entraîne habituellement un réflexe de sursaut dont la réponse diminue progressivement avec l'adaptation aux stimuli (ce test permettant d'évaluer un phénomène indépendant d'une composante sensitive ou motrice). Des modèles de drosophiles pour environ 300 gènes responsables de DI ont montré un déficit dans ce phénomène

d'habituation dans plus de 100 modèles ; les gènes les plus fréquemment responsables de troubles autistiques associés à la DI présentaient plus fréquemment ce type de déficit ²⁵⁹ ;

- étude du rythme circadien : dont le parallèle est fait avec l'Homme par l'observation fréquente de troubles du sommeil chez les patients atteints de DI et/ou de TSA, le syndrome le plus emblématique à ce sujet étant le syndrome de Smith-Magenis dû à une perte de fonction du gène *RAI1* soit par délétion de la région chromosomique 17p11.2 soit par mutation ponctuelle et qui inclut une inversion du rythme nycthéméral parmi les signes cliniques caractéristiques. Ce type de travail des rythmes circadiens sur le modèle drosophile a été récompensé par le prix Nobel de médecine ou physiologie décerné à Jeffrey C. Hall, Michael Rosbash et Michael W. Young pour leurs découvertes sur les mécanismes contrôlant le rythme circadien dont l'horloge est due à l'activité d'un petit groupe de neurones. Une altération des gènes codant pour les protéines d'échafaudage synaptique sont associés à des troubles de cette horloge biologique ; par exemple, les drosophiles mutées pour *Nlg4* ont des troubles du sommeil par altération de la neurotransmission GABA des « neurones horloges » ²⁶⁰.

1.4.2.2. Exemples de modèles drosophiles de maladies humaines

A titre d'exemple, la drosophile mutante pour l'orthologue *fmr1* a pu être explorée pour mieux identifier certaines caractéristiques du syndrome de l'X fragile chez l'Homme. Il a ainsi pu être montré que les rythmes circadiens étaient altérés ²⁶¹, qu'il existait des anomalies histologiques (augmentation de taille des terminaisons synaptiques ²⁶², anomalie de morphogenèse du mushroom body avec notamment une anomalie des lobes qui franchissent la ligne médiane au lieu d'y être délimités ²⁶³), une altération de la mémoire ²⁶⁴, et un temps d'interaction réduit au test des espaces séparés par des mailles ²⁶⁵.

L'ensemble de ces éléments montrent que la drosophile peut s'avérer être un modèle très utile pour accélérer la recherche fondamentale et translationnelle, avec l'avantage de générer un nombre important d'individus rapidement à un moindre coût par rapport aux modèles murins.

1.4.3. Modèle du poisson zèbre

1.4.3.1. Développement de l'espèce et comparaison avec le génome humain

Le poisson zèbre (zebrafish ou *Danio rerio*) est un modèle très utilisé pour étudier le développement des vertébrés grâce au potentiel de screening relativement facile permettant un débit élevé d'études de la fonction des gènes ce qui a permis d'identifier des centaines de gènes impliqués dans l'embryogenèse ²⁶⁶. Cette facilité d'utilisation est expliquée par la grande quantité d'œufs produits par fertilisation externe (ce qui permet la manipulation des gamètes avant fécondation) qui se

retrouvent dans le milieu extérieur sans gestation interne avec une transparence des embryons et des animaux à l'état larvaire permettant l'observation du développement notamment le système nerveux central au microscope, et par des étapes du développement rapides (48-72 heures pour l'embryon, 4 à 12 semaines pour la larve selon les conditions environnementales) et des comportements faciles à étudier. Son nom tient à ses stries provenant de la pigmentation par des cellules dérivées de la crête neurale.

Bien que la divergence d'espèces avec les autres vertébrés soit lointaine (environ 250 millions d'années), le développement du système nerveux central est similaire avec des structures anatomiques équivalentes, les mêmes profils d'expression géniques et une conservation des systèmes de neurotransmission (GABA, glutamate, dopamine, noradrénaline, sérotonine, histamine et acétylcholine) permettant d'envisager des essais pharmacologiques ^{267–270}. Les circuits neuronaux sous-tendant les comportements basiques (réflexe auditif, phénomène d'inhibition à la double stimulation, sommeil, éveil) sont aussi conservés ²⁷¹ et permettent d'envisager des études phénotypiques comportementales similaires à la souris.

La duplication du génome spécifique ayant conduit à la divergence d'espèces vers ce groupe d'animaux nommés téléostes a néanmoins été à l'origine d'une augmentation de la taille des synapses et de leur composition protéique ²⁷² ; bien que la synapse soit moins complexe que chez le mammifère, sa structure est bien conservée.

Le séquençage complet du génome du poisson zèbre a été publié en 2013. Il a une taille d'environ 1,4 milliards de paires de bases et contient plus de 26000 gènes codant, soit le nombre le plus important chez les vertébrés séquencés jusque-là. Environ 70% des gènes humains ont au moins un orthologue chez le poisson zèbre ²⁷³ et il existe de nombreuses régions de synténie ²⁷⁴. De nombreuses régions du génome du poisson zèbre sont dupliquées, dû à des évènements de duplication segmentaire survenus au cours de l'évolution de l'espèce ²⁷⁵ mais 45 à 50% des gènes humains sont aussi présents en une copie (sans duplication) chez le poisson zèbre ²⁷³.

Les duplications conduisent à l'apparition de gènes paralogues ce qui permet une redondance de fonction mais chaque gène peut ensuite évoluer avec des événements mutationnels dans les séquences codantes et régulatrices pouvant tantôt entraîner une dépendance entre les 2 gènes pour assurer la fonction initiale et parfois conduire à des spécifications temporo-spatiales de chaque gène ou à l'apparition de nouvelles fonctions ²⁷⁶.

Ces éléments expliquent la nécessité de bien comparer le génome humain et du poisson zèbre pour créer ce type de modèle animal, nécessitant fréquemment de cibler plusieurs gènes paralogues pour obtenir un modèle de maladie humaine.

1.4.3.2. Moyens utilisables pour manipuler le génome du poisson zèbre

Alors que les souris KO ont pu rapidement être générées à partir de cellules souches embryonnaires, les méthodes étaient plus difficiles à mettre en œuvre jusqu'à récemment :

- une des techniques consistait à induire des mutations au hasard dans le génome par l'agent chimique N-ethyl N-nitrosourée (ENU) (technique de TILLING pour Targeted Induced Local Lesions in Genomes) ce qui nécessitait d'étudier un nombre important d'animaux pour identifier ceux porteurs de mutation au niveau de la région à étudier et du gène d'intérêt, technique chronophage et incertaine ²⁷⁷, d'autant plus du fait de la redondance d'un grand nombre de gènes.

- le développement des technologies de ciblage grâce aux nucléases (zinc finger nucleases et transcription activator-like effector nucleases ou TALEN) a transformé les possibilités de développement de modèles animaux avec un gain d'efficacité conséquent ^{278,279}.

 le système CRISPR-Cas9 a supplanté la méthode précédente par sa flexibilité et son efficacité supérieures, et de nombreuses méthodologies ont été proposées pour optimiser l'efficacité de l'édition du génome du poisson zèbre ^{280,281}

- cependant, un des outils les plus utilisés pour inactiver des gènes chez le poisson zèbre est le morpholino qui correspond à des séquences d'environ 25 bases avec pour but que la complémentarité avec la séquence d'un ARN cible inhibe la production protéique soit en bloquant la maturation de l'ARN au niveau des sites d'épissage soit en bloquant le site d'initiation de la traduction. Ces morpholinos ont un coût très réduit et sont faciles à utiliser. Ils sont injectés dans l'embryon à un stade très précoce et vont se répartir dans les cellules. Les inconvénients sont l'effet transitoire qui empêche d'étudier l'inactivation d'un gène au-delà d'un certain stade de développement et les cibles non désirées (par exemple certains ciblent p53 ce qui entraîne une apoptose des cellules. Les recommandations d'utilisation incluent l'utilisation de 2 morpholinos ciblant différents sites et la démonstration de la possibilité de restaurer le phénotype ²⁸², auquel s'ajoute la nécessité de confirmer le phénotype par mutation au sein du génome ce qui est facilité depuis l'essor de la technologie CRISPR-Cas9 ²⁸³. Le défaut de production protéique a des conséquences phénotypiques dans les jours qui suivent.

1.4.3.3. Applications à la modélisation de la DI

La diminution d'expression (KD / knock-down) par 2 morpholinos ciblant les gènes *shank3a* et *syngap1* dont les mutations perte de fonction sont responsables de DI modérée à sévère avec des troubles autistiques fréquents, a permis d'identifier une diminution d'activité des neurones GABA dans le mésencéphale et glutamatergiques dans le rhombencéphale, avec des anomalies dans la démarcation de ces 2 structures, dans la taille des ventricules et une microcéphalie ²⁸⁴.

Plusieurs études ont exploré des phénotypes comportementaux, comme celui associé à la perte de fonction du gène *mecp2* : les larves avaient une activité motrice réduite, une préférence pour se

placer en périphérie de l'espace de mobilité et une diminution du temps de réaction pour s'enfuir au toucher, indiquant que le gène a un rôle dans le comportement précoce dès le stade larvaire ²⁸⁵.

Certains modèles sont évolutifs comme lorsque l'orthologue de *DYRK1A* est invalidé : le développement est normal dans les stades embryonnaires et larvaires puis il existe une évolution vers une microcéphalie et des troubles du comportement par augmentation de l'apoptose

L'annexe 2 indique plusieurs exemples de modèles de poisson zèbre étudiés dans le cadre de la DI.

Une autre approche pour tester le rôle de nouveaux gènes identifiés en pathologie humaine ou l'implication de variations rares est la surexpression dans les embryons. Par exemple, la surexpression du gène sauvage pk1a (*prickle1*) a un effet plus délétère que l'ajout d'ARNm porteur d'une mutation *PK1A* identifiée chez un patient présentant une épilepsie myoclonique progressive, mais cela suggère néanmoins que l'introduction de la mutation altère la fonction *in vivo* du gène ²⁸⁶.

1.4.4. Modèle Xenope

1.4.4.1. Développement et comparaison avec le génome humain

Le modèle Xenope a été utilisé depuis des décennies pour étudier le développement et les processus biologiques et cellulaires. Cela a permis de bien connaître cet organisme et d'observer une bonne corrélation avec le développement des mammifères. Il a pour avantage d'être peu coûteux, facile à élever, à manipuler, à observer au microscope ²⁸⁷. Les œufs et embryons peuvent être obtenus par centaines. Deux modèles peuvent être utilisables : *X laevis*, bien que son génome soit dupliqué ²⁸⁸, et *X tropicalis* qui est plus simple et plus rapide d'utilisation par son caractère diploïde ²⁸⁹. Le génome partage un haut degré de synténie avec l'Homme et une majorité des gènes responsables de maladies génétiques sont conservés. La base de donnée Xenbase permet une disponibilité large à la communauté et inclut des annotations génétiques, des données développementales et anatomiques ²⁹⁰.

1.4.4.2. Manipulations génétiques et applications

De nombreuses manipulations génétiques peuvent être utilisées pour moduler l'expression génique et étudier leurs conséquences comme CRISPR-Cas9 ou les morpholinos ; ces oligonucléotides pouvant être injectés soit dans l'ensemble de l'organisme soit seulement dans une partie (jusqu'au stade 64 cellules). Cela permet notamment de faire des injections unilatérales et d'observer les différences d'un côté à l'autre. L'utilisation de plusieurs morpholinos ciblant plusieurs gènes représente un bon modèle pour évaluer les conséquences d'une haploinsuffisance dans l'investigation des microdélétions.

Un des avantages est de permettre d'étudier des gènes codant pour des protéines de structure proche comme xFmr1 (dont l'inactivation chez l'Homme est responsable du syndrome de l'X fragile) et xFxr1. Le profil d'expression est similaire aux mammifères mais les isoformes sont moins nombreuses et la diminution d'expression par microinjection de morpholinos est responsable de malformations importantes ce qui indique la tissu-spécificité de chaque protéine au cours du développement et permet d'étudier les capacités de compensation entre l'un et l'autre gène de cette famille ²⁹¹.

Un autre exemple d'études physiologiques concerne le rôle des gènes Trio et Kalrn, 2 paralogues faisant partie de la famille des Guanine nucleotide exchange factors (GEFs) activant les Rho GTPases en accélérant la transition du GDP en GTP. Ils ont une fonction conservée dans la morphogenèse dont le développement neuronal. Les variations pathogènes de *TRIO* (l'orthologue humain) sont responsables de DI et de microcéphalie. Les études d'expression ont montré que les 2 gènes étaient détectés au niveau du mésoderme somitique et plusieurs populations neuronales alors que seul trio était exprimé dans les cellules des crêtes neurales et seul kalrn au niveau des nerfs crâniens ce qui indique des fonctions différentes ²⁹².

Les variations perte de fonction de *MECP2* sont associées au syndrome de Rett mais les duplications et triplications du gène sont aussi responsables d'un trouble neurodéveloppemental dont le modèle *X laevis* a contribué à préciser la physiopathologie. La co-transfection d'un plasmide surexprimant le gène *MECP2* humain dans le mésencéphale en développement a permis d'observer une augmentation du niveau protéique dans les neurones post-mitotiques de la plaque tectale et d'observer une perturbation de l'arborisation dendritique, une diminution du nombre de connexions établies puis une perte de connexions en 48 heures ; les dendrites étaient moins nombreuses et leur morphologie simplifiée mais la morphologie des synapses était inchangée ²⁹³.

Le modèle *X. laevis* a été utilisé pour disséquer la région chromosomique 4p16.3 dont la délétion est responsable du syndrome de Wolf-Hirschhorn (du fait du coût du modèle murin et de l'architecture particulière du génome du poisson zèbre). Plusieurs gènes candidats sont supposés impliqués dans le spectre malformatif et la baisse d'expression de ces gènes (isolé ou associé) permet d'observer l'effet sur le développement, avec des études particulièrement développées sur l'atteinte des cellules des crêtes neurales ²⁹⁴.

1.4.5. Modèles animaux dans l'épilepsie

De nombreux modèles animaux ont été développés dans différents syndrome épileptiques.

Par exemple, le syndrome de Dravet représente une maladie épileptique au pronostic réservé et comprend chez l'Homme des crises fébriles puis afébriles prolongées avec myoclonies, des états de mal, une dégradation neurologique incluant marche ataxique et/ou accroupie, troubles cognitifs et

décès prématuré fréquent. Ce syndrome est dû à des variations hétérozygotes perte de fonction du gène *SCN1A* ²⁹⁵ (cf. chapitre 1.3.2). Plusieurs modèles ont été développés pour mieux comprendre les mécanismes impliqués :

- chez la souris, l'invalidation homozygote du gène entraîne une ataxie et un décès vers 15 jours de vie qui peut être décalé vers 17,5 jours en cas de soins appropriés comme une alimentation assistée. Les souris hétérozygotes présentent des crises spontanées et un décès après 15 jours de vie (avec variabilité selon les caractéristiques génétiques de chaque souris). Les analyses électrophysiologiques ont montré que l'activation ou la désactivation de ces canaux voltage-dépendants est inchangée dans les neurones hippocampiques mais le courant généré est réduit dans les interneurones inhibiteurs mais pas dans les neurones pyramidaux excitateurs à la fois chez les souris homozygotes et hétérozygotes. Ce type d'étude a permis d'évoquer fortement le défaut de courant Na+ dans les neurones inhibiteurs GABAergiques à l'origine de l'hyperexcitabilité neuronale des patients ²⁹⁶. Ce type de données a été reproduit dans d'autres modèles murins ²⁹⁷.

- chez le poisson zèbre, un modèle de la maladie de Dravet a été développé en générant des mutants homozygotes *scn1lab*^{s552/s552}. Ces poissons développent des crises d'épilepsie précoces, une augmentation de l'activité locomotrice pendant les phases nocturnes et une altération des comportements exploratoires lorsque les larves sont placées en champ libre dans un nouvel environnement (de type open field) (diminution de la motricité et temps passé à proximité des parois supérieur aux contrôles). Les analyses neuronales n'ont pas montré de diminution significative de l'activité de neurones inhibiteurs ²⁹⁸.

Les œufs de Xenope ont aussi une utilité pour permettre d'explorer la neurotransmission (récepteurs et/ou canaux ioniques) :

- dans certains syndromes associés à une DI comme le syndrome de Rett avec parfois microtransplantation membranaire à partir de cellules de patients ou issues de modèles animaux (technique du Pr Miledi) : des travaux ont pu caractériser les anomalies de courant GABA (cellules humaines ou de modèle animal moins hyperpolarisées que les contrôles) avec restauration de valeurs normales grâce au bumetanide ²⁹⁹.

- et pour l'exploration des conséquences de la présence de mutation sur la fonction des récepteurs ou des canaux (lorsque des variations génétiques sont identifiées chez des patients atteints d'épilepsie isolée et/ou de DI). Par exemple, le gène *KCNT1* est exprimé notamment dans les neurones embryonnaires murins de l'hippocampe et du cortex et les mutations gain de fonction ont pu être incriminées dans un tableau d'encéphalopathie épileptique de mauvais pronostic avec un profil électroclinique fréquemment reconnaissable à type d'épilepsie partielle migrante ^{300,301}. De même, pour le gène *KCNJ10* qui a un rôle d'homéostasie du K⁺ dans les astrocytes, des mutations gain de fonction ont été associées à un phénotype de DI, autisme et épilepsie (souvent à type de spasmes infantiles) ³⁰².

Ce type de modèle cellulaire permet d'accroître le nombre d'outils de modélisation disponibles permettant des analyses fonctionnelles.

D'autres modèles cellulaires existent, pouvant provenir de cellules humaines (cellules HEK293T, HeLa, cellules dérivées de lignées lymphoblastoïdes) ou non humaines (cellules ovariennes de hamster / CHO par exemple).

1.4.6. Avancées thérapeutiques dans le champ de la DI

La meilleure connaissance des voies physiopathologiques impliquées dans la DI et les TSA et le développement de modèles animaux permettent de proposer des pistes thérapeutiques qui pourraient ensuite être transposables chez l'Homme.

En ce qui concerne le syndrome de l'X fragile, les différentes études réalisées sur modèles animaux ont permis de proposer la théorie mGluR (récepteur métabotropique au glutamate) qui fait l'hypothèse que l'absence de fixation de la protéine FMRP à de multiples ARNm pour réguler la traduction de multiples protéines en particulier post-synaptiques, entraîne, en dehors du défaut de régulation de la synthèse protéique, un excès de stimulation mGluR du groupe 1 (dont l'activité stimule également la traduction de certains ARNm en protéines), un trafic excessif de récepteurs AMPA et une augmentation de la dépression à long terme (phénomène important pour l'élimination de synapses selon l'activité, contrairement à la potentialisation à long terme qui semble être prépondérant chez les nouveau-nés pour le maintien des synapses en cours d'établissement) ³⁰³. La réduction de moitié de l'expression normale du récepteur mGlu5 sur des souris KO Fmr1 a permis de restaurer la synthèse protéique, les épines dendritiques et le phénotype comportemental ³⁰⁴. Cela a été confirmé dans plusieurs modèles murins et plusieurs inhibiteurs mGluR5 ayant une spécificité et une pharmacocinétique différente comme la molécule CTEP qui limite l'augmentation de la dépression à long terme, restaure la synthèse protéique, diminue les crises audiogéniques en traitement aigu, améliore le déficit cognitif, l'hypersensibilité auditive, la densité en épines dendritiques l'hyperactivité des voies ERK et mTOR et diminue partiellement la macro-orchidie en traitement chronique ³⁰⁵.

De même chez la drosophile dont le gène *fmr1* a été invalidé, le traitement par 4 antagonistes indépendants de la voie mGluR ou par lithium permet de restaurer des parades nuptiales normales, de conserver la mémorisation de ce type de comportement dans le temps sans toujours rétablir une morphologie normale du mushroom body, ce qui indique que le rétablissement d'une anatomie normale n'est pas nécessaire pour que cette structure qui est le plus probablement impliquée dans ce type de fonction cognitive puisse fonctionner de façon correcte ³⁰⁶.

D'autres molécules ont été testées comme le lithium utilisé classiquement dans les troubles bipolaires qui a été montré comme restaurant la plasticité synaptique, la synthèse protéique et limitant l'activité anormale GSK-3 β (glycogen synthase kinase-3 β , suspecte d'être à l'origine de l'hyperactivité locomotrice et des crises audiogènes) chez la souris KO *Fmr1*^{307–309}; des inhibiteurs de GSK3 qui ont aussi pu améliorer la plasticité synaptique et les déficits cognitifs ³¹⁰; la minocycline utilisée dans l'acné de l'adolescent qui permettrait de réduire le phénotype hyperactif et les anomalies dendritiques ³¹¹; la lovastatine utilisée dans l'hypercholestérolémie familiale qui inhibe la voie Ras, limite l'augmentation de l'activité basale d'ERK, rabaisse la synthèse protéique à des niveaux normaux et améliore la susceptibilité aux crises audiogènes chez la souris ³¹² (avec effets bénéfiques également trouvés sur des cultures neuronales de rat ³¹³; la sertraline, inhibiteur spécifique de la recapture de sérotonine, qui permet d'augmenter le taux de BDNF dans le système nerveux central et d'augmenter le taux de sérotonine.

Cependant, même si des effets bénéfiques ont pu être identifiés dans certains modèles, la transposition chez les patients atteints de syndrome de l'X fragile n'a pas permis d'obtenir un bénéfice clinique franc sur les critères principaux étudiés lors des études cliniques. Certaines molécules permettaient d'améliorer certains paramètres secondaires, ce qui était insuffisant pour valider l'utilisation de ces molécules en pratique clinique ³¹⁴.

Un autre exemple d'utilisation des modèles pour identifier des molécules thérapeutiques concerne le gène *MECP2* dont les variants perte de fonction sont responsables du syndrome de Rett. Le gène régule l'expression de multiples autres gènes par son action sur la méthylation de l'ADN dont *BDNF* (brain-derived neurotrophic factor). La restauration d'expression du gène *Mecp2* par une construction injectée à des souris transgéniques adultes invalidées à différents temps sous l'effet du système Cre/loxP conditionné par l'apport de tamoxifène permet de restaurer la fonction synaptique, de limiter le déclin moteur et de retarder leur décès prématuré, suggérant une piste thérapeutique même à un stade évolutif de la maladie ³¹⁵. En dehors des expériences de manipulation génétique, la surexpression de *BDNF* améliore la fonction synaptique, l'atteinte motrice et prolonge la survie des souris KO ³¹⁶. Cependant la protéine BDNF ne passe pas la barrière hémato-encéphalique. Les pistes s'orientent donc vers la possibilité d'augmenter sa voie de signalisation comme l'apport d'IGF1 qui a montré des effets d'amélioration similaires ³¹⁷.

Un intérêt similaire d'utilisation de modèles pour identifier des voies thérapeutiques existe chez le poisson zèbre. En effet, les larves sont petites et captent les molécules de leur milieu, ce qui les rend très utiles pour tester à grande échelle des molécules thérapeutiques qui y seraient dissoutes. Le matériel par plaques de 96 permet ainsi de tester un animal par puits donc 96 simultanément ²⁶⁹. Ce type de test n'est pas possible chez le rongeur en raison du coût et du temps nécessaire. La découverte de molécules potentiellement d'intérêt permet d'envisager des essais cliniques. Grâce à ce type de stratégie de recherche, des molécules agissant sur divers aspects physiopathologiques ont pu être mis en évidence : amélioration des comportements et étude des voies impliquées de molécules en
développement ou déjà utilisées dont le mécanisme d'action précis pouvait être encore mal connu ²⁷⁰. Cela a pu nécessiter de développer des tests reproductibles avec des molécules connues comme l'haloperidol³¹⁸, a permis d'identifier des molécules agissant sur le sommeil (identification de phytoestrogène biochanin A comme améliorant l'hyperactivité nocturne des poissons dans un modèle DI-autisme par invalidation de cntnap2 présentant une dérégulation des voies GABA et glutamatergiques ³¹⁹, sur les crises d'épilepsie (étude de 320 molécules sur le poisson zèbre *scn1alab* et identification du clemizole, inhibiteur histaminique, qui améliore les convulsions et les anomalies électro-encéphalographiques ³²⁰ ; puis étude similaire de 1000 molécules permettant d'identifier la fenfluramine qui augmente la signalisation de la sérotonine - voie déjà pointée par des études chez des patients où l'épilepsie était améliorée - et la dimethadione qui bloque les courants calciques des neurones ³²¹). Ces preuves de concept permettent de développer des essais thérapeutiques pour montrer si la molécule est efficace chez les patients, comme pour la fenfluramine, en cours de développement par le laboratoire Zogenix ³²². L'élargissement de la pharmacopée est particulièrement intéressant pour permettre des alternatives thérapeutiques, en particulier à des traitements ayant des effets indésirables (comme le lithium) ou lors d'épuisement des ressources disponibles dans les épilepsies pharmacorésistantes par exemple. L'approfondissement de l'analyse de l'effet des molécules candidates sur les voies biologiques et notamment cérébrales permet de préciser le mécanisme sous-tendant leur action.

1.4.7. Limites des modèles animaux

Il existe des limites dans l'utilisation des différents modèles. Du fait de la nécessité d'une preuve d'efficacité chez un modèle avant transposition à l'Homme, cela peut entraîner l'abandon de molécules thérapeutiques qui pourraient être efficaces chez l'Homme et à l'inverse, ce qui a pu être plus souvent expérimenté, des molécules qui ont fait la preuve d'une efficacité dans un modèle se sont souvent avérées peu efficaces chez l'Homme.

En ce qui concerne l'utilisation d'un modèle murin, le cerveau est plus petit que celui de l'humain, le cortex est proportionnellement plus petit et il n'a pas de circonvolutions comme chez le primate qui permet d'augmenter la surface corticale. A l'échelle du génome, bien qu'il existe des similitudes, des divergences ont aussi été constatées lors des projets visant à étudier le transcriptome dans de multiples tissus : il a pu être noté des différences des séquences impliquées dans la régulation transcriptionnelle, avec une conservation d'environ 20% des facteurs de transcription entre l'Homme et la souris, et dans l'organisation de la chromatine ^{233,234,323}. D'autres études comparatives ont révélé de nombreuses différences dans l'expression génique entre ces 2 mammifères, en particulier *FMR1* a un profil d'expression alternant d'une mini-colonne corticale à l'autre chez l'Homme alors que cela n'est pas retrouvé chez la souris ³²⁴. Les comportements humains sont aussi difficiles à modéliser chez

l'animal tant le cerveau humain, son comportement social, sa communication sont complexes, d'autant que les modèles animaux sont manipulés génétiquement et l'investigateur contrôle ses conditions environnementales. Le développement et la maturation cérébraux et de l'organisme sont aussi différents entre les espèces. Les souris naissent à un équivalent du 3^e trimestre de grossesse chez l'Homme, elles ouvrent les yeux vers 1 semaine de vie, passent à une alimentation solide à 3 semaines, atteignent la maturation sexuelle vers 3-5 semaines de vie. L'étude la mieux conduite chez la souris pour inhiber la voie mGluR a consisté à débuter le traitement à 4 semaines de vie peu de temps après le sevrage ³⁰⁵ ce qui correspond à l'adolescence chez l'Homme.

En ce qui concerne le poisson zèbre, plusieurs points nécessitent une attention particulière dans l'interprétation des résultats. L'évaluation de la réintroduction d'un gène ou la surexpression pour évaluer un phénotype i*n vivo* dans des maladies neurologiques non spécifiques comme la DI n'est pas suffisante pour établir le lien de causalité et ne doit être qu'un complément des études chez l'Homme. Il est aussi nécessaire d'évaluer le phénotype des mutants après la première génération qui est obtenue à l'état de mosaïque lors de certaines constructions comme avec le système CRISPR-Cas9 pour éviter de prendre en compte une perte de fonction liée à la manipulation des embryons. Par ailleurs, une évaluation phénotypique est importante pour valider un modèle donné comme pour un modèle *Fmr1* obtenu par morpholinos qui était décrit comme associé à des malformations cérébrales, crânio-faciales et des anomalies de connexions neuritiques ³²⁵ alors que d'autres modèles obtenus par TILLING/ENU n'avaient pas de phénotype malformatif ³²⁶ ; le phénotype malformatif du 1^{er} modèle était probablement dû à un effet hors-cible.

Tous ces éléments posent donc question quant à la possibilité de transposition des molécules identifiées chez l'animal pour une utilisation chez l'Homme. Il convient donc de s'appuyer sur un faisceau d'arguments. En ce qui concerne l'identification de molécules et leur utilisation dans des essais thérapeutiques, un compromis est probablement nécessaire pour trouver la bonne molécule utilisable dans la bonne période pour espérer une amélioration clinique ³²⁷.

1.5. Interprétation des données de séquençage haut-débit

1.5.1. Principales bases de données utiles à l'interprétation des variants génétiques

Depuis les débuts du séquençage et l'identification des variants de séquence génétique, un grand effort a été déployé pour développer des bases de données (répertoriant les variations pour un gène, un groupe de gènes donné ou des variations tout venant) (annexe 3) et des outils ou logiciels permettant de mieux interpréter leurs impacts et conséquences fonctionnelles (annexe 4). De très nombreuses ressources sont actuellement disponibles ayant leurs avantages et leurs limites.

1.5.2. Vers une harmonisation et une reproductibilité dans l'interprétation des variations génétiques

Vu la multitude de variants génétiques présents dans les données de séquençage d'un patient, plusieurs équipes ont proposé des recommandations pour interpréter les variations qui ont pu être révisées par la suite. Par exemple, le Collège Américain de Génétique Médicale et Génomique (ACMG) a proposé les siennes dès 2007 ³²⁸, puis a proposé des révisions en 2015 ³²⁹. Par ailleurs, l'agrégation de données de séquençage et d'analyses fonctionnelles sur modèles cellulaires ou animaux permet d'enrichir progressivement les capacités d'interprétation. Au vu de ces avancées, l'ACMG a établi des recommandations permettant de donner une pondération des différents arguments de pathogénicité et de bénignité des variants (annexe 5). Un score a été mis au point pour permettre une certaine reproductibilité dans la classification (annexe 6) et des comparaisons intra- et inter-laboratoires ont été menées pour valider l'utilisation de ces critères (figure 26).

Malgré l'existence de critères permettant d'aider à l'interprétation, il a pu être démontré qu'il persiste des discordances d'un biologiste et/ou d'une équipe à l'autre (figure 26).



Figure 26 : Comparaison de la classification des variants entre les laboratoires ³³⁰.

(A) Concordance intra-laboratoire: classification selon la méthode habituellement utilisée au sein du laboratoire puis après application des critères de l'ACMG-AMP : il existe 79% de concordance, 16% de différence n'ayant pas de conséquence sur la prise en charge et 5% de différence ayant des conséquences.

(B) Concordance inter-laboratoire de 97 variants utilisant la méthode du laboratoire puis après application des critères établis par l'ACMG-AMP: le taux de concordance chute à 34%, dans 44% il n'y a pas de conséquences et dans 22% cela entraîne une modification de la prise en charge.

(C) Concordance inter-laboratoire après un effort d'échange entre les équipes pour une meilleure définition des critères de l'ACMG-AMP : 71% de concordance, 24% de différence sans conséquence et 5% de différence associé à une modification de la prise en charge.

1.5.3. Stratégies classiques d'identification des gènes à l'aide des outils pangénomiques

Le séquençage de nouvelle génération a permis une explosion des capacités de débit et de données générées. Les applications en génétique médicale ont rapidement été utilisées permettant une augmentation exponentielle d'identification de nouveaux gènes.

De nombreuses stratégies ont été appliquées selon les designs d'étude (matériel biologique disponible, mode de transmission supposé) (figure 27) :



*Figure 27 : Stratégie pour identifier des variants génétiques rares responsables de maladies génétiques à l'aide du séquençage d'exome*³³¹.

a. En cas de séquençage de plusieurs cas index issus de familles différentes, la recherche de variants dans des gènes communs peut être de grande utilité pour sélectionner les candidats, éventuellement associés au trio.

b. Dans des grandes familles où la transmission est autosomique dominante, la comparaison des données de séquençage de plusieurs individus atteints les plus éloignés génétiquement permet de limiter le nombre de variants communs et de sélectionner un nombre plus restreint de candidats.

c. Dans des familles uniques, le séquençage en trio permet d'analyser les variants selon les différents modes de transmission, autosomique dominant par mutation de novo / autosomique récessif par un statut hétérozygote composite ou homozygote / lié à l'X.

Le séquençage de plusieurs membres d'une fratrie issue de parents consanguins, incluant des individus atteints et sains permet de sélectionner les variants homozygotes chez les atteints et non homozygotes chez les sains et d'exclure les variants homozygotes chez les sains. Cependant, il faut être vigilant dans ces familles puisque même en cas de consanguinité la cause de la maladie peut être une mutation de novo ou liée à l'X.

d. Etude d'un trait phénotypique quantitatif illustrant que les individus avec des variants rares dans un gène donné sont regroupés à une extrémité de la distribution.

Ces éléments théoriques ont pu être mis en pratique dans de nombreux centres dans le monde et ont permis de faire de nombreux diagnostics incluant l'identification de nouveaux gènes responsables de maladies génétiques. Parmi les stratégies discutées le plus fréquemment, se pose la question d'un séquençage patient-cas index seul ou avec ses parents pour une interprétation en trio. Une grande étude a notamment permis de comparer le nombre de variants à considérer selon la stratégie utilisée (figure 28, et cf. paragraphes suivants) ³³².



Figure 28 : Analyse des variants candidats chez les 1133 enfants en solo (rouge) ou en trio (bleu) filtrés selon leur statut clinique et le cas échéant les génotypes parentaux (version DDG2P de Novembre 2013) ³³². (A) Histogramme du nombre de variants candidats. (B) Nombre moyen de variants candidats selon le statut clinique des parents et l'utilisation ou non des données de séquençage des parents.

DDG2P=Developmental Disorders Genotype-to-Phenotype database.

La figure 29 suivante illustre bien l'explosion de la découverte de nouveaux syndromes et gènes responsables de maladies rares ² :



Figure 29 : Augmentation régulière du nombre de maladies pour lesquelles des variants pathogènes dans au moins un gène ont été identifiés (barres violettes) et le nombre de gènes dont des variants sont responsables de maladies génétiques (barres orange). Source : Online Mendelian Inheritance in Man

On retrouve la même croissance rapide de découverte des gènes de déficience intellectuelle isolée ou syndromique (figure 30)²²⁵.



Figure 30 : Evolution du nombre de gènes connus pour être responsables de DI isolée ou syndromique indiquant une augmentation constante du nombre de découverte mais un début de stagnation des gènes du chromosome X. Pointillés rouges et jaunes indiquant le début de l'utilisation de la CGH-array et du séquençage haut débit respectivement ²²⁵.

1.5.4. Stratégie pratique de lecture d'exome pour prioritiser les variants en diagnostic et en recherche et rendement des stratégies

1.5.4.1. Stratégie dans le cadre du diagnostic

Dans le cadre des anomalies du développement, les laboratoires de diagnostic et plus rarement de recherche peuvent adopter la méthode d'un séquençage d'un individu atteint seul (dit en solo) en examinant la liste des variants de séquence pour sélectionner ceux situés dans des gènes associés à un phénotype similaire ou, comme mentionné ci-dessus, les variants dans des gènes communs à plusieurs individus atteints. Le diagnostic moléculaire nécessite ensuite la confirmation de la présence du variant par un test ciblé en testant également les apparentés en commençant par les parents ou les autres membres atteints de la famille. Le laboratoire du CHU de Dijon avait initialement adopté cette stratégie pour les analyses à visée diagnostique puis les exomes négatifs étaient analysés en recherche en étendant la lecture à des gènes non encore associés à des maladies humaines. Il s'agit surtout d'étudier les variants les plus faciles à interpréter : absents des bases de données (le plus souvent pour les formes dominantes de novo, <1% pour les formes récessives, tronquants ou variants homozygotes en cas de consanguinité et de s'appuyer sur les données de Z-score / pLI à partir d'ExAC, d'expression, de fonction de la protéine et de comparer le phénotype à un éventuel modèle cellulaire ou animal (morpholinos, knock out, knock in...) s'il existe. La lecture diagnostique consiste à analyser les variants situés dans des gènes connus en pathologie humaine et de les corréler aux phénotypes connus associés à ces gènes. La stratégie d'analyse des exomes en solo appliquée au CHU de Dijon a permis d'obtenir un rendement initial d'environ 22% à 27,4% dans les années 2013-2016 (moyenne 25% sur la période, 104/416) ³³³. L'expérience au laboratoire d'exome solo dans une cohorte relativement homogène de DI sévère et d'encéphalopathie épileptique était de 14/43 soit 32,5% avec chevauchement entre diagnostic et recherche ³³⁴. Certaines équipes ont pris le parti de prévoir un séquençage en trio d'emblée dès que possible ce qui permettait de filtrer beaucoup plus rapidement les variants candidats et d'étendre rapidement vers la recherche lorsque les analyses diagnostiques étaient négatives.

De nombreuses équipes ont rapporté le rendement diagnostique du séquençage d'exome parfois en solo, parfois en trio et parfois en comparant l'apport du séquençage en trio par rapport au séquençage en solo dans des cohortes de patients souvent hétérogènes dont les âges étaient variables et incluant des anomalies du développement, des maladies neurodégénératives ou autres indications rares comme des maladies ophtalmologiques. Une méta-analyse publiée en 2018 a retenu 37 publications, rassemblant 20.068 enfants atteints de maladies rares, permettant d'évaluer le rendement diagnostique dans des cohortes et de comparer au rendement de la CGH-array ³³⁵. Dans ces études, le rendement diagnostique de la CGH-array était de 10% (intervalle de confiance 95% 8–12%, 13 études, 11429 patients), le rendement de l'exome était de 36% (intervalle de confiance 95% 33–40%, 26 études, 9014 patients) et celui du séquençage de génome était de 41% (intervalle de confiance 95% 34–48%). Il existe donc une différence significative de l'analyse d'exome et de génome par rapport à la CGH-array alors qu'il n'y a pas de différence significative entre exome et génome. Par ailleurs, l'analyse des données des articles permet d'estimer une amélioration du rendement diagnostique de l'exome et du génome de 16%/an entre 2013 et 2017 alors qu'il existait une baisse de rendement de 14%/an de la CGH-array sur la même période, probablement du fait de l'inclusion de patients présentant des phénotypes de moins en moins sévères.

Voici les éléments rapportés dans de grandes études permettant d'apprécier l'évolution du rendement diagnostique en fonction des stratégies en sélectionnant les patients atteints d'anomalies du développement :

- une des premières études publiée en 2013 rapportait une cohorte hétérogène de 250 patients avec un diagnostic positif dans 33% (20/60) des formes non syndromiques de trouble neurodéveloppemental (retard des acquisitions, déficience intellectuelle, retard de langage, trouble du spectre autistique) et 22% (31/140) des formes syndromiques *a priori* selon une approche en solo (pas de notion de séquençage des parents). Par ailleurs, au moins 3 des 4 patients ayant un double diagnostic faisaient partie de la sous-cohorte des troubles neurodéveloppementaux syndromiques (soit 0,015% des patients et 5,9% des diagnostics dans la cohorte anomalies du développement) ³³⁶. La même équipe a rapidement étendu sa cohorte pour rapporter les données de 2000 exomes analysés consécutivement et trouvé des résultats similaires avec 143/526 diagnostics positifs, soit 27,2% de rendement dans la DI et le retard global non syndromique, et 282/1147 diagnostics positifs (25,4%) ; par ailleurs 4,6% des patients diagnostiqués avaient plusieurs maladies génétiques ⁴⁴.

- l'étude publiée par Lee et al en 2014 concerne 814 patients atteints de maladies génétiques variées sans diagnostics. Si on ne prend en compte que les patients présentant une anomalie du développement, les rendements rapportés sont de 28/188 soit 14,9% par exome solo (ou 23/170 soit 13,5% en excluant les troubles de différenciation sexuelle qui représente une catégorie à part de leur étude), et 177/554 soit 31,9% en trio (ou 177/540 soit 32,7% en excluant les troubles de différenciation sexuelle) ³³⁷.

- l'étude publiée par Wright et al en 2015 consistait à interpréter des exomes en trio dès que possible (c'est-à-dire quand les prélèvements parentaux étaient disponibles, soit la majorité des cas), le rendement était de 27,4% (311/1133 patients) avec un diagnostic multiple dans 1,5% (17/1133) ce qui représentait 5,2% (17/328) de l'ensemble des diagnostics ³³².

- l'étude publiée par Farwell et al en 2015 rapporte un rendement global de l'analyse solo de 20,6%, et en trio de 37,3% (cohorte de 500 patients, 141 analysés en solo, 21 en duo et 338 en trio). A noter que

la cohorte semble hétérogène avec des patients inclus présentant potentiellement un phénotype neurogénétique comme ataxie isolée ou spinocérébelleuse, paraparésie spastique, Charcot-Marie-Tooth mais aussi amaurose congénitale de Leber, et que ces chiffres incluent probablement autour de 7% de diagnostic probable après une analyse recherche. Une déficience intellectuelle / un retard neurodéveloppemental était rapporté dans 64,4% des patients avec un diagnostic dans 32,6% des cas lorsque ce signe clinique était inclus dans le phénotype, 28,2% des patients avaient plusieurs malformations congénitales et un diagnostic était porté chez 36,2% d'entre eux. Dans la cohorte anomalie du développement, le rendement indiqué était de 36% sans distinction entre solo et trio et il était noté 11/500 soit 0,022% patients avec un double diagnostic ³³⁸.

- le projet canadien FORGE (Finding Of Rare Disease Genes) a aussi trouvé des rendements diagnostiques similaires avec, en 2016, 105/362 patients ayant un diagnostic par mutation(s) dans au moins un gène connu, soit un rendement diagnostique de 29,0%. En sous-catégorisant les présentations cliniques, il est retrouvé un rendement de 31/98 (31,6%) dans les troubles neurodéveloppementaux, 18/80 (22,5%) dans les syndromes avec particularités physiques et 4/19 (21,1%) dans les syndromes polymalformatifs ³³⁹.

La comparaison des stratégies solo par rapport au trio indique que le trio permet de diviser par 10 le nombre de variants à considérer pour le diagnostic lorsque les parents ne sont pas atteints, par 3 lorsqu'un parent est atteint et par 1,5 lorsque les 2 parents sont atteints, ce qui permet une économie de temps d'interprétation non négligeable, bien que cela entraîne des coûts supplémentaires pour le séquençage des parents (figure 28)³³².

L'intérêt du séquençage d'exome réside dans l'utilisation d'une technique unique permettant d'avoir accès à la séquençage d'exome réside des gènes. Il est donc possible de réanalyser les données de séquençage chez des patients ayant initialement un résultat négatif ou non concluant et l'augmentation rapide des connaissances peut faire que la réinterprétation quelques temps plus tard permet d'obtenir un résultat positif. Plusieurs équipes ont démontré l'intérêt d'une réanalyse diagnostique, sachant que l'analyse en trio permet un tri des variants encore plus efficace que dans la stratégie diagnostique. Dès 2014, une étude faite sur 2000 exomes cliniques constatait que 30% des diagnostics faits concernait des gènes identifiés dans les 3 dernières années ⁴⁴, une autre en 2015 rapportait que 23% des diagnostics faits (152/500 soit 30,4%) concernaient de nouveaux gènes publiés dans les 2 années précédentes ³³⁸. L'équipe britannique du projet DDD a réinterprété les données d'exome (principalement en trio) à 3 ans de la 1^{ère} analyse permettant de passer d'un rendement de 27 à 40% (182 nouveaux patients diagnostiqués, essentiellement des nouvelles maladies permettant d'atteindre 454/1133) avec 4% supplémentaire de variants de signification inconnue ³⁴⁰.

Les données récentes du laboratoire indiquent qu'une réanalyse diagnostique périodique des données d'exome en solo permet d'augmenter le rendement diagnostique de 12/156 soit 7,7% par an sur une période de 3 ans avec 2 réanalyses successives entre 2013 et 2016 ³³³. Dans le groupe des troubles neurodéveloppementaux, le rendement diagnostique était de 20 à 21% dans les 2 premières années d'inclusion et la réanalyse diagnostique permet d'obtenir un diagnostic dans 7,5%/an après 2 analyses successives des patients inclus dans la 1ère année et 10% à 1 an pour les patients inclus dans la 2ème année (figure 8).

Par ailleurs, l'interprétation clinique a aussi montré son efficacité pour aider au classement des variants ³⁴¹. L'existence de critères cliniques parmi ceux permettant de classer les variants renforce l'intérêt de la discussion clinico-biologique. La double lecture par un généticien clinicien en plus du généticien moléculaire permet d'avoir une vision globale indispensable pour un meilleur échange, une meilleure justesse dans l'interprétation et pour optimiser le rendement diagnostique. L'interaction clinico-biologique et la possibilité de faire du phénotypage reverse peut aboutir à réviser la classification de variants. Le rôle de la lecture clinique a ainsi montré une augmentation du taux diagnostique passant de 36 à 43% dans une étude (figure 9).





*Figure 8 : Rendement diagnostique de l'exome dans la cohorte du centre de génétique de Dijon indiquant le rendement global et par sous-catégories cliniques et l'apport de la réanalyse périodique des données*³³³.



*Figure 9 : Illustration du rôle du généticien clinicien dans la lecture des tests génétiques de type exome permettant de confirmer ou reclasser les résultats en positif, incertain ou négatif*³⁴¹.

1.5.4.2. Stratégie en recherche

a. Réanalyse systématique des données en recherche

A la différence du diagnostic, il n'est pas possible de rattacher à des corrélations génotypephénotype lors d'une analyse dans le cadre de la recherche, et de nombreux arguments sont à rechercher pour proposer un gène comme candidat (figure 31).



Interpretation in Research analysis

*Figure 31 : Illustration des arguments pouvant être collecté pour orienter vers la pathogénicité ou non de variants dans une stratégie de réanalyse des exomes en recherche*³⁴².

Grâce à ces informations, une analyse en recherche permet aussi d'identifier des gènes candidats ³⁴². L'expérience issue du laboratoire est que la relecture des exomes solo en recherche a permis d'identifier 84 variants candidats dans 66 gènes différents chez 73/313 patients. L'implication dans la

pathologie a pu être établie pour 40 gènes chez 48/313 patients (15%) dont 30% des cas de syndrome malformatif avec DI, 19% de DI isolée/encéphalopathie épileptique et 6% de syndromes malformatifs sans DI. Parmi les 40 gènes, 17 n'étaient pas connus en pathologie humaine, 22 avaient une faible récurrence et n'avaient pas encore été référencés dans la base OMIM et 1 gène était référencé dans OMIM pour un autre phénotype. Pour 26/313 patients (8%), les résultats sont restés non concluants principalement par défaut de récurrence, concernant surtout des faux sens (81%) (figure 32).



*Figure 32 : Illustration de l'apport des réanalyses successives des exomes pour identifier des gènes candidats*³⁴².

Plusieurs équipes rapportent régulièrement leurs estimations de taux de gènes candidats dans leurs cohortes dont les conclusions sur la pathogénicité ne sont pas encore établies de façon définitive mais un faisceau d'arguments font suspecter leur implication :

- l'étude publiée par Farwell et al en 2015 rapporte 7,5% (31/416) de nouveaux gènes candidats ³³⁸.

- l'étude FORGE publiée en 2016 rapporte 88/362 nouveaux gènes soit 24,3% (cohorte hétérogène canadienne) ³³⁹.

- l'étude publiée par Wright et al en 2018 n'indique pas le nombre de nouveaux gènes candidats mais *a priori* uniquement les nouveaux diagnostics certains pour des gènes identifiés comme associés en pathologie humaine dans l'intervalle de réanalyse avec une augmentation du rendement de 27 à 40% entre 2015 et 2018 (182 nouveaux patients diagnostiqués, essentiellement des nouvelles maladies permettant d'atteindre 454/1133) mais aussi 4% de variants supplémentaire interprétés comme de signification inconnue (cohorte de 1133 patients avec anomalies du développement) ³⁴⁰. Il existe néanmoins probablement un certain degré de chevauchement entre les nouveaux gènes identifiés et des gènes candidats en recherche.

b. Autres stratégies d'identification de nouveaux gènes en recherche

En plus des éléments cités ci-dessus, le séquençage pangénomique ouvre la voie à de multiples stratégies pour optimiser l'utilisation des données et augmenter encore le rendement diagnostique avec

dans certains cas la possibilité d'utiliser des données déjà existantes en appliquant une stratégie complémentaire à l'aide de nouvelles analyses bioinformatiques, mais parfois des coûts additionnels lorsqu'il est nécessaire de séquencer des apparentés.

b1. Réanalyse en trio

Dans les situations où le séquençage des parents n'avait pas été fait et en l'absence de piste identifiée après une lecture solo en recherche, la réanalyse en trio peut être complétée notamment dans les formes sporadiques sans consanguinité (parfois avec), permettant de cibler plus spécifiquement les variants *de novo* ou hétérozygotes composites. Dans certaines familles multiplex permettant de suspecter un mode de transmission lié à l'X ou autosomique dominant, une bonne stratégie permettant de filtrer facilement un maximum de variants consiste à séquencer 2 apparentés atteints les plus éloignés génétiquement pour ne considérer que les variants communs.

Plusieurs équipes ont donc mené des études consistant à relire des exomes restés négatifs après une lecture clinique grâce au séquençage d'autres apparentés et à la réanalyse en trio.

Citons à titre d'exemple :

- l'article publié par Farwell et al en 2015, qui rapporte 31/416 (7,5%) de nouvelles pistes fortement candidates en recherche grâce au séquençage et analyse en trio ³³⁸.

- l'article publié par Eldomery et al en 2017, qui a étudié le rendement du trio après exome solo négatif où la majorité des patients étudiés étaient atteints d'anomalies du développement et/ou DI (et 12-14/74 soit 16-19% étaient atteints d'une maladie génétique appartenant à une autre catégorie, telles qu'une maladie ophtalmologique ou un déficit immunitaire notamment) ³⁴³. Dans cette étude, 16% des diagnostics posés après réanalyse auraient pu l'être par une réanalyse diagnostique périodique du fait de l'avancée de connaissances, 20% correspondait à un réel apport de la stratégie de séquençage d'apparenté et réanalyse et il existait un variant candidat dans 15% de cas supplémentaires (figure 33).



*Figure 33 : Résultats de l'étude pilote de réanalyse trio chez 74 patients menés par l'équipe du Baylor College of Medicine à Houston*³⁴³.

<u>b2</u>. Etudes statistiques et d'agrégation de données d'une cohorte pour faire ressortir des récurrences <u>mutationnelles</u>

Deux études de l'équipe de Nijmegen ont particulièrement mis à profit ce type de stratégie :

- étude de récurrence de variants *de novo* prédits pour être responsable d'une haploinsuffisance dans le même gène dans une grande cohorte de patients séquencés en trio pour rechercher un enrichissement : cohorte testée de 2104 trios dont les cas index ont un trouble neurodéveloppemental, dont 820 déficients intellectuels, avec analyses statistiques en faveur de 10 nouveaux gènes candidats pour la déficience intellectuelle: *DLG4, PPM1D, RAC1, SMAD6, SON, SOX5, SYNCRIP, TCF20, TLK2* et *TRIP12*⁷².

- étude d'agrégation de variants similaires dans le même gène (et principalement faux sens localisés dans le même domaine) dans une grande cohorte de patients séquencés en trio pour rechercher un enrichissement de variants *de novo* ³⁴⁴. Cette stratégie a été appliquée à 4043 variants faux sens *de novo* et a montré un enrichissement dans 15 gènes dont 12 connus en maladies humaines, 11/12 ayant un mécanisme physiopathologique différent de l'haploinsuffisance ; elle a aussi permis d'identifier 3 nouveaux gènes candidats responsables de troubles neurodéveloppementaux. La modélisation 3D de ces faux-sens montrait une perturbation dans la structure protéique. Parmi ces gènes candidats, *PACS2* (même variant faux sens *de novo* chez 3 patients) a ensuite pu être validé comme nouveau gène de DI et l'équipe de Dijon a fait partie de la collaboration permettant de décrire le phénotype ³⁴⁵. *ACTL6B* (variant faux sens *de novo* identique identifié chez 3 patients) a été confirmé par une autre équipe rapportant le même faux sens *de novo* chez 9 patients, une autre mutation *de novo* ainsi que 10 patients

avec mutations bialléliques ³⁴⁶ et *GABBR2* a aussi été confirmé comme muté de façon récurrente dans l'encéphalopathie épileptique ³⁴⁷.

b3. Comparaison des données d'une cohorte donnée à des bases de données existantes

Un des membres du laboratoire a mené une étude visant à utiliser la base de données denovodb répertoriant des variants *de novo* pour chercher des récurrences de variants rapportés dans une cohorte donnée ³⁴⁸ : 67 variants ultra rares dans 67 gènes avec des scores bioinformatiques rendant fiable la présence du variant : 32 variants identiques (n=32), 12 variants différents touchant le même nucléotide, 23 variants touchant le même codon. Le variant concernait un gène OMIM morbide dans 29/67 (43.3%) des cas. De plus, 21 variants (31.3%) étaient annotés pathogènes ou probablement pathogènes dans la base de données ClinVar. Le diagnostic était connu pour 21 variants (31.3%) touchant 23 patients. 3 phénotypes étaient non concordants. 2/23 variants concernaient des gènes non associés à une pathologie humaine. Sur les 46 variants restants, 25 ont été écartés avec des arguments de bénignité ou de pathogénicité non suffisants. Parmi les 21 restants, certains concernaient des gènes récemment identifiés dont *GABBR2* et *ACTL6B* et 3 variants ont aussi été retenus dans des gènes non encore associés à une maladie génétique, nécessitant un partage de données inter-laboratoires et d'éventuelles analyses fonctionnelles (*FEM1B, GNAI2, ZFX*) (figure 34).



*Figure 34 : Illustration de la stratégie utilisée pour sélectionner les variants de novo présents dans la base de données denovodb et retrouvés dans notre cohorte de patients analysés par exome*³⁴⁸. *a. répartition des variants faux sens en comparaison avec la base denovodb b. estimation de la pathogénicité des variants identifiés*

<u>b4. Analyse de variants situés dans des gènes candidats spécifiquement ciblés après interrogation</u> <u>de la base de données d'une cohorte</u>

Certains résultats positifs peuvent également être obtenus lorsque de nouveaux gènes sont rapportés à l'occasion de séminaires ou congrès ou lors de publications puisqu'il est possible d'interroger la base de données de notre laboratoire de recherche GAD et de faire ressortir les patients ayant un type de variant particulier dans un gène donné.

b5. Le recul sur l'analyse basée sur le phénotype

En plus de ces différentes stratégies d'analyse, il faut aussi avoir conscience qu'un patient peut avoir plusieurs maladies et l'intérêt des corrélations génotype-phénotype est de savoir, pour un patient avec anomalie du développement/DI, si la mise en cause d'un gène donné (identification d'un ou 2 variants selon l'hérédité) explique l'ensemble du phénotype, si une différence phénotypique peut être attendue par la variabilité d'expression clinique ou si d'autres facteurs génétiques voire une 2^{ème} maladie est à rechercher pour expliquer certaines atypies dans le tableau clinique habituel. Certaines séries ont montré de 1,5 à 5% de double voire triple diagnostic chez un même patient 44,45,332,336,338. L'étude publiée par Posey et al en 2017 représente la plus grande, avec 2076/7374 (28,2%) patients diagnostiqués entre octobre 2011 et avril 2016 dont 7029 solos et 345 trios. Parmi les patients diagnostiqués, 101/2076 soit 4,9% avaient plusieurs diagnostics combinés (97 double, 3 triple, 1 quadruple), représentant 207 diagnostics moléculaires et 2 patients diagnostiqués avaient aussi un facteur génétique actionnable. Sur 207, 54,1% étaient autosomiques dominants dont 67,8% de novo, 16,4% liés à l'X dont 51,7% de novo. Les estimations statistiques issues de cette cohorte et comparée à la population générale auraient été en faveur de 14 à 26% de diagnostics multiples suggérant plusieurs hypothèses : entité sous-diagnostiquée, apparition non liée au hasard, ou létalité des embryons porteurs d'altérations dans plusieurs gènes⁴⁵. Les analyses pangénomiques ont ainsi l'avantage de ne pas cibler un test sur un *a priori* mais au contraire de pouvoir confronter un phénotype observé à l'ensemble du génotype. La discussion clinico-biologique est importante pour vérifier la concordance et la corrélation entre le génotype et le phénotype.

Si les données bibliographiques et/ou bioinformatiques associées à un gène non connu pour causer une maladie humaine (ou parfois associé à un phénotype différent de celui étudié chez un patient donné) font évoquer son implication dans la pathologie, des arguments complémentaires d'implication et de pathogénicité seront nécessaires. Cela inclut la réplication du type de variant chez d'autres patients ayant un phénotype similaire et/ou des analyses fonctionnelles complémentaires.

1.5.5. Intérêt du partage de données inter-laboratoire

La rareté d'un grand nombre d'affections génétiques fait que la probabilité de diagnostiquer la même maladie chez 2 patients non apparentés au sein d'une même équipe est très faible à moins d'avoir pu collecter un nombre important de données. Dans les affections très rares, cela nécessite également une très large connaissance de la syndromologie/dysmorphologie lorsque certains signes pourraient permettre de poser un diagnostic clinique mais ce dernier est souvent difficile à établir, d'autant que de plus en plus de formes atypiques de syndromes connus et de plus en plus de chevauchement phénotypique entre différents syndromes sont décrits.

Un des moyens permettant de rassembler des patients porteurs de variant(s) dans le même gène conduisant à une réplication des données est de partager anonymement les données cliniques et moléculaires sur une plateforme de partage national ou international. Ce moyen a été concrétisé à partir d'octobre 2013 par la connexion de plusieurs bases de données dans le cadre du « Matchmaker Exchange project ». Un patient peut ainsi être créé dans Phenomecentral et Genematcher ce qui permet d'obtenir l'information si d'autres équipes ont déjà été confrontées à ce gène candidat et à des variants similaires pour le phénotype étudié et d'obtenir leurs coordonnées ^{349,350}. En l'absence de partage préalable, cela permet de se rendre visible à la communauté scientifique pour que d'autres équipes puissent nous contacter au moment où ils se voient confrontés à une situation comparable. Cette réplication des données chez plusieurs patients apporte une grande valeur de confirmation, puisque plus il y a d'éléments concordants, plus l'association entre le phénotype et le type de variant pour le gène candidat est fort. A titre d'exemple, GeneMatcher regroupe 4518 utilisateurs originaires de 71 pays pour 8646 gènes candidats soumis (https://genematcher.org/statistics, accessed March 1st, 2018). Le nombre de comptes utilisateurs augmente de façon importante sur ce type de plateforme de partage.

L'expérience de l'équipe GAD FHU TRANSLAD dans l'utilisation de la plateforme GeneMatcher pendant les 2,5 dernières années a pu être rapportée récemment ³⁵¹. Cela a permis de partager 71 gènes candidats non connus en pathologie humaine. Un lien avec au moins une autre équipe a pu être établi pour 60/71 gènes candidats. Parmi les 228 matches auprès d'équipes de 19 pays différents, 155 étaient en rapport avec une équipe américaine, dont plus de la moitié avec GeneDx. Le nombre de lien allait de 0 à 34 selon les gènes et la médiane était de 2 lorsqu'on prenait en compte les gènes pour lesquels au moins un lien pouvait être établi. Les liens se sont faits par email et le délai d'obtention de réponse pour l'ensemble des gènes allait de 5 minutes à plusieurs mois avec une médiane étant de 31 heures. Neuf parmi 38 soumetteurs de gènes candidats n'ont pas répondu à la demande de partage de données. Au total, le partage de données a permis de confirmer l'implication du gène candidat dans 39% des cas et de l'exclure dans 6% des cas (figures 35 et 36).



Figure 35 : Illustration de l'interprétation des variants pour les 71 gènes candidats après « match » ou « hors match » ³⁵¹.



*Figure 36 : Illustration du délai de contact après soumission d'un gène candidat et le délai de réponse email après une demande de partage de données suite à l'identification d'une équipe ayant déjà soumis un gène candidat*³⁵¹.

Cette stratégie s'était avérée très utile dès la réanalyse d'une série de 156 exomes négatifs au laboratoire puisque 12 nouveaux diagnostics pouvaient être envisagés grâce à l'identification de gènes candidats et la récurrence via le partage international des données ³³³.

Ces données confirment l'utilité essentielle de ce type de plateforme pour une identification rapide de nouveaux gènes responsables d'anomalies du développement. En effet, ces échanges rapides

inter-équipe ont permis de faire exploser l'identification de nouveaux gènes responsables de maladies génétiques humaines. Ceci est d'ailleurs illustré par le fait que le rendement diagnostique en 2012-2014 était estimé à 25% alors qu'il atteint 40% environ actuellement ³⁴⁰.

Au final, et avec les estimations de l'apport du séquençage génome entier, les rendements estimés des différentes techniques génétiques peuvent être estimés ainsi (figure 37) :



Figure 37 : Représentation de l'apport progressif des technologies dans le rendement diagnostique chez les patients atteints d'anomalies du développement

1.5.6. De la stratégie phenotype-first vers le genotype-first (figure 38)

Comme cela a été évoqué ci-dessus, de nombreux gènes ont été identifiés du fait que plusieurs individus étaient atteints dans une même famille, et que plusieurs familles étaient concernées, ce qui permettait l'application d'analyses de liaison et cartographies génétiques (notion d'informativité permettant de faire l'hypothèse d'un mode de transmission mendélien précis). Puis les analyses pangénomiques ont permis de comparer les données de séquençage de plusieurs patients apparentés ou non ayant un phénotype similaire pour trouver une cause commune. Certaines maladies ont aussi un phénotype très particulier permettant de cibler un gène ou un groupe de gènes très déterminé. Ces stratégies sont dites « phenotype-first » (pour phénotypage en première intention).

A l'inverse, de nombreuses maladies sont très hétérogènes génétiquement comme la déficience intellectuelle et de nombreux syndromes sont sporadiques, identifiés chez un seul individu au sein d'une famille donnée. Dans ces situations, le rapprochement par comparaison clinique des patients est limité, d'autant que les bases génétiques sont très hétérogènes. Dans d'autres cas, l'expertise clinique peut être dépassée par la rareté d'un syndrome donné dont certains ne sont actuellement identifiés que chez quelques cas dans le monde et nécessiterait de pouvoir suivre l'ensemble de la littérature scientifique qui s'accroit de façon exponentielle. Il devient donc plus rentable de réaliser une étude pangénomique en première intention comme le séquençage de l'exome (puisqu'aucune cause n'apparait évidente et ne peut être confirmée par un test génétique simple) puis d'interpréter l'examen en corrélant les variants présents au phénotype étudié. Ceci est d'autant plus utile dans les pathologies ultra-rares ou ayant une hétérogénéité génétique très importante comme la déficience intellectuelle peu ou pas syndromique. Il s'agit donc dans ce cas d'appliquer une stratégie dite « genotype-first » (pour génotypage en première intention) puis de constituer des cohortes de patients pour lesquels le mécanisme génétique est commun pour comparer leur phénotype a posteriori et améliorer la reconnaissance du syndrome par la suite, si cela est possible, par l'identification de signes cliniques distinctifs. Lorsque les données sont solides, ceci permet de faire reconnaître un type de variant donné touchant un gène donné comme causal pour un phénotype, souvent par le biais de publications, ce qui permet d'aider à l'interprétation de futurs variants du même gène sur de futures analyses pangénomiques chez d'autres patients. Le partage international des données lorsqu'un gène est suspecté candidat se révèle donc indispensable pour avoir le maximum d'arguments. Cela peut permettre dans certaines situations de s'affranchir d'analyses fonctionnelles qui restent cependant parfois indispensables pour conforter la pathogénicité lorsque le nombre de patients est insuffisant. Ces analyses fonctionnelles étudieront l'expression, la localisation de la protéine, et/ou la perturbation de la voie physiologique dans laquelle la protéine intervient.



Figure 38 : Comparaison entre la stratégie de phénotypage en $1^{\text{ère}}$ intention et génotypage en $1^{\text{ère}}$ intention qui s'accompagne de la nécessité d'un retour au phénotype (phénotypage inversé, fait a posteriori).

Ainsi le génotypage en première intention, le partage international de données et la constitution de cohortes de patients permettent de faire des allers-retours entre génotype et phénotype (phénotypage réverse) et une classification clinico-moléculaire des patients atteints d'anomalie du développement. Cette stratégie a prouvé son efficacité pour l'identification rapide de nouveaux gènes pour lesquels les chances de succès de réplication au sein d'une même équipe voire du même pays auraient pu être limitées. A noter que la notion de phénotypage réverse ne consiste pas en un séquençage sans phénotypage initial. Il s'agit de prescrire un test génétique après un phénotypage initial ne permettant pas d'évoquer un diagnostic clinique puis de reconsidérer le phénotype ou de compléter le phénotypage au vu des résultats du génotype. Parfois, les données phénotypiques initiales permettent de formuler une hypothèse d'un groupe de pathologie ayant une hétérogénéité génétique et l'analyse d'exome peut venir confirmer une hypothèse et préciser le syndrome parmi le groupe suspecté ou être non contributif ou plus rarement conduire à des surprises comme des diagnostics différentiels non initialement suspectés.

1.6. Objectifs du travail de thèse

Ce projet de thèse s'inscrit donc dans la continuité de l'effort d'identification des bases moléculaires des anomalies du développement d'origine génétique. Mon arrivée au sein de l'équipe de recherche GAD INSERM U1231 a permis de poursuivre l'exploitation des travaux du laboratoire dans le domaine du séquençage haut-débit. Suite aux travaux d'analyses par étapes en lecture diagnostique puis en recherche des exomes en solo, j'ai pu m'investir pour ceux qui étaient restés négatifs dans la détermination de l'apport du trio, et donc de l'aide du séquençage des parents du cas index pour filtrer les variants plus efficacement, pour le rendement diagnostique des filtres permettant d'identifier un gène dans lequel se trouvaient des variants tronquants hétérozygotes chez plusieurs patients faisant partie de la cohorte syndrome marfanoïde qui avait été constituée au cours d'un PHRC national ; ce gène avait été précédemment suspecté candidat dans une cohorte de 820 patients. Enfin, l'utilisation de stratégies de partage international a permis d'établir plusieurs collaborations pour incriminer un ou plusieurs variants dans un gène non associé en pathologie humaine ou déjà associé à un phénotype différent, permettant dans ce dernier cas d'étendre son spectre phénotypique associé à un nouveau mécanisme mutationnel.

2. METHODES

La démarche diagnostique actuelle comprend plusieurs étapes et s'appuie sur l'expérience acquise au cours des dernières années localement et par les multiples centres utilisant ces techniques via leurs publications. La stratégie du diagnostic moléculaire évolue avec le temps : initialement, le séquençage d'exome en solo était privilégié puis avec l'abaissement des coûts et l'intérêt dans la facilité d'interprétation, un séquençage en trio avec pools parentaux (ADN du cas index, mélange d'ADN provenant de pères ou de mères). Dans les situations restantes négatives, un séquençage du génome et RNA-seq est ensuite envisagé.

Dans le cadre de mon travail de thèse, j'ai participé à la relecture dans un cadre de recherche d'exomes déjà étudiés préalablement dans l'équipe et restés négatifs. Cela concernait des analyses pour des patients de l'équipe ou d'un autre centre français qui avait adressé l'échantillon au CHU de Dijon dans un but diagnostique via Orphanomix ou dans un but de recherche via des appels à collaboration dans le cadre de programmes de recherche type PHRC.

Une première partie du travail a ciblé la réanalyse des exomes en solo. La première stratégie a consisté à réanalyser et comparer des données d'exomes de patients ayant un phénotype similaire, en l'occurrence associant un déficit intellectuel et un syndrome marfanoïde. Les patients ont été recrutés via un PHRC national dédié. Le but était d'identifier de nouveaux gènes (article 1). La deuxième stratégie a consisté à réanalyser des exomes solo en recherche de patients atteints de déficit intellectuel et/ou syndrome malformatif sans diagnostic syndromique ni argument pour l'implication d'un gène ou groupe de gènes en particulier, préalablement à la poursuite des investigations en trio, en élargissant la lecture à la possibilité d'identifier des variants pathogènes dans des gènes déjà connus en maladie humaine mais pouvant être associés à de nouveaux phénotypes, tout en explorant la liste de variants rares localisés dans des gènes non connus en maladie humaine (article 2).

Une deuxième partie du travail a consisté à poursuivre l'effort de recherche de la cause moléculaire pour les patients non diagnostiqués. Ceux pour lesquels les données n'étaient pas contributives, soit parce qu'aucun gène candidat n'était retenu, soit lorsque les études de ségrégation familiale ne permettaient pas de retenir un gène initialement candidat, ont été poussés vers une réanalyse en trio après séquençage des parents (le plus souvent lorsque la maladie était sporadique) en réanalysant d'abord les gènes connus en diagnostic puis en incluant les gènes non connus en maladie humaine dans un but de recherche translationnelle pour l'implication de nouveaux gènes non morbides (article 3). J'ai pu m'impliquer particulièrement dans un travail collaboratif issu de l'identification d'un variant pathogène *de novo* identifié grâce au trio chez un patient atteint d'un syndrome polymalformatif (article 4).

J'ai également participé dans le cadre de mon activité clinique à adresser des demandes d'exome pour des patients présentant une anomalie du développement suspectée comme étant d'origine génétique et pu participer à la double lecture clinico-biologique de ces analyses, à l'origine de publications collaboratives.

2.1. Séquençage et interprétation

Dans un premier temps, un séquençage haut débit d'exome en solo est réalisé à partir des échantillons d'ADN extrait du sang de chaque cas index. La capture est effectuée par plusieurs kits SureSelect Human All exon (50 Mb pour v5, 60 Mb pour v6, 67,3 Mb pour CRE v2) (Agilent). Le séquençage est réalisé à l'aide d'un séquenceur HiSeq 2000 ou HiSeq 4000 (Illumina) avec une lecture en paired-end et des reads de 75 à 100 pb. Les données générées sont ensuite alignées sur un génome de référence (GRCh37/hg19) à l'aide de l'outil Burrows-Wheeler Aligner (BWA, v0.7.15). Les duplicats sont éliminés par le logiciel picardtools v2.4.1. Le réalignement des indels puis la recalibration des scores de qualité de base est réalisée par le logiciel Genome Analysis Toolkit v3.7. Les variants ayant passé le filtre (un score de qualité >30 et une qualité d'alignement >20) sont annotés avec SeattleSeq SNP annotation v9.05 (http://snp.gs.washington.edu/SeattleSeqAnnotation138/).

Les variants retenus sont ceux qui affectent la séquence codante et les sites d'épissage et qui sont qualifiés de rare, c'est à dire qui ont une fréquence inférieure à 1% dans les bases de données dbSNP150 (http://www.ncbi.nlm.nih.gov/projects/SNP/), NHLBI Exome Sequencing Project - Exome Variant Server (http://evs.gs.washington.edu/EVS/), Exome Aggregation Consortium Browser (http://exac.broadinstitute.org/), gnomAD (http://gnomad.broadinstitute.org). Les données de fréquence sont également comparées à un pool de 69 exomes contrôles locaux. Un filtre pour sélectionner les gènes connus dans la base de données OMIM (Online Mendelian Inheritance in Man, http://www.omim.org/) est aussi utilisé.

Les exomes sont lus selon les différentes hypothèses de modes de transmission possibles. Les variants absents des bases de données sont étudiés plus en détail dans les formes sporadiques alors qu'une atteinte familiale ou une consanguinité permettent de suspecter un mode de transmission particulier et d'orienter les recherches de variants. Certains outils de prédiction bioinformatiques *in silico* sont utilisés pour tenter de hiérarchiser les variants potentiellement pathogènes. Le score PolyPhen v2.2.2 (http://genetics.bwh.harvard.edu/pph2/) permet de prédire l'impact d'une variation faux-sens sur la structure et la fonction de la protéine. Les scores PhastCons et GERP renseignent sur la conservation des bases nucléotidiques au cours de l'évolution. La base de données ExAC apporte des informations très intéressantes sur plus de 60000 individus séquencés dans le cadre de différents projets. Bien que le thème de chaque projet puisse introduire un biais (exemple cohorte de schizophrénie ou de troubles bipolaires lorsqu'on étudie les troubles du spectre autistique avec un chevauchement possible des causes génétiques), des tests statistiques ont été appliqués pour comparer le nombre de variants faux-sens ou tronquants observés par rapport à ce qui devrait être attendu si les

variants survenaient au hasard au cours de l'évolution dans un gène donné (Z-score et pLI respectivement). Cela représente un reflet indirect de la pression de sélection au cours de l'évolution de l'espèce humaine. Les différents variants sélectionnés sont visualisés sur Integrative Genomics Viewer v2.4 (https://www.broadinstitute.org/igv/home). La base de données gnomAD permet d'avoir des informations plus précises qu'ExAC sur la fréquence de chaque variant dans la population générale en intégrant plus d'individus (plus de 120000) séquencés par génome. Cela permet donc d'inclure les variants introniques et intergéniques.

La stratégie de prioritisation des variants lors de la lecture d'exomes et l'arbre décisionnel de leur interprétation sont résumés dans les figures suivantes (figures 39 et 40).



*Figure 39 : Résumé de prioritisation des variants lors des lectures d'exome en diagnostic puis en recherche*³⁴³.



Figure 40 : Arbre décisionnel pour l'interprétation des variants génétiques proposé par l'ACMG en 2007 ³²⁸. Evidence: preuves possibles pour argumenter l'implication d'un variant dans la pathologie.

2.2. Stratégie de réanalyse recherche

En se basant sur les éléments discutés précédemment, une méthodologie a été adaptée pour couvrir l'analyse en diagnostic et la réanalyse en recherche solo et trio (figure 31). La figure 41 suivante représente le circuit d'interprétation des analyses.



Figure 41 : Stratégie d'analyse actuelle des exomes en diagnostic (bleu) puis en recherche (ocre). La stratégie tend de plus en plus à adopter une stratégie en trio que les parents soient séquencés seuls ou après mélange d'ADN provenant soit de plusieurs pères soit de plusieurs mères (stratégie de pools parentaux).

2.3. Stratégie de partage de données

Actuellement, plusieurs situations peuvent amener à soumettre un gène au partage international :

- un diagnostic moléculaire est posé dans un gène connu, qu'il soit fréquemment impliqué ou non, ce qui permet de rendre l'équipe visible à la communauté puisque certaines équipes dans le monde pourraient être intéressées par l'inclusion de patients dans des études en particulier descriptives pour améliorer les corrélations génotype-phénotype concernant des gènes déjà associés à des pathologies génétiques.

- un gène candidat est identifié en recherche, que les variants soient de signification inconnue ou que les arguments de pathogénicité soient forts. Dans cette situation, la visibilité vers les laboratoires extérieurs permet d'initier des échanges avec d'autres équipes qui seraient confrontées aux mêmes difficultés d'interprétation ou de renforcer l'implication d'un gène candidat par la réplication de l'identification de variant(s) dans le même gène aux conséquences attendues similaires et chez des patients ayant un phénotype similaire.

En plus de l'utilisation des plateformes de partage international Phenomecentral et Genematcher du consortium MatchMaker Exchange citées dans le paragraphe 1.6, les variants pathogènes et les variants de signification inconnue sont soumis à la base de données ClinVar permettant de pouvoir apporter à la communauté notre expérience dans le cas où d'autres équipes s'avèreraient confrontées à un variant identique ou similaire. Certains dossiers peuvent également être soumis à la plateforme CPMS (Clinical Patient Management System) de la filière européenne sur les anomalies du développement (ERN ITHACA) (plateforme Plaza) permettant d'obtenir des avis d'autres équipes sur le plan diagnostique ou de la prise en charge. Des appels à collaboration peuvent être proposés via l'ERN ITHACA et la filière nationale de maladies rares AnDDI-rares.

3. RESULTATS

3.1. Identification de mutations du gène *DLG4* comme responsables d'un phénotype de déficience intellectuelle - syndrome marfanoïde (article 1)

Ce travail illustre l'intérêt de l'analyse pangénomique sans *a priori* dans une sous cohorte ayant fait l'objet d'une certaine catégorisation clinique pour permettre un regroupement plus précis des patients selon la cause moléculaire. Cette cohorte comprend 64 patients analysés en solo pour 31 d'entre eux et en trio pour les 33 autres.

Méthodes

Pour les patients pour lesquels aucun diagnostic n'a pu être établi dans les gènes connus en pathologie humaine, une stratégie de réanalyse en recherche des exomes solo et trio a été menée. Après une réanalyse individuelle, nous avons réalisé une méta-analyse sur l'ensemble de la cohorte avec l'aide d'outils de commande bioinformatiques pour identifier des gènes au sein desquels des variants tronquants seraient identifiés chez au moins 2 patients différents, dans des gènes dont la pLI est supérieure à 0,9, c'est-à-dire dont la pression de sélection est importante contre les variants hétérozygotes perte de fonction. Le but de ce travail était essentiellement d'identifier de nouveaux gènes non OMIM-morbides responsables de ce groupe phénotypique.

Résultats

Ceci a permis de confirmer l'implication du gène *EHMT1* comme gène fréquemment responsable de ce spectre phénotypique (initialement décrit dans le syndrome de Kleefstra associant notamment déficience intellectuelle, dysmorphie, anomalies génitales, brachydactylie) et de valider la ligne de commande, puisque 3 patients de la cohorte avaient déjà été diagnostiqués comme porteurs de variants probablement pathogènes dans ce gène et ce gène est bien sorti dans la liste. Cela a surtout permis de faire ressortir que 3 patients de la cohorte portaient un variant hétérozygote prédit comme tronquant dans le gène *DLG4*. Ces variants étaient absents des bases de données (2 frameshifts et un situé au niveau d'un site consensus d'épissage). Les patients avaient entre 23 et 35 ans et présentaient une DI légère à modérée, des signes marfanoïdes dont un visage long, un palais ogival, une dolichosténomélie, des doigts longs et fins, un pectus excavatum et une scoliose. Un patient présentait un nystagmus et les 2 autres un strabisme. Les patients identifiés dans notre étude avaient été explorés en solo et la ségrégation familiale a pu montrer que 2 variants étaient survenus *de novo* alors que le 3^{ème} variant était absent chez la mère, l'échantillon du père n'étant pas disponible. Les études des ARN sur sang total ont montré une haploinsuffisance pour les frameshifts et un épissage aberrant conduisant à un codon stop prématuré pour le variant d'épissage, présentés dans l'article joint.

Discussion

Les données bibliographiques indiquent que le gène DLG4 code pour la protéine PSD-95 (postsynaptic density 95), exprimée dans différents tissus dont le cerveau. Ce gène fait partie de la famille des MAGUK (membrane-associated guanylate kinase) dont les protéines contiennent 3 domaines PDZ (PSD95-DLG-Z01), un domaine SH3 (src homology-3) et un domaine GUK (guanylate kinase). Les domaines PDZ correspondent à des sites d'interaction entre protéines qui permettent des liaisons avec les extrémités C-terminales d'autres protéines (annexe 7 rapportant des exemples de protéines interagissant avec PSD-95 via ce domaine, annexe 8 rapportant d'autres exemples de protéines contenant un domaine PDZ). Ces protéines jouent un rôle d'architecture pour l'assemblage de complexes dans des compartiments cellulaires particuliers, notamment les synapses. L'élément postsynaptique des synapses excitatrices cérébrales sont riches dans ce type de protéine dont PSD-95 est la plus connue. Elle contient de multiples domaines qui permettent la liaison à diverses protéines comme des récepteurs, des canaux ioniques, des molécules d'adhésion cellulaire et des molécules de signalisation cytoplasmiques. Elle permet de moduler l'activité des protéines interagissant avec la membrane cytoplasmique en régulant leur localisation cellulaire y compris leur accès à la surface cellulaire, leur stabilité, l'assemblage de sous-unités et certaines propriétés fonctionnelles. Elle joue donc un rôle dans la transduction du signal lors de la transmission synaptique, en impliquant en particulier les voies de signalisation Ras, Rap et Rac. Les expériences de manipulation génétique de DLG4 dans des cultures de cellules et dans des modèles animaux entraînent des perturbations importantes dans la neurotransmission et la plasticité synaptique. La quantité et l'activité de protéines PSD-95 produites est contrôlé par des modifications post-traductionnelles incluant la palmitoylation, la dégradation et la phosphorylation. Ces modifications, ainsi que l'expression temporo-spatiale et les protéines d'interaction sont différentes d'une protéine à l'autre de la famille MAGUK/DLG/PSD. Certaines protéines de cette famille ont un rôle au niveau des synapses cholinergiques, de la fonction visuelle, de la sensibilité algique et des réponses comportementales aux addictions. Une revue est disponible au sujet de ce groupe de protéines ³⁵².



Figure 42 : Illustration par quelques exemples de la localisation et des interactions des protéines à domaine PDZ au niveau de la densité post-synaptique d'une synapse excitatrice chez les mammifères (domaines PDZ représentés par les cercles violets, cercles verts correspondant aux domaines SH3 et cercles bleus aux domaines GUK, lignes en dent de scie représentant les sites de palmitoylation, flèches indiquant les actions d'interaction ou de régulation des protéines sur le cytosquelette). AKAP79, A-kinase anchor protein 79; AMPAR, AMPA (α-amino-3-hydroxy-5methyl-4-isoxazole propionic acid) receptor; βPIX, PAAK-interactive exchange factor; CaMKIIa, α-subunit of Ca2+/calmodulin-dependent protein kinase II; GK, guanylate kinase-like domain; EphR, ephrin receptor; ErbB2, EGF-related peptide receptor; GKAP, guanylate kinase-associated protein; GRIP, glutamate-receptor-interacting protein; IP3R, IP3 receptor; IRSp53, insulin-receptor substrate p53; K ch, potassium channel; LIN7, lin7 homologue; LIN10, lin10 homologue; mGluR, metabotropic glutamate receptor; NMDAR, NMDA (N-methyl-D-aspartate) receptor; nNOS, neuronal nitric oxide synthase; PICK1, protein interacting with C kinase 1; PSD-95, postsynaptic density protein 95; SER, smooth endoplasmic reticulum; SH3, Src homology 3 domain;

Shank, SH3 and ankyrin repeat-containing potein; SPAR, spine-associated RapGAP; SynGAP, synaptic Ras GTPase-

activating protein. 352

PSD-95 est impliqué dans la signalisation des récepteurs glutamatergiques AMPA et NMDA ³⁵³. Cette protéine pourrait également participer dans la maturation de ces synapses. La déplétion en protéine DLG4 changerait le ratio excitation/inhibition des synapses hippocampiques ³⁵⁴. Des variants hétérozygotes tronquants du gène *DLG4* ont été trouvés enrichis chez 3 patients parmi une cohorte de 820 patients atteints d'anomalie du développement incluant une déficience intellectuelle et analysés dans le cadre d'un séquençage d'exome en trio exomes en comparaison à des contrôles; 2 patients présentaient un déficit visuel. La description clinique restait néanmoins très limitée dans cet article ⁷². Des études d'expression sur ARN extrait du sang des patients a montré une haploinsuffisance pour 2 variants et un épissage aberrant conduisant à une protéine tronquée pour le 3^{ème}. L'association entre la fonction du gène, les données bibliographiques et le type de variant a permis de conclure à leur pathogénicité.

Au moins 2 des 3 autres gènes de la même famille ont été impliqués dans la déficience intellectuelle : les variations perte de fonction de DLG2 et DLG3 ont été associées à une forme autosomique dominante et liée à l'X respectivement ^{70,71}.

Conclusion

Au total, nous décrivons de façon exhaustive 3 nouveaux patients porteurs de variants tronquants du gène DLG4, ce qui permet de confirmer son implication dans la DI et probablement l'habitus marfanoïde, compte tenu de la représentation très élevée de ce gène dans notre cohorte (4,7% contre 0,37% de la cohorte de 820 patients DI).

Exploitation de ce travail :

Poster à l'ASHG (octobre 2017, Orlando) :

Moutton S, Bruel AL, Assoum M, Chevarin M, Sarrazin E, Goizet C, Guerrot AM, Charollais A, Charles P, Heron D, Faudet A, Houcinat N, Vitobello A, Tran-Mau-Them F, Philippe C, Duffourd Y, Thauvin-Robinet C, Faivre L. Truncating variants of the *DLG4* gene are responsible for intellectual disability with marfanoid habitus and visual manifestations.

Poster aux Assises de génétique humaine et médicale (Nantes, janvier 2018) :

Moutton S, Bruel AL, Assoum M, Chevarin M, Sarrazin E, Goizet C, Guerrot AM, Charollais A, Charles P, Heron D, Faudet A, Houcinat N, Vitobello A, Tran-Mau-Them F, Philippe C, Duffourd Y, Thauvin-Robinet C, Faivre L. Les variants tronquants du gène *DLG4* sont responsables de déficience intellectuelle associée à un habitus marfanoïde et des troubles visuels.

Publication d'un article scientifique :

Moutton S, Bruel AL, Assoum M, Chevarin M, Sarrazin E, Goizet C, Guerrot AM, Charollais A, Charles P, Heron D, Faudet A, Houcinat N, Vitobello A, Tran-Mau-Them F, Philippe C, Duffourd Y, Thauvin-Robinet C, Faivre L. Truncating variants of the *DLG4* gene are responsible for intellectual disability with marfanoid features. Clin Genet. 2018 Jun;93(6):1172-1178. doi: 10.1111/cge.13243. Epub 2018 Apr 14. ³⁵⁵

ORIGINAL ARTICLE

WILEY GENETICS

Truncating variants of the DLG4 gene are responsible for intellectual disability with marfanoid features

S. Moutton^{1,2} | A.-L. Bruel² | M. Assoum² | M. Chevarin² | E. Sarrazin³ | C. Goizet⁴ | A.-M. Guerrot⁵ | A. Charollais⁶ | P. Charles⁷ | D. Heron⁷ | A. Faudet⁷ | N. Houcinat^{1,2} | A. Vitobello² | F. Tran-Mau-Them^{1,2} | C. Philippe^{1,2} | Y. Duffourd² | C. Thauvin-Robinet^{1,2} | L. Faivre^{1,2}

¹Reference Center for Developmental Anomalies, Department of Medical Genetics, Dijon University Hospital, Dijon, France

²INSERM U1231, LNC UMR1231 GAD, Burgundy University, Dijon, France

³Caribbean Reference Center for Rare Neurological and Neuromuscular Diseases. Fort de France University Hospital, Fort de France, France

⁴Reference Center for Developmental Anomalies, Medical Genetics Department, CHU Bordeaux and Laboratoire MRGM, INSERM U1211, University of Bordeaux, Bordeaux, France

⁵Department of Genetics, Rouen University Hospital, Rouen, France

⁶Department of Neonatal Medicine and Intensive Care, Neuropediatrics and Reference Centre for Learning Disabilities, Rouen University Hospital, Rouen, France

⁷Reference Center for Rare Intellectual Disability Disorders, AP-HP, Pitié-Salpêtrière Hospital, Paris, France and Clinical Research Group "intellectual disability and autism", UPMC, Paris, France

Correspondence

Moutton Sébastien, Reference Center for Developmental Anomalies, Department of Medical Genetics, Dijon University Hospital, Diion, France. Email: sebastienmoutton@hotmail.fr

Funding information

French Ministry of Health, Grant/Award number: PHRC n°2008-A00515-50; Regional Council of Burgundy and Dijon University hospital, Grant/Award number: PARI 2012

1 | INTRODUCTION

heterogeneous group of overlapping disorders. We performed exome sequencing in 33 trios and 31 single probands to identify novel genes specific to MH-ID. After the search for variants in known disease-causing genes and non-disease-causing genes with classical approaches, we searched for variants in non-disease-causing genes whose pLI was above 0.9 (ExAC Consortium data), in which truncating variants were found in at least 3 unrelated patients. Only DLG4 gene met these criteria. Data from the literature and various databases also indicated its implication in ID. DLG4 encodes post-synaptic density protein 95 (PSD-95), a protein expressed in various tissues, including the brain. In neurons, PSD-95 is located at the post-synaptic density, and is associated with glutamatergic receptor signaling (NMDA and AMPA). PSD-95 probably participates in dendritogenesis. Two patients were heterozygous for de novo frameshift variants and one patient carried a a consensus splice site variant. Gene expression studies supported their pathogenicity through haploinsufficiency and loss-of-function. Patients exhibited mild-tomoderate ID, similar marfanoid features, including a long face, high-arched palate, long and thin fingers, pectus excavatum, scoliosis and ophthalmological manifestations (nystagmus or strabismus). Our study emphasizes the role of DLG4 as a novel post-synaptic-associated gene involved in syndromic ID associated with MH.

Marfanoid habitus (MH) combined with intellectual disability (ID) is a genetically and clinically

KEYWORDS

DLG4, intellectual disability, marfanoid, PSD-95, synaptopathy, whole exome sequencing

Marfan syndrome (MFS) is a multisystem genetic disease, whose main clinical criteria include cardiovascular, skeletal and ophthalmological features. The diagnosis is based on an international classification.¹ The severity of the condition mainly lies in the risk of dilation and subsequent dissection of the ascending aorta. Pathogenic variants in the FBN1 gene are the major cause of MFS,^{2,3} but other overlapping

© 2018 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd

syndromes are caused by pathogenic variations within TGFBR1, TGFBR2, TGFB2, TGFB3, and SMAD3 genes whose protein products are involved in the same physiological signaling pathway,^{4,5} including Loeys-Dietz syndrome. Patients with MFS do not usually present with intellectual disability (ID). A group of disorders named marfanoid habitus (MH)-ID was defined when ID patients displayed clinical features of MFS. Among these patients, it was noted that some did not fulfill the classical MFS criteria and displayed atypical presentations, especially for the skeletal features. MH-ID represents a genetically and clinically heterogeneous group of overlapping developmental disorders. Several chromosomal imbalances were associated with this phenotype.⁶ Additionally, several genes, including MED12, UPF3B and ZDHHC9 have been involved in a very limited number of MH-ID phenotype.⁷⁻⁹ MH-ID has also been associated with more syndromic presentations, including Shprintzen-Goldberg (SKI pathogenic variants, usually de novo)¹⁰ and Snyder-Robinson syndromes (SMS pathogenic variants. X-linked inheritance).¹¹ Examples of digenic inheritance have also raised the question that MH could be a nonspecific feature of ID.⁶ However, the diagnostic yield in MH-ID patients is estimated to reach no more than 20% by combining array-CGH and the analysis of known MFS and related syndrome genes,⁶ and it is therefore expected that other genes could be involved in this phenotype. To further decipher the genetic basis of MH-ID, we performed exome sequencing in 64 affected individuals: a combination of trio-based (33 subjects) and single probands (31 subjects). After 2 molecular analysis steps consisting of searching for probably pathogenic variants in known disease-causing genes and searching for candidate genes by means of trio sequencing data allowing the identification of a molecular diagnosis in about half of the patients (personal data), a third step of analysis was applied. This third strategy relied on searching for truncating variants in genes shared by several patients, in order to identify genes specific to MH-ID and to increase the diagnosis yield in our cohort. Here, we report the clinical and molecular results in 3 patients from this series who were heterozygous for truncating DLG4 variants, confirming the implication of this gene in ID.

2 | MATERIALS AND METHODS

2.1 | Patient recruitment

Patients were recruited through a national collaborative study. All clinicians referring patients worked in one of the French reference centers for developmental anomalies. Referred clinical data, photographs and informed consent were obtained in accordance with French law. This study was approved by the institutional review board of Dijon University Hospital (PARI 2012) and was conducted in accordance with the declaration of Helsinki. The patients were recruited when some of the clinical features could suggest a MH even if the systemic score of 7 according to Loeys et al 2010 was not reached,¹ and in the presence of ID with varying degrees of severity. Regarding MH criteria, at least 3 clinical signs among slender habitus, pectus anomaly, thumb and wrist signs, dolichostenomelia, hyperlaxity, scoliosis, elbow retraction, pes planus, high-arched palate, dysmorphism (including long face, enophtalmia and retrognathism) were required.

2.2 | Molecular investigations

2.2.1 | Exome sequencing, bioinformatics analyses, interpretation and validation

The exome sequencing and subsequent analyses were carried out according to classical procedures detailed in Methods in Appendix S1: several Sureselect capture kits (Agilent technologies, Santa Clara, CA), HiSeg 2000 and 4000 (Illumina, San Diego, California), various softwares and web resources were used. Variants were filtered as previously described¹² and were interpreted according to their potential pathogenicity alone, when solo exome-sequencing was performed, and according to the different inheritance hypotheses, when trio exome data were available. The strategy of the molecular analyses followed 3 steps. First, we searched for pathogenic and probably pathogenic variants in known OMIM disease-causing genes allowing the identification of a molecular diagnosis in 43.8% (28/64) of the patients. Then, we searched for probably pathogenic variants in candidate genes not yet known to be associated with the disease by means of the trio sequencing data: in 9.4% (6/64) of the patients, a candidate gene could be found leading to further investigations to confirm the variant pathogenicity (personal data). The third step methodology is presented below.

2.2.2 | Gene filtering procedure

In order to identify novel genes specific to MH-ID and to improve the diagnosis yield in our cohort, we searched as a third step analysis for genes in which truncating variants had been found in at least 2 or 3 unrelated patients of the entire cohort. Genes were then selected if their pLI was above 0.9 (ExAC Consortium data): pLI refers to the probability of loss-of-function (LoF) intolerance, in other words the evolutionary constraint of selection against LoF variants.¹³ Indeed, ID-causing genes are overrepresented in this category. This strategy was applied regardless of the implication of the genes in human genetic diseases but focused especially on non-disease-causing genes, because it was expected that the molecular diagnosis was previously made for known disease-causing genes. The literature and databases were extensively searched for the selected genes (Pubmed, OMIM, UCSC, dbSNP, NHLBI Exome Sequencing Project Exome Variant Server, ExAC browser, gnomAD browser, ClinVar, Genecards, Uniprot) to confirm their potential implication in MH-ID.

2.2.3 | mRNA analyses

Total RNAs were extracted from blood samples from patients and parents (as controls). Samples were collected in PAXgene tubes using the PAXgene Blood RNA Kit (PreAnalytiX, Gmbh, Hombrechtikon, Switzerland) and RT-PCR was performed with the QuantiTect Reverse Transcription Kit (Qiagen, Gmbh, Hilden, Germany) to obtain cDNA libraries.

Oligonucleotide primers used for the expression study were 5'-AATGATCTTGGTGATGAAAATGG-3' and 5'-CATTGGAAAGGGG-TAACTCAGG-3' along with SsoAdvanced Universal SYBR Green Supermix (Bio-Rad Laboratories, Hercules, California). Experiments
Gene filtering process to select candidate genes among the 64 MH-ID patient cohort	Criteria: truncating variant, gene pLI > 0.9
At least 2 patients carrying a variant fulfilling the criteria in the same gene	Four genes: ATR, DLG4, EHMT1, GPR125
At least 3 patients carrying a variant fulfilling the criteria in the same gene	Two genes: DLG4, EHMT1

Representation of the selected genes according to the applied criteria: at least 2 or 3 patients carrying a truncating (stop gain, frameshift or consensus splice site) variant among our cohort of 64 patients, pLI of the genes above 0.9 according to ExAC Consortium data (http://exac.broadinstitute.org/). *EHMT1* represent a known ID-causing gene with autosomal dominant inheritance. *ATR* pathogenic variants cause Seckel syndrome (microcephalic primordial dwarfism type 1) which is inherited as an autosomal recessive trait. *GPR125* variants were inherited from an unaffected parent. Therefore, we focused on the *DLG4* gene.

were performed twice. Primers used to test the consensus splice site variant were 5'-CAGTCCTTGGTCTTGTCGTAATC-3' and 5'-ATG-GAAGTGCACTATGCTCGCC-3' along with HotStarTaq Plus DNA Polymerase (Qiagen, Gmbh). Regarding the splice site variant, PCR products were gel purified with the QIAquick Gel Extraction Kit (Qiagen, Gmbh) and sequenced on a MiSeq (Illumina, San Diego, California) following the Nextera XT library preparation protocol; similar experiments were conducted to analyze the percentage of mRNA harboring the frameshift variants (primers available upon request).

3 | RESULTS

¹¹⁷⁴ WILEY

3.1 | Gene filtering procedure and variant description

Among the 64 affected probands displaying an MH-ID phenotype participating to this study (43 males and 21 females, median age 23.4 years [7-53 years]), 33 underwent trio-based WES and the 31 remaining underwent single proband WES. Our third-step filtering procedure allowed us to select 4 genes with at least 2 patients carrying a truncating variant in the same gene, and 2 genes with at least 3 patients carrying a truncating variant in the same gene, and 2 genes with at least 3 patients carrying a truncating variant in the same gene (Table 1). This approach allowed to highlight *EHMT1* implication in 3 patients (diagnosis already made in the first step of analysis of our cohort), but also identified *DLG4* (OMIM *602887) as a strong novel candidate gene for MH-ID (Table 1). Overall, 3 sporadic patients who had

undergone solo WES were found to carry a heterozygous potentially truncating variant within the *DLG4* gene (Figure 1), hypothesized to lead to LoF by loss of C-terminal domains or haploinsufficiency through nonsense-mediated mRNA decay (NMD). In 2 patients, the variant was a frameshift found to be de novo, whereas in 1 patient, the variant was located at a donor consensus splice site (intron 16). In the latter patient, the variant was not found in the mother but the father was not available for testing. The search for other types of variations of *DLG4* in the exome data of other patients of the cohort remained negative.

3.2 | Patient description

Table 2 summarizes the clinical data of the 3 patients and detailed clinical description is provided in Appendix S1. The patients presented mild (patient 1) to moderate (patients 2 and 3) ID, autism spectrum disorder (ASD) (patient 3) and similar marfanoid features, including a long face, a high-arched palate, a slender habitus, long and thin fingers, pectus excavatum and scoliosis (Figure 2). They did not exhibit ophthalmological or cardiac manifestations typical of MFS but 1 had nystagmus and the 2 others strabismus (additional description in Appendix S1). Using the criteria of the MFS revised systemic score,¹ the score was 7 in patient 1, and 3 in patients 2 and 3. None of the parents exhibited a slender habitus, which would have suggested a familial morphological feature.



FIGURE 1 Representation of the mutations Genomic organization of the *DLG4* locus at 17p13.1 and location of previously reported (black) and herein described (red and framed) variants. Ex: exon. Patient 1 and 2 carry de novo frameshift variants in exons 19 and 11 respectively. Patient 3 carries a variant located at the 16th consensus donor splice site. Protein representation of functional domains and location of the variants at the protein level (previously reported variants in black and herein described variants in red and framed). PDZ, PSD95-disc large-zonula occludens domains; SH3, src-homology region 3 domain. [Color figure can be viewed at wileyonlinelibrary.com]

 TABLE 2
 Summary of the clinical and molecular data for the 3 patients reported in this study



Patient	P1	P2	P3
Birth year	1993	1996	1981
Sex	М	М	М
Age at last examination	23	21	35
Country	France	France-Portugal	France
Variant NM 001365.3 (cDNA level)	c.1843delG	c.1147 1154delTTTATCCT	c.1672 + 2 T > C
Effect at the mRNA level	Reduced expression by half	Reduced expression by half	r.1672 1673insCATC in half reads
Variant NP 001356.1 (protein)	p.(Glu615Serfs*4)	p.(Phe383Glvfs*31)	p.(Glv558Profs*37)
De novo	+	+	NK (absent from mother, father
			not available)
Familial history	-	-	-
Birth term (weeks)	38	Term	39
Birth parameters			
Weight (g)	3450	3150	2850
Height (cm)	51	50	48
HC (cm)	NK	35	NK
Last examination			
Weight (kg)	68	NK	70
Height (cm)	173	183.5	183
HC (cm)	57	NK	56
Arm span (cm)	188	186	190
ID	Mild (IQ56)	Mild to moderate	Moderate
ASD	-	_	+
Seizures	-	-	+
Brain MRI	Ν	Vermis atrophy	Mild cortico sub cortical atrophy
Marfanoid features	+	+	+
Aortic dilation	-	-	_
Cardiac US	Ν	Ν	FMV
Ectopia lentis	-	_	-
Myopia >3 dp	-	-	-
Other ophthalmological findings	Nystagmus	Strabismus	Strabismus
Facial features 3/5	-	-	-
High-arched palate	+	+	+
Wrist sign	+	-	NK
Thumb sign	+	+	NK
Long and thin fingers	+	+	+
Limited elbow extension	-	-	+
Hyperlaxity	+ (Beighton score = 5)	Mild (Beighton<4)	-
Ratio arm span/height	1.09	1.01	1.04
Dolichostenomelia	+	-	-
Pectus anomaly	-	+	+
Hindfoot deformity /pes planus	-/+	-/-	-/-
Pneumothorax	-	-	-
Dural ectasia	NK	NK	NK
Protrusio acetabuli	NK	NK	NK
Scoliosis/kyphosis	Mild	Moderate	Mild
Skin striae	+	_	_

Abbreviations: ASD, autism spectrum disorders; FMV, floppy mitral valve; HC, head circumference; ID, intellectual disability; N, normal; NK, not known.

3.3 | mRNA studies

mRNA expression analyses showed a reduced expression by half in patients 1 and 2, while no significant difference from controls was observed in patient 3 (Figure 3A). In patient 3, sequencing the

mRNA using primers located within exons surrounding the intron 16 consensus donor splice site variant showed that half of the reads had a 4-nucleotide retention from intron 16, leading to a premature termination codon (PTC) 37 amino acids downstream

1176 WILEY GENETICS



FIGURE 2 Photographs of the patients reported in this study. (A-C) Patient 1. (D-H) Patient 2 (brain MRI in panel H shows vermis atrophy). (I) Patient 3. [Color figure can be viewed at wileyonlinelibrary.com]



FIGURE 3 Effects of the *DLG4* variants at the mRNA level. (A) Results of RT-qPCR experiments. RT-qPCR was performed on total RNAs from leukocytes with primers located in exon 6 (forward) and 7 (reverse). Relative quantification of *DLG4* expression showed a 50% reduction in patients 1 and 2. In these patients, the degradation of the PTC-bearing mRNAs by the NMD probably resulted in haploinsufficiency for the PSD-95 scaffolding protein at glutamatergic synapses. Conversely, the level of *DLG4* expression in patients 3 was not affected by the substitution in the intron 16 donor splice site. P, proband; M, mother; F, father. (B) Effect of the consensus splice site variant c.1672+2T>C on pre-mRNA splicing in leukocytes from patient 3. RT-PCR was performed with primers located in the 14th exon (forward) and the junction between 17th and 18th exons (reverse). Deep cDNA high-throughput sequencing allowed us to characterize the effect of the SNV at the RNA level. IGV visualization of cDNA sequences revealed both physiological splicing (r.[=]) from the wild-type allele and abnormal splicing with a 4-nucleotide retention at the 5' end of intron16 (r.[1672_1673insCATC]) containing the variant. The use of a cryptic donor splice site in intron 16 of *DLG4* leads theoretically to a truncated protein p.(Gly558Profs*37) missing the guanylate kinase domain. [Color figure can be viewed at wileyonlinelibrary.com]

(Figure 3B), thus demonstrating the alternative splicing and the probably LoF pathogenic mechanism of the variant (loss of the guanylate kinase like domain). mRNA deep sequencing in patients 1 and 2 showed that 4% and 15% of the reads harbored the variant respectively (data not shown), supporting the haploinsufficiency by the NMD.

4 | DISCUSSION

Here, we report 3 patients carrying *DLG4* truncating variants in our series of 64 MH-ID probands, representing 4.7% of patients of the overall cohort, and 8.1% of those with a positive molecular diagnosis. These patients displayed mild-to-moderate ID, and shared clinical features, including a long face, a high-arched palate, a slender habitus,

long and thin fingers, scoliosis, inconstant pectus excavatum and dolichostenomelia, and visual manifestations (strabismus or nystagmus). Several data point towards the involvement of these variants in the pathogenesis of the clinical presentations: (1) the *DLG4* gene is intolerant to LoF variants (pLI=1, http://exac.broadinstitute.org/),¹³ (2) none of these variants have ever been described in public databases including ExAC and gnomAD browser, (3) these variants were all truncating, resulting in either haploinsufficiency or a truncated protein lacking the guanylate kinase-like domain, (4) these variants were proved de novo in the 2 patients with available unaffected parents DNA, (5) this gene has been considered a candidate gene in patients carrying 17p13.1 deletions encompassing the whole gene,¹⁴ and (6) de novo *DLG4* truncating variants were found to be significantly enriched in a series of 820 ID trio-based WES because 3 patients were concerned (0.37% of the cohort).¹⁵ Unfortunately,

1177

the clinical description of the probands of this study was rather limited and only mentioned that they presented with ID (n = 3) with motor (n = 3) and language (n = 2) delay and unspecific dysmorphism (n = 2). No details on the skeletal features were available. Besides, 2 of these patients also had visual impairment and one of them had been enrolled in a WES study with the aim to diagnose visually impaired patients due to a neurological/cerebral cause (patient 1 from Ref. 16 presenting with nystagmus, pale optic disc and iris transillumination). Although the latter findings could also be reminiscent of albinism and insufficient visual explorations were performed in our patients, these data suggest that *DLG4* gene could be involved in visual function.

Additionally, several functional and literature data support the implication of these DLG4 variants in ID. DLG4 is a 22-exon gene located at the 17p13.1 locus and encodes post-synaptic density protein 95 (PSD-95), a protein comprising 3 PSD95-disc large-zonula occludens (PDZ) domains, an src-homology region 3 (SH3) domain and a guanylate kinase-like (GUK) domain.¹⁷ Northern blot and Western blot analyses detected different transcripts and multiple proteins with complex distribution patterns, including a ubiquitous 85-kD variant and tissue-specific variants.¹⁸ In particular, a brain-specific transcript lacking exon 3 was suspected to result in a 664-amino acid protein containing all functional domains. DLG4/PSD-95 belongs to the discs large (DLG) subfamily of the membrane-associated guanylate kinase (MAGUK) family, which also includes DLG1 and DLG2. They are located at the membrane-cytoskeleton interface. DLG4 and DLG2 were shown to heteromultimerize with each other and to be recruited into the same NMDA receptor and potassium channel clusters. These findings suggest that these 2 MAGUK proteins could play both structural and functional roles by interacting at postsynaptic sites to form a multimeric scaffold for the clustering of receptors, ion channels, and associated signaling proteins.¹⁹ These proteins were also hypothesized to participate in dendritogenesis and glutamatergic synapse maturation. Indeed, various experiments in cell models of hippocampal neurons of various stages of maturation showed pathological effects of PSD-95 over and underexpression. According to the hypothesis proposed to explain these results, various subsynaptic proteins may be variably affected to these expression modifications (reviewed in Ref. 20). DLG4 depletion is thought to change the ratio of excitatory to inhibitory synapses in hippocampal neurons. One of the potassium channels known to interact with these complexes is KCNA1, for which heterozygous missense or dominant-negative variants were shown to cause episodic ataxia, myokimia and/or seizures.²¹ Moreover, FMRP was shown to increase DLG4 mRNA stability, which could contribute to Fragile X Syndrome pathogenesis.²² FRMPD4 gene whose protein product interacts with PSD-95 with a crucial role for the formation of the excitatory synapses and dendritic spines²³ was recently described as associated with X-linked ID in 10 affected male patients from 4 families and functional data of one of the frameshift pathogenic variant showed decreased spine density in transfected hippocampal neurons.²⁴ Different knock-out mouse models have been described. One of them concerned Dlg4 -/- mice generated by deletion of the GUK domain resulting in undetectable protein with unaffected levels of other MAGUK proteins. These mice displayed increased repetitive behaviors, abnormal communication and social behaviors, impaired motor coordination, and increased stress reactivity and anxiety-related responses. Neuropathological examinations showed subtle dysmorphic amygdala dendritic spine changes and altered forebrain expression of various synaptic genes.²⁵ Otherwise, another team studied a similar Dlg4 –/– mouse model and reported that these animals displayed hypolocomotion, limb clasping, and loss of DARPP-32-positive medium spiny projection neurons.²⁶ A third Dlg4 –/+ mouse model found that mice displayed hypersocial behavior in the dyadic interaction test, mild hypoactivity in the open field but no obvious motor deficit.²⁷ Thus, *DLG4*-related disorder can be classified within the group of synaptopathies comprising different alterations of proteins such as neurologins, neurexins and SHANK proteins, which are responsible for diverse neurological and neurodevelopemental phenotypes, including epilepsy, ID, ASD and schizophrenia.²⁸

Regarding mRNA analyses, we show different consequences between the 2 patients carrying a frameshift variant (reduced expression by NMD) and the patient with a consensus splice site variant (no NMD but intron retention leading to a PTC). It is now widely accepted that NMD functions in at least 2 distinct cellular processes: downregulation of abnormal transcripts carrying a PTC, and maintenance of an appropriate level of gene expression by downregulating physiological mRNAs in response to cellular needs. NMD usually occurs when the PTC is located 50-55 nucleotides upstream the last exon junction complex but this is far from being systematic and multiple factors can influence NMD as reviewed in Ref. 29. From our data, it is not possible to speculate about the factors influencing NMD occurrence in our patients.

Overall, we confirm that *DLG4* is a novel gene that codes for a synapse-associated protein and that truncating variants of this gene causes syndromic ID, associated with several marfanoid features and visual manifestations. Considering that this gene explains 4.7% of our MH-ID cohort, vs 0.37% of a large series of ID patients,¹⁵ we believe that *DLG4* is a strong candidate gene in patients with MH-ID. Additional descriptions are required to better define the clinical spectrum associated with *DLG4* pathogenic variants.

ACKNOWLEDGEMENTS

We are grateful to the patients and their families who accepted to participate to this study. We thank PerkinElmer DNA Sequencing Service for the whole-exome sequencing experiments, and the Université de Bourgogne Centre de Calcul (CCuB, see Web Resources) for technical support and management of the informatics platform. We thank French Ministry of Health who participated in funding by PHRC n°2008-A00515-50, as well as Regional Council of Burgundy/ Dijon University hospital (funding PARI 2012). The Ethics Committee CPP Est I approved the study on June 19th of 2008. The authors thank the subjects and families involved in the study, and the Université de Bourgogne Centre de Calcul (CCuB, see Web Resources) for technical support and management of the informatics platform. The authors also thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at http://gnomad.broadinstitute.org/about. We also thank public databases,

1178 WILEY GENETICS

including the NHLBI GO Exome Sequencing Project (see Web Resources) and its ongoing studies which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010). We thank Philip Bastable (DRCI, Dijon CHU) who revised the English version of the manuscript.

Conflicts of interest

Nothing to declare.

ORCID

S. Moutton b http://orcid.org/0000-0002-8942-6731 A.-L. Bruel b http://orcid.org/0000-0002-0526-465X

REFERENCES

- 1. Loeys BL, Dietz HC, Braverman AC, et al. The revised Ghent nosology for the Marfan syndrome. J Med Genet. 2010;47:476-485.
- Loeys B, De Backer J, Van Acker P, et al. Comprehensive molecular screening of the FBN1 gene favors locus homogeneity of classical Marfan syndrome. *Hum Mutat.* 2004;24:140-146.
- **3.** Faivre L, Collod-Beroud G, Loeys BL, et al. Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study. *Am J Hum Genet*. 2007;81:454-466.
- 4. Stheneur C, Collod-Beroud G, Faivre L, et al. Identification of 23 TGFBR2 and 6 TGFBR1 gene mutations and genotype-phenotype investigations in 457 patients with Marfan syndrome type I and II, Loeys-Dietz syndrome and related disorders. *Hum Mutat.* 2008;29: E284-E295.
- MacCarrick G, Black JH 3rd, Bowdin S, et al. Loeys-Dietz syndrome: a primer for diagnosis and management. *Genet Med.* 2014;16:576-587.
- Callier P, Aral B, Hanna N, et al. Systematic molecular and cytogenetic screening of 100 patients with marfanoid syndromes and intellectual disability. *Clin Genet*. 2013;84:507-521.
- Schwartz CE, Tarpey PS, Lubs HA, et al. The original Lujan syndrome family has a novel missense mutation (p.N1007S) in the MED12 gene. *J Med Genet*. 2007;44:472-477.
- **8.** Tarpey PS, Raymond FL, Nguyen LS, et al. Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. *Nat Genet.* 2007;39: 1127-1133.
- **9.** Raymond FL, Tarpey PS, Edkins S, et al. Mutations in ZDHHC9, which encodes a palmitoyltransferase of NRAS and HRAS, cause X-linked mental retardation associated with a Marfanoid habitus. *Am J Hum Genet*. 2007;80:982-987.
- Carmignac V, Thevenon J, Ades L, et al. In-frame mutations in exon 1 of SKI cause dominant Shprintzen-Goldberg syndrome. Am J Hum Genet. 2012;91:950-957.
- Becerra-Solano LE, Butler J, Castaneda-Cisneros G, et al. A missense mutation, p.V132G, in the X-linked spermine synthase gene (SMS) causes Snyder-Robinson syndrome. Am J Med Genet A. 2009;149A: 328-335.
- 12. Nambot S, Thevenon J, Kuentz P, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genet Med.* Accessed November 02, 2017 [Epub ahead of print].

- Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285-291.
- 14. Zeesman S, Kjaergaard S, Hove HD, Kirchhoff M, Stevens JM, Nowaczyk MJM. Microdeletion in distal 17p13.1: a recognizable phenotype with microcephaly, distinctive facial features, and intellectual disability. Am J Med Genet A. 2012;158A:1832-1836.
- Lelieveld SH, Reijnders MR, Pfundt R, et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci.* 2016;19:1194-1196.
- **16.** Bosch DG, Boonstra FN, de Leeuw N, et al. Novel genetic causes for cerebral visual impairment. *Eur J Hum Genet*. 2016;24:660-665.
- Stathakis DG, Hoover KB, You Z, Bryant PJ. Human postsynaptic density-95 (PSD95): location of the gene (DLG4) and possible function in nonneural as well as in neural tissues. *Genomics*. 1997;44:71-82.
- Stathakis DG, Udar N, Sandgren O, et al. Genomic organization of human DLG4, the gene encoding postsynaptic density 95. J Neurochem. 1999;73:2250-2265.
- Kim E, Cho KO, Rothschild A, Sheng M. Heteromultimerization and NMDA receptor-clustering activity of Chapsyn-110, a member of the PSD-95 family of proteins. *Neuron*. 1996;17:103-113.
- 20. Bustos FJ, Varela-Nallar L, Campos M, et al. PSD95 suppresses dendritic arbor development in mature hippocampal neurons by occluding the clustering of NR2B-NMDA receptors. *PLoS One*. 2014;9:e94037.
- **21.** Eunson LH, Rea R, Zuberi SM, et al. Clinical, genetic, and expression studies of mutations in the potassium channel gene KCNA1 reveal new phenotypic variability. *Ann Neurol.* 2000;48:647-656.
- **22.** Zalfa F, Eleuteri B, Dickson KS, et al. A new function for the fragile X mental retardation protein in regulation of PSD-95 mRNA stability. *Nat Neurosci.* 2007;10:578-587.
- 23. Lee HW, Choi J, Shin H, et al. Preso, a novel PSD-95-interacting FERM and PDZ domain protein that regulates dendritic spine morphogenesis. *J Neurosci.* 2008;28:14546-14556.
- Piard J, Hu JH, Campeau PM, et al. FRMPD4 mutations cause X-linked intellectual disability and disrupt dendritic spine morphogenesis. *Hum Mol Genet*. 2018;27:589–600.
- 25. Feyder M, Karlsson RM, Mathur P, et al. Association of mouse Dlg4 (PSD-95) gene deletion and human DLG4 gene variation with phenotypes relevant to autism spectrum disorders and Williams' syndrome. *Am J Psychiatry*. 2010;167:1508-1517.
- 26. Zhang J, Saur T, Duke AN, et al. Motor impairments, striatal degeneration, and altered dopamine-glutamate interplay in mice lacking PSD-95. J Neurogenet. 2014;28:98-111.
- Winkler D, Daher F, Wustefeld L, et al. Hypersocial behavior and biological redundancy in mice with reduced expression of PSD95 or PSD93. *Behav Brain Res.* Accessed February 09, 2017.
- **28.** Grant SG. Synaptopathies: diseases of the synaptome. *Curr Opin Neurobiol*. 2012;22:522-529.
- Hug N, Longman D, Caceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* 2016;44: 1483-1495.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Moutton S, Bruel A-L, Assoum M, et al. Truncating variants of the *DLG4* gene are responsible for intellectual disability with marfanoid features. *Clin Genet*. 2018;93:1172–1178. https://doi.org/10.1111/cge.13243

SUPPLEMENTARY DATA

Supplementary methods: Exome sequencing, bioinformatics analyses, interpretation, validation of variants

Supplementary results: clinical description of the patient carrying a heterozygous truncating *DLG4* variant

to Review Only

Supplementary Methods: Exome sequencing, bioinformatics analyses, interpretation, validation of variants

Exome sequencing

Genomic DNA was extracted from blood using standard procedures. Paternity-maternity testing of available parental DNA samples was performed by genotyping six highly polymorphic short tandem repeats. All exome capture and sequencing experiments were performed using genomic DNA extracted from whole blood. Three micrograms of genomic DNA per individual were subjected to whole-exome capture using the SureSelect Human All Exon V5 kit (patient 1 and 3) and the SureSelect Human All Exon CRE kit (patient 2) (Agilent). The resulting libraries were amplified and sequenced on a HiSeq 2000 (Illumina) for patients 1 and 3 and HiSeq 4000 (Illumina) for patient 2 as paired-end 75 bp reads.

Bioinformatics analyses

FASTQ files were aligned to a human genome reference sequence (GRCh37/hg19) using BWA (Burrows-Wheeler Aligner; v0.7.15). All aligned read data were subjected to the following steps: (1) duplicate paired-end reads were marked by Picard 2.4.1, (2) indel realignment and (3) base quality score recalibration were done on Genome Analysis Toolkit (GATK; v3.7). Variants with a quality score >30 and alignment quality score >20 were annotated with SeattleSeq SNP Annotation. Rare variants present at a frequency above 1% in dbSNP 150, in the NHLBI GO Exome Sequencing Project, in the ExAC Consortium project, in gnomAD, or present in one of 100 in-house control exomes from unaffected unrelated individuals were excluded (see Web resources).

Interpretation

The in-house pipeline was developed for variant classification considering i) specific variants reported in public databases to be associated with a human disorder; ii) variants in a specific gene associated with a human disorder referenced in OMIM (See URLs) or manually curated via Pubmed (See URLs); iii) variant type consistent with both the clinical presentation and the suspected mode of inheritance (i.e. X-linked, autosomal recessive or autosomal dominant), with a more rapid filtering process when the exome sequencing was part of a trio.

Validation

Candidate variants were confirmed in probands by Sanger sequencing approaches according to the manufacturer's instructions. Subsequent sequencing of parents DNA allowed us to search for inheritance, especially when the parental exomes had not been sequenced. Primers were designed to surround the candidate variants, and PCR reactions were performed using HotStartTaq® Plus DNA Polymerase (Qiagen, Gmbh). PCR products from genomic DNA were sequenced using BigDye dideoxyterminator chemistry and the ABI3130xl Genetic analyzer (Applied Biosystems, Foster City,

2 3	CA, USA). Primer sequences and positions, PCR conditions and product sizes are available upon
4	request.
5	
7	
8	
9 10	
11	
12	
13	
15	
16	
17	
19	
20 21	
22	
23	
24 25	
26	
27	
29	
30	
31	
33	
34 35	
36	
37	
38 39	
40	
41 42	
43	
44	
46	
47	
48 49	
50	
51	
52 53	
54	
55 56	
57	
58	
59 60	

Supplementary results: clinical description of the patient carrying a heterozygous truncating *DLG4* variant

Patient 1 was born to non-consanguineous parents. The parents did not exhibit a marfanoid habitus. The father was 178 cm tall, 78 kg in weight with an OFC of 58 cm; the mother was 158 cm tall, 55 kg in weight with an OFC of 57 cm.

The pregnancy and birth were uneventful. The patient had no particular developmental delay since he could sit unsupported at age 6 months, walk alone at age 12 months, and spoke his first words at age 18 months and his first sentences at age 36 months. He learned to read and write but attended a special needs school and a private vocational high school to obtain a professional high school diploma. He was able to work in a protected environment on graphics. IQ performed at age 17 years was assessed at 56. He was described as anxious.

Examination showed a height of 173 cm, which is consistent with his target height, a long face, a high arched palate with a nasal voice, marked dolichostenomelia, some degree of hyperlaxity and mild unilateral nystagmus. He underwent 2 surgeries for umbilical hernia and was followed for mild scoliosis from age 14 years onwards.

Cardiac ultrasound, ophthalmological examination, brain MRI were normal. Karyotyping, search for fragile X syndrome, array-CGH and homocysteine dosage were not contributive.

Patient 2 was born to non-consanguineous parents, and the family history was not informative. The parents did not display marfanoid features. The father's height was 172 cm, the mother's 157 cm, and the paternal grandfather's 186 cm. The father had large kidneys due to ureteral malimplantation. The pregnancy and birth were uneventful: birth parameters were within normal ranges. He presented with transitory neonatal jaundice. At age 6 months, axial hypotonia was noted. He could walk unassisted at age 18 months and spoke his first word at age 24 months and his first sentences at age $3\frac{1}{2}$ years. Audiometry was normal. He was able to learn 2 native languages with comprehension difficulties in both of them. He was educated in a special needs school from age 6 years onwards. He required a ritualized lifestyle but other behaviors were not consistent for ASD. His autonomy was limited and he needed verbal assistance to perform the main daily activities. At age 17 years, his school abilities were assessed to be equivalent to those of a 7-8-year-old child. He was able to read and write. He was followed for myopia, prevailing on the left side, and scoliosis. He was also treated for urticaria. He did not experience seizures. Morphological examination showed a high arched palate, long and thin fingers and pectus excavatum. Neurological examination revealed mild pyramidal signs, static ataxia, buccal apraxia, dystonia, synkinesis, during a test involving rapid alternating movements of the elbow, decreased hand muscle strength, clumsiness and strabismus. He also had a dental articular disorder but this finding was also present in his mother and his half-sister.

Clinical Genetics

Growth charts indicated height at +1.5SD (standard deviation) (target size -1SD), weight was average and OFC was +0.5SD. The arm span to height ratio was 1.01. Spine X-rays performed at age 17 years revealed a 14° T1-T6 left convexity and 12° T11-L3 right convexity scoliosis with moderate vertebral rotation. The different brain MRI performed at age 9 months, 3 years and 16 years showed mild left occipital subcortical atrophy with very mild FLAIR white matter hyperintensities. The last control also found mild vermis atrophy. Cardiac ultrasound with aortic measurements was normal. Karyotyping, array-CGH, FMR1 CGG repeat count, transferrin electrophoresis and homocysteine dosage were normal.

Patient 3 was born to non-consanguineous parents. The family history was unremarkable and the parents did not present marfanoid features. The mother was 168 cm tall with an OFC of 56 cm. The father's height was estimated at 185 cm. The pregnancy was marked by decreased active fetal movements. The birth was uneventful. Measurements were in normal ranges (2850 g, 48 cm, OFC not available). Developmental delay was noted early during infancy since head control was achieved at 4-5 months. The patient was able to sit unsupported at 12 months and walked alone at 24 months of age. Language skills were severely impaired and the neuropsychological evaluation performed at age 12 years was not possible, even using non-verbal tests, suggesting moderate to severe ID. He has never been able to read or write. He was also described as having problems with relationships and stereotyped activities, and the diagnosis of autism spectrum disorder was made. He went to a special needs school and attended a vocational education institution as an adult. He had limited autonomy and needed help to dress and get washed. He suffered from seizures starting at age 14 years. He was treated with olanzapine and sodium valproate.

He was 183 cm in height (+1.7SD), target height +1.3SD), 70 kg in weight (+0.9SD) with an OFC of 56 cm (-0.6SD). Arm span was 190 cm and the ratio to the height was 1.04. In addition to classical marfanoid habitus features, he exhibited a long face, telecanthus, a short philtrum, a thin vermillion of upper lip (figure 1), a long 2nd toe, increased space between the 1st and 2nd toes, ulna valga, lumbar hyperlordosis and delayed puberty. Mild scoliosis was diagnosed but did not require any specific intervention. The ophthalmological evaluation showed mild myopia, limited convergent and vertical gaze movements. Cardiac ultrasound showed a floppy mitral valve. The neurological examination identified tandem gait difficulties without dysmetria but fine motor anomalies, strabismus and brisk reflexes. Brain MRI performed at age 25 years showed cortical subcortical atrophy and mild ventricle enlargement prevailing on the left side. MR spectroscopy showed a slight decrease in N-acetyl aspartate. The tractography revealed a mild increase in right cingulum volume.

Karyotyping, search for fragile X syndrome, array-CGH and homocysteine dosage were not contributive.

3.2. Implication du gène *KCNQ3* dans un nouveau phénotype comprenant déficience intellectuelle avec épilepsie néonatale potentiellement pharmacodépendante (article 2)

Ce travail illustre l'intérêt de l'analyse pangénomique en solo et de la confirmation de la pathogénicité par des tests fonctionnels lorsqu'un gène est identifié comme candidat pour la pathologie alors qu'il est déjà connu dans un autre phénotype et en l'absence d'un nombre suffisant d'autres familles pour une réplication des données.

Cas clinique et données connues au moment de l'étude

A l'occasion du suivi dans le cadre de mon activité clinique d'une famille demandeuse de conseil génétique dans laquelle la 3^e enfant (d'une fratrie de 3 enfants nés de parents consanguins) présentait une déficience intellectuelle et une épilepsie, l'interprétation de l'exome du cas index en solo en diagnostic puis en recherche a conduit à retenir 2 variations homozygotes, l'une faux-sens du gène *DMBT1*, l'autre tronquante du gène *KCNQ3*.

Le 1^{er} gène, *DMBT1*, a une expression dans de multiples tissus et des délétions et mutations ont été impliquées dans le médulloblastome, le glioblastome de haut grade et l'oligodendrogliome. Cette piste a pu être écartée car un frère sain était aussi porteur de la variation à l'état homozygote.

Le 2^{ème} gène, *KCNQ3*, était connu pour être responsable d'épilepsie familiale néonatale bénigne de type 2 (BFNS2) et d'épilepsie familiale infantile bénigne (BFIS) associée à des mutations faux sens hétérozygotes responsables d'une perte variable de fonction de la protéine, soit par diminution de la conductance du canal K+, soit par effet dominant négatif ³⁵⁶. La forme BFNS2 est caractérisée par des crises débutant en moyenne dans les premiers jours de vie et disparaissant entre 1 et 6-12 mois. Les crises sont habituellement courtes (1-2 minutes) et peuvent être polymorphes dans leurs présentations. La forme BFIS est caractérisée par un début plus tardif après le premier mois de vie mais dans la première année et une résolution entre l'âge d'1 et 2 ans ; les crises sont courtes, peuvent se répéter et sont variables dans leur présentation mais plus souvent focales. Dans ces 2 situations, les patients ont habituellement un développement normal mais certains articles ont rapporté des troubles neurodéveloppementaux associés (I317T ³⁵⁷; R330L ³⁵⁶). Les corrélations génotype-phénotype ne sont pas claires puisque les effets fonctionnels électrophysiologiques sont variables avec des réductions de courant ionique K+ de 20-40% à plus de 60% sans que cela puisse prédire un retard neurodéveloppemental ³⁵⁶.

La famille des gènes KCNQ code pour les protéines Kv7 et comprend 5 membres (KCNQ1-5) avec une expressivité tissulaire et des rôles fonctionnels différents ^{358–360}. Ils sont composés d'une région principale de 6 domaines transmembranaires (S1-S6) et 2 domaines cytoplasmiques dans les régions N- et C-terminales. Les protéines KCNQ2 (Kv7.2) et KCNQ3 (Kv7.3) sont des canaux ioniques K+ voltage dépendants co-exprimés au niveau des corps cellulaires, des dendrites et des

segments initiaux des axones des neurones hippocampiques et corticaux ³⁶¹ où l'association comme hétérodimère a été impliquée dans la physiologie du courant M (nommé pour sa sensibilité à l'inhibition par la muscarine) qui est un courant de K+ qui s'active et s'inactive lentement et en lien avec le voltage supposé stabiliser le potentiel de membrane et contrôler l'excitabilité neuronale en limitant les activations répétées et qui joue un rôle critique dans la détermination du seuil d'excitabilité électrique des neurones et dans la réponse aux afférences synaptiques ^{362,363}. Initialement le courant M a été décrit au niveau des neurones du système nerveux périphérique sympathique ^{364,365}, puis il a été identifié dans de nombreux neurones du système nerveux central ³⁶⁶⁻³⁶⁸. L'inhibition par la muscarine entraîne une augmentation de l'activation des neurones. En dehors de leur localisation somatodendritique du corps cellulaire des neurones corticaux et hippocampiques chacune de ces protéines ont quelques spécificités d'expression ³⁶⁹: KCNQ2 est exprimé au niveau du ganglion cervical supérieur du système sympathique, mais est aussi localisé seul au niveau de la région terminale de l'axone (figure 20) et de l'élément pré-synaptique ^{370,371}, ce qui indique un rôle différent de celui du canal M, plutôt pour la régulation de la propagation du potentiel d'action et de la libération de neurotransmetteur mais KCNQ2 n'est pas exprimé au niveau du cervelet, alors que KCNQ3 est exprimé au niveau du ganglion cervical supérieur mais aussi des ganglions paravertébraux et dans une moindre mesure du cervelet. La voie du phosphatidylinositol (4,5)-bisphosphate (PIP2) a été identifié comme régulateur positif des différentes protéines de la famille KCNQ dont la protéine KCNQ2 humaine et l'hétérodimère KCNQ2 humain et Kcnq3 du rat³⁷².

Le nombre de variations pathogènes identifiées dans le gène *KCNQ2* est relativement important en comparaison à celles du gène *KCNQ3*. Il existe un certain degré de corrélation génotype-phénotype des variations de *KCNQ2* : celles responsables de perte de fonction par haploinsuffisance sont majoritairement héritées, en général responsables d'épilepsie néonatale bénigne (BFNE1) alors que les faux sens *de novo* ont été associés à une encéphalopathie épileptique néonatale souvent à type de suppression burst (ou syndrome d'Ohtahara), parfois des spasmes infantiles (syndrome de West) probablement dû à un effet dominant négatif perturbant la fonction des canaux ^{373,374}.

La ségrégation de la variation homozygote de *KCNQ3* dans la famille suivie dans notre centre était compatible avec la possibilité de son implication et l'hypothèse d'un nouveau phénotype associé à ce gène a donc été proposée.

Méthodes

Comme l'équipe référente de l'Université de Naples disposait déjà d'un modèle pour tester la pathogénicité des variants *KCNQ3*, une collaboration a été établie, des expériences ont pu être menées pour évaluer l'expression des différents gènes KCNQ dans des fibroblastes en culture primaire issus d'une biopsie de peau de la patiente et de son frère sain porteur d'un génotype homozygote sauvage à ce locus. Des études en immunohistochimie ont permis d'évaluer la localisation cellulaire de la

protéine KCNQ3. Des tests électrophysiologiques sur cellules CHO ont été menés après transfection d'un plasmide contenant une version sauvage ou porteuse de la mutation de KCNQ3.

<u>Résultats</u>

Ces expériences ont permis de montrer une diminution conjointe de l'expression de l'ARN messager de KCNQ3 mais aussi KCNQ2 et KCNQ4 chez la patiente, la diminution de l'expression protéique de KCNQ3 avec une protéine de taille réduite suggérant l'implication du système de surveillance des ARNm porteurs de codon stop prématuré (Nonsense Mediated mRNA Decay). Les analyses en immunohistochimie ont montré une localisation cellulaire normale mais dans des quantités réduites. Les tests électrophysiologiques sur cellules CHO après transfection d'un plasmide contenant une version sauvage ou porteuse de la mutation de KCNQ3 ont montré une baisse de la fonction des canaux KCNQ3 lorsqu'ils étaient transfectés seuls mais surtout lorsqu'ils se retrouvaient en situation physiologique en hétérodimères avec KCNQ2 (co-transfection KCNQ3 muté et KCNQ2 sauvage).

Discussion

L'ensemble de ces éléments permet de confirmer l'implication de la mutation homozygote de KCNQ3 dans le phénotype de la patiente, ce qui suggère que les variants tronquants bialléliques de KCNQ3 sont responsables d'un phénotype différent des faux-sens hétérozygotes associés à une altération de la fonction du canal ³⁵⁶. Ceci est d'autant plus vrai qu'au cours de ce travail, une autre famille originaire du Liban a été rapportée dans laquelle 3 membres d'une fratrie issue de parents consanguins présentaient une déficience intellectuelle et une épilepsie active de début néonatal, tous étant porteurs d'un autre variant tronquant du gène à l'état homozygote ; cette famille faisait partie d'une cohorte de patients atteints d'encéphalopathie épileptique précoce sans anomalie corticale à l'IRM cérébrale qui avait été explorée par séquencage d'un panel de 71 gènes ³⁷⁵. Il faut noter qu'aucun des apparentés porteurs hétérozygotes du variant tronquant au sein des 2 familles ne présente des antécédents d'épilepsie. Néanmoins, un oncle maternel de la patiente suivie à Dijon avait été suivi pendant les 6 premiers mois de vie pour épilepsie précoce de résolution rapide et avait eu par la suite des troubles d'apprentissage incluant des difficultés dans l'acquisition de la lecture et l'écriture ; malheureusement, il n'a pas été possible de connaître son statut génétique vis-à-vis du variant familial mais comme ses propres parents ne sont pas consanguins, il est plus probable qu'il existe une autre cause génétique et que la perte de fonction du gène KCNQ3 ne s'exprime phénotypiquement qu'en cas d'atteinte biallélique. Une autre famille a été rapportée dans laquelle une patiente de 7 ans présentait un retard développemental sévère avec tétraparésie spastique, strabisme, atteinte visuelle neurosensorielle et épilepsie de début néonatal qui s'est améliorée avec l'âge mais encore pharmacodépendante sous vigabatrin à 7 ans ; 2 variants faux-sens hétérozygotes composites ont été identifiés dans KCNQ3 et les tests fonctionnels ont montré une perte de fonction importante en cas de co-expression des 2 protéines mutées avec KCNQ2 alors que l'expression d'une seule des 2 protéines KCNQ3 mutées avec KCNQ2 n'avait pas d'effet majeur ; aucun autre antécédent familial n'était rapporté notamment neurologique et épileptique, et les porteurs hétérozygotes de l'un ou l'autre des variants avaient des tests cognitifs normaux ce qui est en faveur d'un mécanisme similaire, à savoir une perte de fonction biallélique ³⁷⁶. *A contrario*, les patients porteurs de variants hétérozygotes perte de fonction (délétion, frameshift, stop, épissage) du gène *KCNQ2* représentent 36% (39/108) des variants associés à l'épilepsie néonatale bénigne de type 1 et il n'a jamais été identifié de variants pathogènes perte de fonction à l'état biallélique pour le gène *KCNQ2* alors que les mutations sont 4 à 10 fois plus fréquentes, suggérant un seuil critique ^{357,377}. Le fait que chez l'homme et les rongeurs, le ratio KCNQ3/KCNQ2 est faible à la naissance et augmente au cours du développement post-natal est une explication plausible ^{378,379}. En particulier, la délétion de Kcnq2 mais pas Kcnq3 des neurones pyramidaux corticaux chez la souris suffit au développement d'une activité électrique cérébrale anormale et entraîne son décès dans les 3 premières semaines de vie ³⁸⁰.

Au cours de notre étude, on peut noter qu'un 3^{ème} phénotype a été associé à des variants fauxsens hétérozygotes *de novo* et parfois récurrents de *KCNQ3* responsables d'un effet gain de fonction en augmentant la sensibilité du canal ce qui entraîne son ouverture pour des stimulations plus faibles et une réponse plus intense aux stimulations: le phénotype associé est un trouble neurodéveloppemental incluant troubles autistiques et retard global des acquisitions marqué (langage absent ou rudimentaire) sans crises convulsives initialement mais avec l'apparition au cours de l'enfance d'épilepsie à pointesondes continues du sommeil ³⁸¹.

L'ensemble de ces éléments indique la nécessité d'une régulation fine de la fonction de ces canaux pour un fonctionnement neuronal normal. Ceci ouvre des perspectives thérapeutiques puisque certains agents antiépileptiques comme la retigabine est plus efficace sur KCNQ3 que KCNQ2, alors que le zinc pyrithome (ZnPy) est plus efficace sur KCNQ2 et sans effet sur KCNQ3³⁸².

Ce travail montre l'intérêt de bien analyser les variants situés dans des gènes connus en pathologie humaine dont des mutations ayant des conséquences attendues différentes peuvent être responsables d'un autre phénotype ; il ne faut donc pas exclure des variants systématiquement sous prétexte que le phénotype n'est pas concordant.

Exploitation de ce travail :

Cette collaboration a permis de rédiger un article scientifique ainsi qu'une communication affichée dans un symposium dédié aux canaux KCNQ :

Publication d'un article scientifique :

Lauritano A*, Moutton S*, Longobardi E*, Tran Mau-Them F, Laudati G, Nappi P, Soldovieri MV, Ambrosino P, Cataldi M, Jouan T, Lehalle D, Maurey H, Philippe C, Miceli F, Vitobello A, Taglialatela M. A novel homozygous KCNQ3 loss-of-function variant causes non-syndromic intellectual disability and neonatal-onset pharmacodependent epilepsy. Epilepsia Open. 2019 Aug 11;4(3):464-475. doi: 10.1002/epi4.12353. eCollection 2019 Sep. PMID:31440727. ³⁸³

* ces auteurs ont contribué au travail de façon égale.

- Poster à l'International Kv7 Channels Symposium, Naples, 12-14/09/2019 :

Elena Longobardi, Anna Lauritano, Sebastien Moutton, Frederic Tran-Mau-Them, Giusy Laudati, Piera Nappi, Maria Virginia Soldovieri, Paolo Ambrosino, Mauro Cataldi, Thibaud Jouan, Daphne lehalle, Helene Maurey, Christophe Philippe, Rita Cicatiello, Lucio Nitsch, Francesco Miceli, Antonio Vitobello, Maurizio Taglialatela.

In-vitro and ex-vivo characterization of a novel homozygous KCNQ3 p.(Phe534Ilefs) variant in a patient diagnosed with a non-syndromic intellectual disability and neonatal-onset pharmacodependent epilepsy.

Received: 9 April 2019 DOI: 10.1002/epi4.12353

FULL-LENGTH ORIGINAL RESEARCH

Open Acces

A novel homozygous KCNQ3 loss-of-function variant causes non-syndromic intellectual disability and neonatal-onset pharmacodependent epilepsy

Anna Lauritano¹ | Sebastien Moutton^{2,3} | Elena Longobardi¹ | Frédéric Tran Mau-Them^{3,4} | Giusy Laudati¹ | Piera Nappi¹ | Maria Virginia Soldovieri⁵ | Paolo Ambrosino⁶ | Mauro Cataldi¹ | Thibaud Jouan^{3,4} | Daphné Lehalle^{2,3} | Hélène Maurey⁷ | Christophe Philippe^{3,4} | Francesco Miceli¹ | Antonio Vitobello^{3,4} | Maurizio Taglialatela¹

¹Division of Pharmacology, Department of Neuroscience, University of Naples "Federico II", Naples, Italy

²Reference Center for Developmental Anomalies, Department of Medical Genetics, Dijon University Hospital, Dijon, France

³INSERM U1231, LNC UMR1231 GAD, Burgundy University, Dijon, France

⁴Laboratoire de Génétique, Innovation en Diagnostic Génomique des Maladies Rares UF6254, Plateau Technique de Biologie, CHU Dijon, Dijon, France

⁵Department of Medicine and Health Science "V. Tiberio", University of Molise, Campobasso, Italy

⁶Division of Pharmacology, Department of Science and Technology, University of Sannio, Benevento, Italy

⁷Service de Neurologie Pédiatrique, APHP, Hôpital Universitaire Bicêtre, Le Kremlin-Bicêtre, France

Correspondence

Maurizio Taglialatela, Department of Neuroscience, University of Naples "Federico II", Via Pansini 5, 80131, Naples, Italy. Email: mtaglial@unina.it.

Funding information

Ministero dell'Istruzione, dell'Università e della Ricerca, Grant/Award Number: PRIN 2017ALCR7C to MT and MVS, PRIN 2017YH3SXK to FM and RBSI1444EM to FM; Fondazione Telethon, Grant/Award Number: GGP15113 to MT; Ministero della Salute, Grant/Award Number: GR-2016-02363337 to MVS; Università degli Studi di Napoli Federico II, Grant/Award Number: 6-CSP-UNINA-120 to FM

Abstract

Objective: Heterozygous variants in *KCNQ2* or, more rarely, *KCNQ3* genes are responsible for early-onset developmental/epileptic disorders characterized by heterogeneous clinical presentation and course, genetic transmission, and prognosis. While familial forms mostly include benign epilepsies with seizures starting in the neonatal or early-infantile period, de novo variants in *KCNQ2* or *KCNQ3* have been described in sporadic cases of early-onset encephalopathy (EOEE) with pharmacoresistant seizures, various age-related pathological EEG patterns, and moderate/severe developmental impairment. All pathogenic variants in *KCNQ2* or *KCNQ3* occur in heterozygosity. The aim of this work was to report the clinical, molecular, and functional properties of a new *KCNQ3* variant found in homozygous configuration in a 9-year-old girl with pharmacodependent neonatal-onset epilepsy and non-syndromic intellectual disability. **Methods:** Exome sequencing was used for genetic investigation. KCNQ3 transcript and subunit expression in fibroblasts was analyzed with quantitative real-time PCR and Western blotting or immunofluorescence, respectively. Whole-cell patch-clamp electrophysiology was used for functional characterization of mutant subunits.

Lauritano, Moutton, and Longobardi contributed equally.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

^{© 2019} University of Naples Federico II. Epilepsia Open published by Wiley Periodicals Inc. on behalf of International League Against Epilepsy.

Results: A novel single-base duplication in exon 12 of *KCNQ3* (NM_004519.3:c.1599dup) was found in homozygous configuration in the proband born to consanguineous healthy parents; this frameshift variant introduced a premature termination codon (PTC), thus deleting a large part of the C-terminal region. Mutant KCNQ3 transcript and protein abundance was markedly reduced in primary fibroblasts from the proband, consistent with nonsense-mediated mRNA decay. The variant fully abolished the ability of KCNQ3 subunits to assemble into functional homomeric or heteromeric channels with KCNQ2 subunits.

Significance: The present results indicate that a homozygous *KCNQ3* loss-of-function variant is responsible for a severe phenotype characterized by neonatal-onset pharmacodependent seizures, with developmental delay and intellectual disability. They also reveal difference in genetic and pathogenetic mechanisms between *KCNQ2*- and *KCNQ3*-related epilepsies, a crucial observation for patients affected with EOEE and/or developmental disabilities.

KEYWORDS

early-onset epileptic encephalopathy, homozygous loss-of-function variant, intellectual disability, KCNQ3, next-generation sequencing, nonsense-mediated mRNA decay

1 | **INTRODUCTION**

Next-generation sequencing (NGS) technologies have revolutionized diagnostic procedures in developmental disorders, intellectual disability (ID), and pediatric epilepsies, allowing early identification of the molecular defects in an ever-growing number of human disease-causing genes.¹ Early genetic diagnosis is critical for prognostic assessment, genetic counseling, and, in some cases, personalized treatment attempts.²

Pathogenic variants in KCNQ2 and KCNQ3 genes coding for Kv7.2 and Kv7.3 neuronal voltage-gated potassium (K⁺) channel subunits cause early-onset epilepsies with wide phenotypic heterogeneity.^{3–5} At the benign end of this spectrum is benign familial neonatal seizures (BFNS), an autosomal dominant self-limiting neonatal epilepsy with seizures beginning in otherwise healthy infants in the first days of life and spontaneously disappearing in the following months, with mostly normal neurocognitive development and unremarkable neuroimaging.^{3,6,7} On the other hand, sporadic cases of early-onset epileptic encephalopathy (EOEE) with cognitive disability, various age-related pathological EEG patterns such as suppression-burst/multifocal epileptic activity/hypsarrhythmia,^{8,9} and distinct neuroradiological features have been more recently described in association to KCNO2 variants.^{10,11}

When compared to *KCNQ2*, pathogenic variants in *KCNQ3* have been more rarely described, mostly in families affected with familial forms of benign epilepsies with

Key Points

- Heterozygous variants in *KCNQ2* or, more rarely, *KCNQ3* genes are responsible for early-onset developmental/epileptic disorders
- We describe a patient with severe epilepsy and intellectual disability carrying a homozygous frameshift loss-of-function variant in KCNQ3
- Familial members who are heterozygous carriers of the *KCNQ3* variant are unaffected
- The present results highlight difference in genetic and pathogenetic mechanisms between *KCNQ2* and *KCNQ3*-related epilepsies

seizures starting in the neonatal $(BFNS)^{6,12,13}$ or early-infantile (benign familial infantile seizures, BFIS) period.^{14,15} Fewer than twenty families with BFNS and three families with BFIS with a heterozygous *KCNQ3* pathogenic variant have been reported to date.¹⁶ In addition, de novo variants in *KCNQ3* have been described in few children with EOEE,^{17–20} ID apparently without epilepsy,^{21,22} and cortical visual impairment.²³

Notably, all pathogenic variants in *KCNQ2* and *KCNQ3*, except one,²⁴ occur in heterozygosity. In the present manuscript, we report the clinical, molecular, and functional properties of a new *KCNQ3* variant found in homozygous configuration

in a 9-year-old girl with pharmacodependent neonatal-onset epilepsy and non-syndromic ID; the variant, a single-base duplication in exon 12 of KCNQ3 (chr8:g.133150233dup in GRCh37; NM 004519.3:c.1599dup), is predicted to result in a frameshift p.(Phe534Ilefs*15) which could lead to either mRNA degradation by the nonsense-mediated mRNA decay (NMD) machinery²⁵ and/or a truncation of a large part of the C-terminus. Ex vivo results showed that KCNQ3 transcript levels were markedly reduced in primary fibroblasts from the proband when compared to those from the unaffected noncarrier brother, and in vitro studies revealed that the variant fully abolished the ability of KCNQ3 subunits to assemble into functional homomeric or heteromeric channels with KCNQ2 subunits. The present results provide the first clinical, genetic, and functional evidence for a severe phenotype associated with a homozygous loss-of-function (LoF) variant in KCNQ3 and highlight previously unrecognized difference in genetic and pathogenetic mechanisms between KCNQ2and KCNQ3-related epilepsies.

2 | MATERIALS AND METHODS

2.1 | Informed patient consent

Written informed consent was obtained from all study participants or from parents/authorized legal guardians. The study was performed within the framework of the Genetique des Anomalies du Développement (GAD) collection performed at Equipe Inserm U1231 of the Université de Bourgogne in Dijon (FR), approved by institutional review board (no DC2011-1332).

2.2 | Exome sequencing

Singleton exome sequencing was performed using an Agilent CRE capture kit (Agilent Technologies) and a HiSeq 4000 sequencer (Illumina); exome data analysis and variant filtering were performed as previously described²⁶ using the following annotation databases and software versions: dbsnp 149, clinvar 2016_11_03, cosmic COSMICv79, omim 2016_11_29, refseq_annotation 2016_11_27, FASTQC 2015_12_15, TRIMMOMATIC 0.35, BWA_0.7.12, PICARD_2.0.1, GATK_3.5, samtools_1.2, IGVTools_2.3.67, reference genome GRCh37/hg19, refseq 2015_07_30. Reads alignment resulted in coverage of at least 10× for 96.5% of the bases on target and an average sequencing depth of 116.01×.

2.3 | Cell culture

Chinese hamster ovary (CHO) cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% fetal bovine serum (FBS), 2 mmol/L L-glutamine, penicillin (50 U/mL), and streptomycin (50 μ g/mL) in a humidified

atmosphere at 37°C with 5% CO₂. Primary fibroblasts were obtained from punch skin biopsies. Tissue was cut into small pieces (1 \times 3 mm), and cells were cultivated in a fibroblast growth medium consisting of AmnioMax (Gibco, Thermo Fisher) supplemented with 20% FBS. For long-term culture, fibroblasts were maintained in DMEM supplemented with 10% FBS.

2.4 | Mutagenesis and heterologous expression of KCNQ2 or KCNQ3 cDNAs

The variant of interest was engineered in KCNQ3 human cDNA cloned into pcDNA3.1 (variant 1; NM_004519.3; 872 aa) by QuikChange site-directed mutagenesis (Agilent Technologies). Channel subunits were expressed in CHO cells by transient transfection using Lipofectamine 2000 (Invitrogen);^{27,28} when two or more cDNAs were cotransfected, their molar ratio was modified such that total cDNA in the transfection mixture was kept constant at 3 μ g. Enhanced green fluorescent protein (1 μ g; Clontech) was used as transfection marker.

2.5 | Western blot experiments

KCNQ3 subunits in total protein lysates from CHO cells obtained 24 hours post-transfection were analyzed by Western blotting on 8% SDS-PAGE using two primary rabbit anti-KCNQ3 polyclonal antibodies: (a) the first directed against a C-terminal epitope (rat aa 668-686; accession number O88944; C-KCNQ3) (clone APC-051, dilution 1:1000; Alomone Labs) and (b) the second raised against an N-terminal epitope (rat aa 1-71; N-KCNQ3) (PA1-930; dilution 1:1000; Thermo Scientific). Both antibodies also recognized human KCNQ3 subunits. Following exposure to primary antibodies, filters were incubated with HRP-conjugated antirabbit secondary antibodies (NA934V; dilution 1:5,000; GE Healthcare) and reactive bands visualized by chemiluminescence. Data acquisition and analysis were performed with the Gel Doc Imaging System (Bio-Rad) and ImageLab software (version 4.1; Bio-Rad), respectively.

2.6 | RNA isolation, reverse transcription, and quantitative PCR

Isolation and purification of RNA was performed using the TriReagent (Sigma). 1 µg of total RNA was retrotranscribed with the High Capacity cDNA RT Kit (Applied Biosystem, Thermo Fisher Scientific). For quantitative PCR, cDNA was amplified with the TaqMan Gene Expression assay in a 7500 Fast Real-Time PCR System thermocycler (Applied Biosystems, Thermo Fisher Scientific). Commercially available probes were used to amplify *KCNQ1*, *KCNQ2*, *KCNQ3*, *KCNQ4*, and *KCNQ5* mRNAs (Applied Biosystem TaqMan

gene expression, codes *KCNQ1*: hs00923522_m1; *KCNQ2*: hs01548339_m1; *KCNQ3*: hs01120412_m1; *KCNQ4*: hs00542548_m1; *KCNQ5*: hs01068536_m1). The comparative $\Delta\Delta$ CT method was used to quantify transcript abundance using the ubiquitin-conjugating enzyme (UBC; hs05002522_g1) gene as control.²⁹ Three separate experiments, each in triplicate, were performed for each probe.

2.7 | Immunofluorescence

Cells were fixed with 4% paraformaldehyde in PBS for 10 minutes at room temperature (RT). After permeabilization with 0.1% Tween-20 for 5 minutes and blocking with 0.5% BSA for 1 hour at RT, cells were incubated overnight at 4°C with the N-KCNQ3 antibody (1:300), followed by a 1 hour incubation with donkey anti-rabbit Cy3-conjugated secondary antibody (Applied Biosystems, Thermo Fisher Scientific) at RT. Nuclei were visualized using Hoechst 33258 (1:5000) in PBS. Coverslips were mounted in Fluoromount G (eBioscience, Hatfield, Hertfordshire, UK); images were acquired with a Zeiss inverted LSM 700 confocal laser scanning microscope (Carl Zeiss) and analyzed with ImageJ (NIH). Slides in which the primary antibody was omitted were used as controls in all experiments.

2.8 | Whole-cell electrophysiology

Macroscopic current recordings from transiently transfected CHO cells, as well as data processing and analysis, were performed as described.^{27,28} Currents from CHO cells were recorded at room temperature with the whole-cell configuration of the patch-clamp technique, using glass micropipettes of 3-5 M Ω resistance. The extracellular solution contained (in mmol/L) the following: 138 NaCl, 5.4 KCl, 2 CaCl₂, 1 MgCl₂, 10 glucose,

and 10 HEPES, pH 7.4 with NaOH. The pipette (intracellular) solution contained (in mmol/L) the following: 140 KCl, 2 MgCl₂, 10 EGTA, 10 HEPES, and 5 Mg-ATP, pH 7.3-7.4 with KOH. The pCLAMP software (version 10.0.2) was used for data acquisition and analysis. Linear cell capacitance (C) and series-resistance (RS) calculation were performed as described previously.³⁰ In the experiments with tetraethylammonium, currents were activated by 3-second voltage ramps from -80 mV to +40 mV at 0.08 Hz frequency. Fast solution exchanges (<1 second) were achieved by means of a cFlow 8 flow controller attached to a cF-8VS eight-valve switching apparatus (Cell MicroControls).

2.9 | Statistics

Data are expressed as the mean \pm SEM. Data reported in Table 1 are average values of at least 9 individual measurements, recorded in at least 3 separate experimental sessions (transfections). Statistically significant differences between the data were evaluated with Student's *t* test (*P* < .05).

3 | RESULTS

3.1 | Clinical features

The proband (individual II-3; Figure 1A) is a French 9-yearold female originary from Morocco born to consanguineous healthy parents after an uneventful pregnancy and delivery, apart from a transient prematurity risk at 35 weeks. She was delivered at term (38 weeks 4 days), and birth parameters were normal: 3040 g weight, 46.5 cm length, and 34 cm occipito-frontal head circumference (OCF). At the age of 2 days, she presented with both focal (affecting either the left or right hemi-body) and generalized convulsions associated with

TABLE 1 B	Biophysical and pharmacological	properties of currents recorded in CHO cel	lls transfected with the indicated plasmid combinations
-----------	---------------------------------	--	---

	cDNA trans- fected (µg)	n	$V_{1/2}$ (mV)	K (mV/efold)	Current density (pA/pF at 0 mV)	Blockade by 3 mmol/L TEA (%)
Nontransfected	-	10	-	-	0.5 ± 0.1	-
KCNQ3	3	9	$-41.9 \pm 1.5^{*,**}$	$8.2 \pm 0.9^{*,**}$	$11.5 \pm 4.8^{****}$	$8.0 \pm 2.1^{*,**}$
KCNQ3 p.(Phe534Ilefs [*] 15)	3	9	-	-	0.3 ± 0.1	-
KCNQ2 + pcDNA3	1.5 + 1.5	11	-21.7 ± 1.9	13.2 ± 0.8	$21.7 \pm 5.1*$	-
KCNQ2	3	13	-23.0 ± 1.5	12.0 ± 0.5	42.2 ± 9.7	$94.0 \pm 1.0^{**}$
KCNQ2 + KCNQ3	1.5 + 1.5	23	-35.1 ± 1.6	13.0 ± 0.7	117.6 ± 15.1	56.1 ± 6.6
KCNQ2 + KCNQ3 p.(Phe534Ilefs [*] 15)	1.5 + 1.5	20	$-23.9 \pm 1.9^{**}$	15.4 ± 1.5	17.5 ± 2.5**	$90.0 \pm 1.5^{**}$
KCNQ2 +KCNQ3 + pcDNA3	1.5 + 0.75 + 0.75	13	$-27.5 \pm 1.0^{**}$	13.0 ± 0.7	$33.6 \pm 6.9^{**}$	$62.0 \pm 4.3^{*}$
KCNQ2 + KCNQ3 + KCNQ3 p.(Phe534Ilefs [*] 15)	1.5 + 0.75 + 0.75	24	$-29.5 \pm 1.8^{**}$	12.8 ± 0.7	39.6 ± 6.1**	$62.2 \pm 3.2^*$

*P < 0.05 vs KCNQ2 (3 μg).

**P < 0.05 vs KCNQ2 + KCNQ3 (1.5 + 1.5 μg).



FIGURE 1 Pedigree of the investigated family and topological models of the mutant KCNQ3 subunit. A, Pedigree of the family investigated. '+' indicates the wild-type *KCNQ3* allele; '-' indicates the mutant *KCNQ3* p.(Phe534Ilefs*15) allele. The arrow indicates the proband. B, Schematic topology of a KCNQ3 subunit: S_1 - S_6 refer to the six transmembrane segments, while boxes labeled from A to D depict the four α -helical regions in the intracellular C-terminus. The p.(Phe534Ilefs*15) variant located in the helix B is indicated by the arrow. The as sequence deleted in the mutant KCNQ3 protein is indicated by a dashed line. The red line indicates the amino acid sequence altered by the frameshift variant. C, Partial alignments of the primary sequences of KCNQ3 and KCNQ3 p.(Phe534Ilefs*15, indicated as KCNQ3^{MUT}) subunits. The B and C helices are highlighted with darker blue boxes.

hypotonia, cyanosis, and clonic movements of the four limbs. Biochemical and metabolic screening was noncontributive. Initial neurological examination and tonus were normal. The first electroencephalograms (EEGs), performed in the following days, revealed electrical seizures characterized by central and temporal slow waves prevailing on the right side not always associated with clinical manifestations. Biotin, pyridoxine, and folinic acid were ineffective. At the age of 2 months, convulsions were controlled with phenytoin and vigabatrin; ocular contact was poor; and the tonus was insufficient. At 7 months of age, sodium valproate monotherapy was effective in controlling seizures; a marked strabismus was noted requiring specialized management including botulinum toxin injection. Interruption of sodium valproate treatment at the age of 3-4 years resulted in seizure recurrence during late night, including febrile episodes, with left hemispheric spikes and waves recorded on the EEG; valproate therapy was therefore reintroduced. Since then, she exhibited rare tonic-clonic seizures, and all subsequent EEG recordings were normal, the last one at age 6 years. She is currently treated with sodium valproate with good response, and her epilepsy shows the characteristics of a pharmacodependent epilepsy.³¹ Brain MRI performed at day 15 revealed a suspected mild cortical dysplasia of the right frontal and temporal lobes, but a further MRI at age 6 years and 5 months was normal, as well as a brain CT scan. The proband acquired head control

at age 6-7 months, could sit unsupported at age 12 months, and walked independently around age 30-36 months. At age 3 years, she could speak 2 words and required speech therapy, including Makaton technique while verbal language was insufficient; she attended specialized educational institution at age 5 years and started to produce short rudimentary word associations since age 6 years, with about 10 words in her vocabulary; sentences were still incompletely produced at age 8 years. At age 6.5 years, her psychomotor development was estimated around 22 months. She did not exhibit behavioral disturbances. At last examination (age 9 years), no abnormal morphological features were noted and neurological examination was normal apart from divergent strabismus increased in superior gaze. She has a moderate intellectual disability with poor vocabulary and little autonomy in daily life; she still attends a medical institute with little schooling abilities. Growth charts showed regular evolution of length, weight, and OFC between median and -1 SD. Familial history revealed that a maternal uncle who had mild cognitive disabilities with some degree of learning (reading and writing) difficulties also suffered from transient neonatal seizures, requiring a specialized pediatric follow-up in the first 6 months of life. However, he has a milder phenotype than his niece as he could achieve a relatively good autonomy in daily life. All other members of the family had normal psychomotor and cognitive development with no history of seizures.

3.2 | Genetic data

Array-CGH, KCNQ2-targeted sequencing, and gene panel sequencing in the proband (individual II-3) were normal; exome sequencing was therefore performed. The analysis on OMIM morbid genes responsible for mental retardation associated or not with epileptic encephalopathy highlighted the presence of two rare heterozygous variants, one in HERC2, associated with an autosomal recessive form of mental retardation (MIM 615516), and one in FRAS1, associated with autosomal recessive Fraser syndrome 1 (MIM 219000). Both variants were not retained as causative because the mode of inheritance was not compatible with the genotype of our patient. A variant of DYNC1H1, gene associated with Mental Retardation autosomal dominant 13 (MIM 614563), was not retained as causal as this variant has been reported 77 times in the healthy population (gnomAD 2.1.1). No further variants were retained using a sporadic mode of inheritance. The homozygosity analysis obtained through HomozygosityMapper³² using default parameters revealed relatively small regions of homozygosity in chr1, chr9, chr10, chr14, chr16, and chr22 compatible with the consanguinity between the parents (the 2 grandmothers of the proband are sisters, see expanded pedigree shown as Figure S1). Within these regions, we identified a rare variant of the DMBT1 gene at the homozygous state, not retained as causal because it was found at the homozygous state in 7 individuals in the healthy population. Outside the regions of homozygosity, the analysis on OMIM morbid genes at the recessive state retained an homozygous variant of the KCNQ3 gene absent from public databases as the only candidate compatible with the clinical presentation of the proband. This variant (ClinVar submission number SUB5837801) is a homozygous single-base duplication (chr8:g133150233dup in GRCh37, NM 004519.3:c.1599dup) in exon 12 (135/138 reads detected the variant) which results in a shift in the open reading frame p.(Phe534Ilefs*15) and the occurrence of a premature termination codon (PTC) at position 549, possibly leading to the synthesis of a truncated protein deleted of a large part of the C-terminal region (Figure 1B,C). The parents and one of the brothers (individuals I-1, I-2, and II-2, respectively) were heterozygous carriers for the KCNQ3 variant; the eldest brother (individual II-1) carried two copies of the wild-type allele (Figure 1A); and the maternal uncle was unavailable for genetic analysis.

3.3 | The *KCNQ3* mutant allele is expressed at lower levels when compared to healthy control

Premature termination codons often result in mRNA instability and precocious degradation by nonsense-mediated Epilepsia Open[®]

mRNA decay (NMD).²⁵ KCNQs expression was previously reported in human fibroblasts³³; therefore, the effect of the NM_004519.3:c.1599dup variant on *KCNQ3* transcript levels was evaluated in primary fibroblasts from the proband (II-3, carrying two copies of the mutant allele) and a control member of the family (II-1, carrying two copies of the wild-type allele), using qRT-PCR. In control cells, the Ct values for the different KCNQs are the following: KCNQ1: 36; KCNQ2: 28; KCNQ3: 28; KCNQ4: 32; and KCNQ5: 31; all these values are considerably higher than that of UBC, the housekeeping gene used as comparator (Ct value of 25), consistent with the described low expression levels for these transcripts.³³

As shown in Figure 2A, in cells from the proband, KCNQ3 transcript levels were reduced to ~22% when compared to those in control cells; a quantitatively similar decrease in transcript levels was also observed for KCNQ4, whereas those for KCNQ2 were markedly increased. No significant difference was instead observed when comparing KCNQ1 and KCNQ5 transcript abundance between proband and control primary fibroblasts.

To investigate whether such decrease in KCNQ3 mRNA levels also led to changes in KCNQ3 protein expression, immunofluorescence experiments using N-KCNO3 primary antibodies were performed in primary fibroblasts from the proband and the noncarrier, unaffected brother. These antibodies were validated in Western blots on lysates from CHO cells transiently transfected with wild-type KCNO3 or KCNO3 p.(Phe534Ilefs*15)-encoding plasmids (KCNQ3^{MUT}; Figure 2B, left panel). In these experimental groups, N-KCNQ3 antibodies specifically recognized a ~100 kDa or a ~60 kDa protein band, respectively. These values are consistent with the molecular masses expected for wild-type or mutant KCNQ3 proteins, respectively. Western blot experiments also revealed that, in this heterologous cellular system, the mutant protein was expressed at lower levels when compared to the wild-type KCNQ3 protein (Figure 2B, left panel). As expected, C-KCNO3 antibodies targeting a C-terminal epitope located downstream the predicted premature termination site introduced by the variant, while detecting the 100kDa band corresponding to the wild-type KCNQ3 protein, failed to recognize any specific signal in cells transfected with the mutant plasmid (Figure 2B, right panel). Western blot experiments performed in primary fibroblasts from proband and control samples with both antibodies failed to detect specific signals corresponding to either wild-type or mutant KCNQ3 proteins, possibly because of the low abundance of the endogenous protein in these cells (data not shown). Immunochemistry experiments using N-KCNQ3 antibodies performed in primary fibroblasts from the proband (individual II-3) and the wild-type brother (individual II-1) (Figure 2C, right and left panels, respectively) revealed a



FIGURE 2 KCNQ transcript and protein expression profile in primary fibroblasts from the proband (individual II-3) and healthy brother (individual II-1). A, *KCNQ*1-5 qRT-PCR data from primary fibroblasts. Data are expressed as cycle threshold values for each *KCNQ* transcript normalized to that of *UBC*; after normalization, data from control fibroblasts were expressed as one (green bars), and data from proband fibroblasts were expressed relative to controls (red bars). Asterisks indicate values statistically different (P < 0.05) from respective controls. B, Western blot performed on protein lysates from transiently transfected CHO cells using N- and C-KCNQ3 antibodies. CHO cells were transfected with wild-type (KCNQ3) or mutant (KCNQ3^{MUT}) *KCNQ3* cDNA, and total lysates were analyzed with N-KCNQ3 (left panel) or C-KCNQ3 (right panel) antibodies. α -tubulin (α -TUB) served as a loading control. C, Confocal images of fibroblasts from the proband (individual II-1) and the healthy brother (individual II-1) stained with N-KCNQ3 primary antibodies (in red) and a nuclear marker (Hoechst, in blue)

KCNQ3-specific signal in both groups, although the intensity was lower in the former when compared to the latter. Notably, in fibroblasts from both individuals, the immunofluorescent signal was mainly cytosolic, showing a subcellular distribution consistent with an endoplasmic reticulum localization.

3.4 | Functional and pharmacological characterization of homomeric and heteromeric channels carrying KCNQ3 p.(Phe534Ilefs*15) mutant subunits

Previously shown data obtained in primary fibroblasts from the proband suggest that the p.(Phe534Ilefs*15) truncating variant led to a significant decrease in KCNQ3 transcript and protein levels, a result consistent with NMD.²⁵ However, transcript expression pattern in fibroblasts might not parallel that in neurons; moreover, a significant, although reduced in abundance, fraction of KCNQ3 protein was still detected in fibroblasts. Therefore, electrophysiological recordings in transiently transfected CHO cells were carried out to evaluate the effects prompted by the p.(Phe534Ilefs*15) variant on KCNQ3 channel function.

Heterologous expression of wild-type KCNQ3 subunits led to the appearance of voltage-dependent K^+ -selective

currents characterized by a slow time course of activation and deactivation and a threshold for current activation around $-50 \text{ mV}.^{5.7}$ By contrast, no currents could be recorded in cells transfected with the KCNQ3 p.(Phe534Ilefs*15) plasmid, consistent with the variant causing a complete loss-of-function (LoF) effect (Figure 3A and Table 1).

In adult neuronal cells, KCNQ3 subunits assemble into heteromeric channels with KCNQ2 to form IKM.⁵ To investigate the functional consequences of mutant KCNQ3 subunits when coexpressed with KCNO2 subunits, CHO cells were cotransfected with KCNQ2 and KCNQ3 cDNAs at a 1:1 ratio (to mimic the genetic balance of the noncarrier Individual II-1) or with KCNO2 and KCNO3 p.(Phe-534Ilefs*15) at 1:1 ratio (to mimic the genetic balance of the proband, Individual II-3). Coexpression of wild-type KCNO3 subunits and KCNO2 subunits markedly increased current size when compared to KCNQ2- or KCNQ3-only expressing cells (Table 1); in addition, KCNO2/3 heteromeric channels display a significant leftward shift in the activation $V_{1/2}$ and a decrease in current sensitivity to blockade by TEA (Table 1).^{5,28,34} By contrast, coexpression of KCNQ3 p.(Phe534Ilefs*15) with KCNQ2 subunits failed to enhance current density and to modify the $V_{1/2}$ (Figure 3B and Table 1). Currents recorded in KCNQ2 + KCNQ3



FIGURE 3 Functional characterization of homomeric or heteromeric channels incorporating KCNQ3 p.(Phe534Ilefs*15) subunits. A, Representative current traces from nontransfected cells (NT; left panel) or from cells transfected with either wild-type KCNQ3- (KCNQ3; middle panel) or KCNQ3 p.(Phe534Ilefs*15)-encoding plasmids (KCNQ3^{MUT}; right panel) in response to the voltage protocol shown. Horizontal scale bar: 100 ms; vertical scale bar: 2 pA/pF. B, Representative current traces recorded in cells expressing the indicated subunits, in response to the same voltage protocol shown in A. Horizontal scale bar: 100 ms; vertical scale bar: 20 pA/pF. C, Representative current traces from cells expressing the indicated subunits in response to the indicated voltage ramp protocol before TEA exposure (CTL, control), during TEA exposure (TEA, 3 mmol/L), and upon drug washout (Wash). Horizontal scale bar: 200 ms; vertical scale bar: 10 pA/pF

p.(Phe534Ilefs*15)-transfected cells displayed a sensitivity to blockade by 3 mmol/L TEA higher than that of KCNQ2/ KCNQ3-transfected cells and identical to that of KCNQ2only transfected cells (Figure 3C, Table 1). These results suggest that mutant KCNQ3 subunits fail to form functional heteromeric channels with KCNQ2 subunits.

To replicate in vitro genetic combination occurring in the family members who are heterozygous carriers of the mutant allele (Individuals I-1, I-2, and II-2), functional studies were also carried out upon transfection of KCNQ2, KCNQ3, and KCNQ3 p.(Phe534Ilefs*15) plasmids in a 1:0.5:0.5 cDNA ratio; for these experiments, cells transfected with an identical (1:0.5:0.5) cDNA ratio of KCNQ2 + KCNQ3+pcDNA plasmids served as controls. The results obtained suggest that the presence of an halved dose of functional/wild-type KCNQ3 allele is sufficient to generate currents displaying the biophysical and pharmacological properties of KCNQ2/ KCNQ3 heteromers, although with a reduced density when compared to that recorded in cells transfected with a full dose of KCNQ3 (1:1 KCNQ2:KCNQ3 cDNA ratio). Notably, the pharmacological and functional properties of the currents recorded in cell transfected with KCNQ2, KCNQ3, and KCNQ3 p.(Phe534Ilefs*15) plasmids (1:0.5:0.5 cDNA ratio) were identical to KCNQ2 + KCNQ3+pcDNA-expressing cells (1:0.5:0.5 cDNA ratio) (Table 1).

4 | DISCUSSION

4.1 | Epilepsy and ID caused by a novel homozygous KCNQ3 frameshift variant: clinical and ex vivo results

We herein report the clinical, ex vivo, and in vitro results from a family carrying the novel frameshift p.(Phe534Ilefs*15) variant in KCNQ3 (NM 004519.3:c.1599dup). The proband is a 9-year-old girl diagnosed with neonatal-onset and pharmacosensitive seizures and non-syndromic ID; she was found to be a homozygous carrier for this variant. While our study was in progress, another family with three siblings affected with neonatal-onset seizures (reported as pharmacosensitive in one of them) and ID of variable severity due to a different homozygous frameshift variant in KCNQ3 (c.1220_1221delCT; p.(Ser407Phefs*27)) has been described in a large cohort of children suffering from epileptic encephalopathy.²⁴ In this family, where no functional analyses were performed, neonatal seizures or neurodevelopmental problems did not occur in consanguineous parents or extended family members who were heterozygous carriers for the KCNQ3 mutant allele. One of the affected sibling in Kothur et al.²⁴ exhibited severe developmental delay, while the two others presented with mild ID; since genetic analysis was only limited to a restricted epileptic encephalopathy gene panel, it is unknown whether additional genetic defects, potentially enabled by consanguinity, could account for the more severe phenotype in this particular sibling.

PTC-containing mRNAs often undergo NMD when the PTC is located upstream the ~50-55th nucleotide before the last exon-intron junction,³⁵ as it occurs with the presently described variant. NMD represents a quality control mechanism to avoid production of truncated proteins with potentially deleterious effects.^{25,36} KCNQ3 transcript levels in primary fibroblasts³³ from the proband were markedly decreased when compared to those from the unaffected noncarrier brother II-1, a result consistent with NMD; the reduction in KCNQ3 mutant transcript levels was accompanied by an increase in KCNQ2, a reduction of KCNQ4, and no change in KCNQ1 or KCNQ5 transcript levels. It remains to be determined whether these changes are due to compensatory effects or uncover a more complex coordination of gene expression. In this regard, whether the transcriptional repressor RE1 silencing transcription factor (REST), which has been shown to regulate KCNQ2 and KCNQ3³⁷ as well as KCNQ4³⁸ expression, participates in this coordination is currently unknown; moreover, it should be highlighted that conditional and selective ablation of Kcng2 or Kcng3 in cortical mouse tissue also modifies the expression of other members of the Kcng subfamily.³⁹ Notably, immunofluorescence experiments also revealed a reduced intensity of the KCNQ3 signal in primary fibroblasts from the proband; in these experiments, the KCNQ3 signal showed a diffuse cytoplasmic staining consistent with endoplasmic reticulum-Golgi localization with no remarkable difference in subcellular localization between fibroblasts from the proband or the control brother. A diffuse staining pattern on both cell surface and intracellular components was also detected for KCNQ3 in rat⁴⁰ and human⁴¹ hippocampal and cortical pyramidal neurons.

4.2 | Functional consequences of the KCNQ3 p.(Phe534Ilefs*15) variant

Our results indicate that transcript and protein levels encoded by the mutant *KCNQ3* allele are detectable, prompting investigation of the functional consequences of the p.(Phe534Ilefs*15) variant on KCNQ3 subunit function. The results obtained clearly suggest the KCNQ3 p.(Phe534Ilefs*15) protein is fully nonfunctional; indeed, homomeric expression of mutant subunits failed to generate detectable voltage-gated K⁺ currents. Moreover, mutant KCNQ3 subunits did not incorporate into functional heteromeric channels when coexpressed with KCNQ2 subunits, a result consistent with the variant-induced deletion of a significant portion of the long C-terminus where critical domains responsible for homomeric and heteromeric subunit assembly and plasmembrane trafficking have been

identified.^{4,42,43} Furthermore, the presence of an halved dose of wild-type KCNQ3 protein, an experimental condition mimicking in vitro the genetic combination of the family members who are heterozygous carriers of the mutant KCNQ3 allele, was sufficient to generate currents displaying the biophysical and pharmacological properties of KCNQ2/ Q3 heteromers, although with a reduced density when compared to that recorded in cells transfected with a full dose of KCNO3. Currents recorded in cells transfected with KCNO2, KCNQ3, and mutant KCNQ3 cDNAs were identical to those in cells transfected with the same cDNA amount of KCNQ2 and KCNQ3 cDNAs only, arguing in favor of the inability of the protein encoded by the mutant allele to heteromerize and interfere with the function of wild-type subunits. This functional result strongly suggests that haploinsufficiency is the main molecular mechanism for the severe disease in the affected proband; this is in sharp contrast to KCNQ2-related EOEE pathogenesis, where mutant subunits carrying heterozygous missense variants are functional and heteromerize with wild-type subunits, thereby poisoning channel function by a dominant-negative mechanism.^{27,44} Additional pathogenic mechanisms for KCNQ2 EOEE include changes in subcellular localization,⁴⁵ and/or in calmodulin-^{46,47} or phosphatidylinositol 4,5-bisphosphate⁴⁸-dependent regulation.

4.3 | Clinical spectrum and genetic mechanisms of KCNQ3-related diseases

Heterozygous pathogenic variants in *KCNQ3* have been associated with neonatal-onset epilepsies showing broad clinical heterogeneity and diverse genetic transmission mechanisms.¹⁶ These range from relatively benign familial phenotypes with seizures starting in the neonatal (BFNS)^{6,12,13} or early-infantile (BFIS)^{14,15} period, to sporadic cases with severe clinical presentations characterized by developmental disabilities with or without refractory seizures^{17,19,21,22,49} or by cortical visual impairment.²³

Notably, in both familial and sporadic cases, KCNO3 pathogenic variants are all single missense variants, either with autosomal dominant inheritance or arising de novo, respectively (34/34; www.rikee.org).¹¹ A notable exception is a recently described EOEE patient who carries two missense variants in compound heterozygosity, each inherited from an asymptomatic parent.²⁰ Interestingly, no heterozygous pathogenic frameshift KCNQ3 variant has ever been associated with a human phenotype, whereas haploinsufficiency due to heterozygous frameshift variants in KCNQ2 is a frequent cause of BFNS;^{16,50} indeed, frameshift/deletion variants account for 36% (39/108) of BFNS-causing variants in the KCNO2 gene^{11,13,51,52} (www.rikee.org). Notably, in both families where KCNQ3 homozygous frameshift variants were identified, that is, the presently described family and the one reported by Kothur et al.²⁴, heterozygous carriers of the *KCNQ3* frameshift variant are unaffected, with no seizures or psychomotor/cognitive impairment. Instead, no homozygous frameshift variant in *KCNQ2* has ever been described in humans, as minimal KCNQ2 residual activity is probably essential under penalty of potential lethality. Although several potential mechanisms may account for the more severe clinical consequences associated with *KCNQ2* variants when compared to *KCNQ3* ones, the fact that, in both rodents and human brains,^{53,54} the ratio of KCNQ3 to KCNQ2 expression is low at birth and increases during postnatal development provides a plausible explanation; notably, deletion of Kcnq2, but not of Kcnq3, from cortical pyramidal neurons in mice is sufficient for the development of aberrant EEG activity and leads to death by the third week of life.³⁹

Noteworthy, about 10 nonsense or frameshift variants leading to KCNQ3 truncation spread throughout the gene are reported in public databases such as ClinVar⁵⁵ or gnomAD.⁵⁶Among these variants, only one, found in an individual in the gnomAD's non-neuro-samples from individuals who were not ascertained for having a neurological condition in a neurological case/control study, was interpreted as pathogenic (although the phenotype is unknown), while the others were associated with an uncertain clinical significance.

5 | CONCLUSIONS

This is the first study exploring the functional consequences of a novel KCNQ3 homozygous LoF variant responsible for a severe phenotype characterized by neonatal-onset pharmacodependent seizures, with developmental delay and ID. Ex vivo and in vitro experiments revealed a decrease in transcript abundance proportional to variant expression levels and an impaired ability of mutant subunits to assemble into functional homomeric or heteromeric channels with KCNQ2. A lesser degree of channel dysfunction occurs when a single copy of the mutant allele is present, a result possibly contributing to the lack of neurodevelopmental phenotype in heterozygous carriers. The described LoF mechanism allows to hypothesize that, in close analogy to LoF variants found in KCNQ2-EOEE-affected patients,^{50,57} KCNQ activators such as retigabine may be useful precision medicines to counteract the channel dysfunction triggered by the herein described novel KCNQ3 variant as well as similar variants that may be identified in the future.

ACKNOWLEDGMENTS

We are deeply indebted to: Dr. Thomas J. Jentsch, Department of Physiology and Pathology of Ion Transport, Leibniz-Institut für Molekulare Pharmakologie (FMP), Berlin (Germany) for sharing KCNQ2 and KCNQ3 cDNAs. The present work was also supported by the following projects: the Telethon Foundation (grant number GGP15113) and the Italian Ministry for University and Research (MIUR) (PRIN 2017ALCR7C) to MT; MIUR (Scientific Independence of Researchers Project 2014 RBSI1444EM and PRIN 2017YH3SXK), PRIN (2017YH3SXK), and the University of Naples "Federico II" and Compagnia di San Paolo within the STAR Program "Sostegno Territoriale alle Attività di Ricerca" (project number 6-CSP-UNINA-120) to FM; and the Italian Ministry of Health (Project GR-2016-02363337) and MIUR (PRIN 2017ALCR7C) to MVS. We thank the family members who participated to the study.

CONFLICT OF INTEREST

The authors declare no competing financial interests. The Authors confirm that they have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

ORCID

Sebastien Moutton b https://orcid.org/0000-0002-8942-6731 Maurizio Taglialatela b https://orcid.org/0000-0002-8202-0560

REFERENCES

- Moller RS, Dahl HA, Helbig I. The contribution of next generation sequencing to epilepsy genetics. Expert Rev Mol Diagn. 2015;15:1531–8.
- Demarest ST, Brooks-Kayal A. From molecules to medicines: the dawn of targeted therapies for genetic epilepsies. Nat Rev Neurol. 2018;14:735–45.
- Jentsch TJ KCNQ potassium channels: physiology and role in disease. Nat Rev Neurosci. 2000;1:21–30.
- Soldovieri MV, Miceli F, Taglialatela M. Driving with no brakes: molecular pathophysiology of Kv7 potassium channels. Physiology (Bethesda). 2011;26:365–76.
- Wang HS, Pan Z, Shi W, Brown BS, Wymore RS, Cohen IS, et al. KCNQ2 and KCNQ3 potassium channel subunits: molecular correlates of the M-channel. Science. 1998;282:1890–3.
- Singh NA, Westenskow P, Charlier C, Pappas C, Leslie J, DillonJ, et al. KCNQ2 and KCNQ3 potassium channel genes in benign familial neonatal convulsions: expansion of the functional and mutation spectrum. Brain. 2003;126:2726–37.
- Soldovieri MV, Castaldo P, Iodice L, Miceli F, Barrese V, Bellini G, et al. Decreased subunit stability as a novel mechanism for potassium current impairment by a KCNQ2 C terminus mutation causing benign familial neonatal convulsions. J Biol Chem. 2006;281:418–28.
- Weckhuysen S, Ivanovic V, Hendrickx R, VanCoster R, Hjalgrim H, Møller RS, et al. Extending the KCNQ2 encephalopathy spectrum: clinical and neuroimaging findings in 17 patients. Neurology. 2013;81(19):1697–703.

⁴⁷⁴ Epilepsia Open[®]

- Millichap JJ, Miceli F, De Maria M, Keator C, Joshi N, Tran B, et al. Infantile spasms and encephalopathy without preceding neonatal seizures caused by KCNQ2 R198Q, a gain-of-function variant. Epilepsia. 2017;58:e10–e15.
- Weckhuysen S, Mandelstam S, Suls A, Audenaert D, Deconinck T, Claes L, et al. KCNQ2 encephalopathy: emerging phenotype of a neonatal epileptic encephalopathy. Ann Neurol. 2012;71:15–25.
- Joshi N, Taglialatela M, Weckhuysen S, Nesbitt G, Cooper EC. An informatics infrastructure for KCNQ2 encephalopahty research including a patient registry, database, curation platform, and website. Annual Meeting of the American Epilepsy Society. 2016;Abstract:3.329.
- Charlier C, Singh NA, Ryan SG, Lewis TB, Reus BE, Leach RJ, et al. A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. Nat Genet. 1998;18:53–5.
- Soldovieri MV, Boutry-Kryza N, Milh M, Doummar D, Heron B, Bourel E, et al. Novel KCNQ2 and KCNQ3 mutations in a large cohort of families with benign neonatal epilepsy: first evidence for an altered channel regulation by syntaxin-1A. Hum Mutat. 2014;35:356–67.
- Heron SE, Ong YS, Yendle SC, McMahon JM, Berkovic SF, Scheffer IE, et al. Mutations in PRRT2 are not a common cause of infantile epileptic encephalopathies. Epilepsia. 2013;54:e86–e89.
- Zara F, Specchio N, Striano P, Robbiano A, Gennaro E, Paravidino R, et al. Genetic testing in benign familial epilepsies of the first year of life: clinical and diagnostic significance. Epilepsia. 2013;54:425–36.
- 16. Miceli F, Soldovieri MV, Joshi N, Weckhuysen S, Cooper EC, Taglialatela M. KCNQ3-related disorders. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean L, Bird TD, Ledbetter N, Mefford HC, Smith R, Stephens K, editors. GeneReviews® [Internet], Seattle (WA): University of Washington, Seattle, 2017.
- Allen AS, Berkovic SF, Cossette P,Delanty N, Dlugos D, Eichler EE, et al. De novo mutations in epileptic encephalopathies. Nature. 2013;501:217–21.
- Miceli F, Striano P, Soldovieri MV, Fontana A, Nardello R, Robbiano A, et al. A novel KCNQ3 mutation in familial epilepsy with focal seizures and intellectual disability. Epilepsia. 2015;56:e15–e20.
- Grozeva D, Carss K, Spasic-Boskovic O, Tejada M-I, Gecz J, Shaw M, et al. Targeted next-generation sequencing analysis of 1,000 individuals with intellectual disability. Hum Mutat. 2015;36:1197–204.
- Ambrosino P, Freri E, Castellotti B, Soldovieri MV, Mosca I, Manocchio L, et al. Compound heterozygous variants in early onset encephalopathy reveal additive contribution of C-terminal residues to PIP2-dependent K+ channel gating. Mol Neurobiol. 2018;55:7009–24.
- Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, et al. Range of genetic mutations associated with severe nonsyndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012;380:1674–82.
- McRae JF, Clayton S, Fitzgerald TW, Kaplanis J, Prigmore E, Rajan D, et al. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542:433–8.
- Bosch D, Boonstra FN, de Leeuw N, Pfundt R, Nillesen WM, de Ligt J, et al. Novel genetic causes for cerebral visual impairment. Eur J Hum Genet. 2016;24:660–5.
- 24. Kothur K, Holman K, Farnsworth E, Ho G, Lorentzos M, Troedson C, et al. Diagnostic yield of targeted massively parallel

sequencing in children with epileptic encephalopathy. Seizure. 2018;59:132-40.

- Lykke-Andersen S, Jensen TH. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol. 2015;16:665–77.
- 26. Thevenon J, Duffourd Y, Masurel-Paulet A, Lefebvre M, Feillet F, El Chehadeh-Djebbar S, et al. Diagnostic odyssey in severe neurodevelopmental disorders: toward clinical whole-exome sequencing as a first-line diagnostic test. Clin Genet. 2016;89:700–7.
- Miceli F, Soldovieri MV, Ambrosino P, Barrese V, Migliore M, Cilio MR, et al. Genotype-phenotype correlations in neonatal epilepsies caused by mutations in the voltage sensor of K(v)7.2 potassium channel subunits. Proc Natl Acad Sci USA. 2013;110:4386–91.
- Miceli F, Soldovieri MV, Ambrosino P, De Maria M, Migliore M, Migliore R, et al. Early-onset epileptic encephalopathy caused by gain-of-function mutations in the voltage sensor of Kv7.2 and Kv7.3 potassium channel subunits. J Neurosci. 2015;35:3782–93.
- Lanzafame M, Botta E, Teson M, Fortugno P, Zambruno G, Stefanini M, et al. Reference genes for gene expression analysis in proliferating and differentiating human keratinocytes. Exp Dermatol. 2015;24:314–6.
- Soldovieri MV, Cilio MR, Miceli F, Bellini G, Miraglia del Giudice E, Castaldo P, et al. Atypical gating of M-type potassium channels conferred by mutations in uncharged residues in the S4 region of KCNQ2 causing benign familial neonatal convulsions. J Neurosci. 2007;27:4919–28.
- Rosati A, De Masi S, Guerrini R. Antiepileptic drug treatment in children with epilepsy. CNS Drugs. 2015;29:847–63.
- Seelow D, Schuelke M, Hildebrandt F, Nürnberg P. HomozygosityMapper–an interactive approach to homozygosity mapping. Nucleic Acids Res. 2009;37:W593–599.
- Poulet C, Künzel S, Büttner E, Lindner D, Westermann D, Ravens U. Altered physiological functions and ion currents in atrial fibroblasts from patients with chronic atrial fibrillation. Physiol Rep. 2016;4.pii:e12681.
- Hadley JK, Noda M, Selyanko AA, Wood IC, Abogadie FC, Brown DA. Differential tetraethylammonium sensitivity of KCNQ1-4 potassium channels. Br J Pharmacol. 2000;129:413–5.
- Hwang J, Sato H, Tang Y, Matsuda D, Maquat LE. UPF1 association with the cap-binding protein, CBP80, promotes nonsense mediated mRNA decay at two distinct steps. Mol Cell. 2010;39:396–409.
- Hug N, Longman D, Cáceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Res. 2016;44:1483–95.
- Mucha M, Ooi L, Linley JE, Mordaka P, Dalle C, Robertson B, et al. Transcriptional control of KCNQ channel genes and the regulation of neuronal excitability. J Neurosci. 2010;30:13235–45.
- Iannotti FA, Barrese V, Formisano L, Miceli F, Taglialatela M. Specification of skeletal muscle differentiation by repressor element-1 silencing transcription factor (REST)-regulated Kv7.4 potassium channels. Mol Biol Cell. 2013;24:274–84.
- Soh H, Pant R, LoTurco JJ, Tzingounis AV. Conditional deletions of epilepsy-associated KCNQ2 and KCNQ3 channels from cerebral cortex cause differential effects on neuronal excitability. J Neurosci. 2014;34:5311–21.
- Shah MM, Mistry M, Marsh SJ, Brown DA, Delmas P. Molecular correlates of the M-current in cultured rat hippocampal neurons. J Physiol. 2002;544:29–37.

- Cooper EC, Aldape KD, Abosch A, Barbaro NM, Berger MS, Peacock WS, et al. Colocalization and coassembly of two human brain M-type potassium channel subunits that are mutated in epilepsy. Proc Natl Acad Sci USA. 2000;97:4914–9.
- 42. Haitin Y, Attali B. The C-terminus of Kv7 channels: a multifunctional module. J. Physiol. 2008;586:1803–10.
- 43. Choveau FS, Zhang J, Bierbower SM, Sharma R, Shapiro MS. The role of the carboxyl terminus helix C-D linker in regulating KCNQ3 K+ current amplitudes by controlling channel trafficking. PLoS ONE. 2015;10:e0145367.
- Orhan G, Bock M, Schepers D, Ilina EI, Reichel SN, Löffler H, et al. Dominant-negative effects of KCNQ2 mutations are associated with epileptic encephalopathy. Ann Neurol. 2014;75:382–94.
- 45. Abidi A, Devaux JJ, Molinari F, Alcaraz G, Michon FX, Sutera-Sardo J, et al. A recurrent KCNQ2 pore mutation causing early onset epileptic encephalopathy has a moderate effect on M current but alters subcellular localization of Kv7 channels. Neurobiol Dis. 2015;80:80–92.
- Ambrosino P, Alaimo A, Bartollino S, Manocchio L, De Maria M, Mosca I, et al. Epilepsy-causing mutations in Kv7.2 C-terminus affect binding and functional modulation by calmodulin. Biochim Biophys Acta. 2015;1852:1856–66.
- 47. Kim EC, Zhang J, Pang W, Wang S, Lee KY, Cavaretta JP, et al. Reduced axonal surface expression and phosphoinositide sensitivity in K_v7 channels disrupts their function to inhibit neuronal excitability in Kcnq2 epileptic encephalopathy. Neurobiol Dis. 2018;118:76–93.
- Soldovieri MV, Ambrosino P, Mosca I, De Maria M, Moretto E, Miceli F, et al. Early-onset epileptic encephalopathy caused by a reduced sensitivity of Kv7.2 potassium channels to phosphatidylinositol 4,5-bisphosphate. Sci Rep. 2016;6:38167.
- Sands TT, Miceli F, Lesca G, Beck AE, Sadleir LG, Arrington DK, et al. Autism and developmental disability caused by KCNQ3 gainof-function variants. Ann Neurol. 2019;86:181–92.
- Millichap JJ, Park KL, Tsuchida T, Ben-Zeev B, Carmant L, Flamini R, et al. (2016) KCNQ2 encephalopathy: features, mutational hot spots, and ezogabine treatment of 11 patients. Neurol Genet. 2016;2:e96.
- Goldberg-Stern H, Kaufmann R, Kivity S, Afawi Z, Heron SE. Novel mutation in KCNQ2 causing benign familial neonatal seizures. Pediatr Neurol. 2009;41:367–70.
- Grinton BE, Heron SE, Pelekanos JT, Zuberi SM, Kivity S, Afawi Z, et al. Familial neonatal seizures in 36 families: Clinical and genetic features correlate with outcome. Epilepsia. 2015;56:1071–80.

- Hadley JK, Passmore GM, Tatulian L, Al-Qatari M, Ye F, Wickenden AD, et al. Stoichiometry of expressed KCNQ2/KCNQ3 potassium channels and subunit composition of native ganglionic M channels deduced from block by tetraethylammonium. J Neurosci. 2003;23:5012–9.
- Kanaumi T, Takashima S, Iwasaki H, Itoh M, Mitsudome A, Hirose S. Developmental changes in KCNQ2 and KCNQ3 expression in human brain: possible contribution to the age-dependent etiology of benign familial neonatal convulsions. Brain Dev. 2008;30:362–9.
- Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46:D1062–D1067. (retrieved from https://www.ncbi.nlm.nih. gov/clinvar/?term=KCNQ3%5Bgene%5D).
- 56. Karczewski KJ, Francioli LC, Tiao G,Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human proteincoding genes. bioRxiv 531210; doi: http://dx.doi. org/10.1101/531210. (retrieved from https://gnomad.broadinstitute.org/gene/ENSG00000184156).
- Weckhuysen S, Ivanovic V, Hendrickx R, Van Coster R, Hjalgrim H, Moller RS, et al. Extending the KCNQ2 encephalopathy spectrum: clinical and neuroimaging findings in 17 patients. Neurology. 2013;81:1697–703.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Lauritano A, Moutton S, Longobardi E, et al. A novel homozygous KCNQ3 loss-of-function variant causes non-syndromic intellectual disability and neonatal-onset pharmacodependent epilepsy. *Epilepsia Open*. 2019;4:464–475. https://doi.org/10.1002/epi4.12353



Supp. Fig 1: Pedigree of the family and *KCNQ3* Sanger sequencing of all available individuals. The chr8:g.133150233dup variation is shown using as a reference the minus DNA strand, from which the mRNA is transcribed. The arrow indicates the proband. In grey, a maternal uncle with ID and epilepsy is indicated.

3.3. Etude de l'apport du séquençage d'exome en trio pour l'identification de nouveaux gènes impliqués dans les anomalies du développement (article 3)

Ce travail porte sur l'étude de la stratégie dans la conduite des analyses pangénomiques d'exome chez les patients atteints d'une forme sporadique d'anomalie du développement avec ou sans DI. Comme cela a été mentionné en introduction, l'analyse d'exome diagnostique consiste à rechercher parmi les variants localisés dans des gènes connus ceux compatibles avec le phénotype du patient. L'analyse en solo est donc tout à fait faisable avec un temps d'analyse légèrement augmenté par rapport au trio. En revanche, la recherche de variants candidats parmi les gènes non connus en pathologie humaine nécessite des recherches bibliographiques poussées pour un nombre important de variant et donc le temps d'investigation est bien plus important. L'analyse en trio permettant de sélectionner beaucoup plus rapidement les variants *de novo* et compatibles avec une forme autosomique récessive est bien plus efficace (en cas d'affection sporadique). Le but de ce travail était donc d'objectiver dans une série l'apport du séquençage en trio pour accélérer la recherche translationnelle chez les patients sans diagnostic après une interprétation de l'exome en solo puisqu'on s'attendait à ce que les patients ayant un diagnostic dans un gène connu étaient déjà identifiés en solo.

Méthodologie

Les données présentées ici portent sur 70 patients (22 filles, 48 garçons) atteints d'une anomalie du développement avec ou sans DI chez qui l'exome solo était négatif. Les patients ont été inclus s'ils présentaient une forme d'anomalie du développement et/ou DI sporadique avec accès aux prélèvements parentaux et accord de ces derniers et/ou du responsable légal de poursuivre les investigations.

<u>Résultats</u>

L'ensemble des résultats est résumé dans le tableau 3.

Trente-deux gènes potentiellement candidats ont pu être identifiés chez 28/70 patients (40%) et pour 3 patients (4%), 2 voire 3 gènes candidats pouvaient être suspectés. Suite à l'interprétation et au partage international de données permettant de collecter des arguments de pathogénicité, les variants ont pu être classés en probablement pathogènes ou de signification indéterminée.

Nous avons retrouvé 21 variants probablement pathogènes situés dans 18 gènes pour 18/70 patients (25,7%). Au total, on retrouvait 10 variants tronquants, 1 variant au niveau d'un site consensus d'épissage, 1 variant touchant un codon d'initiation, et 9 variants faux sens répartis selon le mode de transmission en :

- 13 *de novo* dont 8 probablement associés à une haploinsuffisance (7 stop/frameshifts, 1 site consensus d'épissage), 1 perte de codon d'initiation et 4 faux sens,

- 5 récessifs dont 2 homozygotes (1 tronquant et 1 faux sens) et 3 hétérozygotes composites (2 tronquant/faux sens et 1 double faux sens).

Parmi ces gènes très probablement incriminés, 9 étaient connus comme OMIM morbides et 9 n'étaient pas associés à des maladies humaines. En ce qui concerne les 9 gènes déjà associés à des maladies humaines:

- le gène ACO2 (MIM 100850) n'avait pas été retenu du fait d'un phénotype incomplet;

- il existait des malformations atypiques pour les gènes *KMT5B* (communication inter-auriculaire et hypoplasie rénale), *OTX2* (malformation unilatérale de l'oreille) et *KCNMA1* (retard de développement, hypertrophie gingivale, dilatation aortique, atteinte osseuse, épisodes de pseudo-obstruction intestinale);

- le phénotype était bien plus sévère qu'attendu pour les gènes *MFSD8* (suggérant plusieurs maladies) et *TOR1A* (avec une hérédité différente dans ce cas).

- par ailleurs, 3 gènes rapportés en pathologie humaine dans des publications ont été identifiés chez des patients avec un phénotype concordant mais l'analyse diagnostique n'avait pas retenu ces variants car les gènes n'étaient pas encore compris dans la liste OMIM morbides: *PBX1*, *PPM1D* et *UNC45A* (avec suspicion d'un diagnostic associé dans ce dernier cas).

Les 9 autres gènes n'étaient pas connus en pathologie humaine, un travail collaboratif a pu être mené pour chacun d'entre eux permettant de conclure à une forte probabilité d'implication soit par réplication des données seules du fait d'un nombre de patients concernés importants (*FBXO11*, 20 patients), soit par combinaison de réplication des données et analyses fonctionnelles : *HNRNPR* (3 patients), *IRF2BPL* (11 patients et 7 patients publiés concomitamment par une autre équipe), *PIGK* (8 patients de 6 familles), *PTBP1* (>3), *RALGAPB* (4), *SPEN* (6), *TRAF7* (>10) et *ZMIZ1* (19 patients).

Nous avons aussi retenu 16 variants de signification indéterminée dans 14 gènes : 2 tronquants, 1 variant d'épissage en position -3 et 13 faux sens répartis en :

- 5 de novo (5 faux sens),

- 8 récessifs dont 6 homozygotes (2 tronquants, 4 faux sens) et 2 hétérozygotes composites (1 splice -3/faux sens et 1 double faux sens)

- 1 gène sur le chromosome X (1 faux sens hérité chez un garçon, pas de ségrégation possible chez les apparentés maternels). A noter qu'un autre variant faux sens du gène *KIF4A* avait pu être écarté sur la ségrégation familiale et n'a donc pas été comptabilisé dans ces résultats incertains (diagnostic finalement négatif).

A partir de la nature et du mode de transmission des variants identifiés dans notre étude, on peut retenir que :

- selon le mode de transmission :

- 72.2% des variants probablement pathogènes étaient de novo contre 35.7% des VSI
- 27.8% des variants probablement pathogènes avaient un mode de transmission suspecté autosomique récessif contre 42.9% des VSI

- selon la nature des variants :

- Les variants tronquants (hors site consensus d'épissage) représentent 47.6% des probablement pathogènes contre 12.5% des VSI
- Les variants faux sens représentent 42.9% des probablement pathogènes contre 81.3% des VSI

Par ailleurs :

- Parmi les variants *de novo* : 100% (7/7) des tronquants ont été interprétés comme probablement pathogènes alors que 44% (4/9) des faux sens *de novo* l'ont été.

- Sans prendre en compte la nature des variants, parmi les gènes possiblement AR, 38.5% sont conclus probablement pathogènes contre 72.2% des variants *de novo*.

Pour 2 patients ayant un diagnostic très probable (*UNC45A* et *MFSD8*), le phénotype était incomplètement expliqué et la recherche de variants candidats complémentaires a conduit respectivement à suspecter un autre gène avec faux sens *de novo* (*PBRM1*) et 2 autres gènes avec faux sens homozygotes (*NUP98* et *WTIP*) (consanguinité). Un double diagnostic concernerait potentiellement 2/18 soit 11% des patients pour lequel un gène est déjà fortement en cause. Un $3^{\text{ème}}$ patient était potentiellement concerné par un diagnostic multiple ; né de parents consanguins, il était porteur de 2 VSI à l'état homozygote (*TRAPPC8* et *CBLB*).

Une analyse des 25 premiers trios montrait une différence dans la nécessité de valider des variants par la méthode de référence de Sanger : dans la stratégie solo, 1,47 variants/patient ont été validés alors qu'avec le trio 0,11 variant/patient aurait été testé en 1^{ère} intention puisqu'un certain nombre pouvait être écarté du fait des données de ségrégation parentale apportées par le trio.

Tableau 3 : Description des variants classés par groupe de pathogénicité (probablement pathogènes, candidats, ou de signification indéterminée) identifiés chez 28/70 patients de la cohorte.

Gene	MIM number	Geno- type	genomic position (GRCh37)	c.DNA	protein	CADD	Polyphen2	GERP	misZ	pLI	OMIM disorder	Additioal patients from data sharing	reason
Variants identified in OMIM-morbid genes													
ACO2	100850	htc	chr22:g. 41903904C>T chr22:g	NM_001098.2: c.283C>T NM_001098.2:	p.(Arg95Cys)	34	1	4.66	3.27	0.45	614559 (Infantile cerebellar-retinal		Incomplete
			41895778C>T	c.85C>T	p.(Arg29Trp)	27.9	0	3.91			degeneration)		phenotype
KCNMA 1	600150	ht	chr10:g. 78869939C>T	NM_001014797.2: c.1123G>A	p.(Gly375Arg)	28.1	1	5.49	6,43	1	609446 (Paroxysmal nonkinesigenic dyskinesia, 3, with or without generalized epilepsy)		novel phenotype
KMT5B	610881	ht	chr11:g. 67939009delC	NM_017635.3: c.820+1delG	p.?	NA	NA	NA	2,71	1	617788 (Mental retardation, autosomal dominant 51)	> 10	atypical phenotype
MFSD8 ¹	611124	hm	chr4:g. 128859985C> T	NM_152778.2: c.707G>A	p.(Arg236His)	12.37	0.824	2.8	-0,57	0	610951 (Ceroid lipofuscinosis, neuronal, 7); 616170 (Macular dystrophy with central cone involvement)		atypical phenotype
OTX2	600037	ht	chr14:g.572685 98delC	NM_021728.3: c.749delG	p.(Gly250Asp fs*52)	NA	NA	NA	1,02	0,74	610125 (Microphthalmia, syndromic 5; Retinal dystrophy, early- onset, with or without pituitary dysfunction); 613986 (Pituitary hormone deficiency, combined, 6)		atypical phenotype
PBX1	176310	ht	chr1:g.1645292 04C>T	NM_002585.3: c.145C>T	p.(Gln49*)	14	NA	5,42	3,99	0,91	617641 (Congenital anomalies of kidney and urinary tract syndrome with or without hearing loss, abnormal ears, or developmental delay)		intercurent publication

PPM1D	605100	ht	chr17:g.587403 71delC	NM_003620.3: c.1276delC	p.(Pro426His fs*5)	NA	NA	NA	3,13	0	617450 (Intellectual developmental disorder with gastrointestinal difficulties and high pain threshold)		intercurent publication
TORIA	605204	hm	chr9:g.1325763 88G>A	NM_000113.2: c.862C>T	p.(Arg288*)	29.8	NA	5.27	0,98	0,17	128100 (Dystonia-1, torsion)	2	phenotype
<i>UNC45A</i> ² #	611219	htc	chr15:g.914900 96delinsGCA chr15:g.914964 75G>C	NM_001039675.1: c.1407delinsGCA NM_001039675.1: c.2467G>C	p.(Asp469Glu fs*17) p.(Ala823Pro)	NA 14.9	NA 1	NA 5,63	0,38	0	none	2	4 previously published patients by local team
					Variants identifie	ed in highly	y candidate	genes					
FBXO11 #	607871	ht	chr2:g.4804618 6_48046190 delTTCTC	NM_001190274.1: c.1825_1829 delGAGAA	p.(Glu609*)	NA	NA	NA	4,03	1	618089 (Intellectual developmental disorder with dysmorphic facies and behavioral abnormalities)	> 10	positive dasharing
HNRNPR #	607201	ht	chr1:g.2363724 8_23637249	NM_001102398.1: c.1609_1610insG	p.(Ala537Gly fs*10)	NA	NA	4.97	3,56	1	none	3	positive dasharing
IRF2BPL #	611720	ht	chr14:g.774936 17G>C	NM_024496.3: c.519C>G	p.(Tyr173*)	44	NA	NA	4,73	0,97	618088 (Neurodevelopmental disorder with regression, abnormal movements, loss of speech, and seizures)	> 10	positive dasharing
DICK	(05007	1.	chr1:g.7763241 2T>G	NM_005482.2: c.479A>C	p.(Tyr160Ser)	21.7	1	5.12	0.17	0		5	positive
PIGK	605087	htc	chr1:g.77676 171G>A	NM_005482.2: c.97C>T	p.(Gln33*)	25.2	NA	2.85	-0,17	0	none	5	dasharing
PTBP1	600693	ht	chr19:g.797499 T>C	NM_002819.4: c.2T>C	p.?	12.61	0.622	4.58	2,11	1	none	3	positive dasharing
RALGAP B	none	ht	chr20:g.371282 34T>C	NM_020336.2: c.698T>C	p.(Val233Ala)	27.1	0.99	5.68	2,59	1	none	4	positive dasharing
SPEN	613484	ht	chr1:g.1625882 2_16258823 delAC	NM_015001.2: c.6087_6088delAC	p.(Glu2029Asp fs*5)	NA	NA	NA	2,99	1	none	6	positive dasharing
TRAF7#	606692	ht	chr16:g.2226 351G>A	NM_032271.2: c.1964G>A	p.(Arg655Gln)	18.65	0.997	4.43	3,34	0,46	618164 (Cardiac, facial, and digital anomalies with developmental delay)	> 10	positive dasharing

ZMIZ1#	607159	ht	chr10:g.810520 49C>T	NM_020338.3: c.893C>T	p.(Thr298Ile)	27.9	1	5,16	4,22	1	none	> 10	positive dasharing
Variants of unknown significance													
AADACL 3	none	ht	chr1:g.12785 548G>A	NM_001103169.1: c.428G>A	p.(Gly143Asp)	22.5	1	3.49	-0,35	0	none		
ATP13A1	none	hte	chr19:g.197585 71G>T	NM_020410.2: c.2633-3C>A	p.?	NA	NA	NA	3 1/	0.00	none		
AITISAI	none	inc	chr19:g.197679 10C>T	NM_020410.2: c.799G>A	p.(Val267Ile)	22,8	0,12	4,89	3,14	0,77	none		
CBLB ³	604491	hm	chr3:g.1055724 21G>A	NM_170662.3: c.256C>T	p.(Arg86*)	36	NA	3.72	0,69	0,91	none		
FAM91 Al	none	ht	chr8:g.1247999 55G>A	NM_144963.2: c.1243G>A	p.(Glu415Lys)	36	0.996	5.57	2	0,52	none		
HTATSF 1	none	hmi	chrX:g.135593 936G>A	NM_001163280.1: c.2032G>A	p.(Asp678Asn)	10.07	0.013	2.1	1,80	0,97	none		
NUP98 ¹	601021	hm	chr11:g.3720 545C>T	NM_016320.4: c.3776G>A	p.(Cys1259Tyr)	25.2	1	5.44	1.41	1	none		
PBRM1 ²	606083	ht	chr11:g.3720 545C>T	NM_018313.4: c.1301G>A	p.(Arg434Gln)	32	0.983	5.59	3,48	1	none		
PDGFC	608452	hm	chr4:g.1578919 45_157891948 delTTCC	NM_016205.2: c.108_111delGGA A	p.(Lys36Asn fs*41)	NA	NA	NA	0.76	0.97	none		
DETI	611008	hto	chr3:g.53126 558C>T	NM_052859.3: c.1285G>A	p.(Val429Met)	27.7	1	3.62	3.62	0	612015 (Congenital		
$\mathbf{M}^{T}\mathbf{I}\mathbf{I}$	011908	Inc	chr3:g.53156 410G>A	NM_052859.3: c.436C>T	p.(His146Tyr)	19.6	0.001	5.07	-0,35	0	glycosylation In)		
RPUSD3	617759	hm	chr3:g.9882 413G>A	NM_173659.3: c.484C>T	p.(Arg162Trp)	12.5	1	2.23	0,1	0	none		
SETD2	612778	ht	chr3:g.47147 494T>C	NM_014159.6: c.4832A>G	p.(Asn1611Ser)	24.4	0.98	5.28	3.05	1	616831 (Luscan- Lumish syndrome)		
$\frac{TRAPPC}{8^3}$	614136	hm	chr18:g.294887 66C>T	NM_014939.3: c.1073G>A	p.(Arg358Gln)	34	0.989	5.7	-0,52	1	none	1	
WASF2	605875	ht	chr1:g.27742 522A>G	NM_006990.3: c.494T>C	p.(Leu165Pro)	24.9	1	5,39	0,97	1,7	none		
WTIP ¹	614790	hm	chr19:g.349854 99C>T	NM_001080436.1: c.1040C>T	p.(Ala347Val)	29.2	0.999	4.84	0,93	0	none		

Légende : #: gènes récemment publiés ; ¹ / ² / ³ : correspondance des patients pour qui un double ou triple diagnostic est suspecté; hm: homozygote: hm: hemizygote; ht: heterozygote: htc: heterozygote composite ; NA: non applicable.

Discussion

Nous présentons ainsi le résultat d'une réanalyse en trio de données d'exome non contributives en solo chez 70 patients atteints d'anomalie du développement avec ou sans DI. Le rendement estimé de diagnostics probables ou confirmés est donc de 26% (18/70). De plus, pour 12 gènes chez 10 patients, il existait des variants de signification indéterminée puisqu'il n'a pas été possible de répliquer les données par le partage international ni d'obtenir des arguments fonctionnels. Une seule publication a rapporté le résultat d'une analyse séquentielle d'exome par trio après une analyse en solo non contributive chez 74 patients atteints pour leur majorité 81-84% d'anomalies du développement, montrant un nouveau gène très probablement impliqué pour 10,8% des cas, un gène connu pour 25,7% des cas et un gène potentiellement candidat (variant de signification indéterminée) pour 14,9% des cas ³⁴³.

Quelques exemples permettent d'illustrer certaines situations :

<u>- Variabilité d'expression</u>: exemple du gène <u>OTX2</u>. Un patient adressé pour anomalie des oreilles externes et moyennes et retard développemental avec troubles autistiques alors que le phénotype classique est la micro/anophtalmie possiblement syndromique avec dysplasie septo-optique.

<u>- Intégration de nouveaux gènes associés à une pathologie en lien avec le phénotype avec parfois un délai entre l'analyse d'exome initiale, la publication et l'ajout du gène OMIM morbide.</u> Ce type d'exemple concerne les gènes <u>PPM1D, KMT5B, PBX1, UNC45A</u>. Ces gènes qui n'étaient pas connus initialement comme associés à une maladie génétique et ont été publiés entre le moment de la lecture solo et de la réanalyse en trio. Cela fait partie des exemples qui démontrent l'intérêt de la relecture périodique des données. A noter que le gène UNC45A qui a été décrit dans un phénotype d'anomalie oto-ostéo-hépato-entérique publié en février 2018 n'a toujours pas été rajouté à la liste des gènes OMIM morbides en août 2019 et ne fait donc pas encore partie des gènes en lecture diagnostique.

<u>- Mécanisme mutationnel similaire, mode de transmission différent et élargissement du spectre</u> <u>phénotypique :</u> cela est bien illustré par le gène <u>TOR1A</u>. Il était connu comme associé à la dystonie de type 1 (pénétrance 30%) avec un variant hétérozygote récurrent conduisant à la délétion d'un acide aminé ayant un effet perte de fonction similaire aux quelques variants décrits prédisant l'apparition d'un codon stop prématuré. La présence d'un variant tronquant biallélique a dévoilé le spectre phénotypique allant vers un tableau d'arthrogrypose avec décès néonatal. Un patient porteur du variant récurrent à l'état homozygote a été décrit avec un tableau clinique similaire. Une collaboration a pu être initiée avec l'équipe du Dr. Jan Senderek de la Friedrich-Baur-Institut à Munich qui avait collecté une quinzaine de patients mais n'a finalement plus donné de nouvelles à partir de début 2018 pour une raison inconnue (peut-être du fait que 2 cas cliniques isolés avaient été publiés par d'autres équipes alors que les tests fonctionnels étaient en cours ?). <u>- Force des plateformes de partage international de données permettant de recruter un nombre de patients important :</u> ceci concerne le gène *FBXO11*. Aucune analyse fonctionnelle n'a été nécessaire devant le poids des données moléculaires et le fait que ce gène est soumis à une pression de sélection n'autorisant pas la présence de variants entraînant un codon stop prématuré. Une collaboration a été mise en place avec l'équipe du Prof Christiane Zweier de l'Humangenetisches Institut à Erlangue dans laquelle nous avons pu inclure 2 patients (un issu de l'apport de la réanalyse en trio, l'autre identifié après avoir interrogé la base de données d'exome pour rechercher d'autres variants dans ce gène, ce 2^{ème} patient avait été adressé d'une équipe danoise dans le cadre du recrutement en recherche de patients avec syndrome marfanoïde) ³⁸⁴.

<u>- Validation des mutations candidates par un test fonctionnel ou métabolique unique et relativement</u> <u>simple.</u> Ceci est illustré par les exemples des gènes <u>*PIGK*</u> et <u>*ACO2*</u>.

En ce qui concerne le gène <u>PIGK</u>, un test par cytométrie en flux à partir d'un prélèvement sanguin a permis de confirmer le déficit de synthèse de l'ancre GPI. Ce gène fait partie d'une cascade enzymatique avec de nombreux autres déjà impliqués dans des anomalies du développement par défaut de synthèse du glycosylphosphatidylinositol. Ce test fonctionnel avait déjà été mis au point pour d'autres gènes de la même voie métabolique et a permis de valider la pathogénicité des variants. Une collaboration a pu être établie avec l'équipe du Prof Philippe Campeau, Université de Montréal (en cours de publication).

Concernant <u>ACO2</u>, le gène était connu depuis 2012 comme impliqué dans un phénotype neurodéveloppemental sévère avec atteinte ophtalmologique à type d'atrophie optique et/ou dégénérescence rétinienne dans 2 familles consanguines ³⁸⁵. Le séquençage d'exome solo avait été fait à un âge jeune (consultation à 12 mois) dans un contexte d'hypotonie au 1^{er} plan. La réinterprétation en trio avec un recul clinique supplémentaire a permis de mieux analyser les gènes candidats notamment en sélectionnant les variants hétérozygotes composites et les données bibliographiques étaient concordantes avec le phénotype. Une collaboration avec Steven Kushner, Femke de Vrij, Christian Bouwkamp et Anniek Meesters du département de psychiatrie neurobiologique de l'université de Rotterdam a permis de confirmer la pathogénicité des 2 variants puisque le test fonctionnel consistant en un dosage de l'activité enzymatique sur fibroblastes a montré des résultats similaires à des patients déjà testés.

<u>- Utilité de coupler la description de patients à des tests fonctionnels plus complexes.</u> Cela a été particulièrement important pour le gène <u>HNRNPR</u> qui n'était pas connu comme responsable de maladie génétique. Le faible nombre de patients (3) décrits a d'autant plus nécessité de tester les conséquences fonctionnelles des variants notamment en testant l'expression des gènes sur des cellules de patients puisque la protéine intervient dans la régulation de l'épissage de multiples ARNm. Une collaboration a été établie avec l'équipe du Dr Alyson W. MacInnes de l'Amsterdam University Medical Centers. Le RNA-seq sur fibroblastes a montré que l'haploinsuffisance du gène
dérégulait de nombreux gènes homeobox et facteurs de transcription du fait de son rôle dans la régulation de l'épissage ³⁸⁶.

Le partage international des données a ainsi permis d'initier des collaborations pour 9 nouveaux gènes (*FBXO11, HNRNPR, IRF2BPL, PIGK, PTBP1, RALGAPB, SPEN, TRAF7* et *ZMIZ1*), pour un gène associé à un nouveau phénotype par mécanisme physiopathologique différent (*KCNMA1*), pour un gène associé à un mécanisme similaire mais dans un mode de transmission différent (*TOR1A*), et pour un gène déjà décrit dans un phénotype similaire et un mécanisme similaire mais très peu de patients décrits jusqu'à présent (*UNC45A*, une seule publication ³⁸⁷). L'apport supplémentaire de cette stratégie serait au minimum de 18/70 soit 25,7%.

Cette étude démontre qu'en plus de l'intérêt que le séquençage en trio aurait sur le plan diagnostique pour filtrer rapidement les variants selon les différentes hypothèses de transmission mendélienne, il permet d'accélérer la recherche translationnelle et la découverte de nouveaux gènes par rapport à la relecture recherche en solo bien plus chronophage. Cependant, l'interprétation des variants faux sens reste la difficulté principale puisque les outils de prédiction bioinformatiques ont leurs limites et qu'il est souvent nécessaire de réaliser des tests fonctionnels à moins que plusieurs variants soient situés au niveau du même codon, ou très proches les uns des autres au sein du même domaine fonctionnel de la protéine, ou encore que le nombre de patients avec phénotype concordant soit relativement élevé.

Ces résultats confirment aussi qu'il existe un biais d'interprétation avec une plus grande facilité d'incriminer un variant *de novo* que des variants bialléliques d'un gène et des variants tronquants par rapport aux faux sens.

Une des questions émanant de ce travail est la balance entre :

1) le séquençage en solo permettant d'épargner le coût du séquençage des parents dans les 30 à 35% des cas où un diagnostic moléculaire pourra être retenu dans un gène connu, mais la nécessité de validations des variants par la méthode de référence de Sanger, par rapport à

2) séquençage en trio d'emblée permettant de gagner du temps humain d'analyse de la liste des variants touchant des gènes connus comme associés à des pathologies humaines et évitant la validation inutile de variants qui sont exclus par une hérédité non compatible.

Bien que cette différence puisse être considérée comme très acceptable si on compare au coût additionnel du séquençage des parents d'un cas index pour bénéficier de l'interprétation en trio, il faudrait pouvoir mettre en perspective le coût humain de tri des variants en solo et le temps nécessaire à la validation de ces variants additionnels. Plusieurs points peuvent être discutés quant au choix de la meilleure stratégie : - Avec l'augmentation des connaissances et des gènes associés à une ou plusieurs maladies génétiques, la liste des gènes à explorer en diagnostic augmente ce qui entraîne un allongement du temps d'interprétation

- Une autre publication issue du consortium du Royaume Uni (Deciphering Developmental Disorders) a comparé l'apport du trio par rapport au solo et indique que le solo conduit à considérer 10 fois plus de variants candidats que le trio lorsque les parents ne sont pas atteints, 3 fois plus lorsqu'un parent est atteint et 1,5 fois plus lorsque les 2 parents sont atteints, ce qui ajoute un temps d'interprétation non négligeable (figure 28)³³².

- Un compromis dans ces stratégies permettant d'avoir un rendement coût-efficacité correct et dont l'utilisation va probablement prendre de l'ampleur est de rassembler un mélange d'ADN issus de plusieurs pères, un mélange d'ADN issu de plusieurs mères des cas index et de séquencer ces mélanges (séparément pools de pères ou pools de mères) pour permettre une interprétation en trio en prenant en compte que cela crée artificiellement une mosaïque et que le nombre de séquences porteuses d'un variant est plus faible et proportionnel au nombre d'ADN parentaux dilués mais que compte tenu de l'exploration de variants rares, la présence d'un variant sur plusieurs lectures dans un des 2 pools permet de conclure qu'il est hérité. Cela permet de limiter les coûts de séquençage des parents tout en obtenant une information de qualité suffisante pour l'interprétation (c'est-à-dire information sur le caractère hérité, et si oui de quel parent, ou non du variant). Cette stratégie est maintenant en place au sein de l'équipe.

Exploitation de ce travail :

Poster à l'ESHG (Juin 2017, Copenhague) :

Moutton S^{*}, Tran Mau-Them F^{*}, Vitobello A, Bruel AL, Thevenon J, Lehalle D, Jean N, Carmignac V, Kuentz P, St-Onge J, Masurel A, El Chehadeh S, Callier P, Mosca AL, Marle N, Vabres P, Poe C, Jouan T, Chevarin M, Houcinat N, Riviere JB, Philippe C, Duffourd Y, Faivre L, Thauvin C. When solo exomes fail: added-value of parents sequencing

Poster à l'ESHG (Juin 2018, Barcelone) :

S Moutton^{*}, F Tran Mau-Them^{*}, C Philippe, A Vitobello, AL Bruel, J Thevenon, D Lehalle, N Jean-Marçais, V Carmignac, C Poe, T Jouan, M Chevarin, N Houcinat, Kuentz, J St-Onge, A Masurel, S El Chehadeh, AL Mosca, N Marle, JB Riviere, P Vabres, P Callier, Y Duffourd, L Faivre, C Thauvin-Robinet. Trio whole exome sequencing as an efficient second step strategy to decipher molecular basis of developmental disorders after negative first-tier solo clinical WES Article en cours de rédaction :

Tran Mau-Them F^{*}, Moutton S^{*}, Philippe C, Vitobello A, Bruel AL, Thevenon J, Lehalle D, Jean N, Carmignac V, Poe C, Jouan T, Chevarin M, Houcinat N, Kuentz P, J St-Onge, A Masurel, S El Chehadeh, P Callier, AL Mosca, N Marle, JB Riviere, P Vabres, Duffourd Y, Faivre L, Thauvin-Robinet C^{*}. Trio whole exome sequencing: an efficient second step strategy to decipher molecular basis in developmental disorders.

* ces auteurs ont contribué au travail de façon égale.

BRIEF COMMUNICATION

2	
3	Second-tier trio exome sequencing after negative solo clinical exome sequencing: an
4	efficient strategy to increase diagnostic yield and decipher molecular bases in
5	developmental disorders
6	
7	Frederic Tran Mau-Them MD ^{1,2,*} , Sebastien Moutton MD ^{1,3,*} , Caroline Racine ³ , Antonio
8	Vitobello PhD ^{1,2} , Ange-Line Bruel PhD ^{1,2} , Sophie Nambot MD ^{1,3} , Steven A. Kushner MD
9	PhD ⁶ , Femke M.S. de Vrij PhD ⁶ , Daphné Lehalle MD PhD ^{1,3} , Nolwenn Jean-Marçais MD ³ ,
10	François Lecoquierre MD ¹ , Julian Delanne ¹ , Julien Thevenon MD PhD ³ , Charlotte Poe ² ,
11	Thibaut Jouan ¹ , Martin Chevarin ² , David Geneviève MD PhD ^{4,5} , Marjolaine Willems MD
12	PhD ⁴ , Christine Coubes MD ⁴ , Nada Houcinat MD ^{1,3} , Alice Masurel-Paulet MD ³ , Anne-Laure
13	Mosca-Boidron MD ^{1,2} , Emilie Tisserant ¹ , Patrick Callier MD PhD ^{1,2,} , Arthur Sorlin MD ^{1,2,3} ,
14	Yannis Duffourd ¹ , Laurence Faivre MD PhD ^{1,3} , Christophe Philippe PhD ^{1,2} , Christel
15	Thauvin-Robinet MD PhD ^{1,2,3,7,*}
16	
17	1. Inserm- Université de Bourgogne UMR1231 GAD, FHU-TRANSLAD, F-21000, Dijon,
18	France
19	2. Unité Fonctionnelle d'Innovation diagnostique des maladies rares, Pôle de Biologie, FHU-
20	TRANSLAD, CHU Dijon Bourgogne, F-21000, Dijon, France
21	3. Centre de Référence maladies rares « Anomalies du Développement et syndrome
22	malformatifs » de l'Est, Hôpital d'Enfants, FHU-TRANSLAD, CHU Dijon Bourgogne, F-
23	21000, Dijon, France

- 24 4. Département de Génétique Médicale, Maladies Rares et Médecine Personnalisée, Centre de
- 25 Référence Maladies Rares Anomalies du Développement et Syndromes Malformatifs
- 26 ADSOOR, CHU Montpellier, Montpellier, France
- 27 5. Université Montpellier, Unité Inserm U1183, Montpellier, France
- 28 6. Department of psychiatry, Erasmus MC Goederenontvangst, Rotterdam, The Netherlands
- 29 7. Centre de Référence Déficiences Intellectuelles de Causes Rares, Hôpital d'Enfants, CHU
- 30 Dijon Bourgogne, F-21000, Dijon, France
- 31 * these authors contributed equally to this work
- 32
- 33 Corresponding author
- 34 Dr. Tran Mau-Them Frederic
- 35 Inserm Université de Bourgogne UMR 1231 GAD
- 36 UFR des Sciences de Santé
- 37 Bâtiment B3
- 38 15 avenue du maréchal Delattre de Tassigny
- 39 21000 Dijon, France
- 40 Tel : +33 3 80 39 66 59
- 41 Email: <u>frederic.tran-mau-them@u-bourgogne.fr</u>

42 Abstract

disorders (DD), 43 **Purpose**: Developmental which are characterized by malformations/dysmorphism and/or intellectual disability and affect around 3% of the 44 worldwide population, are mostly linked to genetic anomalies. Despite clinical exome 45 sequencing (cES) centered on the genes involved in human genetic disorders, the majority of 46 patients affected by DD remain undiagnosed after solo-cES. The trio-based strategy is 47 expected to facilitate variant selection thanks to rapid parental segregation. Methods: We 48 performed second-tier trio-ES analysis in 70 patients with negative results after solo-cES. All 49 candidate variants were shared with a MatchMaking exchange system to identify additional 50 patients carrying variants in the same genes and with a similar phenotype. Results: In 18/70 51 patients (26%), we confirmed the causal implication of 18 genes including 9 OMIM-morbid 52 genes and 9 new strong candidate genes. Moreover, 11 genes in 10/70 patients (14%) 53 54 harbored variants of unknown significance. Conclusion: In DD, a second step trio-based ES analysis appears to be an efficient strategy for identifying strong candidate genes and 55 improving diagnostic yield in diagnostic and translational research. 56

57

58

- 59
- 60
- 61
- 62

Key words: Clinical Exome Sequencing, solo strategy, trio strategy, data sharing,
developmental disorders

65 **Introduction**

Developmental disorders (DD) affect around 3% of the worldwide population and are recorded in approximately 2 to 5% of births¹. These disorders are characterized by malformations/dysmorphism and/or intellectual disability (ID). The numerous etiologies of DD include environmental (e.g. maternal condition, delivery complications, trauma, or infection) and genetic causes, resulting in a heterogeneous group of diseases with possible clinical overlap. Genetic causes likely account for a large majority of patients with DD, since up to 80% are estimated to be of genetic origin².

Since the introduction of Next-Generation Sequencing (NGS), numerous Mendelian disorders 73 have been explained following the identification of genes newly involved in human disorders 74 or the implication of known disease-causing genes in novel phenotypes. Clinical exome 75 sequencing (cES) targeting genes involved in human genetic disorders (OMIM-morbid genes) 76 77 has become a highly powerful first-tier genetic tool to identify the genetic cause of DD and reduce the diagnostic odyssey of affected patients. Clinical ES can be used in solo (proband 78 79 only) or in trio strategies (proband and parents). Genetic laboratories tend to opt for trio-cES strategy since rapid filtering can be done based on parental segregation (i.e. compound 80 heterozygotes, heterodisomy or uniparental disomy) and because they can easily select de 81 *novo* variants, whose enrichment is a well-known cause of DD^1 . Current solo-ES usually 82 vields a diagnostic rate of 30% in DD³. In trio-ES, analysis can also easily be extended to non 83 OMIM-morbid and non OMIM genes, allowing researchers to identify putative candidates 84 and to implicate them in novel human disorders, after recurrence is confirmed by data sharing 85 and genotype-phenotype correlation⁴. 86

For economic reasons (*i.e.* to be able to offer cES to the majority of patients with suspected genetic conditions), some teams have preferred to focuse on solo-cES, and others have followed up with annual re-analysis of solo-cES data, resulting in a causal diagnosis in 8% of 90 the patients whose initial results were negative⁴. Only one pilot study has described the 91 benefit of a second step trio-based ES after solo-cES in a heterogeneous cohort, mainly 92 regarding developmental disorder patients. They reported a diagnostic yield of 36% when 93 pathogenic variant were considered, and the yield rose to 51% when candidate genes were 94 included⁵. Complementary studies are required to confirm the value of this interesting two-95 step strategy.

Accordingly, we present here the results of second-tier trio-based ES performed subsequent to
a first-tier negative solo-cES in 70 patients with DD.

98

99 **Patients and methods**

100 **Patients**

From June 2012 to March 2018, we performed solo-cES for 816 patients affected by DD. A 101 102 positive molecular diagnosis was obtained in 239/816 patients (29.2%), and an uncertain molecular diagnosis was obtained in 81/816 patients (9.9%)³. After thorough clinical re-103 104 evaluation, we selected 70 of the 496/816 patients with negative solo-cES. Before performing 105 trio-ES, we did an extensive reanalysis of their solo-cES data, focused on SNVs and CNVs implicating OMIM-morbid genes, with an updated bioinformatics pipeline and databases⁴. 106 Cases of negative solo-cES were then selected for inclusion and families were contacted to 107 obtain their consent to perform these additional tests. Clinical data were gathered through 108 PhenomeCentral, a Human Phenotype Ontology (HPO) term platform. Among the different 109 HPO terms, the main term was abnormality of the nervous system (31.1%) and, in this 110 category, the main HPO term was neurodevelopmental abnormality (37.2%). HPO terms from 111 70 index cases are summarized in figure 1A and 1B. 112

113 Methods

ES was performed in the healthy asymptomatic parents of all 70 patients, as previously 114 described⁴. Variant interpretation was first focused on *de novo* variants, homozygous or 115 compound heterozygous variants within OMIM-morbid genes and secondarily extended to all 116 other genes (non OMIM-morbid and non OMIM). Candidate variants were selected as 117 previously described and required further functional studies and/or reccurence³. Incidental and 118 secondary findings were not investigated. Candidate variants were confirmed with a second 119 independent method *i.e.* Sanger sequencing or real-time quantitative PCR (primers and 120 conditions available on request). Candidate variants were shared through international 121 collaborative platforms (GeneMatcher and MatchMaker Exchange⁶) to gather additional 122 patients and to strengthen genotype-phenotype correlations. 123

124

125 Results

126 Among the 70 patients, second-tier trio-based ES identified 37 potential causative variants in 32 different genes in 28/70 patients (40%). After data sharing and/or genotype-phenotype 127 128 interpretation, we confirmed 21 variants as strong candidates in 18 different genes in 18/70 129 patients (26%). We found 13 de novo occurrences and 5 variants with recessive inheritance, including 2 patients with a confirmed molecular diagnosis who also carried one or two 130 variants of unknown significance (VOUS) in other genes, suspecting double or triple hit 131 (Table 1). VOUS were also identified in 11 different genes (13 variants) in 10/70 patients 132 (14%) with a *de novo* occurrence in 4 patients, autosomal recessive inheritance in 5 patients, 133 and X-linked in 1 patient (Table 1, figure 1C). 134

Among the 18 highly suspected causal genes, 9 were OMIM-morbid genes that were not retained in the solo analysis (figure 1D). For most genes, the phenotype or the mode of inheritance were atypical compared to what was previously known: young age at first analysis for *ACO2* (MIM 100850); unilateral ear malformation for *OTX2* (MIM 600037);

developmental delay, gingival hypertrophy, aortic dilatation, bone anomaly, and intestinal 139 pseudo-obstruction episodes for KCNMA1 (MIM 600150); much more severe phenotype than 140 usually reported for MFSD8 (MIM 611124) (double or triple hit hypothesis) and atypical 141 inheritance for TOR1A (MIM 605204). Otherwise, a molecular diagnosis was obtained in 4 142 genes found to be disease-causing during the trio-ES reanalysis in patients presenting with 143 phenotypes similar to ours: KMT5B (MIM 610881), PBX1 (MIM 176310), PPM1D (MIM 144 605100) and UNC45A (MIM 611219), with a double hit hypothesis in the last case. The 9 145 other candidate genes were newly associated with a human phenotype or led to the initiation 146 of collaborative work: FBX011 (MIM 607871), HNRNPR (MIM 607201), IRF2BPL (MIM 147 611720), PIGK (MIM 605087), PTBP1 (MIM 600693), RALGAPB, SPEN (MIM 613484), 148 TRAF7 (MIM 606692) and ZMIZ1 (MIM 607159) (Table 1 and Figure 1D). The most 149 common type of variant was sporadic truncating (4 frameshift, 3 nonsense and 1 canonical 150 151 splice site; figure 1C, Table 1). The patient with pathogenic compound heterozygous variants in UNC45A (1 truncating and 1 missense) also harbored a heterozygous missense VOUS in 152 PBRM1 (MIM 606083). The patient with a pathogenic homozygous missense variant in 153 MFSD8 also harbored homozygous missense VOUS in WTIP (MIM 614790) and NUP98 154 (MIM 601021). One patient born from consanguineous parents had 2 homozygous VOUS, 155 one stop gain in CBLB and one missense in TRAPPC8. 156

Trio-ES did not identify any candidate variant in 42/70 patients (60%). Strategies for
detecting copy number variants (CNV) found no anomalies.

159

160 **Discussion**

161 This study presents the results of a two-step ES analysis strategy for the diagnosis and 162 exploration of the genetic origins of DD. In patients with negative solo-cES, we used a second step trio-based ES strategy which resulted in the identification of pathogenic variants in 26%of cases and VOUS and 14% of cases.

One of the advantages of second-tier trio-ES was that it led us to suspect the implication of 9 165 known OMIM-morbid genes, which, for various reasons, were initially not considered 166 pathogenic in the solo strategy. First, 3 of the genes (*PBX1*, *PPM1D* and *UNC45A*) were only 167 168 recently described in DD and were not related to OMIM-disorders at the time of solo-cES analysis⁷⁻⁹. Second, the lack of a phenotype-genotype correlation made variants interpretation 169 difficult in the solo strategy: an OTX2 heterozygous truncating variant (p.Gly250Aspfs*52) 170 was identified in a patient with unilateral ear anomalies associated with ID and without ocular 171 or pituitary anomalies when OTX2 variants are usually associated with syndromic 172 microphtalmia (MIM 610125)¹⁰. However, one report has highlighted the implication of 173 OTX2 in a similar clinical presentation¹⁰. Compound heterozygous missense variants in ACO2174 were identified in a young patient referred to our center at 12 months of age for DD and 175 176 severe hypotonia. Exome data were analyzed when he was 18 months, and the phenotype was enriched with the usual ophthalmological features when the patient consulted again at age 2 177 years 3 months. Patients with aconitase-2 deficiency usually display severe ID and/or optic 178 atrophy, but only few patients with compound heterozygous missense variants have been 179 described¹¹. Enzymatic activity testing confirmed variant pathogenicity, as previously 180 demonstrated in other patients. Third, an unusual mode of inheritance initially led to the 181 erroneous exclusion of pathogenic variants. While a TOR1A homozygous truncating variant 182 was identified in 2 neonatally-deceased siblings presenting with severe arthrogryposis and 183 184 respiratory distress, pathogenic TOR1A heterozygous variants have been previously involved only in an autosomal dominant dystonia with incomplete penetrance (MIM128100). Only 185 recently, homozygous truncating variants were implicated in severe arthrogryposis with 186 autosomal recessive transmission¹². 187

In addition, clinicians and molecular geneticists should be aware that patients can be affected 188 by multiple disorders which can complicate the interpretation. Pangenomic analyses promote 189 the identification of putative double/triple hit events. In our cohort, this means that 3/70 190 patients (4.3%) are potentially concerned, which coincides with previous publications (about 191 1.5 to 5%)¹³. The first example was a patient born to healthy consanguineous parents with 192 atypical severe early-onset epilepsy and microcephaly, carrying a pathogenic homozygous 193 MFSD8 missense variant, responsible for late-onset ceroid lipofuscinosis (MIM 610951). 194 195 Trio-ES analysis led us to suspect 2 homozygous missense variants in 2 potential candidate genes (WTIP and NUP98), which could account for this atypical correlation. The other 196 example was the patient with ID carrying biallelic UNC45A variants; ID was not reported as 197 part of the O2HE/O2CD syndrome, leading us to search for an additional explanation 198 (possible de novo PBRM1 heterozygous missense variant). 199

200 The use of the trio strategy after negative solo-cES yielded major gains for translational 201 research through the identification of genes highly likely to be involved in novel human disorders (10/70 patients - 14%) (Table 1). These genes, suspected because of compatible 202 familial segregation, were mostly de novo variants in sporadic disease (FBXO11, HNRNPR, 203 IRF2BPL, KCNMA1, PIGK, PTBP1, RALGAPB, SPEN, TRAF7 and ZMIZ1). Their 204 implication was strengthened by the replication of identical variants in patients with 205 overlapping phenotypes, and our cases were used to support the implication of 5 recently 206 published disease-causing genes (FBX011, HNRNPR, IRF2BPL, TRAF7 and ZMIZ1)¹⁴⁻¹⁸ and 207 a novel phenotype associated with $KCNMA1^{19}$. The second-tier trio-ES also focused on 13 208 209 additional VOUS in 11 different genes for 10 patients, with enrichment in missense variants (figure 1C). These remain VOUS seeing as no other affected patients were identified despite 210 international data sharing. Recommendations suggest that the gene-disease link in autosomal 211 212 recessive and dominant/sporadic diseases should be confirmed in 3 and 5 similarly affected

patients, respectively ²⁰. Moreover, despite *in-silico* scores suggesting pathogenicity, sporadic
missense variants often require more investigation before pathogenicity can be confirmed.

When we compared the variants in the 18 patients with causal variants and the 10 patients 215 with VOUS, we noted a bias towards de novo variants and likely disruptive variants in 216 positives patients, which were more prone to be interpreted as likely pathogenic. Altogether, 217 218 with the identification of pathogenic and highly candidate variants in 18/70 patients, the second step trio-ES strategy increased the initial diagnostic yield of solo-cES from about 30% 219 to 49% in our selected cohort. This confirms the results of the pilot study performed in 74 220 patients mainly suffering from DD (81-84%), which also underlined the interest of a second 221 step trio-based ES after solo-cES with a diagnostic yield of 36% considering only pathogenic 222 variants⁵. Interestingly, this publication reported a potential diagnostic increase of 11% for 223 genes newly identified in human disorders⁵, similar to the present results (15%). 224

The contribution of ES is now undisputed in the identification of causal variants in Mendelian 225 disorders. Nevertheless, questions remain about the most efficient approach in terms of 226 227 diagnostic yield and cost. One of the major advantages of the trio-cES strategy is that it speeds up the selection of candidate variants according to the different inheritance 228 possibilities, especially de novo and compound heterozygous variants¹. In addition, trio 229 analysis can easily be extended to non OMIM and OMIM non-disease-causing genes, leading 230 to identify putative candidate variants. Meanwhile, research analysis of solo-ES is time-231 232 consuming because of the numerous missense variants and the missense enrichment among 233 VOUS (Fig. 1D). The main disadvantage of trio-ES is the cost of additional parental ES. Nevertheless, this cost is expected to decrease and can also be counterbalanced since trio-ES 234 235 drastically reduces the biological (variant interpretation) and technical (primer design, sanger validation) costs, especially for the interpretation of missense variants. Therefore, a two-step 236 strategy with solo-cES followed by trio-ES in negative patients would lower costs by 30% 237

(*i.e.* the approximate diagnostic rate of solo-cES) and could thus provide a good compromisebetween diagnostic yield and economic concerns.

In conclusion, a second-step trio-based strategy after negative solo-cES is an efficient strategy for accelerating translational research and increasing diagnostic yield, with a good financial compromise. A novel approach involving first-tier trio-like ES, with pooled DNA from the parents of affected patients, could be used to reduce sequencing costs further while maintaining diagnostic power, but the efficiency of this strategy needs to be evaluated.

245

246 Acknowledgments

We thank the patients and their families for their participation. We thank the Integragen society (Evry) and CNRGH (CEA - Evry) for ES. This work was supported by grants from the Regional Council of Burgundy (Plan d'Actions Régional pour l'Innovation - PARI) and the "Fonds Européen de développement regional" (FEDER).

We also like Christian Bouwkamp 251 to thank Dr (252 Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands) and Miss Anniek Meesters (Department of Psychiatry, Erasmus University 253 Medical Center, Rotterdam, The Netherlands). 254

1 Deciphering Developmental Disorders Study. Prevalence and architecture of de novo
mutations in developmental disorders. Nature. 2017 Feb 23;542(7642):433-438.

258

2 Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in
children. Nat Rev Genet. 2018 May;19(5):253-268.

261

3 Sawyer SL, Hartley T, Dyment DA, et al. Utility of whole-exome sequencing for those near
the end of the diagnostic odyssey: time to address gaps in care. Clin Genet. 2016
Mar;89(3):275-84.

265

4 Nambot S, Thevenon J, Kuentz P, et al. Clinical whole-exome sequencing for the diagnosis
of rare disorders with congenital anomalies and/or intellectual disability: substantial interest
of prospective annual reanalysis. Genet Med. 2017 Nov 2.

269

5 Eldomery MK, Coban-Akdemir Z, Harel T, et al. Lessons learned from additional research
analyses of unsolved clinical exome cases. Genome Med. 2017 Mar 21;9(1):26.

272

6 Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for
connecting investigators with an interest in the same gene. Hum Mutat. 2015 Oct;36(10):92830.

276

7 Slavotinek A, Risolino M, Losa M, et al. De novo, deleterious sequence variants that alterthe transcriptional activity of the homeoprotein PBX1 are associated with intellectual

disability and pleiotropic developmental defects. Hum Mol Genet. 2017 Dec 15;26(24):48494860.

281

8 Jansen S, Geuer S, Pfundt R, et al. De Novo Truncating Mutations in the Last and
Penultimate Exons of PPM1D Cause an Intellectual Disability Syndrome. Am J Hum Genet.
2017 Apr 6;100(4):650-658.

285

9 Esteve C, Francescatto L, Tan PL, et al. Loss-of-Function Mutations in UNC45A Cause a
Syndrome Associating Cholestasis, Diarrhea, Impaired Hearing, and Bone Fragility. Am J
Hum Genet. 2018 Mar 1;102(3):364-374.

289

290 10 Gerth-Kahlert C, Williamson K, Ansari M, et al. Clinical and mutation analysis of 51
291 probands with anophthalmia and/or severe microphthalmia from a single center. Mol Genet
292 Genomic Med. 2013 May;1(1):15-31.

293

11 Srivastava S, Gubbels CS, Dies K, Fulton A, Yu T, Sahin M. Increased Survival and
Partly Preserved Cognition in a Patient With ACO2-Related Disease Secondary to a Novel
Variant. J Child Neurol. 2017 Aug;32(9):840-845.

297

12 Kariminejad A, Dahl-Halvarsson M, Ravenscroft G, et al. TOR1A variants cause a severe
arthrogryposis with developmental delay, strabismus and tremor. Brain. 2017 Nov
1;140(11):2851-2859.

301

302 13 Posey JE, Harel T, Liu P, et al. Resolution of Disease Phenotypes Resulting from
303 Multilocus Genomic Variation. N Engl J Med. 2017 Jan 5;376(1):21-31.

304

305 14 Gregor A, Sadleir LG, Asadollahi R, et al. De Novo Variants in the F-Box Protein
306 FBXO11 in 20 Individuals with a Variable Neurodevelopmental Disorder. Am J Hum Genet.
307 2018 Aug 2;103(2):305-316.

308

- 309 15 Duijkers FA, McDonald A, Janssens GE, et al. HNRNPR Variants that Impair Homeobox
 310 Gene Expression Drive Developmental Disorders in Humans. Am J Hum Genet. 2019 Apr 24.
 311
- 16 Tran Mau-Them F, Guibaud L, Duplomb L, et al. De novo truncating variants in the
 intronless IRF2BPL are responsible for developmental epileptic encephalopathy. Genet Med.
 2018 Aug 31.

315

316 17 Tokita MJ, Chen CA, Chitayat D, et al. De Novo Missense Variants in TRAF7 Cause
317 Developmental Delay, Congenital Anomalies, and Dysmorphic Features. Am J Hum Genet.
318 2018 Jul 5;103(1):154-162.

319

18 Carapito R, Ivanova EL, Morlon A, et al. ZMIZ1 Variants Cause a Syndromic
Neurodevelopmental Disorder. Am J Hum Genet. 2019 Feb 7;104(2):319-330.

322

19 Liang L, Li X, Moutton S, De novo loss-of-function KCNMA1 variants are associated
with a new multiple malformation syndrome and a broad spectrum of developmental and
neurological phenotypes. Hum Mol Genet. 2019 Jun 1. pii: ddz117.

326

- 20 Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Disease gene identification
- 328 strategies for exome sequencing. Eur. J. Hum. Genet. 20, 490–497 (2012).

329

330 **Online Database**

- 331 GeneMatcher: <u>https://www.genematcher.org/</u>
- 332 Genome Aggregation Database: <u>http://gnomad.broadinstitute.org/</u>
- 333 PubMed: <u>https://www.ncbi.nlm.nih.gov/pubmed/</u>
- 334 UCSC Genome Browser: <u>https://www.genome.ucsc.edu/</u>
- 335 PolyPhen-2 : <u>http://genetics.bwh.harvard.edu/pph2/</u>

Figure legends

Figure 1: A) Two-level partial sunburst showing the percentage of phenotype category of the first level according to HPO terms in the 70 included patients. Abnormality of the nervous system was the main HPO category of this cohort. B.) Three-level partial sunburst showing percentage of abnormality of the nervous system category according to HPO terms of the 70 included patients. C) Percentage and type of variants in the 28/70 patients. D) Percentage of the gene categories in the 28/70 patients. VOUS: variants of unknown significance.

Table 1: Molecular data from the 28 patients with pathogenic, candidate or variants of unknown significance. #: recently published OMIM-morbid genes; *: double or triple hit suspected; hm: homozygous: hmi: hemizygous; ht: heterozygous: htc: compound heterozygous; NA: not available.





(35%)

VOUS

	18 causal variants in the 18/70 patients	11 VOUS in the 10/70 patients
Sporadic truncating	39	0
Sporadic missense	22.3	36.3
Start loss	11.1	0
Compound heterozygous (truncating / missense)	11.1	9.1
Compound heterozygous (missense/ missense)	5.5	9.1
Homozygous missense	5.5	18.2
Homozygous truncating	5.5	18.2
X-linked	0	9.1

3.4. Implication de variants hétérozygotes perte de fonction du gène *KCNMA1* dans un syndrome malformatif avec déficit intellectuel et épilepsie (article 4)

Ce travail représente l'illustration de l'apport du séquençage de l'exome en trio pour améliorer le rendement diagnostique en accélérant la recherche translationnelle chez les patients atteints d'une forme sporadique d'anomalie du développement et/ou DI.

Introduction

La réinterprétation du séquençage d'exome en trio suite à un solo négatif a permis de mettre en évidence un variant faux-sens *de novo* du gène *KCNMA1* chez un patient atteint d'un syndrome malformatif avec retard du développement psychomoteur. Il présentait une mégavessie et une omphalocèle détectés aux échographies fœtales, des épisodes de pseudo-obstruction intestinale chronique, des dilatations artérielles au niveau aortique et des gros axes artériels (cérébraux, mésentériques, céliaque), une hypertrophie gingivale et une dysplasie osseuse avec hyperostose relativement diffuse et marquée au niveau des arcs costaux.

L'interrogation des données de séquençage d'exome de notre base de données a permis de trouver qu'une autre patiente suivie dans notre centre et ayant un tableau clinique similaire était porteuse du même variant. Il a pu être validé par Sanger avec confirmation du caractère *de novo* également.

Le partage international des données sur la plateforme Genematcher a permis d'être mis en contact avec 2 équipes proposant de collecter des patients pour décrire le phénotype associé et prévoir des analyses fonctionnelles :

- Laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Center for Human Genome Research, Cardio-X Institute, Huazhong University of Science and Technology, Wuhan, Hubei Province, P. R. China ;

- Department of Genetics and Genome Science, Case Western Reserve University, Cleveland, OH, USA; Center for Cardiovascular Genetics, Department of Molecular Cardiology, Lerner Research Institute, Department of Cardiovascular Medicine, Cleveland Clinic; Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH, USA.

Le canal BK est un canal K+ activé à la fois par le Ca2+ et par l'état de polarisation transmembranaire. Il consiste en un homotétramère de 4 sous-unités α codées par *KCNMA1* (10q22.3). Chaque sous-unité comprend 7 domaines transmembranaires (S0-S6) dans sa partie N-terminale, et 2 domaines RCK (regulator of conductance of K domains, RCK1 et RCK2) responsables de la sensibilité au calcium dans sa partie C-terminale ^{388,389}. Le domaine S4 correspond à la région sensible au potentiel de membrane et la région entre S5 - S6 joue le rôle de pore permettant le passage des ions K+. Les 4 domaines RCK1-RCK2 du tétramère forment une structure en anneau qui interagit avec la région sensible au voltage, permettant l'ouverture du pore. L'ouverture est activée soit par une dépolarisation soit par la [Ca2+] intracellulaire. Les propriétés physiologiques des canaux dépendent de l'interaction des sous-unités alpha et des sous-unités auxiliaires (optionnelles) comme β 1-4 ou γ 1-4 ^{388,389}. Ces canaux ont une expression large dans de multiples tissus dont le cerveau (différentes populations neuronales), les cellules musculaires lisses, les cellules ciliées de la cochlée et dans différentes cellules tumorales ³⁹⁰. Il joue un rôle dans de nombreux processus physiologiques comme la repolarisation du potentiel de membrane, le contrôle de l'excitabilité neuronale, la libération des neurotransmetteurs, le contrôle du tonus des cellules musculaires lisses, la régulation du potentiel de membrane des cellules ciliées cochléaires non associée à leurs propriétés de déclenchement de potentiels d'action, et l'immunité innée ^{391–397}. L'ARNm est aussi sujet à épissage alternatif qui contribue à une certaine diversité de protéines pouvant être produites.

En 2005, une mutation faux-sens hétérozygote du gène KCNMA1 responsables d'un gain de fonction (augmentation de l'activation du canal et de la sensibilité au Ca2+ entraînant une repolarisation plus rapide et plus importante sous l'effet de la dépolarisation) a été associée au phénotype d'épilepsie généralisée avec dyskinésies paroxystiques dans une grande famille multiplex (13 individus atteints) après analyse de liaison et séquençage des gènes candidats ³⁹⁸. Deux autres articles ont confirmé cette constatation avec des variants faux-sens de novo chez des patients avec ou sans dyskinésie et parfois associé à un retard neurodéveloppemental ^{399,400}; dans un cas, les tests fonctionnels ont montré une augmentation de l'ouverture du canal sans qu'il y ait d'anomalie de sensibilité au Ca2+ ⁴⁰⁰. Dans une autre étude, le séquençage d'exome combiné à la cartographie par homozygotie dans une famille dans laquelle 2 filles issues de parents consanguins et atteintes de retard neurodéveloppemental et d'épilepsie, avec atrophie cérébelleuse sévère a conduit à suspecter un variant prédit frameshift à l'état homozygote (pas d'études fonctionnelles)⁴⁰¹; les 2 arguments principaux de pathogénicité était la fréquence faible dans les bases de données avec absence de contrôles homozygotes et que le phénotype de la souris KO -/- partageait des signes cliniques similaires comprenant une dysfonction motrice sévère, une ataxie cérébelleuse sous la forme d'un tremblement d'intention, des enjambées courtes et un schéma de marche irrégulier ⁴⁰². Enfin, des variants faux-sens ont été rapport dans différents phénotypes épileptiques mais les analyses fonctionnelles ne montraient pas d'altération ce qui indique l'importance de confirmation fonctionnelle⁴⁰⁰, même s'il ne peut être exclu une divergence entre des résultats in vitro et une physiopathologie in vivo.

Méthodes

Suite à l'identification de patients porteurs de variants du gène *KCNMA1*, une étude a été menée pour décrire le spectre phénotypique associé et des analyses fonctionnelles ont été conduites pour explorer le mécanisme physiopathologique en cause. Pour cela, un plasmide contenant la séquence

codante de KCNMA1 marquée par un tag et où les mutations ont été introduites a été transfecté dans des cellules HEK293, ce qui a permis d'explorer l'efficacité de la transfection par Western blot et de tester les conséquences des mutations selon leur sensibilité au voltage et à la [Ca2+].

<u>Résultats</u>

Le partage international des données a permis de regrouper 7 patients additionnels porteurs de variants faux-sens. Chez un patient, il existait un statut hétérozygote composite (faux-sens p.(Cys413Tyr) et frameshift p.(Asn449fs)) hérité de parents asymptomatiques. Les variants étaient répartis dans les différents domaines fonctionnels de la protéine et un seul était récurrent p.Gly375Arg, associé au phénotype le plus sévère avec atteinte multisystémique. Les analyses fonctionnelles ont montré une perte de fonction des canaux BK en abolissant l'activation au voltage (variants p.(Ser351Tyr), p.(Gly356Arg), p.(Gly375Arg), p.(Asn449fs), p.(Ile663Val)) ou en la diminuant soit avec perturbation de la cinétique et activation pour des potentiels supérieurs (p.(Cys413Tyr) et p.(Pro805Leu)), soit sans altération de cinétique (p.(Asp984Asn)). Il a aussi pu être montré que le variant p.(Cys413Tyr) altérait la sensibilité au Ca2+. Ces données fonctionnelles n'ont pas permis d'établir des corrélations génotype phénotype solides puisque la sévérité du phénotype ne dépendait pas de l'importance d'altération de la fonction du canal.

Discussion

La diversité des manifestations cliniques peut s'expliquer par l'expression de KCNMA1 dans divers tissus. La perte de fonction du canal est associée à une diminution de prolifération et de minéralisation par les ostéoblastes et une diminution d'expression des marqueurs de différenciation de ces cellules ⁴⁰³. Tout comme pour les neurones, dans les cellules musculaires lisses, les canaux BK sont impliqués dans la régulation du potentiel de membrane au repos selon le taux de Ca2+ ce qui régule leur excitabilité ⁴⁰⁴. Ce rôle pourrait expliquer les signes digestifs, vasculaires et urinaires observés chez certains patients.

Différentes molécules ont été identifiées pour activer ou inhiber le canal (revue ⁴⁰⁵). L'acide isoprimarique est un activateur du canal BK qui améliore la mémoire des modèles murins de maladie d'Alzheimer en jouant sur la transmission synaptique et la potentialisation à long terme ^{389,406}. BMS-204352 ([3S]-[+]-[5-chloro-2-methoxyphenyl]-1,3-dihydro-3-fluoro-6-[trifluoromethyl]-2H-indol-2-one) améliore l'homéostasie glutamatergique au niveau hippocampique et traite les altérations de reconnaissance sociale, l'anxiété et la mémoire spatiale des modèles murins de syndrome de l'X fragile ⁴⁰⁷. Ces pistes thérapeutiques pourraient donc être intéressantes à tester in vitro puis éventuellement chez les patients porteurs de mutation perte de fonction de *KCNMA1*.



*Figure 43 : synthèse de l'expression du gène KCNMA1 dans les tissus humains, des mutations répertoriées et des corrélations génotype-phénotype*⁴⁰⁸.

Exploitation de ce travail : participation à la rédaction d'un article scientifique

Lina Liang#, Xia Li#, Sébastien Moutton#, Samantha A. Vergano#, Bertrand Isidor #, Anne de Saint-Martin#, Anna C. E. Hurst#, Yushuan Hu#, Olaf Bodamer#, Julien Thevenon#, Christina Y. Hung#, Bénédicte Gerard#, Adelaide Rega, Sophie Nambot, Daphné Lehalle, Yannis Duffourd, Christel Thauvin-Robinet, Laurence Faivre, Benjamin Cogné, Qiuyun Chen, Grazia M.S. Mancini,*, Antonio Vitobello* & Qing Kenneth Wang*. De novo loss-of-function KCNMA1 variants are associated with a new multiple malformation syndrome and a broad spectrum of developmental and neurological phenotypes. Hum Mol Genet. 2019 Jun 1. pii: ddz117. doi: 10.1093/hmg/ddz117. [Epub ahead of print]⁴⁰⁹. #, * : ces auteurs ont contribué au travail de façon égale.

doi: 10.1093/hmg/ddz117 Advance Access Publication Date: 15 May 2019 General Article

GENERAL ARTICLE

OXFORD

De novo loss-of-function KCNMA1 variants are associated with a new multiple malformation syndrome and a broad spectrum of developmental and neurological phenotypes

Lina Liang^{1,†}, Xia Li^{1,18,†}, Sébastien Moutton^{2,3,4,†}, Samantha A. Schrier Vergano^{5,†}, Benjamin Cogné^{6,†}, Anne Saint-Martin de^{7,†}, Anna C. E. Hurst^{8,†}, Yushuang Hu^{1,†}, Olaf Bodamer^{9,10,†}, Julien Thevenon^{2,3,4,†}, Christina Y. Hung^{9,†}, Bertrand Isidor^{6,†}, Bénédicte Gerard^{11,†}, Adelaide Rega¹², Sophie Nambot^{2,3,4}, Daphné Lehalle^{2,3,4}, Yannis Duffourd^{2,3,4}, Christel Thauvin-Robinet^{2,3,4}, Laurence Faivre^{2,3,4}, Stéphane Bézieau⁶, Leon S Dure¹³, Daniel C. Helbling¹⁴, David Bick¹⁴, Chengqi Xu¹, Qiuyun Chen¹⁵, Grazia M.S. Mancini^{16,*}, Antonio Vitobello^{4,*} and Qing Kenneth Wang^{1,15,17,*}

¹Key Laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Center for Human Genome Research, Cardio-X Institute, Huazhong University of Science and Technology, Wuhan, Hubei 430074, P. R. China, ²Centre de Référence Anomalies du Développement et Syndromes Malformatifs, Hôpital d'Enfants, Dijon 21079, France, ³Centre de Référence Déficiences Intellectuelles de Causes Rares, Hôpital d'Enfants, Dijon 21079, France, ⁴Inserm UMR 1231 GAD team, Genetics of Developmental Disorders, Université de Bourgogne Franche-Comté, Dijon 21070, France, ⁵Medical Genetics and Metabolism, Children's Hospital of The King's Daughters, Eastern Virginia Medical School, Norfolk, VA 23507, USA, ⁶Service de Génétique Médicale, CHU de Nantes, Nantes 44093, France, ⁷Neuropédiatrie, Centre de Référence des Epilepsies Rares, Hôpitaux Universitaires de Strasbourg, Strasbourg 67098, France, ⁸Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA, ⁹Division of Genetics and Genomics, Boston Children's Hospital/Harvard Medical School, Boston, MA 02115, USA, ¹⁰The Broad Institute of Harvard and MIT, Boston, MA 02115, USA, ¹¹Institut de Génétique Médicale d'Alsace, Laboratoires de Diagnostic Génétique, Unité de Génétique Moléculaire, Nouvel Hôpital Civil, Strasbourg 67000, Franc, ¹²Pediatric Radiologist, Département de Radiologie et Imagerie Diagnostique et Thérapeutique, CHU, Dijon 21079, France, ¹³Department of Pediatrics and Neurology, University of Alabama at Birmingham, Birmingham, AL 35233, USA, ¹⁴Clinical Services Laboratory, HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA,

 $^{^{\}dagger} \text{These}$ authors contributed equally to this work.

Received: February 4, 2019. Revised: April 17, 2019. Accepted: May 21, 2019

[©] The Author(s) 2019. Published by Oxford University Press. All rights reserved.

For Permissions, please email: journals.permissions@oup.com

¹⁵Department of Cardiovascular and Metabolic Sciences, Lerner Research Institute, Department of Cardiovascular Medicine, Cleveland Clinic, Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH 44195, USA, ¹⁶Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam 3015, The Netherlands, ¹⁷Department of Genetics and Genome Science, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA and ¹⁸Present address: Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109, USA

*To whom correspondence should be addressed at: Qing Kenneth Wang, Department of Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Ave, Cleveland, OH 44195, USA and Center for Human Genome Research, HUST, Wuhan 430074, China. Tel: (216) 4450570; Fax: (216) 4458204; Email: wangq2@ccf.org or qkwang@hust.edu.cn; Antonio Vitobello, Genetics of Developmental Disorders, Université de Bourgogne Franche-Comté, Dijon 21070, France. Tel: 03 80 39 32 38; Fax: 03 80 29 32 66; Email: Antonio.Vitobello@u-bourgogne.fr; Grazia Mancini, Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam 3015, The Netherlands. Tel: 31-10-7036915; Fax: 31-10-7043072; Email: g.mancini@erasmusmc.nl

Abstract

KCNMA1 encodes the large-conductance Ca²⁺ - and voltage-activated K⁺ (BK) potassium channel α-subunit, and pathogenic gain-of-function variants in this gene have been associated with a dominant form of generalized epilepsy and paroxysmal dyskinesia. Here, we genetically and functionally characterize eight novel loss-of-function (LoF) variants of KCNMA1. Genome or exome sequencing and the participation in the international Matchmaker Exchange effort allowed for the identification of novel KCNMA1 variants. Patch clamping was used to assess functionality of mutant BK channels. The KCNMA1 variants p.(Ser351Tyr), p.(Gly356Arg), p.(Gly375Arg), p.(Asn449fs) and p.(Ile663Val) abolished the BK current, whereas p.(Cys413Tyr) and p.(Pro805Leu) reduced the BK current amplitude and shifted the activation curves toward positive potentials. The p.(Asp984Asn) variant reduced the current amplitude without affecting kinetics. A phenotypic analysis of the patients carrying the recurrent p.(Gly375Arg) *de novo* missense LoF variant revealed a novel syndromic neurodevelopmental disorder associated with severe developmental delay, visceral and cardiac malformations, connective tissue presentations with arterial involvement, bone dysplasia and characteristic dysmorphic features. Patients with other LoF variants presented with neurological and developmental symptoms including developmental delay, intellectual disability, ataxia, axial hypotonia, cerebral atrophy and speech delay/apraxia/dysarthria. Therefore, LoF KCNMA1 variants are associated with a new syndrome characterized by a broad spectrum of neurological phenotypes and developmental disorders. LoF variants of KCNMA1 cause a new syndrome distinctly different from gain-of-function variants in the same gene.

Introduction

The large-conductance Ca^{2+} - and voltage-activated K^+ (BK) channel is a tetramer consisting of four α -subunits encoded by the KCNMA1 gene on chromosome 10q22.3. Each α -subunit spans 1236 amino acids and has an estimated molecular weight of 138 kDa (1–3). The α -subunit contains seven transmembrane domains (S0-S6) at its N terminus and a large C-terminal domain comprising of two RCK domains (regulator of conductance of K domains RCK1 and RCK2) responsible for calcium sensing through a high-affinity Ca^{2+} binding site (2,3). The S4 domain acts as the voltage sensor, and a region between S5 and S6 serves as the pore (2,3). The BK channel has a tetrameric structure composed of four pore-forming α -subunits, in which the four RCK1-RCK2 tandems form the 'gating ring' structure that is believed to interact with the voltage-sensing region, leading to the opening of the pore (2,3). The BK channel can be allosterically activated by both changes in the membrane voltage (voltage-dependent activation pathway) and intracellular [Ca²⁺] concentration (calcium-dependent activation pathway) (2,3). The physiological properties of BK channels are further influenced by the interaction of their α -subunits with optional auxiliary subunits such as β 1–4 subunits or γ 1–4 subunits (2,3).

The BK channel is expressed in many organs and tissues, including the postnatal central nervous system (2,3). It plays a pleiotropic role in many physiological processes, such as in the repolarization of the membrane potential, the control of neuronal excitability, neurotransmitter release, the control of smooth muscle tone, the tuning of hair cells in the cochlea and in innate immunity (4–8). Recently, the BK channel was found to be involved in the maintenance, migration and differentiation of mesenchymal stromal cell populations and in the SH-SY5Y neuroblastoma cell cycle and proliferation (9,10).

In 2005, BK channel abnormalities were linked to human disease for the first time through the identification of a KCNMA1 gain-of-function mutation, which causes a coexistent syndrome of generalized epilepsy and paroxysmal dyskinesia (GEPD) (1). By genome-wide linkage analysis of a large family with 13 affected individuals and through subsequent DNA sequence analyses, Du et al. (1) had found the p.(Asp434Gly) pathogenic variant in KCNMA1, which co-segregated with GEPD in the family. Patch-clamping studies showed that the p.(Asp434Gly) mutation increased the BK current by enhancing channel activation and increasing the calcium sensitivity of the BK channel (1). Recently, Li et al. (11) identified a de novo KCNMA1 variant p.(Asn995Ser) in two independent patients who presented with epilepsy but not with paroxysmal dyskinesia. The p.(Asn995Ser) variant also increased the BK current by enhancing the BK channel activation through increases in the single-channel open probability and single-channel open dwell time. However, the variant did not alter the calcium sensitivity of the channel (11). As both p.(Asp434Gly) and p.(Asn995Ser) variants enhance the BK channel current and activation, they elicit a



Figure 1. Identification and electrophysiological characterization of a recurrent *de novo* KCNMA1 variant p.(Gly375Arg) in the transmembrane segment S6 of the BK channel. (A) Pedigree of three families with genotyping data for the KCNMA1 variant p.(Gly375Arg) showing the *de novo* nature of the variant. The affected male and females are shown as filled squares or circles. Normal individuals are shown as empty symbols. (B) Schematic structure of the BK channel. The location of the p.(Gly375Arg) variant is indicated by an arrow. (C) The p.(Gly375Arg) variant occurs at an evolutionally conserved amino acid residue. (D) Electrophysiological characterization of the p.(Gly375Arg) variant. Representative macroscopic currents of WT and p.(Gly374Arg) mutant BK channels from inside-out patch experiments in the presence of 10 μ M Ca²⁺ using the protocol indicated at the top (n = 6-10/group).

gain-of-function effect (1,11). Tabarki et al. (12) reported two Saudi sisters with epilepsy, developmental delay and severe cerebellar atrophy carrying a homozygous 1 bp duplication [c.2026dupT, p.(Tyr676Leufs*7)] in KCNMA1. The variant was seen once in the Exome Aggregation Consortium (ExAC) database but was absent from the Genome Aggregation Database (gnomAD). No functional studies of the variant were performed. Zhang et al. (13) reported two Chinese boys with paroxysmal non-kinesigenic dyskinesia and developmental delay without generalized epilepsy but de novo heterozygous p.(Glu884Lys) and p.(Asn1053Ser) variants in KCNMA1, respectively. Functional studies of these variants were not performed. Li et al. (11) reported three additional KCNMA1 variants p.(Lys518Asn), p.(Asp656Ala) and p.(Asn1159Ser) in three unrelated patients with divergent epileptic phenotypes. However, functional studies established the variants as benign polymorphisms, indicating the necessity of functional characterization of KCNMA1 variants for accurate variant classification.

Here, we report both genetic and functional analyses of eight novel KCNMA1 variants, including six de novo variants, which were detected in eight unrelated patients and two compound heterozygous variants found to be present in one patient. Electrophysiological data suggest that all variants act through a loss-of-function (LoF) mechanism. Therefore, this study expands the mutational spectrum of KCNMA1-related disorders and describes a new clinical entity elicited by the recurrent LoF variant p.(Gly375Arg). This new syndrome is associated with developmental delay, visceral and cardiac malformations, epilepsy, connective tissue symptoms with arterial involvement, bone dysplasia and characteristic dysmorphic features in three independent patients. Furthermore, this research offers comprehensive functional analyses of the largest cohort of BK channel LoF variants and reveals the novel pathogenic potential of KCNMA1 LoF variation in human disease.

Results

Identification of a new syndrome associated with a recurrent *de novo* missense variant in KCNMA1

Exome sequencing identified a *de novo* KCNMA1 missense variant [NM_002247.3: c.1123G>A, p.(Gly375Arg)] in three unrelated families (Fig. 1A). The p.(Gly375Arg) variant is located in the S6 transmembrane domain of the BK channel (Fig. 1B) and occurs at an amino acid residue, which is highly conserved among different species during evolution (Fig. 1C). This variant is absent from gnomAD and has a single submission as a VUS (variant of uncertain significance) in both ClinVar (RCV000623526.1) and the denovo-db database (denovo-db.gs.washington.edu, sample DDD4K.008888, associated with developmental disorder). Patchclamping experiments revealed that the p.(Gly375Arg) variant abolishes the function of the BK channel and blocks generation of potassium current (Fig. 1D), suggesting a LoF effect of this variant.

The p.(Gly375Arg) variant was first identified in a 2-yearold boy (Patient 1 in Table 1 and Supplementary Material, Fig. S1) presenting with a polymalformation syndrome characterized by severe developmental delay, visceral and cardiac malformations, connective tissue symptoms with arterial involvement, generalized seizures and dysmorphic features including coarse facial features, macroglossia with two-lobed tongue, gingival hypertrophy, synophrys, hypertelorism, down-slanting palpebral fissures, broad nasal root, thin superior lip, wide mouth with downturned corners, everted lower lip, frontal hemangioma, mild hirsutism and skin infiltrations. Strabismus and vertical nystagmus were also noted at clinical examinations (Table 1). At the age of 3 years and 4 months, his sitting position was still unstable/unsteady. He could walk some steps with support, the tone progressively improved, the sleep quality was better after amygdala and adenoid surgery improved babbling at the same time. There was no feeding difficulty. At the age of 3 years and 6 months, snoring recurrence, decreased vocal production and tonsil hypertrophy led to another amygdala and adenoid surgery. In the post-surgery period, it was not possible to remove ventilation support due to severe laryngomalacia. Five days after the surgery, he presented with degradation of respiratory functions, occlusive syndrome and oligo-anuria. Palliative measures were established in accordance with parent's choice and after approval of the local ethics committee. He deceased at the sixth day after surgery.

Similar clinical features were observed in a second unrelated patient (Patient 2 in Table 1 and Supplementary Material, Fig. S1), who was previously identified to carry the same p.(Gly375Arg) variant by singleton exome sequencing by the same diagnostic laboratory. Subsequent Sanger sequencing revealed the de novo occurrence of this variant also in Patient 2. She is a 27-yearold female patient, who has been followed since birth for developmental delay, intellectual disability (ID), speech delay, severe axial hypotonia, mild cerebellar and cerebral atrophy identified by magnetic resonance imaging (MRI), absence epilepsy, megalocornea, intestinal atresia and malrotation, patent ductus arteriosus and dysmorphic features (Table 1). Patients 1 and 2 have undergone multiple imaging studies [X-rays, computer tomography (CT) scans and brain MRIs] during their clinical care, which show generalized bone thickening, sclerosis and dysplasia, bowel loop dilatation during episodes of intestinal pseudoobstruction, and dilatation, ectasia and tortuosity of large and medium-sized arteries in both individuals (Table 1; Supplementary Material, Fig. S2). A medical record review of Patient 2 identified additional clinical symptoms such as chronic Staphylococcus epidermidis bacteremia and frequent fevers. A CT scan at age 26 years documented hepatomegaly (20 cm), splenomegaly (17 cm) and several vertebrae collapses (Table 1; Supplementary Material, Fig. S2). A cardiac ultrasound at age 26 years measured an enlarged ascending aorta diameter of 38-39 mm with no evidence of an infectious endocarditis (Table 1). Through the international Matchmaker Exchange initiative (14), we identified one additional unrelated female patient bearing the same de novo p.(Gly375Arg) variant who presented with an overlapping phenotype (Patient 3 in Table 1; Supplementary Material, Fig. S1). Together, these data suggest that this recurrent de novo p.(Gly375Arg) LoF variant of KCNMA1 is associated with a multiple malformation syndrome.

De novo variants in the pore domain also abolish BK channel function

Two additional *de novo* missense variants, p.(Ser351Tyr) and p.(Gly356Arg), were identified in the pore domain of the BK channel (Fig. 2A and B). Patient 4, who carries the p.(Ser351Tyr) variant, is a 10-year-old female patient with ataxia, tremor, apraxia, hypertelorism, bilateral medial deviation of the distal phalanges of the third finger and lateral deviation of the toes (Table 1; Supplementary Material, Fig. S3). Brain MRI did not detect any structural abnormalities in this individual. She is not affected by seizures or paroxysmal dyskinesia (Table 1). Patient 5 carries the p.(Gly356Arg) variant and is a 12-year-old male affected by mild cognitive delay, axial hypotonia, ataxia and dysarthria (Table 1). Serial MRI detected cerebellar atrophy. The patient does not exhibit epilepsy.

Both p.(Ser351Tyr) and p.(Gly356Arg) variants occur at amino acid residues, which are highly conserved among different species and are located in a motif important for the selectivity for potassium (Fig. 2C). Patch-clamp recordings showed that BK channels with either p.(Ser351Tyr) or p.(Gly356Arg) variants fail to elicit any potassium current under voltage stimulus ranging from -160 mV-160 mV (Fig. 2D). These data suggest that the p.(Ser351Tyr) and p.(Gly356Arg) variants are also LoF variants.

Electrophysiological characterization of compound heterozygous variants p.(Asn449fs) and p.(Cys413Tyr) identified in a patient

We identified two compound heterozygous variants in a patient (Patient 6 in Table 1). Both, the maternally inherited missense variant p.(Cys413Tyr) and the frameshift variant p.(Asn449fs), which is inherited from the father, are located in the C-terminal RCK1 domain (Fig. 3A and B) important for the regulation of K conductance. The patient presented with multiple congenital abnormalities, developmental delay, ID, axial hypotonia, ataxia, generalized mild cerebral atrophy documented by MRI and strabismus (Table 1). Both parents are clinically asymptomatic at the present time. The p.(Cys413Tyr) variant occurs at an amino acid residue, which is highly conserved among different species during evolution (Fig. 3C), and markedly reduces the amplitude of the BK current and shifts the G-V curve to the positive voltage direction by 38 mV at the 10 µM calcium concentration (Fig. 3D). The mean macroscopic current amplitude of the p.(Cys413Tyr) mutant channel was much smaller than that of the wild-type (WT) channels (n = 11 for mutant and n = 6 for WT) (Fig. 3D). Thus, the p.(Cys413Tyr) variant is also a LoF variant that inhibits the function of the BK channel substantially. To test the calcium dependency of the p.(Cys413Tyr) mutant channel, we further analyzed the G-V relationship at 1 μM and nominal 0 μM calcium concentrations and found that the G-V curves of the mutant channels also shifted to the positive voltage direction by 26.2 mV and 33.7 mV, respectively (Fig. 3D). Thus, the p.(Cys413Tyr) variant likely reduces the activation and macroscopic current amplitude of the BK channel significantly at calcium concentrations ranging from nominal 0 µм–10 µм.

The frameshift variant p.(Asn449fs) did not elicit any potassium current under the voltage stimulus from -160 mV-160 mV at a 10 μ M calcium concentration (Fig. 3D). These data suggest that the p.(Asn449fs) variant is also a LoF variant.

De novo variant p.(Ile663Val) in the RCK1 domain abolishes the function of the BK channel

KCNMA1 variant p.(Ile663Val) was identified in a female patient (Patient 7) who suffers from developmental delay, ID, axial hypotonia, ataxia and strabismus (Table 1). The variant was found to occur de novo and is located in the RCK1 domain of the BK channel (Fig. 4A and B). This substitution occurs at an amino acid residue, which is highly conserved among different species during evolution (Fig. 4C). Patch-clamp recordings showed that p.(Ile663Val) mutant BK channels failed to elicit any potassium current under the voltage stimulus from -160 mV-160 mV (Fig. 4D). Interestingly, western blot analyses revealed that the p.(Ile66Val) mutated BK channel moved much slower than the WT or other mutant channels through the gel (Fig. 4E). The 3D structure of the BK channel shows that the amino acid Ile663 is a hydrophobic residue buried inside the structure. Thus, the substitution of isoleucine by valine at codon 663 may alter the 3D structure of the BK channel dramatically, which may explain the altered migration pattern on the western blot. These data suggest that the p.(Ile663Val) variant is a functional variant, which leads to LoF of the BK channel.

Patient ID	4	5	1	2	ŝ	9	7	8	6
Genomic variant	p.Ser351Tyr	p.Gly356Arg	p.Gly375Arg	p.Gly375Arg	p.Gly375Arg	p.Asn449fs/ p.Cys413Tyr	p.Ile663Val	p.Pro805Leu	p.Asp984Asn
De novo variant	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Consanguinity	No	No	No	No	No	No	No	No	No
Gender	F	Μ	М	Ъ	Ъ	M	Ь	M	М
Gestational Age	40 + 1/7 w	At term	36 w	37 w	40 w	At term	NA	40 w	40 w
(WG) Birth						(40 + 4/7 w)			
parameters									
Weight	3.2 kg (30%)	NA	3.80 kg (98%)	2.95 kg (50%)	3.97 kg (93%)	3.640 kg (60%)	2.55 kg	5.58 kg (99.9%)	2.13 kg (0.1%)
Length	53 cm /90%)	NA	48 cm (55%)	50 cm (85%)	48.5 cm /36%)	NA	NA	55 cm (99%)	50 cm (30%)
OFC	NA	NA	35 cm (85%)	33.5 cm	35 cm (83%)	NA	NA	35.5 cm	34 cm (20%)
				(20%)				(55%)	
Growth parameters	10 y	10 y	3 y 4 m	21 y	12 y	2 y	3 y 10 m	4 y	4 y
(age)									
Weight	32.3 kg (39%)	%06	13 kg (20%)	46 kg (25%)	26.7 kg (0.4%)	75%	12.2 kg (5%)	35.1 kg (99.99%)	19.3 kg (97.5%)
Height	138 cm	75%	93 cm (30%)	151 cm	136 cm	75%	92 cm (5%)	117.3 cm	109 cm
	(42%)	2		(2.5%)	(2.12%)	2		(%6.66)	(97.5%)
OFC	53 cm (62%)	50%	50 cm (45%)	56 cm (15%)	NA	50%	46.5 cm (0.1%)	52.4 cm (80%)	50 cm (70%)
Facial	Yes	No	Yes	Yes	Yes	Mild brachy-	No	No	No
dysmorphism						cephalic			
Synophrys	I		+	I	I	I			
Hypertelorism	+ (Inner		+	+	+	I			
	cantrius 95%ile)								
Down-slanting nalnehral	I		+	+	+	+ (Short nalnehral			
fissures						fissures,			
						epicanthic			
-						(snini			
Broad nasal root	I		+	+	(+)	+			
									(Continued)

Table 1. Continued									
Patient ID Genomic variant	4 p.Ser351Tyr	5 p.Gly356Arg	1 p.Gly375Arg	2 p.Gly375Arg	3 p.Gly375Arg	6 p.Asn449fs/	7 p.Ile663Val	8 p.Pro805Leu	9 p.Asp984Asn
						p.Cys413Tyr			
Thin superior	1		+	+	(+)	+			
чир тт: 1									
with	I		+	+	I	I			
downturned									
corners									
Prominent	I		NA	NA	+	I			
incisors									
Diastemas of the teeth	Ι		NA	NA	+	I			
Everted inferior	I		+	No	+	I			
up Miccourl	M.C.	NIC	Control of the office	Moderate	Mild contin	M.O.	NO	Obcoite.	NI.
Visceral	No	NO	severe aortic	Moderate	Mild aortic	No	No	Obesity	No
malformations			root dilatation, omphalo- cele, megabladder	aortic root dilatation, common mesentery, umbilical	root dilatation megacystis				
Eve/hearing	NO	NO	Strahismus	Megalocornea	NO	Strahismus	Strahismus/right	NO	NO
features		2	nystagmus				macular coloboma, left sen- sorineural deafness	2	
Neurological/ developmental features									
Sitting age	8 m	8.5 y	NA	2 y	2.5 y	NA	NA	Unknown	7 m
Walking age	20 m	NA	NA	NA	No	NA	NA	15 m	12 m
Speech delay	Yes,	Yes,	NA	Yes	Few words	NA	24 m	Severe,	Severe
	apraxia	dysarthria, No ID: TIQ						apraxia	
DD/ID	Mild	85, Yes	Severe	Severe	Severe	Yes	Yes	Yes	Yes
Axial	Mild	No	Mild	Severe	Severe	Yes	NA	No	No
hypotonia									1
Dystonia	No	Yes +++	No	No	No	No	NA	No	No
									(Continued)

Table T. Condition									
Patient ID	4	Ŋ	1	2	3	9	7	8	6
Genomic variant	p.Ser351Tyr	p.Gly356Arg	p.Gly375Arg	p.Gly375Arg	p.Gly375Arg	p.Asn449fs/ p.Cys413Tyr	p.Ile663Val	p.Pro805Leu	p.Asp984Asn
Ataxia	Yes	Spasticity of the lower extremi- ties	No	oZ	No	Yes	Yes	N	QN
Others	Tremor, dysmetria		No	Areflexia	No	No (normal DTR)			Autistic features
Epilepsy details	°N	No seizures, no EEG	Ŷ	Absence	Absence	°Z	AN	Ŷ	Generalized and focal dystonic seizures, status epilepticus
Anticonvulsant treatment		No		Sodium valproate (depakine)	Lamotrigine				CLB, ESM, LTG, ZNS, VPA, PB, LCM
Cerebral MRI	Normal	Progressive atrophy of cerebellar hemi- spheres and vermis	Normal	Mild cerebellar and cerebral atrophy	Normal	Generalized mild cerebral atrophy, thin corpus callosum, mild atrophy of cerebellar vermis	Normal	White matter signal abnormality frontal lobes, pineal cyst	Normal
EEG features	NA	NA	NA	Normal	Normal	Normal	NA	Normal	Variable foci Global EEG slowing
									(Continued)

Continued Table 1

Genomic variant p.St	er351Tyr	5 p.Gly356Arg	1 p.Gly375Arg	2 p.Gly375Arg	3 p.Gly375Arg	6 p.Asn449fs/ p.Cys413Tyr	7 p.Ile663Val	8 p.Pro805Leu	9 p.Asp984Asn
Additional Mer clinical features dev fing pha latt dev (esf hall toe ove toe bila	dial third ger distal ger distal diange trally. fiation of toes toes a 3 terally terally	Ž	Neonatal diabetes, laryngoma- lacia, feeding difficulties, kyphosis thick ribs	High-arched palate, short neck, hypertelorism, long and smooth philtrum, massive and progressive gingival hypertrophy, abnormal dental implantation, mandibular exostosis, macroglossia, abnormal bowel motility (frequent digestive occlusions, gastroparesis), recurrent pyelonephritis, bilateral valgus foot deformity, elbows flessum, genu valgun, dorsal kyphosis, fingers hypertrophy hypertrophy	Abnormal bowel motility (chronic use of flagyl), megacystis (diagnosed prenatally), recurrent urrinary tract infections and chronic retention, gingival hyperplasia required four reduction surgeries	Ž	Congenital torticoli	9 _Z	°Z

Table 1. Continued



Figure 2. Identification and electrophysiological characterization of KCNMA1 variants p.(Ser351Tyr) and p.(Gly356Arg) in the pore region of the BK channel. (A) Pedigree structure and genotyping data for each KCNMA1 variant showing the *de novo* nature of the variants. (B) Schematic structure of the BK channel with the two variants indicated by arrows. (C) The two variants occur at an evolutionarily conserved amino acid residues. (D) Electrophysiological characterization of the two variants. Representative macroscopic currents of WT and mutant BK channels with p.(Ser351Tyr) and p.(Gly356Arg) from inside-out patch experiments in the presence of 10 μM Ca²⁺ using the protocol indicated at the top (n = 6–10/group).



Figure 3. Identification and electrophysiological characterization of compound heterozygous KCNMA1 variants p.(Asn449fs) and p.(Cys413Tyr) found in a patient. (A) Pedigree structure of the family and genotyping data for two KCNMA1 variants. The affected male is shown as a filled square. Individuals with uncertain phenotype or without medical examination are shown as slashed symbols. (B) Schematic structure of the BK channel with the two variants in the RCK domains indicated with arrows. (C) The Cys413 residue occurs at an evolutionarily conserved amino acid residue. (D) Electrophysiological characterization of variants p.(Asn449fs) and p.(Cys413Tyr). Representative macroscopic currents of WT and mutant BK channels with p.(Asn449fs) and p.(Cys413Tyr) variants from inside-out patch experiments in the presence of 10 μ M Ca²⁺ using the protocol indicated at the top. G-V curves of WT and p.(Cys413Tyr) mutant BK channels at nominal 0 μ M Ca²⁺, 1 μ M Ca²⁺ and 10 μ M Ca²⁺. All G-V curves are fitted by Boltzmann function (solid lines) with V_{1/2} and slope factor at nominal 0 μ M Ca²⁺ [149.0 ± 9.5 mV, 17.9 ± 1.9 for WT and 182.7 ± 5.2 mV, 19.4 ± 2.0 for p.(Cys413Tyr)] and at 10 μ M Ca²⁺ [31.7 ± 4.1 mV, 13.2 ± 1.3 for WT and 146.9 ± 12.6 mV, 14.2 ± 2.5 for p.(Cys413Tyr)] and at 10 μ M Ca²⁺ [31.7 ± 4.1 mV, 13.2 ± 1.3 for WT and 70.2 ± 9.6 mV, 12.6 ± 2.0 for p.(Cys413Tyr)]. The data are presented as mean ± SD (n = 6-11/group).



Figure 4. Identification and electrophysiological characterization of KCNMA1 variant p.(Ile663Val) in the RCK1 domain of the BK channel. (**A**) Pedigree structure and genotyping data for KCNMA1 variant p.(Ile663Val) showing the *de novo* occurrence of the variant. (**B**) Schematic structure of the BK channel with the location of variant p.(Ile663Val) indicated by an arrow. (**C**) The p.(Ile663Val) variant occurs at an evolutionarily conserved amino acid residue. (**D**) Electrophysiological characterization of variant p.(Ile663Val). Representative macroscopic currents of WT and p.(Ile663Val) mutant BK channels from inside-out patch experiments in the presence of 10 μM Ca²⁺ using the protocol indicated at the top (*n*=12–15/group). (**e**) Western blot analysis showing that variant p.(Ile663Val) altered the mobility of the BK channel through SDS-polyacrylamide gels. Tubulin was used as a loading control.

De novo variants in the RCK2 domain markedly reduce the amplitude of the BK current

Two KCNMA1 variants were identified at two evolutionarily highly conserved amino acid residues in the RCK2 domain of the BK channel (Fig. 5A–C). KCNMA1 variant p.(Pro805Leu) was found *de novo* (Fig. 5A) in a male patient (Patient 8) presenting with severe speech delay, development delay, ID, apraxia and abnormal cerebral MRI (Table 1). The p.(Pro805Leu) variant markedly reduces the amplitude of the BK current (Fig. 5D). A western blot analysis showed that p.(Pro805Leu) significantly reduced the expression level of the BK channel (Fig. 5E). Moreover, the p.(Pro805Leu) variant shifts the G-V curve to the positive voltage direction (Fig. 5D) similar to p.(Cys413Tyr). Thus, the p.(Pro805Leu) variant is also a LoF variant.

KCNMA1 variant p.(Asp984Asn) is also a *de novo* variant (Fig. 5A), which was identified in a male patient (Patient 9) affected with non-syndromic moderate ID, severe language impairment, pharmacoresistant multifocal epilepsy but without paroxysmal dyskinesia (Table 1). The p.(Asp984Asn) variant occurs at an amino acid residue, which is highly conserved among different species during evolution (Fig. 5C). Patch-clamp recordings showed that the p.(Asp984Asn) variant markedly reduces the activation of the BK channel by decreasing the mean macroscopic current amplitude of BK potassium current (Fig. 5D). In contrast to the p.(Cys413Tyr) and p.(Pro805Leu) variants, the p.(Asp984Asn) variant does not alter the G-V curves (Fig. 5D). These data suggest that the p.(Asp984Asn) variant is a variant that leads to LoF of the BK channel.

Discussion

In this study, we identified and characterized eight LoF variants of KCNMA1, including p.(Ser351Tyr) and p.(Gly356Arg) in the pore region; p.(Gly375Arg) in the S6 transmembrane

domain; p.(Asn449fs), p.(Cys413Tyr) and p.(Ile663Val) in RCK1; and p.(Pro805Leu) and p.(Asp984Asn) in RCK2. None of the variants have previously been reported in public databases such as the ExAC, gnomAD, 1000Genomes and Exome Variant Server. The p.(Gly375Arg) variant was reported once as VUS in ClinVar and in the denovo-db database (denovo-db.gs. washington.edu). All eight variants showed profound effects on BK channel function exerting their effects through a LoF mechanism (Fig. 1-5). This study presents in-depth functional electrophysiological characterization of the largest cohort of LoF variants identified in the BK channel to date. The deleterious effects of the eight identified variants on the BK channel suggest that the affected amino acid residues are indeed critical to the structure and function of the BK channel. In addition, the observed impairments of the BK channel function in the patch-clamping experiments on missense variants in RCK1 [p.(Cys413Tyr) and p.(Ile663Val)] as well as in RCK2 [p.(Pro805Leu) and p.(Asp984Asn)] suggest that both domains (RCK1 and RCK2) are critical to the generation and conductance of BK potassium currents

Among patients with LoF variants in the BK channel, six patients shared the clinical features of ID and developmental delay, four individuals presented with ataxia and axial hypotonia and three affected were found to have cerebral atrophy (Table 1). These neurological and developmental abnormalities may be related to the roles of BK channels in calcium influx and neurotransmitter release. The etiology of ID mainly involves four biological functions, including (1) presynaptic vesicle cycling and transport, (2) cytoskeleton dynamics, (3) cell-adhesion and transsynaptic signaling and (4) protein degradation and turnover (15). Specifically, presynaptic vesicle cycling and freats the electrical signal transmission across synapses. On rapid timescales, which are relevant to information processing, the



Figure 5. Identification and electrophysiological characterization of KCNMA1 variants p.(Pro805Leu) and p.(Asp984Asn) in the RCK2 domain of the BK channel. (A) Pedigree structure and genotyping data for KCNMA1 variants p.(Pro805Leu) and p.(Asp984Asn) showing the *de novo* nature of the variants. (B) Schematic structure of the BK channel with the location of the two variants p.(Pro805Leu) and p.(Asp984Asn) indicated by arrows. (C) The p.(Pro805Leu) and p.(Asp984Asn) variants occur at an evolutionally conserved amino acid residue. (D) Electrophysiological characterization of variants p.(Pro805Leu) and p.(Asp984Asn). Representative macroscopic currents of WT and mutant BK channels with variants p.(Pro805Leu) and p.(Asp984Asn) from inside-out patch experiments in the presence of 10 μ M Ca²⁺, 1 μ M Ca²⁺ and 10 μ M Ca²⁺. The G-V curves are fitted by Boltzmann function (solid lines) with V_{1/2} and slope factor at nominal 0 μ M Ca²⁺ [195.6± 15.8 mV, 25.0± 2.5 for WT and 238.2± 7.9 mV, 17.4±2.9 for p.(Pro805Leu)], at 1 μ M Ca²⁺ [124.9± 15.3 mV, 13.5±3.1 for WT and 17.3.1±12.9 mV, 15.2± 3.3 for p.(Pro805Leu)] and a 10 μ M Ca²⁺ [41.2± (27.4± 6.3 mV, 14.7±2.1 for WT and 60.5± 9.6 mV, 15.4±2.2 for p.(Pro805Leu)]. The data are presented as mean ± SEM (n=7-18/group). Bottom left panel: the G-V curves of WT and p.(Asp984Asn) mutant BK channels are shown at 10 μ M Ca²⁺ [41.2± 10.5 mV, 13.5±1.9 for WT and 41.9± 12.6 mV, 13.7±1.7 for p.(Asp984Asn)]. The data are presented as mean ± SEM (n=7-18/group). (e) Western blot analysis showing the significant effect of variant p.(Pro805Leu) on the BK channel expression level (P=0.0081, n=3/group). Tubulin was used as a loading control.

release of neurotransmitters is predominantly determined by the shape, frequency and pattern of presynaptic action potentials (16). Action potential duration is an especially important determinant of neurotransmitter release and the amount of presynaptic calcium influx, which translates approximately into the fourth power to the release magnitude (16). Modulation of action potential duration thus represents a precise and powerful mechanism to control and regulate neurotransmitter release, which is primarily controlled by the activity of voltage-gated K+ channels (16). Furthermore, BK channels are among the major determinants of action potential duration during repetitive activity in central neurons, owing to their activation being both voltage and calcium regulated (17). In addition, BK channels are responsible for the repolarization of the action potential (17). LoF of BK channels can slow the repolarization process and prolong the action potential duration, which may increase the calcium influx and neurotransmitter release. Therefore, we hypothesize that the molecular pathogenesis of ID associated with BK channel LoF variants is mediated by abnormal neurotransmitter release.

It is also noteworthy that pathogenic CRBN variants associated with mild to severe ID (18,19) were thought to play a role in the assembly and surface expression of the BK channel (20,21). These pathomechanistic links underscore the association of KCNMA1 LoF variants to ID, which can also easily be observed in the clinical presentation of the described individuals in this study.

Five of the eight patients with LoF variants in the BK channel also shared clinical features of speech delay/dysarthria/apraxia and strabismus. These findings may implicate a role of the BK channel in neuronal control of speech and extraocular muscles or in developmental abnormalities of the same and other muscles. Collectively, the clinical data from the eight patients suggest a remarkable phenotypic variability associated with LoF variants of the BK channel. The patients carrying the p.(Gly375Arg) missense variant located in the transmembrane segment S6 present with a particularly severe phenotype including recurrent episodes of intestinal pseudoobstruction, arterial dilations and tortuosities as well as bone thickening and dysplasia. No other obvious genotype-phenotype correlation could be established between the location of variants and the presenting clinical features. The additional manifestations associated with the p.(Gly375Arg) variant could be attributed to the role of BK channels in tissues other than neurons. For instance, LoF of the BK channel was found to decrease the ability of proliferation and mineralization of osteoblasts through decreased expression of osteoblast differentiation marker genes (22). Moreover, the BK channel is also involved in smooth muscle physiology. It regulates the resting membrane potential through its activation by Ca²⁺ and hyperpolarizes plasma membranes as it does in neurons. Thus, BK channels regulate neuronal and smooth muscle excitabilities by limiting cell depolarization and contractions (23). In line with this, the BK channel is found to be a target of flavonoid naringenin, accounting for its relaxant effect on rat colonic smooth muscle contractility (23). The observed arterial dilation could be explained by KCNMA1 expression in smooth muscle cells of the aorta (24) and its involvement in blood pressure control, although existing data are spare (25). Subsequently, other organs in which smooth muscle cell function plays a major role may be affected as well. For instance, BK channels were found to also be crucial for urinary bladder function by controlling its filling-emptying cycle (26).

We noticed a high phenotypic heterogeneity among patients carrying KCNMA1 LoF variants (Table 1), which may be caused by variable expressivity, reduced penetrance or by the longitudinal evolution of the clinical phenotype and of different clinical features. Moreover, the p.(Asn449fs) frameshift variant was inherited from the healthy father (Fig. 3A) and thus appears to be a hypomorphic allele, which on its own is unlikely to be sufficient to cause a clinical phenotype and therefore has to occur in combination with another hypomorphic LoF variant [in our patient p.(Cys413Tyr)] to elicit a clinical presentation. These data are consistent with a previous report, which described two affected sibs with an autosomal recessive disorder consisting of generalized hypotonia, severe delayed milestones and myoclonic seizures beginning at 12 months of age. The seizures in those siblings were responsive to therapy in one and evolved to a Lennox-Gastaut pattern in the other. Brain MRI showed non-progressive severe cerebellar atrophy, vermis atrophy (more pronounced in the superior part) and slight paucity of the periventricular white matter in (12). A homozygous frameshift variant p.(Tyr676Leufs*7) was identified in the two sibs, but no functional studies were performed (12). As the heterozygous parents of these two affected sibs were healthy, the frameshift variant p.(Tyr676Leufs*7) only causes clinical disease in a homozygous but not in a heterozygous state. This raises an interesting possibility of a different pathophysiological mechanism between predicted truncating and haploinsufficient variants and missense variants. It is however also possible that some KCNMA1 missense LoF variants confer a higher degree of penetrance by a dominant negative mechanism by impacting multiple subunits and binding proteins of the BK channel α-subunit (2,20).

Interestingly, some of the clinical features associated with the p.(Gly375Arg) variant are reminiscent of Zimmermann– Laband syndrome (ZLS), which is characterized by facial dysmorphism with gingival enlargement, hypoplasia or aplasia of nails and terminal phalanges, hypertrichosis, join hyperextensibility, hepato (spleno) megaly and ID with or without epilepsy and caused by heterozygous *de novo* gain-of-function mutations in the voltage-gated K+ channel Eag1 (Kv10.1) encoded by the KCNH1 gene (27). The clinical features shared by ZLS and the syndrome described here to be associated with the p.(Gly375Arg) variant suggest existing interplay between Kv10.1 and BK channels in tissues where both contribute to the regulation of potassium homeostasis and current in response to stimuli.

The KCNMA1 gene appears to harbor some hotspots for recurrent *de novo* variants. In particular, the p.(Gly375Arg) LoF variant described in this study occurred *de novo* in three independent patients (Table 1). We previously reported another *de novo* variant in KCNMA1, p.(Asn995Ser), which is a gain-of-function variant identified in two independent patients affected with epilepsy without paroxysmal dyskinesia (11).

Of note, KCNMA1 knockout (KO) mice present with several abnormalities, including ataxia, weak grip, hearing loss, circadian imbalances and urinary bladder incontinence (28-31). In addition, a recent study reported slower weight gain for KCNMA1 KO mice compared with WT littermates after weaning. Moreover, the body composition determined by quantitative magnetic resonance indicated a higher fat proportion in KCNMA1 KO mice compared with WT littermates (32). Ataxia and weak grip in the KCNMA1 KO mice are reminiscent of the ataxia and axial hypotonia identified in four patients with KCNMA1 LoF variants, particularly in Patient 6 (Table 1) and in the patient carrying compound heterozygous pathogenic variants (12). Following this study, it would be interesting to determine whether KCNMA1 KO mice also express other phenotypes present in our cohort affecting cognition/behavior, development and oral, ocular and other muscle function. Other phenotypic abnormalities such as intestinal pseudoobstruction, bone dysplasia, aortic and arterial dilations and tortuosities should be looked for in the KO mouse model

BK channel-associated disease may be amenable to BK channel activators as BK channel LoF variants are found to abolish or markedly reduce the potassium current. Isoprimaric acid, for instance, is a BK channel activator that was found to enhance non-spatial memory in mice with Alzheimer's disease and restore the basic synapse transmission and longterm potentiation (3,33). BK channel opener BMS-205352 was found to restore hippocampal glutamate homeostasis and treat abnormalities in social recognition and interaction, non-social anxiety and spatial memory in FMR1 KO mice (34). BMS-205352 was also found to restore impaired habituation associated with some psychiatric disorders (35). Evaluation of the therapeutic potential of these compounds and other BK channel activators in appropriate KCNMA1 animal models and eventually in human patients with LoF variants of BK channels would be of clinical interest. Similarly, strategies aimed to increase the function or activity of the WT allele of KCNMA1 in heterozygous carriers by enhancing trafficking of the WT allele to the cell surface or its stability could have therapeutic potential. Obviously, gene therapy to correct the disease-causing variants using various technologies such as CRISPR-Cas9-based gene editing may also be considered for treatment of patients with LoF variants in BK channels.

In summary, we identified and functionally characterized eight different LoF variants in the BK channel [p.(Ser351Tyr),
p.(Gly356Arg), p.(Gly375Arg), p.(Asn449fs), p.(Cys413Tyr), p.(Ile66 3Val), p.(Pro805Leu) and p.(Asp984Asn)] affecting nine patients and demonstrated that impairment of BK channel function is associated with a variable phenotypic presentation ranging from ID and developmental delay alone to ataxia, axial hypotonia, apraxia/dysarthria/speech delay, cerebral atrophy, strabismus and dysmorphism. Our report substantially expands the spectrum of clinical phenotypes and diseases associated with BK channel variants.

Materials and Methods

WES, WGS and study subjects

Whole exome sequencing (WES) and data analyses were performed as previously described (37-39). Whole genome sequencing (WGS) was performed using the Illumina HiSeq X sequencing platform with an average $40 \times$ coverage (37–39). Sequence variants were uploaded to a custom analysis software application called Codicem for interpretation. Observed variants of interest were confirmed using an orthogonal sequencing technology [Sanger (dideoxy) sequencing]. The p.(Ser351Tyr) variant was identified by WGS while all other KCNMA1 variants were identified by WES. All KCNMA1 variants were annotated based on the GenBank transcript NM_002247.3. This study was approved by the Institutional Review Board (IRB) on Human Subject Research at the Cleveland Clinic, the Ethics Committee on Human Subject Research at Huazhong University of Science and Technology and other local IRBs. Written informed consent was obtained from all study subjects according to the appropriate IRB policies. The study abides by the Declaration of Helsinki principles.

Plasmids and mutagenesis

The expression plasmid for FLAG-tagged KCNMA1 (GenBank accession number U23767), KCNMA1-pcDNA3.1, was described previously (11). Mutations were generated in the KCNMA1-pcDNA3.1 plasmid using polymerase chain reaction-based site-directed mutagenesis and verified by direct DNA sequence analysis as described by us previously (40,41).

Cell culture and transfection

HEK293T cells were cultured on coverslips in 24-well plates with Dulbecco's Modified Eagle Medium (DMEM) (Gibco, Gaithersburg, MD) supplemented with 10% fetal bovine serum (Gibco, Gaithersburg, MD) at 37°C and 5% CO₂. HEK293T were co-transfected with either the WT or mutant KCNMA1 expression plasmid and the pEGFP-N1 plasmid using Lipofectamine 2000 (Invitrogen, Carlsbad, CA). The transfected cells showing a green EGFP signal (marker for successful transfection) were selected for patch-clamp recordings.

Western blot analysis

HEK293T cells were transfected with either the WT or mutant KCNMA1 expression plasmid as described above, lysed and used for western blot analysis using an anti-FLAG antibody (MBL, Woburn, MA) as described previously (36).

Patch-clamp recording

Electrophysiological experiments were performed using a patchclamping work station with a Multiclamp 700B amplifier and a 1440A Digital Analog Converter (Axon Instruments, San Jose, CA). All data was collected using the Pclamp software as described (1,11,40). Recorded data were analyzed with Clampfit (Axon Instruments, San Jose, CA), Origin 8.5 and SigmaPlot (SPSS) software programs as described previously (1,11,40). Patchclamp recordings were carried out 24 h after transfection. The experiments were performed at room temperatures (22-25°C). After a $G\Omega$ seal was established, the pipette was moved to drag the patch from the cell membrane. All recordings were carried out with inside-out configuration. The patch pipettes were made of borosilicate glass capillaries and pulled with a P-97 instrument (Sutter Instrument, Novato, CA). The resistances of the patch pipettes were 3–5 $M\Omega$ when filled with the pipette solution. The recordings were digitized at 50 kHz and low-passfiltered at 5 kHz. During the recordings, solutions with different concentrations of calcium were applied onto the patches by gravity via a perfusion pipette containing eight solution channels. The solutions used for patch clamping included the following:

- Pipette solution containing the following (in mM): 160 $\rm MeSO_3K, 2~MgCl_2~and$ 10 HEPES.
- Nominal 0 μ M Ca²⁺ solution containing the following (in mM): 160 MeSO₃K, 5 EGTA and 10 HEPES.
- 1 μm Ca^{2+} solution containing the following (in mm): 160 MeSO_3K, 5 EGTA, 3.25 CaCl_2 and 10 HEPES.
- 10 μ M Ca²⁺ solution containing the following (in mM): 160 MeSO₃K, 5 HEDTA, 2.988 CaCl₂ and 10 HEPES.

Statistical analysis

A two-tailed Student's t-test was used for determination of statistically significant differences. Data are shown as mean \pm standard deviation (SD) or mean \pm standard error of mean (SEM) as noted in the text. A P-value of 0.05 or less was considered to be statistically significant.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank Gang Yu for the analysis of the variant p.(Ile663Val) against the 3D structure of the BK channel and members of the Wang laboratory and the Center for Human Genome Research for scientific discussions and assistance.

Conflict of Interest statement. None declared.

Funding

National Natural Science Foundation of China (31430047); Hubei Province's Innovative Team (2017CFA014); National Institutes of Health/National Heart, Lung and Blood Institute (R01 HL121358 and R01 HL126729).

Author contributions

Experimental design: Q.K.W., G.M.S.M., Q.C. and A.V.; clinical studies and data analysis: A.V., G.M.S.M., S.M., S.A.S.V., B.I., A.S.M., A.C.E.H., O.B., J.T., C.H., B.C., B.G., A.R., S.N., D.L., Y.D., C.T.R., L. F., S. B, L.D., D.H. and D.B; basic experiments and data analysis: L.L., X.L., Y.H., C.X., Q.C. and Q.K.W.; draft of manuscript: L.L., X.L., Y.H., Q.C., A.V. and Q.K.W.; revision of manuscript: all

authors; project supervision: Q.K.W., G.M.S.M., Q.C. and A.V.; approval of submission: all authors.

References

- Du, W., Bautista, J.F., Yang, H., Diez-Sampedro, A., You, S.A., Wang, L., Kotagal, P., Luders, H.O., Shi, J., Cui, J. et al (2005) Calcium-sensitive potassium channelopathy in human epilepsy and paroxysmal movement disorder. Nat. Genet., 37, 733–738.
- Latorre, R., Castillo, K., Carrasquel-Ursulaez, W., Sepulveda, R.V., Gonzalez-Nilo, F., Gonzalez, C. and Alvarez, O. (2017) Molecular determinants of BK channel functional diversity and functioning. Physiol. Rev., 97, 39–87.
- Kshatri, A.S., Gonzalez-Hernandez, A. and Giraldez, T. (2018) Physiological roles and therapeutic potential of Ca²⁺ activated potassium channels in the nervous system. Front. Mol. Neurosci., 11, 258.
- 4. Petersen, O.H. and Maruyama, Y. (1984) Calcium-activated potassium channels and their role in secretion. *Nature*, **307**, 693–696.
- 5. Robitaille, R. and Charlton, M.P. (1992) Presynaptic calcium signals and transmitter release are modulated by calcium-activated potassium channels. *J. Neurosci.*, **12**, 297–305.
- 6. Murrow, B.W. and Fuchs, P.A. (1990) Preferential expression of transient potassium current (IA) by 'short' hair cells of the chick's cochlea. Proc. Biol. Sci., **242**, 189–195.
- Brayden, J.E. and Nelson, M.T. (1992) Regulation of arterial tone by activation of calcium-dependent potassium channels. Science, 256, 532–535.
- 8. Wu, Y.C., Ricci, A.J. and Fettiplace, R. (1999) Two components of transducer adaptation in auditory hair cells. J. Neurophysiol., 82, 2171–2181.
- Zhang, Y.Y., Yue, J., Che, H., Sun, H.Y., Tse, H.F. and Li, G.R. (2014) BKCa and hEag1 channels regulate cell proliferation and differentiation in human bone marrow-derived mesenchymal stem cells. J. Cell. Physiol., 229, 202–212.
- Maqoud, F., Curci, A., Scala, R., Pannunzio, A., Campanella, F., Coluccia, M., Passantino, G., Zizzo, N. and Tricarico, D. (2018) Cell cycle regulation by Ca²⁺-activated K⁺ (BK) channels modulators in SH-SY5Y neuroblastoma cells. Int. J. Mol. Sci., 19, 2442.
- Li, X., Poschmann, S., Chen, Q., Fazeli, W., Oundjian, N.J., Snoeijen-Schouwenaars, F.M., Fricke, O., Kamsteeg, E.J., Willemsen, M. and Wang, Q.K. (2018) De novo BK channel variant causes epilepsy by affecting voltage gating but not Ca²⁺ sensitivity. *Eur. J. Hum. Genet.*, **26**, 220–229.
- 12. Tabarki, B., AlMajhad, N., AlHashem, A., Shaheen, R. and Alkuraya, F.S. (2016) Homozygous KCNMA1 mutation as a cause of cerebellar atrophy, developmental delay and seizures. Hum. Genet., **135**, 1295–1298.
- Zhang, Z.B., Tian, M.Q., Gao, K., Jiang, Y.W. and Wu, Y. (2015) De novo KCNMA1 mutations in children with early-onset paroxysmal dyskinesia and developmental delay. *Mov. Dis*ord., **30**, 1290–1292.
- Sobreira, N., Schiettecatte, F., Valle, D. and Hamosh, A. (2015) GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.*, 36, 928–930.
- Srivastava, A.K. and Schwartz, C.E. (2014) Intellectual disability and autism spectrum disorders: causal genes and molecular mechanisms. *Neurosci. Biobehav. Rev.*, 46, 161–174.
- Bean, B.P. (2007) The action potential in mammalian central neurons. Nat. Rev. Neurosci., 8, 451–465.

- Salkoff, L., Butler, A., Ferreira, G., Santi, C. and Wei, A. (2006) High-conductance potassium channels of the SLO family. Nat. Rev. Neurosci., 7, 921–931.
- Sheereen, A., Alaamery, M., Bawazeer, S., Al Yafee, Y., Massadeh, S. and Eyaid, W. (2017) A missense mutation in the CRBN gene that segregates with intellectual disability and self-mutilating behaviour in a consanguineous Saudi family. J. Med. Genet., 54, 236–240.
- Higgins, J.J., Pucilowska, J., Lombardi, R.Q. and Rooney, J.P. (2004) A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation. *Neurology*, 63, 1927–1931.
- Jo, S., Lee, K.H., Song, S., Jung, Y.K. and Park, C.S. (2005) Identification and functional characterization of cereblon as a binding protein for large-conductance calciumactivated potassium channel in rat brain. J. Neurochem., 94, 1212–1224.
- Higgins, J.J., Hao, J., Kosofsky, B.E. and Rajadhyaksha, A.M. (2008) Dysregulation of large-conductance Ca2+-activated K+ channel expression in nonsyndromal mental retardation due to a cereblon p.R419X mutation. *Neurogenetics*, 9, 219–223.
- Hei, H., Gao, J., Dong, J., Tao, J., Tian, L., Pan, W., Wang, H. and Zhang, X. (2016) BK knockout by TALEN-mediated gene targeting in osteoblasts: KCNMA1 determines the proliferation and differentiation of osteoblasts. *Mol. Cells*, **39**, 530–535.
- Yang, Z., Pan, A., Zuo, W., Guo, J. and Zhou, W. (2014) Relaxant effect of flavonoid naringenin on contractile activity of rat colonic smooth muscle. J. Ethnopharmacol., 155, 1177–1183.
- Gueguinou, M., Chantome, A., Fromont, G., Bougnoux, P., Vandier, C. and Potier-Cartereau, M. (2014) KCa and Ca(2+) channels: the complex thought. *Biochim. Biophys. Acta*, 1843, 2322–2333.
- 25. Kohler, R., Kaistha, B.P. and Wulff, H. (2010) Vascular KCa-channels as therapeutic targets in hypertension and restenosis disease. *Expert Opin. Ther. Targets*, **14**, 143–155.
- Petkov, G.V. (2014) Central role of the BK channel in urinary bladder smooth muscle physiology and pathophysiology. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, **307**, R571–R584.
- Kortum, F., Caputo, V., Bauer, C.K., Stella, L., Ciolfi, A., Alawi, M., Bocchinfuso, G., Flex, E., Paolacci, S., Dentici, M.L. et al. (2015) Mutations in KCNH1 and ATP6V1B2 cause Zimmermann–Laband syndrome. Nat. Genet., 47, 661–667.
- Sausbier, M., Hu, H., Arntz, C., Feil, S., Kamm, S., Adelsberger, H., Sausbier, U., Sailer, C.A., Feil, R., Hofmann, F. et al. (2004) Cerebellar ataxia and Purkinje cell dysfunction caused by Ca2+-activated K+ channel deficiency. Proc. Natl. Acad. Sci. USA, 101, 9474–9478.
- Meredith, A.L., Thorneloe, K.S., Werner, M.E., Nelson, M.T. and Aldrich, R.W. (2004) Overactive bladder and incontinence in the absence of the BK large conductance Ca2+activated K+ channel. J. Biol. Chem., 279, 36746–36752.
- Meredith, A.L., Wiler, S.W., Miller, B.H., Takahashi, J.S., Fodor, A.A., Ruby, N.F. and Aldrich, R.W. (2006) BK calcium-activated potassium channels regulate circadian behavioral rhythms and pacemaker output. Nat. Neurosci., 9, 1041–1049.
- Ruttiger, L., Sausbier, M., Zimmermann, U., Winter, H., Braig, C., Engel, J., Knirsch, M., Arntz, C., Langer, P., Hirt, B. et al. (2004) Deletion of the Ca2+-activated potassium (BK) alphasubunit but not the BKbeta1-subunit leads to progressive hearing loss. Proc. Natl. Acad. Sci. U. S. A., 101, 12922–12927.
- Halm, S.T., Bottomley, M.A., Almutairi, M.M., Di Fulvio, M. and Halm, D.R. (2017) Survival and growth of C57BL/6J mice lacking the BK channel, Kcnma1: lower adult body weight

occurs together with higher body fat. Physiol. Rep., 5, pii: e13137.

- Wang, L., Kang, H., Li, Y., Shui, Y., Yamamoto, R., Sugai, T. and Kato, N. (2015) Cognitive recovery by chronic activation of the large-conductance calcium-activated potassium channel in a mouse model of Alzheimer's disease. *Neuropharmacology*, 92, 8–15.
- Hebert, B., Pietropaolo, S., Meme, S., Laudier, B., Laugeray, A., Doisne, N., Quartier, A., Lefeuvre, S., Got, L., Cahard, D. et al. (2014) Rescue of fragile X syndrome phenotypes in Fmr1 KO mice by a BKCa channel opener molecule. Orphanet. J. Rare Dis., 9, 124.
- Zaman, T., De Oliveira, C., Smoka, M., Narla, C., Poulter, M.O. and Schmid, S. (2017) BK channels mediate synaptic plasticity underlying habituation in rats. J. Neurosci., 37, 4540–4551.
- Chakrabarti, S., Wu, X., Yang, Z., Wu, L., Yong, S.L., Zhang, C., Hu, K., Wang, Q.K. and Chen, Q. (2013) MOG1 rescues defective trafficking of Na(v)1.5 mutations in Brugada syndrome and sick sinus syndrome. Circ. Arrhythm. Electrophysiol., 6, 392–401.
- 37. Thevenon, J., Duffourd, Y., Masurel-Paulet, A., Lefebvre, M., Feillet, F., El Chehadeh-Djebbar, S., St-Onge, J., Steinmetz,

A., Huet, F., Chouchane, M. *et al.* (2016) Diagnostic odyssey in severe neurodevelopmental disorders: toward clinical whole-exome sequencing as a first-line diagnostic test. *Clin. Genet.*, **89**, 700–707.

- Wang, C., Wu, M., Qian, J., Li, B., Tu, X., Xu, C., Li, S., Chen, S., Zhao, Y., Huang, Y. et al. (2016) Identification of rare variants in TNNI3 with atrial fibrillation in a Chinese GeneID population. Mol. Genet. Genomics, 291, 79–92.
- Li, S., Xi, Q., Zhang, X., Yu, D., Li, L., Jiang, Z., Chen, Q., Wang, Q.K. and Traboulsi, E.I. (2018) Identification of a mutation in CNNM4 by whole exome sequencing in an Amish family and functional link between CNNM4 and IQCB1. Mol. Genet. Genomics, 293, 699–710.
- Chen, Q., Kirsch, G.E., Zhang, D., Brugada, R., Brugada, J., Brugada, P., Potenza, D., Moya, A., Borggrefe, M., Breithardt, G. et al. (1998) Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. Nature, 392, 293–296.
- 41. Zhang, X., Chen, S., Yoo, S., Chakrabarti, S., Zhang, T., Ke, T., Oberti, C., Yong, S.L., Fang, F., Li, L. et al. (2008) Mutation in nuclear pore component NUP155 leads to atrial fibrillation and early sudden cardiac death. Cell, 135, 1017–1027.

Supplementary Materails

De novo loss-of-function *KCNMA1* variants are associated with a new multiple malformation syndrome and a broad spectrum of developmental and neurological phenotypes Liang et al

Figure 1. Characteristic clinical features of three patients carrying the p.(Gly375Arg) variant (i-ii, Patient 1; iii-vi, Patient 2; vii, Patient 3).



Figure 2. X-ray, CT and MRI images of Patient 1 (i-x) and Patient 2 (xi-xvi) showing the similar radiological findings, including generalized increased bone density with thickened and dysplastic bones involving both axial and appendicular skeleton affecting mainly the spongiosa; vascular malformations with ectasia and tortuosity of major arteries and ascending aorta; dilatation of gastro-intestinal tract related to chronic intestinal pseudo-obstruction. Skull, rib cage, lower and upper limbs X-rays (i-v) at the age of 7 weeks showing generalized sclerosing dysplasia. Note the mandible and ribs sclerosis and enlargement. A supine radiograph of the abdomen (vi) at the age of 32 months demonstrates abnormal dilatation of stomach (asterisk) during an acute episode of pseudo-obstruction, note the ribs enlargement (arrowhead). CT scan axial, sagittal and coronal images (vii-x) at the age 28 months. Thickened bones of the skull base and face (vi and vii) with thinning of the internals

auditory meats (vi), in opposition of carotid canal enlargement (vi); Ectasia and tortuosity of the brain arteries (vii) and cervical arteries (ix-x) with retropharyngeal internal carotids; Ascending aortic aneurism appears on image(ix). Brain MRI T2 coronal image (xi): thichening of skull base with normal bone marrow signal in T2, enlarged right middle cerebral artery; Supine radiographs of the rib cage and abdomen (xii-xiii) at the age of 22 years showing increased bone density. Note the clavicle and rib enlargement (xii-xii, arrowheads). Abnormal dilatation of bowel loops (xiii, star) during an acute episode of chronic intestinal pseudo-obstruction. CT scan coronal, sagittal and axial images (xiv-xvi) at the age of 24 years: Ascending aortic aneurism, ectasia of superior mesenteric and celiac trunk arteries (xiv-xv, arrows). Note the dense endplate sclerosis (xv, curved arrow). Thickening of the skull base and face bones (xv, star) with thinning of the internals auditory canals (xvi, arrow) in opposition of carotid canal enlargement (xvi, curved arrow).



Figure 3. Characteristic clinical features of the patient carrying the p.(Ser351Tyr) variant. Patient 4 at 10 years of age who displays the subtle appearance of widely spaced eyes (as her inner canthal and interpupillary distances are at the 95% ile) and some transient left upper eyelid ptosis, but otherwise no overtly dysmorphic findings. On her hands, note the distal phalanx of the third finger is medially deviated bilaterally. Her toes are laterally deviated (especially the hallux) and toe 2 overlaps toe 3 bilaterally.

3.5. Participation aux autres travaux de l'équipe

Dans le cadre des multiples projets menés par le centre de génétique de Dijon, je participe à l'effort d'identification de nouveaux variants pathogènes, nouveaux gènes et nouveaux mécanismes mutationnels au sein de l'équipe de recherche. Plusieurs stratégies ont été utilisées pour permettre d'augmenter le rendement diagnostique chez les patients atteints d'anomalie du développement et/ou DI auxquelles je suis amené à participer sur le plan du recrutement clinique:

- description de l'intérêt de l'outil de partage international des données GeneMatcher ³⁵⁰ pour l'identification de nouveaux gènes : bilan des collaborations et du partage réalisé, délai entre le premier contact et l'aboutissement ou non d'un projet (mené par le Dr Ange-Line Bruel).

- utilisation de la base de données denovodb permettant d'identifier de nouveaux gènes candidats dans notre cohorte sur la base du principe de la récurrence de mutations *de novo* (mené par François Lecoquierre)³⁴⁸.

4. DISCUSSION / PERSPECTIVES

4.1. Synthèse du travail

Ce travail de thèse illustre l'utilité de la poursuite de l'analyse en recherche des données d'exome lorsqu'aucun diagnostic n'a pu être obtenu ou lorsqu'une partie du phénotype reste inexpliquée malgré l'identification de variant(s) (probablement) pathogène(s) :

- Réanalyse des exomes en solo en recherche et optimisation par des stratégies bioinformatiques et de filtration permettant de faire ressortir des variants dans des gènes communs à plusieurs patients ayant des phénotypes similaires (exemple *DLG4*).

- Prise en compte des variants situés dans des gènes OMIM morbides même si le phénotype est différent de ce qui est habituellement décrit, en ayant à l'esprit que différents types de variants aux conséquences physiopathologiques différentes peuvent être responsables de plusieurs phénotypes (exemple *KCNQ3* retenu lors d'une analyse en solo nécessitant des analyses fonctionnelles complémentaires en l'absence de réplication chez d'autres patients et *KCNMA1* retenu après réanalyse en trio avec combinaison d'arguments par réplication des données chez plusieurs patients et analyses fonctionnelles montrant un mécanisme physiopathologique similaire chez les patients et différent de ce qui était connu).

- Réanalyse des exomes, négatifs après l'interprétation en solo, à l'aide du séquençage des parents et réinterprétation des données en trio permettant dans notre expérience d'identifier un variant probablement pathogène pour 26% (18/70) des patients et un ou plusieurs variants de signification inconnue chez 14% (10/70) des patients (incluant certains patients chez qui un diagnostic est déjà établi mais chez qui est suspecté un autre diagnostic combiné). A noter que les variants faux sens sont plus difficiles à interpréter et peuvent nécessiter des coûts supplémentaires conséquents pour permettre de faire la preuve de leur implication sur le plan fonctionnel ; il existe donc un biais de publication et d'interprétation vers la pathogénicité des variants prédits comme disruptifs (responsables de perte de fonction par haploinsuffisance: délétions complètes ou intragéniques, frameshift, non sens, altération de l'épissage).

L'ensemble des éléments indiqués ci-dessus justifie de pouvoir disposer de données pangénomiques sans *a priori* pour les patients atteints de maladie génétique à forte hétérogénéité génétique ou ayant des présentations cliniques atypiques et de partager les résultats et gènes candidats sur des plateformes d'échange international. Cela a fait la preuve de son efficacité et permis d'accroître rapidement les découvertes dans les dernières années en rassemblant les données de multiples équipes, notamment du fait de la rareté des patients porteurs de variant(s) pathogène(s) dans certains gènes.

L'analyse des données d'exome chez les patients atteints de pathologie par anomalie du développement permet de poser un diagnostic dans le champ clinique (en l'état actuel des connaissances) dans environ 30 à 40% des cas. L'ensemble de ces travaux permet d'illustrer la nécessité d'établir des corrélations génotype-phénotype de bonne qualité pour incriminer un gène candidat et ainsi vérifier la concordance entre un type de variant donné et une maladie génétique. Il faut être d'autant plus critique après identification de variant(s) convaincant(s) dans un gène candidat pour expliquer la pathologie recherchée que plusieurs diagnostics génétiques coexistent chez environ 5% des patients ayant un diagnostic moléculaire 44,45,332,336,338, et il faut donc se poser la question si l'hypothèse diagnostique proposée peut expliquer l'ensemble du phénotype ou s'il convient de rechercher d'autres candidats potentiels associés; cela est probablement plus fréquent dans les familles consanguines. Tout cela en ayant à l'esprit la possibilité d'observer un spectre phénotypique qui peut être très large pour des variants aux conséquences attendues similaires dans un gène donné en inter-familial, voire en intra-familial dans les maladies autosomiques dominantes. L'analyse de l'ensemble des variants présents dans les gènes OMIM morbides constitue donc le minimum nécessaire à l'interprétation d'un exome diagnostique dans les anomalies du développement. En cas de doute entre variabilité phénotypique associé à un variant pathogène et association de plusieurs maladies chez le même patient, il pourra être pertinent de prévoir des réanalyses dans le champ de la recherche, soit à partir des données déjà disponibles, soit par des stratégies complémentaires comme l'analyse en trio si cela n'a pas été fait. Une bonne collaboration entre les services cliniques et les laboratoires, incluant le partage international des données, est donc nécessaire pour décrire précisément le spectre phénotypique des maladies génétiques.

4.2. Perspectives d'évolution des pratiques

4.2.1. Evolution en diagnostic

Sur un versant purement diagnostique, la méta-analyse publiée par Clark et al 2018 indique l'apport clair des technologies de séquençage par exome et génome par rapport aux stratégies classiques incluant CGH-array, analyses biochimiques et métaboliques, que ce soit en soins intensifs néonataux ou dans un cadre moins urgent. En particulier, 11 études publiées en 2017 indiquent un apport significatif de l'utilité diagnostique du séquençage du génome par rapport à la CGH-array (0.42 vs. 0.005; P < 0.0001). A partir de 4 études, il a pu être calculé une probabilité d'identifier le diagnostic par exome ou génome 8,3 fois plus importante que par CGH-array. Deux études rapportent l'absence de différence significative entre séquençage d'exome et de génome ³³⁵. Il existe donc un intérêt à développer des circuits d'urgence dans certaines conditions particulières. Cette procédure est

en cours de développement dans l'équipe avec une étude pilote pour 28 enfants en soins intensifs néonataux et un PHRC-I qui débutera prochainement visant à inclure une cinquantaine d'enfants supplémentaires de plusieurs centres de génétique français (étude FASTGEN menée par Dr Arthur Sorlin).

4.2.2. Evolution en recherche

4.2.2.1. Séquençage du génome et regroupement de cohortes de patients atteints de syndromes caractérisés sans base moléculaire connue

Afin d'améliorer le rendement diagnostique dans les affections génétiques notamment les anomalies du développement/DI, plusieurs approches complémentaires sont utiles et sont en cours de déploiement dans plusieurs équipes, avec en particulier le séquençage génome entier et éventuellement des analyses ciblées ou pangénomiques complémentaires explorant l'expression génique, les modifications épigénétiques pour permettre d'interpréter les données de génome (variants introniques, intergéniques, de structure...). Ces approches sont considérées dans les perspectives de ce travail, pour lutter contre l'errance diagnostique :

- concernant les syndromes bien caractérisés pour lesquels les recherches de bases moléculaires sont restées négatives malgré des initiatives pangénomiques répétées à travers le monde, la combinaison des données de séquençage et d'approches Omics à la recherche de mécanismes génétiques similaires permettront de se donner le maximum de chances d'aboutir à une cause moléculaire. Cela correspond à l'initiative Solve-RD ⁴¹⁰ à laquelle nous participons dans le centre de génétique de Dijon, menée par le réseau européen des anomalies du développement ERN-ITHACA pour les syndromes d'Aicardi, Gomez-Lopez-Hernandez, dysplasie fronto-nasale, Pai, Goldenhar, Wildervanck, VACTERL, MURCS, OEIS en particulier. Parfois, les analyses seront proposées sur tissus atteints du fait d'hypothèse de mosaïcisme génétique.

- concernant les affections aspécifiques et à grande hétérogénéité génétique telle que la déficience intellectuelle peu ou pas syndromique, les données pangénomiques plus poussées comme WGS (short read, long read), RNA-seq, methyl-seq... pourront permettre de révéler des variants non détectés par le séquençage d'exome ou des mécanismes que l'exome n'est pas capable d'identifier (variants de structure, modifications épigénétiques). Une étude pionnière visant à évaluer l'apport du séquençage WGS en trio dans la déficience intellectuelle sévère a inclus 50 patients chez qui les analyses par CGH et WES s'étaient avérées non contributives ⁴¹¹ : dans 21/50 (42%), un diagnostic a pu être établi dont 13 diagnostics incriminant des variants ponctuels dans les régions codantes pour lesquels il n'est pas précisé si les variants étaient présents dans les données d'exome et le motif de l'absence de diagnostic en WES. Certaines équipes suspectent d'ailleurs qu'une partie des

diagnostics manqués de l'exome se trouvent à portée d'identification mais que le gène en cause n'a pas été retenu par l'analyse, ou n'est pas encore connu en pathologie humaine, ou encore le variant a été filtré par les étapes bioinformatiques et n'apparait plus dans le rapport de variant ³³⁹. Néanmoins, le WGS permet d'obtenir une profondeur de lecture plus homogène et une couverture complète puisqu'il n'y a pas d'étape de capture des régions d'ADN cible comme pour le séquençage d'exome ou de panels de gènes. Par ailleurs, dans la même série, chez 8/50 patients atteints de DI sévère, les variants en cause n'étaient pas détectables par les techniques précédentes : 5 délétions, une duplication en tandem, une duplication inter-chromosomique et un remaniement complexe par inversion-duplication-délétion ⁴¹¹. Une autre étude rapporte des exemples de plusieurs événements génétiques pathogènes situés dans les régions non codantes ⁴¹².

4.2.2.2. Intérêt du RNA-seq

Quelques études montrant l'intérêt d'une approche par RNA-seq émergent, la preuve de principe ayant initialement été faite sur l'analyse ciblée en cas d'hypothèse d'un diagnostic clinique mais ce type d'analyse est vouée à prendre de l'ampleur dans un but d'analyse pangénomique dans des maladies génétiquement hétérogènes :

- Gonorazky et al 2016: étude par analyse conventionnelle incluant RT-PCR et RNA-seq sur biopsie musculaire d'un patient atteint de DMD : les 2 techniques ont permis d'identifier un variant intronique profond responsable d'une altération d'épissage montrant l'intérêt du RNA-seq ⁴¹³.

- Kremer et al 2017 : étude de 48 patients suspects de mitochondriopathie sans cause moléculaire identifiée et exploration par RNA-seq sur biopsie cutanée : 5/48 soit environ 10% de diagnostic obtenu, une hypothèse pouvant être formulée pour les patients restant ; il a pu être identifié une médiane d'un gène ayant une altération d'expression, 5 gènes ayant des anomalies d'épissage et 6 gènes ayant une expression mono-allélique par patient ⁴¹⁴.

- Cummings et al 2017 : étude par RNA-seq provenant de tissu musculaire de 50 patients atteints de myopathie sans diagnostic moléculaire malgré le séquençage de panel ou d'exome : taux diagnostique de 35% dont un variant intronique *de novo* récurrent du gène *COL6A1*⁴¹⁵.

- Hamanaka et al 2018 : recherche de la caractérisation moléculaire de 6 patients étiquetés myopathie à némaline sans identification de mutation du gène *NEB* après exome : le RNA-seq sur tissu musculaire a permis d'identifier un variant d'épissage intronique profond c.1569+339A>G chez un patient et un variant synonyme exonique altérant l'épissage c.24684G>C (p.Ser8228Ser) chez 3 patients, ce dernier variant représentant la mutation la plus fréquente dans la population japonaise ⁴¹⁶.

- Kernohan et al 2017: identification d'un variant hétérozygote du gène *ASAH1* (c.458A>G;p.Tyr153Cys) sur un panel de 15 gènes responsables de pathologie du motoneurone

chez un patient ayant un phénotype compatible avec cette pathologie mais il s'agit d'une pathologie autosomique récessive. Le RNA-seq sur sang a permis d'identifier une altération d'épissage par saut d'exon 6 non expliqué par le variant précédent conduisant à identifier par séquençage Sanger le 2^{ème} variant (c.504A>C;p.Lys168Asn) non prédit comme pathogène. Le laboratoire n'a pas fourni l'information si le variant était ressorti de l'analyse du panel ou non ⁴¹⁷.

- Frésard et al 2019 : étude par RNA-seq provenant du sang d'une cohorte hétérogène de 94 patients atteints principalement de pathologies génétiques neurologiques (40), musculosquelettiques (8), hématologiques (8), ophtalmologiques (6) avec analyses préalables dont séquençage d'exome non contributives : identification de la cause génétique dans 7,5% des cas et hypothèse solide dans 16,7% des cas ⁴¹⁸.

Ces études montrent l'intérêt du RNA-seq mais posent aussi la question du tissu le plus pertinent à étudier du fait de l'expression différentielle des gènes selon les tissus ou même les types cellulaires dans un tissu donné. Cela est d'autant plus problématique en ce qui concerne la déficience intellectuelle même si une bonne partie des gènes exprimés dans les cellules sanguines et dans la peau sont communs au cerveau ^{419,420}.

4.2.2.3. Autres mécanismes à prendre en considération pour de futures perspectives

D'autres mécanismes peuvent aussi être en cause, comme certains variants cytogénétiques, qui pourraient conduire à perturber l'expression génique :

a. Altération de l'expression génique et/ou en perturbant l'organisation chromatinienne via l'altération des TADs (domaines topologiquement associés)

Les TAD correspondent à des unités de régulation génique pouvant être différents d'un type cellulaire à l'autre. Une équipe a pu estimer que parmi 922 délétions chromosomiques identifiées chez des patients ayant un syndrome malformatif, environ 11,8% avaient pour mécanisme physiopathologique une altération des régions régulatrices pour tout ou partie ⁴²¹.

Une autre équipe a aussi rapporté un rendement diagnostique certain de 11% de la CGH array dans la DI/retard neurodéveloppemental chez 1015 patients. Il existait aussi 18% de VSI dont 52 probablement pathogènes (5%) de par la région chromosomique impactée et le rôle des gènes situés dans ces régions. De plus, 2 délétions ont été suspectes de perturber la frontière de domaines chromatiniens (0,2% de la cohorte et 1,1% des VSI) ⁴²². Quelques autres équipes ont décrits des cas porteurs de variants chromosomiques pouvant expliquer une anomalie du développement. Un des modèles caractéristiques implique la région chromosomique où les frontières chromatiniennes délimitent les régions régulatrices des gènes *WNT6, IHH, EPHA4, PAX3* ⁴²³.

La comparaison de remaniements chromosomiques apparemment équilibrés et de microdélétions à proximité de gènes connus en pathologie humaine a permis de préciser les régions fonctionnelles de chromatine comme pour le gène *FOXG1*⁴²⁴.

b. Altération de l'expression génique et/ou en perturbation de l'organisation chromatinienne via l'interruption de gènes ou d'interaction gène(s) - élément(s) régulateur(s) du fait de la mobilité de séquences répétées notamment des rétrotransposons

La localisation de ces séquences répétées pourrait entraîner des anomalies d'expression en perturbant les éléments régulateurs mais aussi une perte de fonction du gène, par exemple par altération de l'épissage en cas d'insertion intronique ou par frameshift/non sens en cas d'insertion dans un exon, ou en altérant une séquence de polyadénylation ⁴²⁵.

c. Anomalies épigénétiques altérant la méthylation de l'ADN et/ou les modifications posttraductionnelles des histones

Cela représente aussi un mécanisme mutationnel pouvant être responsable de maladies génétiques principalement par anomalies d'expression génique. Elles résultent soit de mutations dans des gènes ayant un rôle de lecture, d'appositions ou d'effacement de marques épigénétiques ^{426,427}, soit de facteurs environnementaux comme les troubles neurodéveloppementaux associés à l'exposition fœtale à certains toxiques comme l'alcool ⁴²⁸. Ces mutations ou expositions ont pour conséquence d'entraîner l'observation d'épi-signatures et peuvent aussi permettre d'aider à l'interprétation de certains VSI. Des anomalies épigénétiques spontanées et localisées pourraient éventuellement être en cause dans certaines anomalies du développement. Les progrès des technologies de methyl-seq et les outils bioinformatiques permettent d'améliorer les connaissances sur le mécanisme physiopathologique de certains syndromes connus mais pourront probablement expliquer certains patients sans diagnostic, d'autant plus que de nombreuses modifications épigénétiques sont identifiées dans le système nerveux, et notamment l'acétylation des histones a été associée au contrôle transcriptionnel de gènes impliqués dans les apprentissages et le processus de mémorisation ⁴³⁰.

4.2.2.4. Projets menés dans l'équipe GAD INSERM U1231

Sur la base de ces données, le centre de génétique de Dijon a plusieurs projets visant à tenter d'identifier la cause moléculaire chez des patients touchés par des anomalies du développement sans diagnostic. Je participe à ces projets principalement dans le cadre du recrutement clinique :

- projet OMIXCARE de multiOMICS pour le séquençage de génome entier (WGS) en trio et comparaison d'une stratégie de séquençage de fragments courts et longs couplée au RNA-seq (mené par Dr Antonio Vitobello, financement Labex GeneMed).

- projet DIWA : combinaison de WGS short read en trio et RNA-seq solo avec lectures en aveugle en plusieurs stratégies pour évaluer l'apport de ces technologies (mené par Dr Frédéric Tran-Mau-Them, financement PHRCI).

Ainsi, les projets OMIXCARE, DIWA, Solve-RD⁴¹⁰ permettront d'appréhender l'apport des nouvelles technologies en recherchant notamment des mécanismes mutationnels pathogènes moins habituels car encore peu reconnus.

Enfin, il existe la possibilité que des mécanismes multifactoriels et un oligogénisme puissent être en cause pour une proportion de patients (comme dans l'autisme associé ou non à la DI ^{54,431–435}).

4.2.3. Enjeux de l'interprétation des variations génétiques

Les techniques d'analyses pangénomiques conduisent à identifier de multiples variations de séquence, de structure, de nombre de copies chromosomiques qu'il convient d'interpréter. Les outils *in silico* qui sont développés permettent d'aider à filtrer ces variations pour les prioritiser et de retenir celle(s) la(les) plus pertinente(s) par rapport à la question clinique posée. Des outils intégratifs permettant de combiner les différents scores *in silico* seraient particulièrement intéressants.

Un tel outil nommé Exomiser a été développé pour classer les variations candidates en intégrant des données phénotypiques, des données génétiques (notamment en considérant les gènes connus en pathologie humaine), des données de modèles cellulaires et animaux et des données phénotypiques de l'interactome de la protéine pour laquelle le gène correspondant porte une variation. Cette analyse intégrative permettait de classer les variations génétiques par ordre de probabilité de leur implication dans la pathologie. Dans un premier temps, l'analyse de 1000 exomes simulés avec des variations pathogènes issues de la base de données HGMD a permis de faire ressortir la variation pathogène en haut de liste dans 94 à 96% des cas en comparaison à 20-77% des cas en utilisant les filtres habituels. Dans un 2^{ème} temps, l'analyse de 11 familles déjà diagnostiquées par exome permettait d'obtenir le diagnostic dans les 7 premières lignes dans tous les cas dont dans 6 cas à la première ligne. Parmi ces 11 situations, 5/11 diagnostics étaient prioritisés grâce à l'interactome de la protéine codée par le gène en cause. Enfin, parmi 21 familles sans diagnostic, le diagnostic a pu être posé dans 2 familles et un gène candidat intéressant était détecté dans 17 parmi les 19 familles restantes. Ce type d'étude illustre l'intérêt des analyses intégrant les données de séquençage, les données phénotypiques et les bases de données et outils bioinformatiques ⁴³⁶.

Des outils prédisant les anomalies d'épissage ont également été développés pour améliorer la sensibilité et la spécificité de ces prédictions, comme par exemple l'outil SPiCE (Splicing Prediction in Consensus Elements) qui permet d'atteindre des taux de sensibilité et spécificité de 99,5 et 95,%

respectivement ⁴³⁷, ou l'outil UMD-predictor qui permet de combiner les prédictions de pathogénicité des variations faux sens avec une efficacité similaire voire supérieure aux outils existants en apportant par ailleurs des informations sur les conséquences sur l'épissage des variations situées dans les régions codantes ⁴³⁸.

Cependant, malgré le développement d'outils *in silico* de plus en plus performants, des limites persistent et le caractère bénin ou pathogène des variations ne peut être prédit systématiquement. D'une part le partage inter-laboratoire des données pour un groupe de gènes donné peut avoir une utilité mais cela concerne surtout des gènes bien connus, relativement fréquemment responsables de maladies génétiques et très étudiés comme *DMD* dans le spectre des dystrophinopathies ⁴³⁹, ou *BRCA1/2* analysés de longue date dans les prédispositions aux cancers du sein et de l'ovaire ⁴⁴⁰. Des arguments fonctionnels peuvent donc être très utiles grâce au développement de modèles cellulaires ou animaux. En recherche, cela peut permettre de disséquer les voies physiopathologiques et d'impliquer un gène candidat et un type de variant particulier, les délais sont souvent longs et nécessitent un nombre minimal de patients analysés ; en diagnostic, les délais doivent être compatibles avec la routine et il peut être utile de s'appuyer sur des laboratoires ayant déjà un modèle à disposition ou d'utiliser un modèle relativement facilement exploitable.

Par ailleurs, concernant les gènes impliqués dans la régulation épigénétique et de l'expression génique, des comparaisons d'épi-signatures ou de profils d'expression pourraient être faites entre des patients porteurs de variations de signification indéterminée et des patients porteurs de variations pathogènes (cf. chapitre 1.2.3.2) ^{131–134}. Le consortium EpiSign vise à développer l'identification d'épi-signatures pour de multiples autres maladies, ce qui permettra d'avoir un profil type par maladies et permettre ce type de comparaisons.

L'utilisation de modèles animaux peut aussi être efficace dans certaines situations.

Le poisson zèbre représente un outil très utile pour l'étude à grande échelle de variations génétiques rares dans des gènes connus ou de nouveaux gènes de DI ou autres troubles neurodéveloppementaux avec ou sans épilepsie. Ces études évaluent notamment la possibilité de restaurer un phénotype normal après apport d'un ARNm sauvage par rapport à un ARNm porteur d'une variation à tester. Par exemple, la baisse d'expression de *stx1b* par morpholinos entraîne un phénotype de crises d'épilepsie ce qui a permis d'incriminer le gène orthologue humain avec le phénotype présenté par certains patients ; l'apport d'ARNm sauvage permet de réduire ces anomalies, contrairement à un ARNm porteur d'une mutation testée ce qui apporte un fort argument pour sa pathogénicité ⁴⁴¹. De même, la baisse d'expression de *trappc6b* par morpholino entraîne une augmentation de l'activité neuronale à l'état basal et le déclenchement de crises induites par le PTZ (pentylenetetrazol, antagoniste du GABA) ; les variations bialléliques perte de fonction chez

l'orthologue humain sont responsables de DI, TSA, microcéphalie et épilepsie, principalement identifiées dans des familles consanguines ⁴⁴².

D'autres modèles comme le Xenope peuvent être utiles. Par exemple, un modèle de *Xenopus laevis* a été utilisé pour évaluer les conséquences de variations du gène *dyrk1a* sur le développement rénal puisque certains patients porteurs de mutations hétérozygotes perte de fonction ont des malformations rénales en plus d'une déficience intellectuelle et d'une microcéphalie (11/15 patients dans une série). La baisse d'expression du gène par morpholino entraîne des anomalies de développement des néphrons chez l'embryon, conduisant à des malformations génito-urinaires. Ces malformations peuvent être évitées en co-injectant l'ARNm humain normal mais pas un ARNm porteur d'une mutation (testé pour un non sens p.Arg205* et un faux sens p.LeuL245Arg)⁴⁴³.

Une équipe a aussi pu montrer la pathogénicité de variations hétérozygotes composites (p.Tyr503Cys et p.Gln485*) de *PIBF1* responsables d'une forme de syndrome de Joubert (ciliopathie autosomique récessive incluant retard neurodéveloppemental et DI, ataxie, anomalie des paires crâniennes touchant notamment l'oculomotricité, troubles respiratoires et malformation pontocérébelleuse). L'injection de morpholino dans un modèle Xenope a permis d'observer un déficit mucociliaire par réduction du nombre et de la mobilité des cils et l'analyse de tissu cutané a montré une localisation anormale après expression de la mutation non sens ne permettant pas de restaurer un phénotype normal alors que la mutation faux sens permet une restauration partielle. Ce modèle permet donc assez rapidement d'établir la pathogénicité de tels variants ⁴⁴⁴.

L'utilisation de modèles cellulaires peut aussi permettre d'établir des corrélations génotype phénotype, notamment en ce qui concerne les canaux ioniques comme cela a pu être identifié pour *KCNMA1* et comme cela a déjà été décrit pour *SCN2A* : en effet, en ce qui concerne *SCN2A*, les mutations perte de fonction responsables d'une perte de courant du canal ionique causent une DI pouvant être sans épilepsie alors que les variations pathogènes faux sens sont responsables d'encéphalopathie épileptique précoce ou d'épilepsie néonatale bénigne selon que le gain de fonction est majeur avec de grosses perturbations dans les propriétés électrophysiologiques ou qu'il a un effet plus modeste ⁴⁴⁵.

Les modèles animaux permettent de tester des mécanismes mutationnels comme des variations perte de fonction mais l'enjeu dans l'interprétation représente surtout les variants faux sens. Les nouvelles technologies CRISPR-Cas9 permettent d'envisager de produire rapidement tous ces types de variations génétiques spécifiquement identifiées chez des patients pour tester leurs conséquences ce qui sera très informatif, en particulier pour les faux sens dont l'interprétation peut être délicate.

4.3. Conclusion

En conclusion, ce travail a permis de contribuer à l'identification de plusieurs gènes encore non associés en pathologie humaine mais aussi à de nouveaux mécanismes physiopathologiques de gènes déjà connus comme associés à d'autres phénotypes. Les efforts complémentaires d'étude de cohorte dans des phénotypes spécifiques ou très hétérogènes avec des outils plus performants (WGS, RNA-seq, methyl-seq, multi-OMICS), des regroupements de données (cliniques de patients ayant des phénotypes similaires et/ou génétiques) pour renforcer la puissance statistique et mieux interpréter certains variants devraient permettre d'augmenter encore le rendement des tests ¹ ou l'interprétation de certains variants ⁴⁴⁶ et les découvertes de nouveaux mécanismes mutationnels et physiopathologiques. Tous ces efforts permettront ainsi d'augmenter le rendement diagnostique et de diminuer la proportion de patients en impasse diagnostique. L'évaluation du potentiel oligogénisme et des causes multifactorielles chez les patients pour lesquels aucune explication ne peut être obtenue est une hypothèse de travail d'autres équipes.

ANNEXES

Annexe 1 : Liste des gènes impliqués dans la régulation des marques épigénétiques et dont les variations pathogènes sont responsables de maladies génétiques, pour la plupart avec DI¹²⁷.

Gene symbol	OMIM gene	Phenotype	OMIM Morbid	Function (chromatin)	Inheritance
1. Writers					
DNA methyl	ation				
DNIMT1	106275	Hereditary sensory neuropathy type IE (HSN1E)	<u>614116</u>	DNA methylation	AD
DINMIT	120373	Cerebellar ataxia, deafness, and narcolepsy,	<u>604121</u>		AD
DNMT3B	602900	ICF syndrome	<u>242860</u>	DNA methylation	AR
FTO	610966	Growth retardation, developmental delay, coarse facies, and early death	<u>612938</u>	RNA demethylation	AR
Histone mod	lification				
CREBBP	600140	Rubinstein–Taybi syndrome 1 (RSTS1)	<u>180849</u>	Histone acetyltransferase (HAT)	AD, de novo
CUL4B	300304	Mental retardation, X-linked, syndromic 15 (Cabezas type)	<u>300354</u>	Histone ubiquitination (H2AK119Ub1)	XL
EHMT1	607001	Kleefstra syndrome	<u>610253</u>	Histone methyltransferase (KMT1D; H3K9me1/2)	AD, de novo
EP300	602700	Rubinstein–Taybi syndrome 2 (RSTS2)	<u>613684</u>	Histone acetyltransferase (HAT)	AD, de novo
EZH2	601573	Weaver syndrome	<u>277590</u>	Lysine N-methyltransferase 6 (KMT6A; H3K27me3)	AD, de novo
HLCS	609018	Holocarboxylase synthetase deficiency	<u>253270</u>	Histone biotinylation (H4K16bio)	AR
HUWE1	300697	Mental retardation, X-linked syndromic, Turner type	<u>300706</u>	Histone ubiquitination	XL
KAT6B	605880	Ohdo syndrome, SBBYS variant	<u>603736</u>	Histone acetyltransferase (MYST4)	AD, de novo
		Genitopatellar syndrome	<u>606170</u>		AD, de novo
KMT2A	159555	Wiedemann-Steiner syndrome	<u>605130</u>	Histone methylation (MLL)	AD, de novo
KMT2D	602113	Kabuki syndrome	<u>147920</u>	Histone methylation (MLL2)	AD, de novo
KMT2C	607001	Kleefstra Syndrome spectrum	<u>610253</u>	Histone methylation (MLL3)	AD, de novo
NSD1	606681	Sotos syndrome	<u>117550</u>	Histone methylation (KMT3B)	AD, mostly de novo

Gene symbol	OMIM gene	Phenotype	OMIM Morbid	Function (chromatin)	Inheritance			
		Beckwith-Wiedemann syndrome	<u>130650</u>		AD, de novo			
WHSC1	602952	Wolf–Hirschhorn syndrome	<u>194190</u>	Histone methyltransferase (NSD2) (H3K36me3, H4K20me2)	AD, deletions			
UBE2A	312180	X-Linked ID, Nascimento type	<u>300860</u>	Histone ubiquitination	XL			
2. Erasers								
HDAC4	605314	Brachydactyly-mental retardation syndrome	<u>600430</u>	Histone deacetylase	AD, de novo			
	300260	Wilson–Turner syndrome	<u>309585</u>	Histone deacetylase	XL			
HDAC0	300209	Cornelia de Lange syndrome 5	<u>300882</u>		XL			
KDM5C	314690	X-linked syndromic mental retardation; Claes-Jensen type	<u>300534</u>	Histone demethylase (H3K4 tridemethylase)	XL			
KDM6A	300128	Kabuki syndrome 2	<u>300867</u>	Histone demethylase (H3K27me3/2)	XL			
PHF8	300560	Siderius X-Linked Mental Retardation Syndrome	<u>300263</u>	Histone demethylase (H4K20me1, H3K9me1/me2)	XL			
3. Chromati	in remode	lers (DEAD/H ATPase family)						
	102620	Baraitser-Winter syndrome	<u>243310</u>	SWI/SNF, INO80 and ISWI complex	AD, de novo			
ACID	102050	Dystonia, juvenile onset	<u>607371</u>		AD* (single family, caution)			
ARID1A	603024	Mental retardation, Autosomal dominant 14	<u>614607</u>	DEAD/H ATPase helicase family, SWI/SNF subfamily	AD, de novo			
		Coffin-Siris syndrome	<u>135900</u>		AD, de novo			
ARID1B	614556	Mental retardation, Autosomal dominant 12	<u>614562</u>	DEAD/H ATPase helicase family, SWI/SNF subfamily	AD, de novo			
		Coffin-Siris syndrome	<u>135900</u>		AD, de novo			
ATRX	300032	Alpha thalassemia mental retardation syndrome, X-linked (ATRX)	<u>301040</u>	DEAD/H ATPase helicase family	XL			
	500032	Mental retardation-hypotonic facies syndrome	<u>309580</u>		XL			
CHD2	602119	Epileptic encephalopathy, childhood-onset	<u>615369</u>	DEAD/H ATPase helicase family, CHD subfamily	AD, de novo			
CHD7	608892	CHARGE syndrome	<u>214800</u>	DEAD/H ATPase helicase family, CHD subfamily	AD, de novo			
CHD8	610528	Autism, susceptibility to, AUTS18	<u>615032</u>	DEAD/H ATPase helicase	AD, de novo			

Gene symbol	OMIM gene	Phenotype	OMIM Morbid	Function (chromatin)	Inheritance
				family, CHD subfamily	
SMARCA2	600014	Nicolaides-Baraitser syndrome	<u>601358</u>	DEAD/H ATPase helicase family, SWI/SNF subfamily	AD, de novo
		Coffin-Siris syndrome	<u>135900</u>		AD, de novo
SMARCA4	603254	Mental retardation, Autosomal dominant 16	<u>614609</u>	DEAD/H ATPase helicase family, SWI/SNF subfamily	AD, de novo
		Coffin-Siris	<u>135900</u>		
SMARCB1	601607	Mental retardation, Autosomal dominant 15	<u>614608</u>	DEAD/H ATPase helicase family, SWI/SNF subfamily	AD
		Coffin-Siris syndrome	<u>135900</u>		
SMARCE1	603111	Coffin-Siris syndrome		DEAD/H ATPase helicase family, SWI/SNF subfamily	
SRCAP	611421	Floating-Harbor syndrome	<u>136140</u>	DEAD/H ATPase helicase family, INO80/SWR subfamily	AD, de novo
SS18L1	606472	Amyotrophic lateral sclerosis (ALS)		SWI/SNF complex	AD, de novo
4. Other rea	ders and o	chromatin remodelers			
ASXL1	612990	Bohring–Opitz syndrome	<u>605039</u>	PR-DUB complex, histone H2A deubiquitination	AD, de novo
BCOR	300485	Microphthalmia, syndromic	<u>300166</u>	Polycomb complexes containing Ring1B	XL
CHMP1	164010	pontocerbellar hypoplasia 8	<u>614961</u>	Targets polycomb protein BMI to condensed chromatin	AR
CTCF	604167	ID, microcephaly and growth retardation		Chromatin binding factor, insulator	AD, de novo
GATAD2B	614998	Mental retardation, autosomal dominant 18	<u>615074</u>	NuRD complex	AD, de novo
		Mental retardation, X-linked; MRX3/	<u>309541</u>	Found in repressor and activator complexes	XL
HCFC1	300019	Methylmalonic acidemia and homocysteinenia (Cobalamin disorder)			XL
KANSL1	612452	Koolen–De Vries syndrome	<u>610443</u>	NSL1 histone acetyltransferase complex	AD, de novo
MBD5	611472	Mental retardation, autosomal dominant 1 (2q deletion syndrome)	<u>156200</u>	Associated with heterochromatin	AD, de novo
		Kleefstra Syndrome spectrum	<u>610253</u>		AD, de novo
MECP2	300005	RETT syndrome; RTT	<u>312750</u>	Binds to methylated DNA	XL, mainly

Gene symbol	OMIM gene	Phenotype	OMIM Morbid	Function (chromatin)	Inheritance
					females
		RETT syndrome, atypical; Angelman syndrome-like	<u>105830</u>		XL
		Autism, susceptibility to, X-linked 3; AUTSX3	<u>300496</u>		XL
		Encephalopathy, neonatal severe	<u>300673</u>		Xl
		Mental retardation, X-linked, syndromic 13; MRXS13	<u>300055</u>		XL
		Duplication MECP2; Lubs X-linked mental retardation syndrome	<u>300260</u>		XL
PHF6	300414	Borjeson–Forssman–Lehmann syndrome	<u>301900</u>	NuRD complex	XL
		Coffin–Siris-like	<u>135900</u>		XL
POGZ	614787	Autism spectrum disorder		Pogo transposable element with ZNF domain	AD, de novo
SKI	164780	Shprintzen-Goldberg syndrome	<u>182212</u>	HDAC recruiting complexes	AD
		Lujan-Fryns syndrome	<u>309520</u>	Mediator complex	XL
MED12	300188	Opitz-Kaveggia syndrome	<u>305450</u>	Mediator complex	XL
		OHDO syndrome Maat–Kievit– Brunner	<u>300895</u>	Mediator complex	XL
MED17	603810	Microcephaly, postnatal progressive, with seizures and brain atrophy	<u>613668</u>	Mediator complex	AR
MED23	605042	Mental retardation, autosomal recessive 18	<u>614249</u>	Mediator complex	AR
NIPBL	608667	Cornelia de Lange syndrome 1; CDLS1	<u>122470</u>	Cohesin complex	AD
RAD21	606462	Cornelia de Lange syndrome 4; CDLS4	<u>614701</u>	Cohesin complex	AD, de novo
SALL1	602218	Townes-Brocks syndrome	<u>107480</u>	Member of NuRD histone deacetylase complex	AD
SMC1A	300040	Cornelia de Lange syndrome 2; CDLS2	<u>300590</u>	Cohesin complex	XL
SMC3	606062	Cornelia de Lange syndrome 3; CDLS3	<u>610759</u>	Cohesin complex	AD, de novo

Human Gene(s)	Zebrafish Gene(s)	Disorder*	Method**	Phenotype(s)	Rescue	Reference
CHD7	chd7	DS/ID (* in this study) ASD	MO	Morphological abnormalities: pericardial edema, microcephaly, body curvature, no swim bladder, decreased growth (≥48 hpf) Abnormal movement (4 dpf) Epileptiform discharges in tectal field-potential recordings (4 dpf)	-	Suls et al. (2013)
PK1	pk1a	EP	OE	Gastrulation defects: Reduced anterior-posterior length (20 hpf); lateral expansion of somites (8–10 somite stage) OE zebrafish <i>pk1a</i> mRNA with human mutation (R104Q) causes less severe defects	-	Bassuk et al. (2008)
			MO	 horeased sensitivity to drug-induced selzures (48 hpf) Retinal inner plexiform layer defects (76–78 hpf) Normal visual startle response (5 dpf) 	Zebrafish mRNA partially rescues retinal defects; mRNA with human R150H, Y465H mutations do not	Mei et al. (2013)
PSMD12	psmd12	ID	F0 CRISPR	 Decreased size of optic tecta (3 dpf) Proximal renal tubule defects, decreased proximal tubule convolution area (4 dpf) Craniofacial abnormalities (3 dpf) 	-	Kury et al. (2017)
RHEB	rheb	ID	F0 CRISPR; OE	FO CRISPR: Microcephaly (decreased head area excluding eyes; 5 dpf) OE human RHEB and RHEB carrying human mutations: Macrocephaly (5 dpf)	Rapamyoin rescues macrocephaly	Reijnders et al. (2017)
SCN1A	scn1lab	DS EP ASD	ENU Mutant; MO	 Abnormal optokinetic response (OKR; 5 dpf; MO and mutant) Darkened pigmentation (MO and mutant) Death by 14 dpf (mutant) 	Stimulation of hindbrain saccade generator rescues OKR (morphants)	Schoonheim et al. (2010)
			ENU Mutant	 Spontaneous seizure-like activity Abnormal forebrain electrographic activity (3–7 dpf) 	Clemizole identified in drug screen rescues seizure activity	Baraban et al. (2013)
			MO	 Hyperactivity (3–5 dpf): spontaneous electrographic activity (5 dpf) Increased sensitivity to hyperthermia (5–7 dpf) 	Fenfluramine and valproate rescue seizure activity	Zhang et al. (2015)
			ENU Mutant	Seizure activity	Fenfluramine and dimethadione rescue seizure activity	Dinday and Baraban (2015)
				Lower levels of serotonin in <i>scn1lab</i> —/— head homogenates at 7 dpf Zebrafish express orthologs of human serotonin (5HT) receptor subtypes at 5 dpf	5HT _{1D} , 5HT _{2C} , 5HT _{2A} -agonists, fenfluramine rescue seizure activity	Sourbron et al. (2016)
				Nighttime hyperactivity (5 dpf) Open field: Increased thigmotaxis, decreased movement (5 dpf) No differences in GABAergic neurons (5 dpf)	Clemizole, diazepam rescue nighttime hyperactivity, thigmotaxis	Grone et al. (2017)
				 5-HT_{2A} (<i>htt2aa/b</i>) and 5-HT_{2C} (<i>htt2c1</i>) orthologs expressed in larval heads in wild-type and <i>scn1lab</i> –/ – mutants and in adult wild-type brains Clemizole binds to serotonin receptors, 5HT_{2A} and 5HT_{2B}, in radioligand binding assay 	Trazodone and lorcaserin identified in screen of 5HT compounds rescue seizure activity	Griffin et al. (2017)
STXBP1	stxbp1a/b	EP ID	CRISPR Mutant	 stxbp1a-/-: Abnormal craniofacial development; failure to hatch; death by 10 dpf Darkened pigmentation Immobility; no response to dark flash (5-6 dpf) No spontaneous seizure-like behaviors or electrographic seizures Reduced heart rate and metabolism stxbp1b-/-: Darkened pigmentation No spontaneous seizure-like behaviors Electrographic seizure events in forebrain field potential recordings Reduced response to dark flash (5dpf) 	-	Grone et al. (2016)
STX1B	stx1b	EP	МО	 No swim bladder, pericardial edema, tail curvature (5 dpf) Lack of touch response in ~40% of embryos (4 dpf) Abnormal behaviors: repetitive fin, increased orofacial, myoclonic-like movement (5 dpf) Spontaneous epileptitorim activity on local field potentials of optic tecta: polyspiking discharges, high frequency oscillations (5 dpf) 	CNS-specific expression of human mRNA rescues epileptiform activity; mRNA with a human mutation (V216E) does not rescue	Schubert et al. (2014)
TRAPPC	B trappc6b	EP ASD	MO	 Microcephaly (2 dpf, 5 dpf) Increased apoptosis (24 hpf) Decreased baseline activity (5 dpf) Increased sensitivity to drug-induced seizures Increased neuronal activity at baseline and in response to seizure-inducing drug (5 dpf) 	Human mRNA rescues motility deficit	Marin-Valencia et al. (2018)

Annexe 2 : modèles de déficience intellectuelle chez le poisson zèbre 447.

Key: hpf, hours post fertilization; dpf, days post fertilization; * Disorder: ASD, autism spectrum disorder; DS, Dravet syndrome; EP, Epilepsy; ID, intellectual disability; ** Methods of Risk Gene Disruption: CRISPR, clustered regularly hterspaced short palindromic repeats; CRISPR F0, Mosaic CRISPR-injected embryo; ENU, N-ethyl N-nitrosourea-induced mutant; MO, Morpholino; OE, overexpression. Annexe 3 : Revue des principales bases de données utilisées pour aider à l'interprétation des tests génétiques en « population générale », associées à des maladies génétiques humaines et rapportant les variants de séquence ³²⁹.

Population Databases	
Exome Aggregation Consortium http://exac.broadinstitute.org/	Database of variants found during exome sequencing of 61,486 unrelated individuals sequenced as part of various disease-specific and population genetic studies. Pediatric disease subjects as well as related individuals were excluded.
Exome Variant Server http://evs.gs.washington.edu/EVS	Database of variants found during exome sequencing of several large cohorts of individuals of European and African American ancestry. Includes coverage data to inform the absence of variation.
1000 Genomes http://browser.1000genomes.org	Database of variants found during low-coverage and high-coverage genomic and targeted sequencing from 26 populations. Provides more diversity compared to EVS but also contains lower quality data and some cohorts contain related individuals.
dbSNP http://www-ncbi-nlm-nih-gov.gate2.inist.fr/snp	Database of short genetic variations (typically 50 bp or less) submitted from many sources. May lack details of originating study and may contain pathogenic variants.
dbVar http://www-ncbi-nlm-nih-gov.gate2.inist.fr/dbvar	Database of structural variation (typically greater than 50 bp) submitted from many sources.
Disease Databases	
ClinVar http://www-ncbi-nlm-nih-gov.gate2.inist.fr/clinvar	Database of assertions about the clinical significance and phenotype relationship of human variation.
OMIM http://www.omim.org	Database of human genes and genetic conditions that also contains a representative sampling of disease-associated genetic variants.
Human Gene Mutation Database http://www.hgmd.org	Database of variant annotations published in the literature. Requires fee-based subscription for much of the content.
Locus/Disease/Ethnic/Other-Specific Databases http://www.hgvs.org/dblist/dblist.html http://www.lovd.nl	The HGVS site developed a list of thousands of different databases that provide variant annotations on specific subsets of human variation. A large percentage of databases are built in the LOVD system.
DECIPHER	A molecular cytogenetic database for clinicians and

http://decipher.sanger.ac.uk	researchers linking genomic microarray data with phenotype using the Ensembl genome browser.
Sequence Databases	
NCBI Genome http://www-ncbi-nlm-nih-gov.gate2.inist.fr/genome	Source of full human genome reference sequences.
RefSeqGene http://www-ncbi-nlm-nih-gov.gate2.inist.fr/refseq/rsg and Locus Reference Genomic (LRG) http://www.lrg-sequence.org	Medically relevant gene reference sequence resource
MitoMap http://www.mitomap.org/MITOMAP/HumanMitoSeq	Revised Cambridge reference sequence (rCRS) for the Human Mitochondrial DNA

Annexe 4 : Outils et logiciels de prédiction in silico permettant d'aider à l'interprétation des variants de séquence faux sens, d'épissage ou de conservation ³²⁹.

Missense prediction	Name	Website	Basis
	ConSurf	http://consurftest.tau.ac.il	Evolutionary conservation
	FATHMM	http://fathmm.biocompute.org.uk	Evolutionary conservation
	MutationAss essor	http://mutationassessor.org	Evolutionary conservation
	PANTHER	http://www.pantherdb.org/tools/csnpScoreForm.jsp	Evolutionary conservation
	PhD-SNP	http://snps.biofold.org/phd-snp/phd-snp.html	Evolutionary conservation
	SIFT	http://sift.jcvi.org	Evolutionary conservation
	SNPs&GO	http://snps-and-go.biocomp.unibo.it/snps-and-go	Protein structure/function
	Align GVGD	http://agvgd.iarc.fr/agvgd_input.php	Protein structure/function and evolutionary conservation
	MAPP	http://mendel.stanford.edu.gate2.inist.fr/SidowLab/do wnloads/MAPP/index.html	Protein structure/function and evolutionary conservation
	MutationTas ter	http://www.mutationtaster.org	Protein structure/function and evolutionary conservation
	MutPred	http://mutpred.mutdb.org	Protein structure/function and evolutionary conservation
	PolyPhen-2	http://genetics.bwh.harvard.edu/pph2	Protein structure/function and evolutionary conservation
	PROVEAN	http://provean.jcvi.org/index.php	Alignment and measurement of similarity between variant sequence and protein sequence homolog
	nsSNPAnaly zer	http://snpanalyzer.uthsc.edu	Multiple sequence alignment and protein structure analysis
	Condel	http://bg.upf.edu/condel/home	Combines SIFT, PolyPhen-2 and MutationAssessor

Missense prediction	Name	Website	Basis
	CADD	http://cadd.gs.washington.edu	Contrasts annotations of fixed/nearly fixed derived alleles in humans with simulated variants
Splice site prediction			
	GeneSplicer	http://www.cbcb.umd.edu/software/GeneSplicer/gene	Markov models
	Human Splicing Finder	http://www.umd.be/HSF	Position-dependent logic
	MaxEntScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_s coreseq.html	Maximum entropy principle
	NetGene2	http://www.cbs.dtu.dk/services/NetGene2	Neural networks
	NNSplice	http://www.fruitfly.org/seq_tools/splice.html	Neural networks
	FSPLICE	http://www.softberry.com/berry.phtml?topic=fsplice &group=programs&subgroup=gfind	
Nucleotide conserva- tion prediction			
	GERP	http://mendel.stanford.edu.gate2.inist.fr/sidowlab/do wnloads/gerp/index.html	
	PhastCons	http://compgen.bscb.cornell.edu/phast/	
	PhyloP	http://compgen.bscb.cornell.edu/phast/	

Rule Description **Clarifications and/or Suggestions** variant predicted null where LOF is a do not apply to variants that are near the 3' end of the PVS1 mechanism of disease gene and escape nonsense-mediated decay does not include the same variant being assessed variant with the same amino acid change because it is not yet pathogenic, and the rule is PS1 as a previously established pathogenic intended for variants with a different nucleotide variant, regardless of nucleotide change change apply this rule as moderate or supporting if the variant de novo variant with confirmed PS2 is mosaic and its frequency in tissue is consistent with maternity and paternity the phenotype variant shown to have a deleterious reduce the strength for assays that are not as well PS3 effect by a well-established functional validated or linked to the phenotype studv variant located in a mutational hotspot not meant for truncations: more clarification is needed and/or critical and well-established PM1 for applying this rule functional domain cannot assume longer indels would be detected by next-generation sequencing use a published control dataset if its size is at least variant absent in population databases or PM2, with an allele frequency too high for the 1,000 individuals BS1 cannot be applied for low-quality calls or non-covered disease regions must define the condition and inheritance pattern invoke this rule as supporting if the phase is not for recessive disorders, variant in *trans* established PM3 with a pathogenic variant can upgrade if more than one proband is reported applicable for in-frame deletions, insertions, or stop-PM4 protein-length-changing variant loss variants, but not frameshifts, nonsense, and splice variants ensure pathogenicity of previously reported variant novel missense variant at amino acid suggest changing "novel" to "different" because some PM5 with different pathogenic missense variants that are not novel might require assessment change with this rule PP3, variant with multiple lines of all lines must agree BP4 computational evidence not meant to be used for genetically heterogeneous the patient's phenotype or family history conditions or conditions with unsolved etiology PP4 is highly specific to the genotype not typically applied for an analysis of incidental findings, but it could be applied for prior observations PP5. variant called pathogenic or benign by a only applicable when evidence is not available (e.g., BP6 Sharing Clinical Reports Project) reputable source variant observed in a healthy adult for a populations might not have been screened or excluded BS2 disorder with full penetrance at an early for the phenotype age variant in a gene in which truncations BP1 clarify the meaning of "primary"; suggest >90% primarily cause disease variant in *trans* with a dominant clarify that one should apply BP2 when the BP2, pathogenic variant (BP2) or in an pathogenic variant is seen in the same gene as the BP5 variant being evaluated and apply BP5 when the

pathogenic variant is in a different gene

individual with an alternate molecular

basis for disease (BP5)

Annexe 5 : règles de classification des variants proposes par l'ACMG-AMP et suggestions de modifications ³³⁰.

Abréviations : BS, benign strong (forte probabilité de bénignité); BP, benign supporting (argument faible en faveur du caractère bénin); FH, family history; LOF, loss-of-function; MAF, minor allele frequency; path., pathogenic; PM, pathogenic moderate (probabilité modérée); PP, pathogenic supporting (argument faible en faveur du caractère pathogène); PS, pathogenic strong (forte probabilité de pathogénicité); PVS, pathogenic very strong (très forte probabilité de pathogénicité).

Annexe 6 : règles utilisées pour combiner les critères et aider à la classification des variants ³²⁹.

Pathog	genic
1. 2. 3.	 Very Strong (PVS1) AND a. ≥1 Strong (PS1–PS4) OR b. ≥2 Moderate (PM1–PM6) OR c. 1 Moderate (PM1–PM6) and 1 Supporting (PP1–PP5) OR d. ≥2 Supporting (PP1–PP5) ≥2 Strong (PS1–PS4) OR 1 Strong (PS1–PS4) AND a. ≥3 Moderate (PM1–PM6) OR b. 2 Moderate (PM1–PM6) AND ≥2 Supporting (PP1–PP5) OR c. 1 Moderate (PM1–PM6) AND ≥4 Supporting (PP1–PP5)
Likely	Pathogenic
1. 2. 3.	1 Very Strong (PVS1) AND 1 Moderate (PM1–PM6) OR 1 Strong (PS1–PS4) AND 1–2 Moderate (PM1–PM6) OR 1 Strong (PS1–PS4) AND ≥2 Supporting (PP1–PP5) OR

- 4. \geq 3 Moderate (PM1–PM6) *OR*
- 5. 2 Moderate (PM1–PM6) $AND \ge 2$ Supporting (PP1–PP5) OR
- 6. 1 Moderate (PM1–PM6) $AND \ge 4$ Supporting (PP1–PP5)

Benign

- 1. 1 Stand-Alone (BA1) OR
- 2. ≥ 2 Strong (BS1–BS4)

Likely Benign

- 1. 1 Strong (BS1–BS4) and 1 Supporting (BP1–BP7) OR
- 2. \geq 2 Supporting (BP1–BP7)

Uncertain Significance

- 1. other criteria shown above are unmet OR
- 2. the criteria for benign and pathogenic are contradictory

Interacting protein	Comments on the interacting proteins	References
PDZ domains		
NR2A-D	Subunits of NMDA receptors	131
GluR6	Subunit of kainate receptors	132
δ2 GluR	Subunit of δ -ionotropic glutamate receptors	133
β1-adrenergic receptor	G-protein-coupled receptor	25
nAChRc	Subunit of neuronal nicotinic acetylcholine receptor	76,77
5-HT2A and 5-HT2C Rc	Subunits of 5-HT (serotonin) receptors	134
ErbB4	A receptor tyrosine kinase for neuregulin	135,136
Kv1	Voltage-gated potassium channel	14
Kir2, Kir3, Kir4 and Kir5	Inward-rectifying potassium channels	32,124
Neuroligin	A postsynaptic membrane protein that binds to β-neurexins and regulates synaptic adhesion and development	26,28,30
Stargazin family proteins	Tetra-spanning transmembrane proteins required for surface and synaptic expression of AMPA receptors	64
nNOS	Neuronal nitric oxide synthase	3,33
SynGAP	An abundant RasGAP of the PSD that regulates synaptic plasticity	35,36,39,40
Kalirin-7	A guanine nucleotide exchange factor for Rac1 that regulates spine morphogenesis	53
Fyn, Lyn, Src and Yes	Src family non-receptor protein tyrosine kinases; might also interact with the SH3 domain of PSD-95	48,49
Cypin	A cytosolic protein that regulates dendrite patterning by promoting microtubule assembly	137
CRIPT	A microtubule-binding protein	18
Sec8	A subunit of the exocyst complex involved in protein and vesicle trafficking	115
KIF1Bα	A motor of the kinesin superfamily	111
SH3 domain		
Pyk2	A non-receptor tyrosine kinase regulated by calcium and PKC and required for LTP induction	50,138
GK domain		
GKAP/SAPAP	An abundant multi-domain scaffold of PSD that links PSD-95 with Shank	139
SPAR	A postsynaptic RapGAP that regulates spine morphogenesis	57
SH3 and GK dom	ains	
KA2 GluR	Subunit of kainate receptors	132
AKAP79/150	An anchoring protein that binds to protein kinase A and protein phosphatase 1	41
L27 domain		
CASK	Mammalian homologue of LIN2	140,141
Myosin VI	A minus-end-directed actin-based motor	113

Annexe 7 : Protéines interagissant avec les domaines PDZ de PSD-95³⁵².

Only proteins that interact directly with PSD-95 family scaffolds are listed. These interactions might not apply to all members of the PSD-95 family. Owing to space limitations, this list is not comprehensive and not all relevant references are cited. AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid; GK, guanylate kinase-like domain; LTP, long-term potentiation; NMDA, *N*-methyl-*D*-aspartate; PKC, protein kinase C; PSD-95, postsynaptic density protein 95; Rac, Rap and Ras, small monomeric G-proteins; RapGAP, Rap GTPase-activating protein; RasGAP, Ras GTPase-activating protein; SH3 domain, Src homology 3 domain; Shank, SH3 and ankyrin repeat-containing protein.

Annexe 8 : Autres protéines contenant des domaines PDZ³⁵².

PDZ protein	Interacting protein(s)	References					
Neurabin, spinophilin/neurabin-II							
Localized in spines; modulates synaptic transmission and spine morphology	Protein phosphatase 1 F-actin	142					
Afadin							
Involved in synapse adhesion and development	Nectin (cell adhesion molecule) F-actin Eph receptors (receptor tyrosine kinas	143 ses)					
Densin-180							
Abundant PSD protein; member of LAP (leucine-rich repeat and PDZ) family of proteins	CaM kinase ΙΙα α-Actinin (F-actin-binding protein) δ-Catenin (N-cadherin-interacting pro	144 tein)					
Erbin							
LAP protein; suppresses the Ras–MAPK signalling pathway	ErbB2 (receptor tyrosine kinase for neuregulin) PSD-95 δ-Catenin	145					
S-SCAM							
Synaptic multi-PDZ scaffold; might regulate assembly and trafficking of synaptic proteins	NMDAR Neuroligin KIF1Bα β-Catenin (cadherin-associated protei nRapGEF (guanine nucleotide exchan factor for Rap1)	2 in) ge					
Shank							
Important scaffold protein of the PSD; promotes morphological and functional maturation of synapse and dendritic spine	GKAP Homer Cortactin (actin regulatory protein) CIRL (calcium-independent receptor for α-latrotoxin) IRSp53 (actin regulatory protein that binds Rac1 and Cdc42) ABP1 (F-actin-binding protein) βPIX (guanine nucleotide exchange factor for Rac1 and Cdc42) Sharpin (multimeric PSD protein)	59,146					
Syntenin							
Small scaffold protein that binds to phosphatidylinositol 4,5-bisphosphate	AMPA, kainate and metabotropic glutamate receptor subunits Syndecan (transmembrane proteoglyd Neurexin (neuronal surface proteins) SynCAM (synaptic cell adhesion mole Ephrin B Neurofascin (neural cell adhesion mole Merlin (product of the causal gene for neurofibromatosis type II)	94 can) cule) ecule)					
Tamalin							
Possibly involved in trafficking of mGLURs	Group I metabotropic glutamate receptor subunits Cytohesin (guanine nucleotide exchar factor for ARF small GTPases) GKAP S-SCAM	147 Ige					

Only proteins that directly interact with the indicated PDZ proteins are described. Owing to space limitations, this list is not comprehensive and not all relevant references are cited. AMPA, α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid; Cdc42, Rac, Rap and Ras, small monomeric G-proteins; GKAP, guanylate kinase-associated protein; KIF1B α , kinesin family member 1B α ; NMDAR, N-methylp-aspartate receptor; PSD-95, postsynaptic density protein 95; Ras–MAPK, Ras mitogen activated protein kinase; S-SCAM, synaptic scaffolding molecule.

REFERENCES

- 1. Wright, C. F., FitzPatrick, D. R. & Firth, H. V. Paediatric genomics: diagnosing rare disease in children. *Nat. Rev. Genet.* **19**, 253–268 (2018).
- Amberger, J. S., Bocchini, C. A., Scott, A. F. & Hamosh, A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* 47, D1038–D1043 (2019).
- 3. La définition des maladies rares | Fondation maladies rares. https://fondation-maladiesrares.org/lesmaladies-rares/les-maladies-rares-bis/la-definition-des-maladies-rares/.
- 4. Yoon, P. W. *et al.* Contribution of birth defects and genetic diseases to pediatric hospitalizations. A population-based study. *Arch. Pediatr. Adolesc. Med.* **151**, 1096–1103 (1997).
- 5. Dodge, J. A. *et al.* The importance of rare diseases: from the gene to society. *Arch. Dis. Child.* **96**, 791–792 (2011).
- 6. Rode, J. Rare Diseases: Understanding this Public Health Priority. Rare Dis. 14 (2005).
- 7. Lejeune, J., Turpin, R. & Gautier, M. [Chromosomic diagnosis of mongolism]. Arch. Fr. Pediatr. 16, 962–963 (1959).
- 8. Lubs, H. A. A marker X chromosome. Am. J. Hum. Genet. 21, 231–244 (1969).
- 9. Pieretti, M. *et al.* Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**, 817–822 (1991).
- 10. Verkerk, A. J. *et al.* 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**, 905–914 (1991).
- 11. Lubs, H. A., Stevenson, R. E. & Schwartz, C. E. Fragile X and X-linked intellectual disability: four decades of discovery. *Am. J. Hum. Genet.* **90**, 579–590 (2012).
- 12. Metzker, M. L. Sequencing technologies the next generation. Nat. Rev. Genet. 11, 31-46 (2010).
- 13. Venter, J. C. et al. The sequence of the human genome. Science 291, 1304–1351 (2001).
- 14. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
- 15. Lupski, J. R. *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* **362**, 1181–1191 (2010).
- 16. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).
- Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* 42, 790–793 (2010).
- 18. Mashimo, T. *et al.* Progressive Purkinje cell degeneration in tambaleante mutant mice is a consequence of a missense mutation in HERC1 E3 ubiquitin ligase. *PLoS Genet.* **5**, e1000784 (2009).
- 19. Nguyen, L. S. *et al.* A nonsense variant in HERC1 is associated with intellectual disability, megalencephaly, thick corpus callosum and cerebellar atrophy. *Eur. J. Hum. Genet. EJHG* **24**, 455–458 (2016).
- 20. Définition : les déficiences intellectuelles. http://www.euro.who.int/fr/health-topics/noncommunicablediseases/mental-health/news/news/2010/15/childrens-right-to-family-life/definition-intellectual-disability (2019) doi:http://www.euro.who.int/fr/health-topics/noncommunicable-diseases/mentalhealth/news/news/2010/15/childrens-right-to-family-life/definition-intellectual-disability.
- 21. Definition. https://aaidd.org/intellectual-disability/definition.
- 22. Defining intellectual disability: Finally we all agree... almost. *https://www.apa.org* https://www.apa.org/pi/disability/resources/publications/newsletter/2016/09/intellectual-disability.
- Borkowski, J. G., Carothers, S. S., Howard, K., Schatz, J. & Farris, J. R. Intellectual Assessment and Intellectual Disability. in *Handbook of Intellectual and Developmental Disabilities* (eds. Jacobson, J. W., Mulick, J. A. & Rojahn, J.) 261–277 (Springer US, 2007). doi:10.1007/0-387-32931-5_14.

- Campbell, J. M., Brown, R. T., Cavanagh, S. E., Vess, S. F. & Segall, M. J. Evidence-based assessment of cognitive functioning in pediatric psychology. *J. Pediatr. Psychol.* 33, 999–1014; discussion 1015-1020 (2008).
- 25. WISC-IV Clinical Use and Interpretation 1st Edition. https://www.elsevier.com/books/wisc-iv-clinical-use-and-interpretation/prifitera/978-0-12-564931-5.
- 26. Diagnostics & évaluations « Defiscience. https://www.defiscience.fr/diagnostic/.
- 27. Boulet, S. L., Schieve, L. A. & Boyle, C. A. Birth weight and health and developmental outcomes in US children, 1997-2005. *Matern. Child Health J.* **15**, 836–844 (2011).
- 28. Swanson, E. C. & Schleiss, M. R. Congenital cytomegalovirus infection: new prospects for prevention and therapy. *Pediatr. Clin. North Am.* **60**, 335–349 (2013).
- 29. Miranda-Filho, D. de B. *et al.* Initial Description of the Presumed Congenital Zika Syndrome. *Am. J. Public Health* **106**, 598–600 (2016).
- Malm, G., Forsgren, M., el Azazi, M. & Persson, A. A follow-up study of children with neonatal herpes simplex virus infections with particular regard to late nervous disturbances. *Acta Paediatr. Scand.* 80, 226–234 (1991).
- 31. Wilhelm, C. J. & Guizzetti, M. Fetal Alcohol Spectrum Disorders: An Overview from the Glia Perspective. *Front. Integr. Neurosci.* 9, (2016).
- 32. Embryo-Foetopathie au Valproate. *Haute Autorité de Santé* https://www.hassante.fr/jcms/c_2751900/fr/embryo-foetopathie-au-valproate.
- 33. Levy, H. L. & Ghavami, M. Maternal phenylketonuria: a metabolic teratogen. *Teratology* **53**, 176–184 (1996).
- 34. Waisbren, S. E. *et al.* Maternal Phenylketonuria: Long-term Outcomes in Offspring and Post-pregnancy Maternal Characteristics. *JIMD Rep.* **21**, 23–33 (2015).
- 35. Mendola, P., Selevan, S. G., Gutter, S. & Rice, D. Environmental factors associated with a spectrum of neurodevelopmental deficits. *Ment. Retard. Dev. Disabil. Res. Rev.* **8**, 188–197 (2002).
- 36. Miller, D. T. *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* **86**, 749–764 (2010).
- 37. Doco-Fenzy, M. et al. Version 1.0 Novembre 2010. 23.
- Driscoll, D. J., Miller, J. L., Schwartz, S. & Cassidy, S. B. Prader-Willi Syndrome. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 39. Dagli, A. I., Mueller, J. & Williams, C. A. Angelman Syndrome. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 40. Chiurazzi, P. & Pirozzi, F. Advances in understanding genetic basis of intellectual disability. *F1000Research* **5**, 599 (2016).
- 41. Tarpey, P. S. *et al.* A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat. Genet.* **41**, 535–543 (2009).
- 42. Piton, A., Redin, C. & Mandel, J.-L. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am. J. Hum. Genet.* **93**, 368–383 (2013).
- 43. Srivastava, A. K. & Schwartz, C. E. Intellectual disability and autism spectrum disorders: causal genes and molecular mechanisms. *Neurosci. Biobehav. Rev.* **46 Pt 2**, 161–174 (2014).
- 44. Yang, Y. *et al.* Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA* **312**, 1870 (2014).
- 45. Posey, J. E. *et al.* Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. *N. Engl. J. Med.* **376**, 21–31 (2017).
- 46. Yavarna, T. *et al.* High diagnostic yield of clinical exome sequencing in Middle Eastern patients with Mendelian disorders. *Hum. Genet.* **134**, 967–980 (2015).
- 47. Vissers, L. E. L. M. et al. A de novo paradigm for mental retardation. Nat. Genet. 42, 1109–1112 (2010).

- 48. Vissers, L. E. L. M. & Stankiewicz, P. Microdeletion and microduplication syndromes. *Methods Mol. Biol. Clifton NJ* 838, 29–75 (2012).
- 49. Vaillend, C., Poirier, R. & Laroche, S. Genes, plasticity and mental retardation. *Behav. Brain Res.* **192**, 88–105 (2008).
- 50. Hu, W. F., Chahrour, M. H. & Walsh, C. A. The diverse genetic landscape of neurodevelopmental disorders. *Annu. Rev. Genomics Hum. Genet.* **15**, 195–213 (2014).
- 51. van Bokhoven, H. Genetic and Epigenetic Networks in Intellectual Disabilities. *Annu. Rev. Genet.* **45**, 81–104 (2011).
- 52. Sweatt, J. D. Neural plasticity and behavior sixty years of conceptual advances. *J. Neurochem.* **139**, 179–199 (2016).
- 53. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234.e4 (2019).
- 54. Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 551–563 (2015).
- 55. Mignogna, M. L. & D'Adamo, P. Critical importance of RAB proteins for synaptic function. *Small GTPases* 9, 145–157 (2018).
- 56. Giannandrea, M. *et al.* Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. *Am. J. Hum. Genet.* **86**, 185–195 (2010).
- 57. Wilson, G. R. *et al.* Mutations in RAB39B cause X-linked intellectual disability and early-onset Parkinson disease with α-synuclein pathology. *Am. J. Hum. Genet.* **95**, 729–735 (2014).
- 58. Hughson, F. M. Neuroscience. Chaperones that SNARE neurotransmitter release. *Science* **339**, 406–407 (2013).
- 59. Ma, C., Su, L., Seven, A. B., Xu, Y. & Rizo, J. Reconstitution of the vital functions of Munc18 and Munc13 in neurotransmitter release. *Science* **339**, 421–425 (2013).
- 60. Saitsu, H. *et al.* De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. *Nat. Genet.* **40**, 782–788 (2008).
- 61. Baker, K. *et al.* Identification of a human synaptotagmin-1 mutation that perturbs synaptic vesicle cycling. *J. Clin. Invest.* **125**, 1670–1678 (2015).
- 62. Hackett, A. *et al.* CASK mutations are frequent in males and cause X-linked nystagmus and variable XLMR phenotypes. *Eur. J. Hum. Genet. EJHG* **18**, 544–552 (2010).
- 63. Bergmann, C. *et al.* Oligophrenin 1 (OPHN1) gene mutation causes syndromic X-linked mental retardation with epilepsy, rostral ventricular enlargement and cerebellar hypoplasia. *Brain J. Neurol.* **126**, 1537–1544 (2003).
- 64. Piton, A. *et al.* Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Hum. Mol. Genet.* **17**, 3965–3974 (2008).
- 65. Molinari, F. *et al.* Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation. *Science* **298**, 1779–1781 (2002).
- 66. Schaaf, C. P. *et al.* Phenotypic spectrum and genotype-phenotype correlations of NRXN1 exon deletions. *Eur. J. Hum. Genet. EJHG* **20**, 1240–1247 (2012).
- 67. Alarcón, M. *et al.* Linkage, association, and gene-expression analyses identify CNTNAP2 as an autismsusceptibility gene. *Am. J. Hum. Genet.* **82**, 150–159 (2008).
- 68. Zweier, C. *et al.* CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. *Am. J. Hum. Genet.* **85**, 655–666 (2009).
- 69. Smogavec, M. *et al.* Eight further individuals with intellectual disability and epilepsy carrying bi-allelic CNTNAP2 aberrations allow delineation of the mutational and phenotypic spectrum. *J. Med. Genet.* **53**, 820–827 (2016).
- 70. Tarpey, P. *et al.* Mutations in the DLG3 gene cause nonsyndromic X-linked mental retardation. *Am. J. Hum. Genet.* **75**, 318–324 (2004).

- 71. Reggiani, C. *et al.* Novel promoters and coding first exons in DLG2 linked to developmental disorders and intellectual disability. *Genome Med.* **9**, 67 (2017).
- 72. Lelieveld, S. H. *et al.* Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **19**, 1194–1196 (2016).
- 73. Paoletti, P., Bellone, C. & Zhou, Q. NMDA receptor subunit diversity: impact on receptor properties, synaptic plasticity and disease. *Nat. Rev. Neurosci.* **14**, 383–400 (2013).
- 74. Lemke, J. R. *et al.* Delineating the GRIN1 phenotypic spectrum: A distinct genetic NMDA receptor encephalopathy. *Neurology* **86**, 2171–2178 (2016).
- 75. Lemke, J. R. *et al.* GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. *Ann. Neurol.* **75**, 147–154 (2014).
- 76. Platzer, K. *et al.* GRIN2B encephalopathy: novel findings on phenotype, variant clustering, functional consequences and treatment aspects. *J. Med. Genet.* **54**, 460–470 (2017).
- 77. Li, D. *et al.* GRIN2D Recurrent De Novo Dominant Mutation Causes a Severe Epileptic Encephalopathy Treatable with NMDA Receptor Channel Blockers. *Am. J. Hum. Genet.* **99**, 802–816 (2016).
- 78. Lesca, G. *et al.* GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat. Genet.* **45**, 1061–1066 (2013).
- Lemke, J. R. *et al.* Mutations in GRIN2A cause idiopathic focal epilepsy with rolandic spikes. *Nat. Genet.* 45, 1067–1072 (2013).
- 80. Carvill, G. L. *et al.* GRIN2A mutations cause epilepsy-aphasia spectrum disorders. *Nat. Genet.* **45**, 1073–1076 (2013).
- 81. Hollmann, M. & Heinemann, S. Cloned glutamate receptors. Annu. Rev. Neurosci. 17, 31-108 (1994).
- 82. Wu, Y. *et al.* Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18163–18168 (2007).
- 83. Philips, A. K. *et al.* X-exome sequencing in Finnish families with intellectual disability--four novel mutations and two novel syndromic phenotypes. *Orphanet J. Rare Dis.* **9**, 49 (2014).
- 84. Hackmann, K. *et al.* Partial deletion of GLRB and GRIA2 in a patient with intellectual disability. *Eur. J. Hum. Genet. EJHG* **21**, 112–114 (2013).
- 85. Geisheker, M. R. *et al.* Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains. *Nat. Neurosci.* **20**, 1043–1051 (2017).
- 86. Martin, S. *et al.* De Novo Variants in GRIA4 Lead to Intellectual Disability with or without Seizures and Gait Abnormalities. *Am. J. Hum. Genet.* **101**, 1013–1020 (2017).
- 87. Glatt, K., Glatt, H. & Lalande, M. Structure and organization of GABRB3 and GABRA5. *Genomics* **41**, 63–69 (1997).
- 88. Bittiger, H., Froestl, W., Mickel, S. & Olpe, H. R. GABAB receptor antagonists: from synthesis to therapeutic applications. *Trends Pharmacol. Sci.* **14**, 391–394 (1993).
- 89. Goei, V. L. *et al.* Human gamma-aminobutyric acid B receptor gene: complementary DNA cloning, expression, chromosomal location, and genomic organization. *Biol. Psychiatry* **44**, 659–666 (1998).
- 90. Carvill, G. L. *et al.* GABRA1 and STXBP1: novel genetic causes of Dravet syndrome. *Neurology* **82**, 1245–1253 (2014).
- 91. Lachance-Touchette, P. *et al.* Novel α 1 and γ 2 GABAA receptor subunit mutations in families with idiopathic generalized epilepsy. *Eur. J. Neurosci.* **34**, 237–249 (2011).
- 92. Yoo, Y. *et al.* GABBR2 mutations determine phenotype in rett syndrome and epileptic encephalopathy. *Ann. Neurol.* **82**, 466–478 (2017).
- Louros, S. R. & Osterweil, E. K. Perturbed proteostasis in autism spectrum disorders. J. Neurochem. 139, 1081–1092 (2016).
- 94. Bhakar, A. L., Dölen, G. & Bear, M. F. The pathophysiology of fragile X (and what it teaches us about synapses). *Annu. Rev. Neurosci.* **35**, 417–443 (2012).

- 95. Darnell, J. C. & Klann, E. The translation of translational control by FMRP: therapeutic targets for FXS. *Nat. Neurosci.* **16**, 1530–1536 (2013).
- 96. Devys, D., Lutz, Y., Rouyer, N., Bellocq, J. P. & Mandel, J. L. The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat. Genet.* **4**, 335–340 (1993).
- 97. Schmidt, M. & Finley, D. Regulation of proteasome activity in health and disease. *Biochim. Biophys. Acta* **1843**, 13–25 (2014).
- 98. Ehlers, M. D. Activity level controls postsynaptic composition and signaling via the ubiquitin-proteasome system. *Nat. Neurosci.* **6**, 231–242 (2003).
- Kishino, T., Lalande, M. & Wagstaff, J. UBE3A/E6-AP mutations cause Angelman syndrome. *Nat. Genet.* 15, 70–73 (1997).
- 100. Matsuura, T. *et al.* De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nat. Genet.* **15**, 74–77 (1997).
- Riederer, B. M. Microtubule-associated protein 1B, a growth-associated and phosphorylated scaffold protein. *Brain Res. Bull.* 71, 541–558 (2007).
- 102. von Bohlen Und Halbach, O. Dendritic spine abnormalities in mental retardation. *Cell Tissue Res.* **342**, 317–323 (2010).
- Grossman, A. W., Elisseou, N. M., McKinney, B. C. & Greenough, W. T. Hippocampal pyramidal cells in adult Fmr1 knockout mice exhibit an immature-appearing profile of dendritic spines. *Brain Res.* 1084, 158–164 (2006).
- Chechlacz, M. & Gleeson, J. G. Is mental retardation a defect of synapse structure and function? *Pediatr. Neurol.* 29, 11–17 (2003).
- Newey, S. E., Velamoor, V., Govek, E.-E. & Van Aelst, L. Rho GTPases, dendritic structure, and mental retardation. *J. Neurobiol.* 64, 58–74 (2005).
- Govek, E.-E., Hatten, M. E. & Van Aelst, L. The role of Rho GTPase proteins in CNS neuronal migration. *Dev. Neurobiol.* 71, 528–553 (2011).
- 107. Martin-Vilchez, S. *et al.* RhoGTPase Regulators Orchestrate Distinct Stages of Synaptic Development. *PloS One* **12**, e0170464 (2017).
- 108. Banihani, R. *et al.* Cognitive and Neurobehavioral Profile in Boys With Duchenne Muscular Dystrophy. *J. Child Neurol.* **30**, 1472–1482 (2015).
- 109. Battini, R. *et al.* Cognitive profile in Duchenne muscular dystrophy boys without intellectual disability: The role of executive functions. *Neuromuscul. Disord. NMD* **28**, 122–128 (2018).
- 110. Wingeier, K. *et al.* Neuropsychological impairments and the impact of dystrophin mutations on general cognitive functioning of patients with Duchenne muscular dystrophy. *J. Clin. Neurosci. Off. J. Neurosurg. Soc. Australas.* **18**, 90–95 (2011).
- 111. Felisari, G. *et al.* Loss of Dp140 dystrophin isoform and intellectual impairment in Duchenne dystrophy. *Neurology* **55**, 559–564 (2000).
- 112. Taylor, P. J. *et al.* Dystrophin gene mutation location and the risk of cognitive impairment in Duchenne muscular dystrophy. *PloS One* **5**, e8803 (2010).
- 113. Juan-Mateu, J. *et al.* P.2.14 Isolated cognitive abnormalities associated to DMD mutations. *Neuromuscul. Disord.* **23**, 753 (2013).
- 114. de Brouwer, A. P. M. *et al.* A 3-base pair deletion, c.9711_9713del, in DMD results in intellectual disability without muscular dystrophy. *Eur. J. Hum. Genet. EJHG* **22**, 480–485 (2014).
- 115. Srivastava, S. & Naidu, S. Alexander Disease. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- Aoki, Y., Niihori, T., Inoue, S. & Matsubara, Y. Recent advances in RASopathies. J. Hum. Genet. 61, 33– 39 (2016).
- 117. Krab, L. C., Goorden, S. M. I. & Elgersma, Y. Oncogenes on my mind: ERK and MTOR signaling in cognitive diseases. *Trends Genet. TIG* 24, 498–510 (2008).
- Zhu, J. J., Qin, Y., Zhao, M., Van Aelst, L. & Malinow, R. Ras and Rap control AMPA receptor trafficking during synaptic plasticity. *Cell* 110, 443–455 (2002).
- Thomas, G. M. & Huganir, R. L. MAPK cascade signalling and synaptic plasticity. *Nat. Rev. Neurosci.* 5, 173–183 (2004).
- 120. Cui, Y. *et al.* Neurofibromin regulation of ERK signaling modulates GABA release and learning. *Cell* **135**, 549–560 (2008).
- 121. Kushner, S. A. *et al.* Modulation of presynaptic plasticity and learning by the H-ras/extracellular signalregulated kinase/synapsin I signaling pathway. *J. Neurosci. Off. J. Soc. Neurosci.* **25**, 9721–9734 (2005).
- 122. Davis, H. P. & Squire, L. R. Protein synthesis and memory: a review. Psychol. Bull. 96, 518-559 (1984).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).
- 124. Kundaje, A. *et al.* Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res.* 22, 1735–1747 (2012).
- Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.* 20, 259–266 (2013).
- 126. de Graaf, C. A. & van Steensel, B. Chromatin organization: form to function. *Curr. Opin. Genet. Dev.* 23, 185–190 (2013).
- 127. Kleefstra, T., Schenck, A., Kramer, J. M. & van Bokhoven, H. The genetics of cognitive epigenetics. *Neuropharmacology* **80**, 83–94 (2014).
- 128. Kleefstra, T. *et al.* Disruption of an EHMT1-Associated Chromatin-Modification Module Causes Intellectual Disability. *Am. J. Hum. Genet.* **91**, 73–82 (2012).
- 129. Kleefstra, T. & de Leeuw, N. Kleefstra Syndrome. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 130. Tatton-Brown, K. *et al.* Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability. *Am. J. Hum. Genet.* **100**, 725–736 (2017).
- Choufani, S. *et al.* NSD1 mutations generate a genome-wide DNA methylation signature. *Nat. Commun.* 6, 10207 (2015).
- 132. Schenkel, L. C. *et al.* Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome. *Epigenetics Chromatin* **10**, 10 (2017).
- 133. Schenkel, L. C. *et al.* Peripheral blood epi-signature of Claes-Jensen syndrome enables sensitive and specific identification of patients and healthy carriers with pathogenic mutations in KDM5C. *Clin. Epigenetics* **10**, 21 (2018).
- 134. Aref-Eshghi, E. *et al.* BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. *Nat. Commun.* **9**, 4885 (2018).
- Karpova, N. N., Sales, A. J. & Joca, S. R. Epigenetic Basis of Neuronal and Synaptic Plasticity. *Curr. Top. Med. Chem.* 17, 771–793 (2017).
- 136. Moretti, P. & Zoghbi, H. Y. MeCP2 dysfunction in Rett syndrome and related disorders. *Curr. Opin. Genet. Dev.* 16, 276–281 (2006).
- 137. Illingworth, R. S. & Bird, A. P. CpG islands--'a rough guide'. FEBS Lett. 583, 1713–1720 (2009).
- 138. Fuks, F. DNA methylation and histone modifications: teaming up to silence genes. *Curr. Opin. Genet. Dev.* **15**, 490–495 (2005).
- Samaco, R. C. & Neul, J. L. Complexities of Rett syndrome and MeCP2. J. Neurosci. Off. J. Soc. Neurosci. 31, 7951–7959 (2011).
- 140. Medrihan, L. *et al.* Early defects of GABAergic synapses in the brain stem of a MeCP2 mouse model of Rett syndrome. *J. Neurophysiol.* **99**, 112–121 (2008).
- Maezawa, I., Swanberg, S., Harvey, D., LaSalle, J. M. & Jin, L.-W. Rett syndrome astrocytes are abnormal and spread MeCP2 deficiency through gap junctions. *J. Neurosci. Off. J. Soc. Neurosci.* 29, 5051–5061 (2009).

- 142. Maezawa, I. & Jin, L.-W. Rett syndrome microglia damage dendrites and synapses by the elevated release of glutamate. *J. Neurosci. Off. J. Soc. Neurosci.* **30**, 5346–5356 (2010).
- Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 320, 1224–1229 (2008).
- Ballas, N., Lioy, D. T., Grunseich, C. & Mandel, G. Non-cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. *Nat. Neurosci.* 12, 311–317 (2009).
- 145. Nowaczyk, M. J. Smith-Lemli-Opitz Syndrome. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 146. Kaback, M. M. & Desnick, R. J. Hexosaminidase A Deficiency. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 147. Gomez-Ospina, N. Arylsulfatase A Deficiency. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 148. Orsini, J. J., Escolar, M. L., Wasserstein, M. P. & Caggana, M. Krabbe Disease. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 149. Wasserstein, M. P. & Schuchman, E. H. Acid Sphingomyelinase Deficiency. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 150. Matalon, R., Delgado, L. & Michals-Matalon, K. Canavan Disease. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 151. Regier, D. S. & Tifft, C. J. GLB1-Related Disorders. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 152. Clarke, L. A. Mucopolysaccharidosis Type I. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- Leroy, J. G., Cathey, S. S. & Friez, M. J. GNPTAB-Related Disorders. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 154. Mole, S. E. & Williams, R. E. Neuronal Ceroid-Lipofuscinoses. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 155. Steinberg, S. J., Raymond, G. V., Braverman, N. E. & Moser, A. B. Zellweger Spectrum Disorder. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 156. Sparks, S. E. & Krasnewich, D. M. Congenital Disorders of N-Linked Glycosylation and Multiple Pathway Overview. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 157. Marklová, E. & Albahri, Z. Screening and diagnosis of congenital disorders of glycosylation. *Clin. Chim. Acta Int. J. Clin. Chem.* **385**, 6–20 (2007).
- 158. Sanz-Nebot, V., Balaguer, E., Benavente, F., Neusüss, C. & Barbosa, J. Characterization of transferrin glycoforms in human serum by CE-UV and CE-ESI-MS. *Electrophoresis* **28**, 1949–1957 (2007).
- Sparks, S. E. & Krasnewich, D. M. PMM2-CDG (CDG-Ia). in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 160. Willems, A. P. *et al.* Mutations in N-acetylglucosamine (O-GlcNAc) transferase in patients with X-linked intellectual disability. *J. Biol. Chem.* **292**, 12621–12631 (2017).
- 161. Vaidyanathan, K. *et al.* Identification and characterization of a missense mutation in the O-linked β-N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. *J. Biol. Chem.* 292, 8948–8963 (2017).
- 162. Nguyen, T. T. M. *et al.* Mutations in PIGS, Encoding a GPI Transamidase, Cause a Neurological Syndrome Ranging from Fetal Akinesia to Epileptic Encephalopathy. *Am. J. Hum. Genet.* **103**, 602–611 (2018).
- 163. Murakami, Y. *et al.* Mutations in PIGB Cause an Inherited GPI Biosynthesis Defect with an Axonal Neuropathy and Metabolic Abnormality in Severe Cases. *Am. J. Hum. Genet.* **105**, 384–394 (2019).
- 164. Knaus, A. *et al.* Mutations in PIGU Impair the Function of the GPI Transamidase Complex, Causing Severe Intellectual Disability, Epilepsy, and Brain Anomalies. *Am. J. Hum. Genet.* **105**, 395–402 (2019).

- Harlalka, G. V. *et al.* Mutations in B4GALNT1 (GM2 synthase) underlie a new disorder of ganglioside biosynthesis. *Brain J. Neurol.* 136, 3618–3624 (2013).
- 166. Boccuto, L. *et al.* A mutation in a ganglioside biosynthetic enzyme, ST3GAL5, results in salt & pepper syndrome, a neurocutaneous disorder with altered glycolipid and glycoprotein glycosylation. *Hum. Mol. Genet.* 23, 418–433 (2014).
- 167. Lazaro, L. *et al.* Phenotypic and molecular variability of the holoprosencephalic spectrum. *Am. J. Med. Genet. A.* **129A**, 21–24 (2004).
- 168. Lacbawan, F. *et al.* Clinical spectrum of SIX3-associated mutations in holoprosencephaly: correlation between genotype, phenotype and function. *J. Med. Genet.* **46**, 389–398 (2009).
- 169. Mercier, S. *et al.* New findings for phenotype-genotype correlations in a large European series of holoprosencephaly cases. *J. Med. Genet.* **48**, 752–760 (2011).
- 170. Dubourg, C. et al. Recent advances in understanding inheritance of holoprosencephaly. Am. J. Med. Genet. C Semin. Med. Genet. 178, 258–269 (2018).
- 171. Sun, L. *et al.* Characterization of the human oncogene SCL/TAL1 interrupting locus (Stil) mediated Sonic hedgehog (Shh) signaling transduction in proliferating mammalian dopaminergic neurons. *Biochem. Biophys. Res. Commun.* **449**, 444–448 (2014).
- 172. Xavier, G. M. *et al.* Hedgehog receptor function during craniofacial development. *Dev. Biol.* **415**, 198–215 (2016).
- 173. OMIM Entry # 607432 LISSENCEPHALY 1; LIS1. https://www.omim.org/entry/607432.
- 174. Verloes, A. *et al.* Baraitser-Winter cerebrofrontofacial syndrome: delineation of the spectrum in 42 cases. *Eur. J. Hum. Genet. EJHG* **23**, 292–301 (2015).
- 175. OMIM Entry # 236670 MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 1; MDDGA1. https://www.omim.org/entry/236670.
- 176. Alcantara, D. & O'Driscoll, M. Congenital microcephaly. Am. J. Med. Genet. C Semin. Med. Genet. 166C, 124–139 (2014).
- 177. OMIM Entry # 251200 MICROCEPHALY 1, PRIMARY, AUTOSOMAL RECESSIVE; MCPH1. https://www.omim.org/entry/251200.
- 178. Chen, M. H. & Walsh, C. A. FLNA-Related Periventricular Nodular Heterotopia. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 179. Robertson, S. Otopalatodigital Spectrum Disorders. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- Mircsof, D. *et al.* Mutations in NONO lead to syndromic intellectual disability and inhibitory synaptic defects. *Nat. Neurosci.* 18, 1731–1736 (2015).
- Eng, L. F., Ghirnikar, R. S. & Lee, Y. L. Glial fibrillary acidic protein: GFAP-thirty-one years (1969-2000). *Neurochem. Res.* 25, 1439–1451 (2000).
- Messing, A. & Brenner, M. GFAP: functional implications gleaned from studies of genetically engineered mice. *Glia* 43, 87–90 (2003).
- 183. Stutterd, C. A., Dobyns, W. B., Jansen, A., Mirzaa, G. & Leventer, R. J. Polymicrogyria Overview. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 184. Rivière, J.-B. *et al.* De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat. Genet.* **44**, 934–940 (2012).
- Plaisier, E. & Ronco, P. COL4A1-Related Disorders. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 186. Stumpel, C. & Vos, Y. J. L1 Syndrome. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 187. Shaheen, R. *et al.* The genetic landscape of familial congenital hydrocephalus. *Ann. Neurol.* **81**, 890–897 (2017).

- Flores-Sarnat, L., Sarnat, H. B., Dávila-Gutiérrez, G. & Alvarez, A. Hemimegalencephaly: part 2. Neuropathology suggests a disorder of cellular lineage. J. Child Neurol. 18, 776–785 (2003).
- 189. Gowda, S. *et al.* Surgery for catastrophic epilepsy in infants 6 months of age and younger. *J. Neurosurg. Pediatr.* **5**, 603–607 (2010).
- 190. Ribierre, T. *et al.* Second-hit mosaic mutation in mTORC1 repressor DEPDC5 causes focal cortical dysplasia–associated epilepsy. *J. Clin. Invest.* **128**, 2452–2458 (2018).
- 191. Northrup, H., Koenig, M. K., Pearson, D. A. & Au, K. S. Tuberous Sclerosis Complex. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
- 192. French, J. A. *et al.* Adjunctive everolimus therapy for treatment-resistant focal-onset seizures associated with tuberous sclerosis (EXIST-3): a phase 3, randomised, double-blind, placebo-controlled study. *The Lancet* **388**, 2153–2163 (2016).
- 193. Zanni, G. *et al.* Oligophrenin 1 mutations frequently cause X-linked mental retardation with cerebellar hypoplasia. *Neurology* **65**, 1364–1369 (2005).
- 194. OMIM Entry # 213300 JOUBERT SYNDROME 1; JBTS1. https://www.omim.org/entry/213300.
- 195. OMIM Entry # 216360 COACH SYNDROME. https://www.omim.org/entry/216360.
- 196. Tarpey, P. S. *et al.* Mutations in the gene encoding the Sigma 2 subunit of the adaptor protein 1 complex, AP1S2, cause X-linked mental retardation. *Am. J. Hum. Genet.* **79**, 1119–1124 (2006).
- 197. Seidahmed, M. Z. *et al.* Ritscher-Schinzel (cranio-cerebello-cardiac, 3C) syndrome: report of four new cases with renal involvement. *Am. J. Med. Genet. A.* **155A**, 1393–1397 (2011).
- 198. OMIM Entry # 607596 PONTOCEREBELLAR HYPOPLASIA, TYPE 1A; PCH1A. https://www.omim.org/entry/607596.
- Paushkin, S. V., Patel, M., Furia, B. S., Peltz, S. W. & Trotta, C. R. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation. *Cell* 117, 311–321 (2004).
- 200. Brouwer, R. et al. Three novel components of the human exosome. J. Biol. Chem. 276, 6177-6184 (2001).
- 201. Chen, C. Y. *et al.* AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* **107**, 451–464 (2001).
- 202. Renbaum, P. *et al.* Spinal Muscular Atrophy with Pontocerebellar Hypoplasia Is Caused by a Mutation in the VRK1 Gene. *Am. J. Hum. Genet.* **85**, 281–289 (2009).
- 203. OMIM Entry # 312080 PELIZAEUS-MERZBACHER DISEASE; PMD. https://www.omim.org/entry/312080.
- 204. Bugiani, M., Vuong, C., Breur, M. & van der Knaap, M. S. Vanishing white matter: a leukodystrophy due to astrocytic dysfunction. *Brain Pathol. Zurich Switz.* **28**, 408–421 (2018).
- 205. Autism Diagnosis Criteria: DSM-5. Autism Speaks https://www.autismspeaks.org/autism-diagnosiscriteria-dsm5.
- 206. Wiggins, L. D. *et al.* DSM-5 criteria for autism spectrum disorder maximizes diagnostic sensitivity and specificity in preschool children. *Soc. Psychiatry Psychiatr. Epidemiol.* **54**, 693–701 (2019).
- 207. Loomes, R., Hull, L. & Mandy, W. P. L. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 466–474 (2017).
- 208. Rylaarsdam, L. & Guemez-Gamboa, A. Genetic Causes and Modifiers of Autism Spectrum Disorder. *Front. Cell. Neurosci.* **13**, 385 (2019).
- 209. Iakoucheva, L. M., Muotri, A. R. & Sebat, J. Getting to the Cores of Autism. Cell 178, 1287–1298 (2019).
- 210. Deneault, E. *et al.* Complete Disruption of Autism-Susceptibility Genes by Gene Editing Predominantly Reduces Functional Connectivity of Isogenic Human Neurons. *Stem Cell Rep.* **11**, 1211–1225 (2018).
- Torres, F., Barbosa, M. & Maciel, P. Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. *J. Med. Genet.* 53, 73–90 (2016).

- England, M. J., Liverman, C. T., Schultz, A. M. & Strawbridge, L. M. Epilepsy across the spectrum: promoting health and understanding. A summary of the Institute of Medicine report. *Epilepsy Behav. EB* 25, 266–276 (2012).
- 213. Hildebrand, M. S. *et al.* Recent advances in the molecular genetics of epilepsy. *J. Med. Genet.* **50**, 271–279 (2013).
- 214. Berg, A. T. *et al.* Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005-2009. *Epilepsia* **51**, 676–685 (2010).
- 215. Commission on Classification and Terminology // International League Against Epilepsy. https://www.ilae.org/about-ilae/topical-commissions/yes/commission-on-classification-and-terminology?amp;&=amp%3B.
- 216. Myers, C. T. & Mefford, H. C. Advancing epilepsy genetics in the genomic era. *Genome Med.* **7**, 91 (2015).
- 217. McTague, A., Howell, K. B., Cross, J. H., Kurian, M. A. & Scheffer, I. E. The genetic landscape of the epileptic encephalopathies of infancy and childhood. *Lancet Neurol.* **15**, 304–316 (2016).
- 218. von Deimling, M., Helbig, I. & Marsh, E. D. Epileptic Encephalopathies-Clinical Syndromes and Pathophysiological Concepts. *Curr. Neurol. Neurosci. Rep.* **17**, 10 (2017).
- 219. Orsini, A., Zara, F. & Striano, P. Recent advances in epilepsy genetics. Neurosci. Lett. 667, 4-9 (2018).
- 220. Mefford, H. C. *et al.* Rare copy number variants are an important cause of epileptic encephalopathies. *Ann. Neurol.* **70**, 974–985 (2011).
- 221. Olson, H. *et al.* Copy number variation plays an important role in clinical epilepsy. *Ann. Neurol.* **75**, 943–958 (2014).
- 222. Lal, D. *et al.* Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. *PLoS Genet.* **11**, e1005226 (2015).
- Axeen, E. J. T. & Olson, H. E. Neonatal epilepsy genetics. Semin. Fetal. Neonatal Med. 23, 197–203 (2018).
- 224. Leung, H. T. T. & Ring, H. Epilepsy in four genetically determined syndromes of intellectual disability: Epilepsy in genetic ID syndromes. *J. Intellect. Disabil. Res.* **57**, 3–20 (2013).
- Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat. Rev. Genet.* 17, 9–18 (2016).
- 226. Loeys, B. L. *et al.* The revised Ghent nosology for the Marfan syndrome. *J. Med. Genet.* **47**, 476–485 (2010).
- 227. Loeys, B. *et al.* Comprehensive molecular screening of the FBN1 gene favors locus homogeneity of classical Marfan syndrome. *Hum. Mutat.* **24**, 140–146 (2004).
- 228. Faivre, L. *et al.* Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study. *Am. J. Hum. Genet.* **81**, 454–466 (2007).
- Callier, P. *et al.* Systematic molecular and cytogenetic screening of 100 patients with marfanoid syndromes and intellectual disability: Systematic molecular and cytogenetic screening. *Clin. Genet.* 84, 507–521 (2013).
- 230. Kalueff, A. V., Wheaton, M. & Murphy, D. L. What's wrong with my mouse model? *Behav. Brain Res.* **179**, 1–18 (2007).
- 231. Stewart, A. M. & Kalueff, A. V. Developing better and more valid animal models of brain disorders. *Behav. Brain Res.* **276**, 28–31 (2015).
- 232. Mouse Genome Sequencing Consortium *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355–364 (2014).
- 234. Stergachis, A. B. *et al.* Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* **515**, 365–370 (2014).

- Bouabe, H. & Okkenhaug, K. Gene Targeting in Mice: A Review. in *Virus-Host Interactions* (eds. Bailer, S. M. & Lieber, D.) vol. 1064 315–336 (Humana Press, 2013).
- 236. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).
- 237. Lander, E. S. The Heroes of CRISPR. Cell 164, 18-28 (2016).
- 238. Simon, P., Dupuis, R. & Costentin, J. Thigmotaxis as an index of anxiety in mice. Influence of dopaminergic transmissions. *Behav. Brain Res.* **61**, 59–64 (1994).
- 239. Singer, P. *et al.* Prepulse inhibition predicts working memory performance whilst startle habituation predicts spatial reference memory retention in C57BL/6 mice. *Behav. Brain Res.* 242, 166–177 (2013).
- 240. Rabaneda, L. G., Robles-Lanuza, E., Nieto-González, J. L. & Scholl, F. G. Neurexin Dysfunction in Adult Neurons Results in Autistic-like Behavior in Mice. *Cell Rep.* **8**, 338–346 (2014).
- 241. Sørensen, E. M. *et al.* Hyperactivity and lack of social discrimination in the adolescent Fmr1 knockout mouse: *Behav. Pharmacol.* **26**, 733–740 (2015).
- 242. Koekkoek, S. K. E. *et al.* Deletion of FMR1 in Purkinje Cells Enhances Parallel Fiber LTD, Enlarges Spines, and Attenuates Cerebellar Eyelid Conditioning in Fragile X Syndrome. *Neuron* **47**, 339–352 (2005).
- 243. Amiri, A. *et al.* Analysis of *Fmr1* Deletion in a Subpopulation of Post-Mitotic Neurons in Mouse Cortex and Hippocampus: *Fmr1* deletion in cortex and hippocampus. *Autism Res.* **7**, 60–71 (2014).
- 244. Mientjes, E. J. Fxr1 knockout mice show a striated muscle phenotype: implications for Fxr1p function in vivo. *Hum. Mol. Genet.* **13**, 1291–1302 (2004).
- 245. Bontekoe, C. J. M. *et al.* Knockout mouse model for Fxr2: a model for mental retardation. *Hum. Mol. Genet.* **11**, 487–498 (2002).
- Spencer, C. M. *et al.* Exaggerated behavioral phenotypes in Fmr1/Fxr2 double knockout mice reveal a functional genetic interaction between Fragile X-related proteins. *Hum. Mol. Genet.* 15, 1984–1994 (2006).
- 247. Zhang, J. *et al.* Fragile X-related proteins regulate mammalian circadian behavioral rhythms. *Am. J. Hum. Genet.* **83**, 43–52 (2008).
- 248. Lewis, J. D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905–914 (1992).
- 249. Chen, R. Z., Akbarian, S., Tudor, M. & Jaenisch, R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.* **27**, 327–331 (2001).
- 250. Oortveld, M. A. W. *et al.* Human Intellectual Disability Genes Form Conserved Functional Modules in Drosophila. *PLoS Genet.* **9**, e1003911 (2013).
- 251. Tian, Y., Zhang, Z. C. & Han, J. Drosophila Studies on Autism Spectrum Disorders. *Neurosci. Bull.* **33**, 737–746 (2017).
- 252. Coll-Tané, M., Krebbers, A., Castells-Nobau, A., Zweier, C. & Schenck, A. Intellectual disability and autism spectrum disorders 'on the fly': insights from *Drosophila*. *Dis*. *Model*. *Mech.* **12**, dmm039180 (2019).
- 253. Pardo, C. A. & Eberhart, C. G. The Neurobiology of Autism. Brain Pathol. 17, 434-447 (2007).
- 254. Stephenson, J. R. *et al.* A Novel Human *CAMK2A* Mutation Disrupts Dendritic Morphology and Synaptic Transmission, and Causes ASD-Related Behaviors. *J. Neurosci.* **37**, 2216–2233 (2017).
- 255. Ori-McKenney, K. M. *et al.* Phosphorylation of β-Tubulin by the Down Syndrome Kinase, Minibrain/DYRK1a, Regulates Microtubule Dynamics and Dendrite Morphogenesis. *Neuron* **90**, 551–563 (2016).
- 256. Campbell, R. A. A. & Turner, G. C. The mushroom body. Curr. Biol. 20, R11-R12 (2010).
- 257. Farris, S. M. Are mushroom bodies cerebellum-like structures? *Arthropod Struct. Dev.* **40**, 368–379 (2011).

- 258. Tomer, R., Denes, A. S., Tessmar-Raible, K. & Arendt, D. Profiling by Image Registration Reveals Common Origin of Annelid Mushroom Bodies and Vertebrate Pallium. *Cell* **142**, 800–809 (2010).
- 259. Fenckova, M. et al. Integrative cross-species analyses identify deficits in habituation learning as a widely affected mechanism in Autism. http://biorxiv.org/lookup/doi/10.1101/285981 (2018) doi:10.1101/285981.
- Li, Y. *et al. Drosophila* Neuroligin 4 Regulates Sleep through Modulating GABA Transmission. J. Neurosci. 33, 15545–15554 (2013).
- 261. Dockendorff, T. C. *et al.* Drosophila Lacking dfmr1 Activity Show Defects in Circadian Output and Fail to Maintain Courtship Interest. *Neuron* **34**, 973–984 (2002).
- 262. Zhang, Y. Q. *et al.* Drosophila Fragile X-Related Gene Regulates the MAP1B Homolog Futsch to Control Synaptic Structure and Function. *Cell* **107**, 591–603 (2001).
- Michel, C. I., Kraft, R. & Restifo, L. L. Defective neuronal development in the mushroom bodies of Drosophila fragile X mental retardation 1 mutants. *J. Neurosci. Off. J. Soc. Neurosci.* 24, 5798–5809 (2004).
- 264. Bolduc, F. V., Bell, K., Cox, H., Broadie, K. S. & Tully, T. Excess protein synthesis in Drosophila Fragile X mutants impairs long-term memory. *Nat. Neurosci.* **11**, 1143–1145 (2008).
- 265. Bolduc, F. V., Valente, D., Nguyen, A. T., Mitra, P. P. & Tully, T. An assay for social interaction in Drosophila fragile X mutants. *Fly (Austin)* **4**, 216–225 (2010).
- Granato, M. & Nüsslein-Volhard, C. Fishing for genes controlling development. *Curr. Opin. Genet. Dev.* 6, 461–468 (1996).
- Burgess, H. A. & Granato, M. Sensorimotor Gating in Larval Zebrafish. J. Neurosci. 27, 4984–4994 (2007).
- 268. Renier, C. *et al.* Genomic and functional conservation of sedative-hypnotic targets in the zebrafish: *Pharmacogenet. Genomics* **17**, 237–253 (2007).
- 269. Guo, S. Using zebrafish to assess the impact of drugs on neural development and function. *Expert Opin. Drug Discov.* **4**, 715–726 (2009).
- 270. Rihel, J. *et al.* Zebrafish Behavioral Profiling Links Drugs to Biological Targets and Rest/Wake Regulation. *Science* **327**, 348–351 (2010).
- 271. Lovett-Barron, M. *et al.* Ancestral Circuits for the Coordinated Modulation of Brain State. *Cell* **171**, 1411-1423.e17 (2017).
- 272. Bayés, À. *et al.* Evolution of complexity in the zebrafish synapse proteome. *Nat. Commun.* **8**, 14613 (2017).
- 273. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
- 274. Barbazuk, W. B. *et al.* The syntenic relationship of the zebrafish and human genomes. *Genome Res.* **10**, 1351–1358 (2000).
- 275. Lu, J., Peatman, E., Tang, H., Lewis, J. & Liu, Z. Profiling of gene duplication patterns of sequenced teleost genomes: evidence for rapid lineage-specific genome expansion mediated by recent tandem duplications. *BMC Genomics* **13**, 246 (2012).
- 276. Postlethwait, J., Amores, A., Cresko, W., Singer, A. & Yan, Y.-L. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet. TIG* **20**, 481–490 (2004).
- 277. Moens, C. B., Donn, T. M., Wolf-Saxon, E. R. & Ma, T. P. Reverse genetics in zebrafish by TILLING. *Brief. Funct. Genomic. Proteomic.* **7**, 454–459 (2008).
- 278. Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. & Wolfe, S. A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* **26**, 695–701 (2008).
- 279. Sander, J. D. *et al.* Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.* **29**, 697–698 (2011).
- Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 227–229 (2013).

- 281. Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **12**, 982–988 (2015).
- 282. Eisen, J. S. & Smith, J. C. Controlling morpholino experiments: don't stop making antisense. *Development* 135, 1735–1743 (2008).
- 283. Stainier, D. Y. R. et al. Guidelines for morpholino use in zebrafish. PLOS Genet. 13, e1007000 (2017).
- 284. Kozol, R. A. *et al.* Two knockdown models of the autism genes SYNGAP1 and SHANK3 in zebrafish produce similar behavioral phenotypes associated with embryonic disruptions of brain morphogenesis. *Hum. Mol. Genet.* **24**, 4006–4023 (2015).
- 285. Pietri, T. *et al.* The first mecp2-null zebrafish model shows altered motor behaviors. *Front. Neural Circuits* **7**, (2013).
- 286. Bassuk, A. G. *et al.* A Homozygous Mutation in Human PRICKLE1 Causes an Autosomal-Recessive Progressive Myoclonus Epilepsy-Ataxia Syndrome. *Am. J. Hum. Genet.* **83**, 572–581 (2008).
- 287. Erdogan, B., Ebbert, P. T. & Lowery, L. A. Using Xenopus laevis retinal and spinal neurons to study mechanisms of axon guidance in vivo and in vitro. *Semin. Cell Dev. Biol.* **51**, 64–72 (2016).
- 288. Session, A. M. *et al.* Genome evolution in the allotetraploid frog Xenopus laevis. *Nature* **538**, 336–343 (2016).
- 289. Hellsten, U. *et al.* The Genome of the Western Clawed Frog Xenopus tropicalis. *Science* **328**, 633–636 (2010).
- James-Zorn, C. *et al.* Navigating Xenbase: An Integrated Xenopus Genomics and Gene Expression Database. in *Eukaryotic Genomic Databases* (ed. Kollmar, M.) vol. 1757 251–305 (Springer New York, 2018).
- 291. Huot, M.-E., Bisson, N., Moss, T. & Khandjian, E. W. Manipulating the fragile X mental retardation proteins in the frog. *Results Probl. Cell Differ.* 54, 165–179 (2012).
- 292. Kratzer, M.-C., England, L., Apel, D., Hassel, M. & Borchers, A. Evolution of the Rho guanine nucleotide exchange factors Kalirin and Trio and their gene expression in Xenopus development. *Gene Expr. Patterns GEP* **32**, 18–27 (2019).
- 293. Marshak, S., Meynard, M. M., De Vries, Y. A., Kidane, A. H. & Cohen-Cory, S. Cell-autonomous alterations in dendritic arbor morphology and connectivity induced by overexpression of MeCP2 in Xenopus central neurons in vivo. *PloS One* **7**, e33153 (2012).
- 294. Lasser, M., Pratt, B., Monahan, C., Kim, S. W. & Lowery, L. A. The Many Faces of Xenopus: Xenopus laevis as a Model System to Study Wolf–Hirschhorn Syndrome. *Front. Physiol.* **10**, 817 (2019).
- 295. SCN1A Seizure Disorders GeneReviews® NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK1318/.
- 296. Yu, F. H. *et al.* Reduced sodium current in GABAergic interneurons in a mouse model of severe myoclonic epilepsy in infancy. *Nat. Neurosci.* **9**, 1142–1149 (2006).
- 297. Ogiwara, I. *et al.* Nav1.1 Localizes to Axons of Parvalbumin-Positive Inhibitory Interneurons: A Circuit Basis for Epileptic Seizures in Mice Carrying an Scn1a Gene Mutation. *J. Neurosci.* **27**, 5903–5914 (2007).
- 298. Grone, B. P., Qu, T. & Baraban, S. C. Behavioral Comorbidities and Drug Treatments in a Zebrafish *scn1lab* Model of Dravet Syndrome. *eneuro* **4**, ENEURO.0066-17.2017 (2017).
- 299. Ruffolo, G. *et al.* Rare Diseases of Neurodevelopment: Maintain the Mystery or Use a Dazzling Tool for Investigation? The Case of Rett Syndrome. *Neuroscience* (2019) doi:10.1016/j.neuroscience.2019.06.015.
- 300. Barcia, G. *et al.* De novo gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy. *Nat. Genet.* **44**, 1255–1259 (2012).
- McTague, A. *et al.* Clinical and molecular characterization of KCNT1-related severe early-onset epilepsy. *Neurology* 90, e55–e66 (2018).
- 302. Sicca, F. *et al.* Gain-of-function defects of astrocytic Kir4.1 channels in children with autism spectrum disorders and epilepsy. *Sci. Rep.* **6**, 34325 (2016).

- 303. Bear, M. F., Huber, K. M. & Warren, S. T. The mGluR theory of fragile X mental retardation. *Trends Neurosci.* **27**, 370–377 (2004).
- 304. Dölen, G. et al. Correction of fragile X syndrome in mice. Neuron 56, 955–962 (2007).
- 305. Michalon, A. *et al.* Chronic pharmacological mGlu5 inhibition corrects fragile X in adult mice. *Neuron* **74**, 49–56 (2012).
- 306. McBride, S. M. J. *et al.* Pharmacological Rescue of Synaptic Plasticity, Courtship Behavior, and Mushroom Body Defects in a Drosophila Model of Fragile X Syndrome. *Neuron* **45**, 753–764 (2005).
- 307. Yuskaitis, C. J. *et al.* Lithium ameliorates altered glycogen synthase kinase-3 and behavior in a mouse model of fragile X syndrome. *Biochem. Pharmacol.* **79**, 632–646 (2010).
- 308. Liu, Z.-H., Chuang, D.-M. & Smith, C. B. Lithium ameliorates phenotypic deficits in a mouse model of fragile X syndrome. *Int. J. Neuropsychopharmacol.* **14**, 618–630 (2011).
- 309. Liu, Z.-H., Huang, T. & Smith, C. B. Lithium reverses increased rates of cerebral protein synthesis in a mouse model of fragile X syndrome. *Neurobiol. Dis.* **45**, 1145–1152 (2012).
- 310. Franklin, A. V. *et al.* Glycogen synthase kinase-3 inhibitors reverse deficits in long-term potentiation and cognition in fragile X mice. *Biol. Psychiatry* **75**, 198–206 (2014).
- 311. Bilousova, T. V. *et al.* Minocycline promotes dendritic spine maturation and improves behavioural performance in the fragile X mouse model. *J. Med. Genet.* **46**, 94–102 (2009).
- 312. Osterweil, E. K. *et al.* Lovastatin corrects excess protein synthesis and prevents epileptogenesis in a mouse model of fragile X syndrome. *Neuron* **77**, 243–250 (2013).
- 313. Cerezo-Guisado, M. I. *et al.* Lovastatin inhibits the extracellular-signal-regulated kinase pathway in immortalized rat brain neuroblasts. *Biochem. J.* **401**, 175–183 (2007).
- 314. Erickson, C. A. *et al.* Fragile X targeted pharmacotherapy: lessons learned and future directions. *J. Neurodev. Disord.* **9**, 7 (2017).
- 315. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of Neurological Defects in a Mouse Model of Rett Syndrome. *Science* **315**, 1143–1147 (2007).
- 316. Chang, Q., Khare, G., Dani, V., Nelson, S. & Jaenisch, R. The disease progression of Mecp2 mutant mice is affected by the level of BDNF expression. *Neuron* **49**, 341–348 (2006).
- 317. Tropea, D. *et al.* Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. *Proc. Natl. Acad. Sci.* **106**, 2029–2034 (2009).
- 318. Bruni, G. *et al.* Zebrafish behavioral profiling identifies multitarget antipsychotic-like compounds. *Nat. Chem. Biol.* **12**, 559–566 (2016).
- 319. Hoffman, E. J. *et al.* Estrogens Suppress a Behavioral Phenotype in Zebrafish Mutants of the Autism Risk Gene, CNTNAP2. *Neuron* **89**, 725–733 (2016).
- 320. Baraban, S. C., Dinday, M. T. & Hortopan, G. A. Drug screening in Scn1a zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. *Nat. Commun.* **4**, 2410 (2013).
- 321. Dinday, M. T. & Baraban, S. C. Large-Scale Phenotype-Based Antiepileptic Drug Screening in a Zebrafish Model of Dravet Syndrome. *eneuro* **2**, ENEURO.0068-15.2015 (2015).
- 322. ZX008. Zogenix https://www.zogenix.com/pipeline/zx008/.
- 323. Lin, S. *et al.* Comparison of the transcriptional landscapes between human and mouse tissues. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17224–17229 (2014).
- 324. Kwan, K. Y. *et al.* Species-dependent posttranscriptional regulation of NOS1 by FMRP in the developing cerebral cortex. *Cell* **149**, 899–911 (2012).
- 325. Tucker, B., Richards, R. I. & Lardelli, M. Contribution of mGluR and Fmr1 functional pathways to neurite morphogenesis, craniofacial development and fragile X syndrome. *Hum. Mol. Genet.* 15, 3446–3458 (2006).
- 326. den Broeder, M. J. *et al.* Generation and Characterization of Fmr1 Knockout Zebrafish. *PLoS ONE* **4**, e7910 (2009).

- 327. Veenstra-VanderWeele, J. & Warren, Z. Intervention in the context of development: pathways toward new treatments. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* **40**, 225–237 (2015).
- 328. Richards, C. S. *et al.* ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet. Med.* **10**, 294–300 (2008).
- 329.; on behalf of the ACMG Laboratory Quality Assurance Committee *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–423 (2015).
- Amendola, L. M. *et al.* Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am. J. Hum. Genet.* 98, 1067– 1076 (2016).
- Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* 12, 745–755 (2011).
- 332. Wright, C. F. *et al.* Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet Lond. Engl.* **385**, 1305–1314 (2015).
- 333. Orphanomix Physicians' Group *et al.* Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genet. Med.* 20, 645–654 (2018).
- Thevenon, J. *et al.* Diagnostic odyssey in severe neurodevelopmental disorders: toward clinical wholeexome sequencing as a first-line diagnostic test: Diagnostic odyssey in severe neurodevelopmental disorders. *Clin. Genet.* 89, 700–707 (2016).
- 335. Clark, M. M. *et al.* Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *Npj Genomic Med.* **3**, 16 (2018).
- Yang, Y. *et al.* Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *N. Engl. J. Med.* 369, 1502–1511 (2013).
- Lee, H. *et al.* Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *JAMA* 312, 1880 (2014).
- 338. Farwell, K. D. *et al.* Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet. Med.* 17, 578–586 (2015).
- 339. Sawyer, S. L. *et al.* Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care: Whole-exome sequencing for rare disease diagnosis. *Clin. Genet.* 89, 275– 284 (2016).
- 340. Wright, C. F. *et al.* Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **20**, 1216–1223 (2018).
- 341. Baldridge, D. *et al.* The Exome Clinic and the role of medical genetics expertise in the interpretation of exome sequencing results. *Genet. Med.* **19**, 1040–1048 (2017).
- 342. Bruel, A.-L. *et al.* Increased diagnostic and new genes identification outcome using research reanalysis of singleton exome sequencing. *Eur. J. Hum. Genet. EJHG* (2019) doi:10.1038/s41431-019-0442-1.
- 343. Eldomery, M. K. *et al.* Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome Med.* **9**, 26 (2017).
- 344. Lelieveld, S. H. *et al.* Spatial Clustering of de Novo Missense Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes. *Am. J. Hum. Genet.* **101**, 478–484 (2017).
- 345. Olson, H. E. *et al.* A Recurrent De Novo PACS2 Heterozygous Missense Variant Causes Neonatal-Onset Developmental Epileptic Encephalopathy, Facial Dysmorphism, and Cerebellar Dysgenesis. *Am. J. Hum. Genet.* **103**, 631 (2018).
- 346. Bell, S. *et al.* Mutations in ACTL6B Cause Neurodevelopmental Deficits and Epilepsy and Lead to Loss of Dendrites in Human Neurons. *Am. J. Hum. Genet.* **104**, 815–834 (2019).

- 347. Hamdan, F. F. *et al.* High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies. *Am. J. Hum. Genet.* **101**, 664–685 (2017).
- 348. Lecoquierre, F. *et al.* Variant recurrence in neurodevelopmental disorders: the use of publicly available genomic data identifies clinically relevant pathogenic missense variants. *Genet. Med. Off. J. Am. Coll. Med. Genet.* (2019) doi:10.1038/s41436-019-0518-x.
- 349. Philippakis, A. A. *et al.* The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum. Mutat.* **36**, 915–921 (2015).
- 350. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928–930 (2015).
- 351. Orphanomix Physicians' Group *et al.* 2.5 years' experience of GeneMatcher data-sharing: a powerful tool for identifying new genes responsible for rare diseases. *Genet. Med.* **21**, 1657–1661 (2019).
- 352. Kim, E. & Sheng, M. PDZ domain proteins of synapses. Nat. Rev. Neurosci. 5, 771-781 (2004).
- 353. Stathakis, D. G., Hoover, K. B., You, Z. & Bryant, P. J. Human postsynaptic density-95 (PSD95): location of the gene (DLG4) and possible function in nonneural as well as in neural tissues. *Genomics* 44, 71–82 (1997).
- 354. Bustos, F. J. *et al.* PSD95 suppresses dendritic arbor development in mature hippocampal neurons by occluding the clustering of NR2B-NMDA receptors. *PloS One* **9**, e94037 (2014).
- 355. Moutton, S. *et al.* Truncating variants of the DLG4 gene are responsible for intellectual disability with marfanoid features. *Clin. Genet.* **93**, 1172–1178 (2018).
- 356. Miceli, F. *et al.* A novel KCNQ3 mutation in familial epilepsy with focal seizures and intellectual disability. *Epilepsia* **56**, e15-20 (2015).
- 357. Soldovieri, M. V. *et al.* Novel KCNQ2 and KCNQ3 mutations in a large cohort of families with benign neonatal epilepsy: first evidence for an altered channel regulation by syntaxin-1A. *Hum. Mutat.* **35**, 356–367 (2014).
- 358. Jentsch, T. J. Neuronal KCNQ potassium channels: physiology and role in disease. *Nat. Rev. Neurosci.* **1**, 21–30 (2000).
- 359. Haitin, Y. & Attali, B. The C-terminus of Kv7 channels: a multifunctional module. *J. Physiol.* **586**, 1803–1810 (2008).
- 360. Soldovieri, M. V., Miceli, F. & Taglialatela, M. Driving with no brakes: molecular pathophysiology of Kv7 potassium channels. *Physiol. Bethesda Md* **26**, 365–376 (2011).
- Chung, H. J., Jan, Y. N. & Jan, L. Y. Polarized axonal surface expression of neuronal KCNQ channels is mediated by multiple signals in the KCNQ2 and KCNQ3 C-terminal domains. *Proc. Natl. Acad. Sci. U. S.* A. 103, 8870–8875 (2006).
- Wang, H. S. *et al.* KCNQ2 and KCNQ3 potassium channel subunits: molecular correlates of the Mchannel. *Science* 282, 1890–1893 (1998).
- Cooper, E. C. & Jan, L. Y. M-channels: neurological diseases, neuromodulation, and drug development. *Arch. Neurol.* 60, 496–500 (2003).
- 364. Brown, D. A. & Adams, P. R. Muscarinic suppression of a novel voltage-sensitive K+ current in a vertebrate neurone. *Nature* **283**, 673–676 (1980).
- Constanti, A. & Brown, D. A. M-Currents in voltage-clamped mammalian sympathetic neurones. *Neurosci. Lett.* 24, 289–294 (1981).
- 366. Brown, D. A. M currents. Ion Channels 1, 55-94 (1988).
- Storm, J. F. An after-hyperpolarization of medium duration in rat hippocampal pyramidal cells. *J. Physiol.* 409, 171–190 (1989).
- 368. Womble, M. D. & Moises, H. C. Muscarinic inhibition of M-current and a potassium leak conductance in neurones of the rat basolateral amygdala. *J. Physiol.* **457**, 93–114 (1992).
- 369. Cooper, E. C. *et al.* Colocalization and coassembly of two human brain M-type potassium channel subunits that are mutated in epilepsy. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4914–4919 (2000).

- 370. Cooper, E. C., Harrington, E., Jan, Y. N. & Jan, L. Y. M channel KCNQ2 subunits are localized to key sites for control of neuronal network oscillations and synchronization in mouse brain. *J. Neurosci. Off. J. Soc. Neurosci.* **21**, 9529–9540 (2001).
- 371. Devaux, J. J., Kleopa, K. A., Cooper, E. C. & Scherer, S. S. KCNQ2 is a nodal K+ channel. *J. Neurosci. Off. J. Soc. Neurosci.* **24**, 1236–1244 (2004).
- 372. Zhang, H. *et al.* PIP(2) activates KCNQ channels, and its hydrolysis underlies receptor-mediated inhibition of M currents. *Neuron* **37**, 963–975 (2003).
- 373. Miceli, F. *et al.* Genotype-phenotype correlations in neonatal epilepsies caused by mutations in the voltage sensor of K(v)7.2 potassium channel subunits. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4386–4391 (2013).
- 374. Orhan, G. *et al.* Dominant-negative effects of KCNQ2 mutations are associated with epileptic encephalopathy. *Ann. Neurol.* **75**, 382–394 (2014).
- 375. Kothur, K. *et al.* Diagnostic yield of targeted massively parallel sequencing in children with epileptic encephalopathy. *Seizure* **59**, 132–140 (2018).
- Ambrosino, P. *et al.* Kv7.3 Compound Heterozygous Variants in Early Onset Encephalopathy Reveal Additive Contribution of C-Terminal Residues to PIP2-Dependent K+ Channel Gating. *Mol. Neurobiol.* 55, 7009–7024 (2018).
- 377. Fister, P., Soltirovska-Salamon, A., Debeljak, M. & Paro-Panjan, D. Benign familial neonatal convulsions caused by mutation in KCNQ3, exon 6: a European case. *Eur. J. Paediatr. Neurol. EJPN Off. J. Eur. Paediatr. Neurol. Soc.* 17, 308–310 (2013).
- Hadley, J. K. *et al.* Stoichiometry of expressed KCNQ2/KCNQ3 potassium channels and subunit composition of native ganglionic M channels deduced from block by tetraethylammonium. *J. Neurosci. Off. J. Soc. Neurosci.* 23, 5012–5019 (2003).
- Kanaumi, T. *et al.* Developmental changes in KCNQ2 and KCNQ3 expression in human brain: possible contribution to the age-dependent etiology of benign familial neonatal convulsions. *Brain Dev.* 30, 362–369 (2008).
- 380. Soh, H., Pant, R., LoTurco, J. J. & Tzingounis, A. V. Conditional deletions of epilepsy-associated KCNQ2 and KCNQ3 channels from cerebral cortex cause differential effects on neuronal excitability. *J. Neurosci. Off. J. Soc. Neurosci.* 34, 5311–5321 (2014).
- Sands, T. T. *et al.* Autism and developmental disability caused by *KCNQ3* gain-of-function variants. *Ann. Neurol.* 86, 181–192 (2019).
- Zhou, P. *et al.* Phosphatidylinositol 4,5-bisphosphate alters pharmacological selectivity for epilepsycausing KCNQ potassium channels. *Proc. Natl. Acad. Sci. U. S. A.* 110, 8726–8731 (2013).
- 383. Lauritano, A. *et al.* A novel homozygous KCNQ3 loss-of-function variant causes non-syndromic intellectual disability and neonatal-onset pharmacodependent epilepsy. *Epilepsia Open* **4**, 464–475 (2019).
- 384. Gregor, A. *et al.* De Novo Variants in the F-Box Protein FBXO11 in 20 Individuals with a Variable Neurodevelopmental Disorder. *Am. J. Hum. Genet.* **103**, 305–316 (2018).
- 385. Spiegel, R. *et al.* Infantile Cerebellar-Retinal Degeneration Associated with a Mutation in Mitochondrial Aconitase, ACO2. *Am. J. Hum. Genet.* **90**, 518–523 (2012).
- 386. Duijkers, F. A. *et al.* HNRNPR Variants that Impair Homeobox Gene Expression Drive Developmental Disorders in Humans. *Am. J. Hum. Genet.* **104**, 1040–1059 (2019).
- 387. Esteve, C. *et al.* Loss-of-Function Mutations in UNC45A Cause a Syndrome Associating Cholestasis, Diarrhea, Impaired Hearing, and Bone Fragility. *Am. J. Hum. Genet.* **102**, 364–374 (2018).
- 388. Latorre, R. *et al.* Molecular Determinants of BK Channel Functional Diversity and Functioning. *Physiol. Rev.* **97**, 39–87 (2017).
- 389. Kshatri, A. S., Gonzalez-Hernandez, A. & Giraldez, T. Physiological Roles and Therapeutic Potential of Ca2+ Activated Potassium Channels in the Nervous System. *Front. Mol. Neurosci.* **11**, 258 (2018).
- 390. Guéguinou, M. *et al.* KCa and Ca(2+) channels: the complex thought. *Biochim. Biophys. Acta* **1843**, 2322–2333 (2014).

- Petersen, O. H. & Maruyama, Y. Calcium-activated potassium channels and their role in secretion. *Nature* 307, 693–696 (1984).
- 392. Murrow, B. W. & Fuchs, P. A. Preferential expression of transient potassium current (IA) by 'short' hair cells of the chick's cochlea. *Proc. Biol. Sci.* 242, 189–195 (1990).
- 393. Robitaille, R. & Charlton, M. P. Presynaptic calcium signals and transmitter release are modulated by calcium-activated potassium channels. *J. Neurosci. Off. J. Soc. Neurosci.* **12**, 297–305 (1992).
- 394. Brayden, J. E. & Nelson, M. T. Regulation of arterial tone by activation of calcium-dependent potassium channels. *Science* **256**, 532–535 (1992).
- 395. Wu, Y. C., Ricci, A. J. & Fettiplace, R. Two components of transducer adaptation in auditory hair cells. *J. Neurophysiol.* **82**, 2171–2181 (1999).
- Weiger, T. M., Hermann, A. & Levitan, I. B. Modulation of calcium-activated potassium channels. J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol. 188, 79–87 (2002).
- 397. Faber, E. S. L. & Sah, P. Calcium-activated potassium channels: multiple contributions to neuronal function. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* **9**, 181–194 (2003).
- 398. Du, W. *et al.* Calcium-sensitive potassium channelopathy in human epilepsy and paroxysmal movement disorder. *Nat. Genet.* **37**, 733–738 (2005).
- 399. Zhang, Z.-B., Tian, M.-Q., Gao, K., Jiang, Y.-W. & Wu, Y. De novo KCNMA1 mutations in children with early-onset paroxysmal dyskinesia and developmental delay. *Mov. Disord. Off. J. Mov. Disord. Soc.* 30, 1290–1292 (2015).
- 400. Li, X. *et al.* De novo BK channel variant causes epilepsy by affecting voltage gating but not Ca2+ sensitivity. *Eur. J. Hum. Genet. EJHG* **26**, 220–229 (2018).
- Tabarki, B., AlMajhad, N., AlHashem, A., Shaheen, R. & Alkuraya, F. S. Homozygous KCNMA1 mutation as a cause of cerebellar atrophy, developmental delay and seizures. *Hum. Genet.* 135, 1295–1298 (2016).
- 402. Sausbier, M. *et al.* Cerebellar ataxia and Purkinje cell dysfunction caused by Ca2+-activated K+ channel deficiency. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9474–9478 (2004).
- 403. Hei, H. *et al.* BK Knockout by TALEN-Mediated Gene Targeting in Osteoblasts: KCNMA1 Determines the Proliferation and Differentiation of Osteoblasts. *Mol. Cells* **39**, 530–535 (2016).
- 404. Yang, Z., Pan, A., Zuo, W., Guo, J. & Zhou, W. Relaxant effect of flavonoid naringenin on contractile activity of rat colonic smooth muscle. *J. Ethnopharmacol.* **155**, 1177–1183 (2014).
- 405. Köhler, R., Kaistha, B. P. & Wulff, H. Vascular K _{Ca} -channels as therapeutic targets in hypertension and restenosis disease. *Expert Opin. Ther. Targets* **14**, 143–155 (2010).
- 406. Wang, L. *et al.* Cognitive recovery by chronic activation of the large-conductance calcium-activated potassium channel in a mouse model of Alzheimer's disease. *Neuropharmacology* **92**, 8–15 (2015).
- 407. Hébert, B. *et al.* Rescue of fragile X syndrome phenotypes in Fmr1 KO mice by a BKCa channel opener molecule. *Orphanet J. Rare Dis.* **9**, 124 (2014).
- 408. Bailey, C. S., Moldenhauer, H. J., Park, S. M., Keros, S. & Meredith, A. L. *KCNMA1* -linked channelopathy. *J. Gen. Physiol.* jgp.201912457 (2019) doi:10.1085/jgp.201912457.
- 409. Liang, L. *et al.* De novo loss-of-function KCNMA1 variants are associated with a new multiple malformation syndrome and a broad spectrum of developmental and neurological phenotypes. *Hum. Mol. Genet.* **28**, 2937–2951 (2019).
- 410. Solve-RD Solving the Unsolved Rare Diseases. http://solve-rd.eu/.
- 411. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).
- 412. Scacheri, C. A. & Scacheri, P. C. Mutations in the noncoding genome: *Curr. Opin. Pediatr.* 27, 659–664 (2015).
- 413. Gonorazky, H. *et al.* RNAseq analysis for the diagnosis of muscular dystrophy. *Ann. Clin. Transl. Neurol.* 3, 55–60 (2016).

- 414. Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* **8**, 15824 (2017).
- 415. Cummings, B. B. *et al.* Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* **9**, eaal5209 (2017).
- 416. Hamanaka, K. *et al.* RNA sequencing solved the most common but unrecognized NEB pathogenic variant in Japanese nemaline myopathy. *Genet. Med.* **21**, 1629–1638 (2019).
- 417. Kernohan, K. D. *et al.* Whole-transcriptome sequencing in blood provides a diagnosis of spinal muscular atrophy with progressive myoclonic epilepsy: KERNOHAN et al. *Hum. Mutat.* **38**, 611–614 (2017).
- 418. Undiagnosed Diseases Network *et al.* Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat. Med.* **25**, 911–919 (2019).
- 419. Liew, C.-C., Ma, J., Tang, H.-C., Zheng, R. & Dempsey, A. A. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J. Lab. Clin. Med.* **147**, 126–132 (2006).
- 420. Rollins, B., Martin, M. V., Morgan, L. & Vawter, M. P. Analysis of whole genome biomarker expression in blood and brain. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* **153B**, 919–936 (2010).
- 421. Ibn-Salem, J. *et al.* Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* **15**, 423 (2014).
- 422. Di Gregorio, E. *et al.* Copy number variants analysis in a cohort of isolated and syndromic developmental delay/intellectual disability reveals novel genomic disorders, position effects and candidate disease genes. *Clin. Genet.* **92**, 415–422 (2017).
- 423. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).
- 424. Mehrjouy, M. M. *et al.* Regulatory variants of FOXG1 in the context of its topological domain organisation. *Eur. J. Hum. Genet.* **26**, 186–196 (2018).
- 425. Hancks, D. C. & Kazazian, H. H. Roles for retrotransposon insertions in human disease. *Mob. DNA* 7, 9 (2016).
- 426. Sadikovic, B., Aref-Eshghi, E., Levy, M. A. & Rodenhiser, D. DNA methylation signatures in mendelian developmental disorders as a diagnostic bridge between genotype and phenotype. *Epigenomics* **11**, 563–575 (2019).
- 427. Larizza, L. & Finelli, P. Developmental disorders with intellectual disability driven by chromatin dysregulation: Clinical overlaps and molecular mechanisms. *Clin. Genet.* **95**, 231–240 (2019).
- 428. Chokroborty-Hoque, A., Alberry, B. & Singh, S. M. Exploring the Complexity of Intellectual Disability in Fetal Alcohol Spectrum Disorders. *Front. Pediatr.* **2**, (2014).
- 429. Angarica, V. E. & del Sol, A. Bioinformatics Tools for Genome-Wide Epigenetic Research. in *Neuroepigenomics in Aging and Disease* (ed. Delgado-Morales, R.) vol. 978 489–512 (Springer International Publishing, 2017).
- 430. Hwang, J.-Y., Aromolaran, K. A. & Zukin, R. S. The emerging field of epigenetics in neurodegeneration and neuroprotection. *Nat. Rev. Neurosci.* **18**, 347–361 (2017).
- 431. Jiang, Y.-H. *et al.* A mixed epigenetic/genetic model for oligogenic inheritance of autism with a limited role for UBE3A. *Am. J. Med. Genet. A.* **131**, 1–10 (2004).
- 432. Schaaf, C. P. *et al.* Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum. Mol. Genet.* **20**, 3366–3375 (2011).
- 433. Tabet, A.-C. *et al.* A framework to identify contributing genes in patients with Phelan-McDermid syndrome. *Npj Genomic Med.* **2**, 32 (2017).
- 434. iPSYCH-Broad ASD Group *et al.* Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med.* **10**, 19 (2018).
- 435. Guo, H. *et al.* Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. *Mol. Autism* **9**, 64 (2018).

- 436. Bone, W. P. *et al.* Computational evaluation of exome sequence data using human and model organism phenotypes improves diagnostic efficiency. *Genet. Med.* **18**, 608–617 (2016).
- 437. Leman, R. *et al.* Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Res.* **46**, 7913–7923 (2018).
- 438. Salgado, D. *et al.* UMD-Predictor: A High-Throughput Sequencing Compliant System for Pathogenicity Prediction of any Human cDNA Substitution. *Hum. Mutat.* **37**, 439–446 (2016).
- 439. The UMD-DMD French Knowledgebase. http://www.umd.be/DMD/.
- 440. Béroud, C. *et al.* BRCA Share: A Collection of Clinical BRCA Gene Variants. *Hum. Mutat.* **37**, 1318–1328 (2016).
- 441. EuroEPINOMICS RES Consortium *et al.* Mutations in STX1B, encoding a presynaptic protein, cause fever-associated epilepsy syndromes. *Nat. Genet.* **46**, 1327–1332 (2014).
- Marin-Valencia, I. *et al.* A homozygous founder mutation in TRAPPC6B associates with a neurodevelopmental disorder characterised by microcephaly, epilepsy and autistic features. *J. Med. Genet.* 55, 48–54 (2018).
- 443. Blackburn, A. T. M. *et al.* DYRK1A-related intellectual disability: a syndrome associated with congenital anomalies of the kidney and urinary tract. *Genet. Med. Off. J. Am. Coll. Med. Genet.* (2019) doi:10.1038/s41436-019-0576-0.
- 444. Ott, T. *et al.* The Frog Xenopus as a Model to Study Joubert Syndrome: The Case of a Human Patient With Compound Heterozygous Variants in PIBF1. *Front. Physiol.* **10**, 134 (2019).
- 445. Begemann, A. *et al.* Further corroboration of distinct functional features in SCN2A variants causing intellectual disability or epileptic phenotypes. *Mol. Med. Camb. Mass* **25**, 6 (2019).
- 446. Wright, C. F. *et al.* Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am. J. Hum. Genet.* **104**, 275–286 (2019).
- 447. Sakai, C., Ijaz, S. & Hoffman, E. J. Zebrafish Models of Neurodevelopmental Disorders: Past, Present, and Future. *Front. Mol. Neurosci.* **11**, 294 (2018).

RESUME

Les anomalies du développement (AD) regroupent les syndromes malformatifs et les pathologies neurodéveloppementales comme la déficience intellectuelle (DI) et les troubles du spectre autistique. Ces manifestations peuvent être isolées ou non, et touchent environ 3% de la population. Il s'agit d'une cause importante de morbi-mortalité, d'hospitalisation notamment pédiatrique, de handicap, et donc d'un problème de santé publique. Une cause génétique est fortement suspectée dans environ 80% des cas. Depuis 2004, plusieurs plans nationaux maladies rares ont été déployés et reconduits, dont le dernier (3^{ème} PNMR 2018-2022) se décline en 10 axes visant entre autres à améliorer le rendement diagnostique, réduire l'errance et l'impasse diagnostiques, améliorer le parcours de soin, favoriser l'émergence et l'accès à l'innovation, et impulser un nouvel élan à la recherche.

Le taux diagnostique chez les patients atteints de maladies rares est tributaire de l'amélioration des connaissances cliniques et des innovations technologiques, qui évoluent en connexion permanente. Le rassemblement de cohortes de patients ayant un phénotype similaire a permis de rechercher une cause moléculaire commune pendant que l'identification de variations génétiques similaires a permis de décrire le spectre phénotypique de syndromes qui n'avaient pas été identifiés cliniquement jusque-là ou de démembrer certaines pathologies très hétérogènes génétiquement. Dans les AD/DI, le rendement diagnostique a pu être estimé à 5-15% grâce aux analyses chromosomiques (principalement sur puce à ADN), de 10 à 25% selon la taille des panels de gènes séquencés et jusqu'à 40% grâce au séquençage d'exome, qui couvre la majorité des régions codantes des gènes. Malgré ces avancées, l'origine de la pathologie reste inexpliquée chez de nombreux patients, en partie du fait de mécanismes mutationnels non détectables par les tests utilisés, mais aussi du fait des limites dans l'interprétation des données de séquençage alors que les variants génétiques pathogènes sont bien détectés.

Cette thèse s'inscrit dans l'effort d'identification de nouveaux types de mutations et nouveaux gènes impliqués dans les pathologies du spectre des AD lorsqu'une cause génétique est fortement suspectée mais que les analyses habituelles en procédure diagnostique n'ont pas permis de l'identifier. Ce travail a donc consisté en la réanalyse de données d'exome dans une cohorte phénotypiquement reconnaissable (DI et syndrome marfanoïde) mais hétérogène génétiquement pour rechercher des mécanismes similaires chez certains patients, et dans une cohorte de patients atteints de DI peu ou pas syndromique. Par ailleurs, à partir d'une cohorte de patients restés négatifs malgré une réanalyse des données d'exome solo en recherche, un séquençage des parents a été complété et une réinterprétation en trio a été menée. Dans tous les cas, lorsqu'un gène candidat a été identifié, la base de données locale d'exomes était interrogée et le gène partagé sur une plateforme internationale pour tenter d'augmenter le nombre de patients concernés, confirmer son implication et en décrire un spectre phénotypique ; des analyses fonctionnelles prouvant l'effet délétère d'un type de variant génétique ont été menées pour apporter des arguments supplémentaires lorsque cela a été nécessaire.

Cette stratégie a permis d'identifier plusieurs gènes candidats dans le cadre de cette thèse, dont l'implication a pu être confirmée pour une grande proportion et conduire à des publications, soit à partir de travaux menés intégralement au sein de l'équipe (variant tronquants de *DLG4* identifiés chez 3 patients ayant un syndrome DI marfanoïde), soit à partir de collaborations internationales (nouveau mécanisme mutationnel et nouveau mode d'hérédité associé au gène *KCNQ3* confirmé par analyses fonctionnelles (DI, épilepsie, et consanguinité parentale) ; réinteprétation en trio dans une série de 70 patients ayant permis d'identifier un gène fortement candidat chez 26% d'entre eux dont *KCNMA1*, *FBXO11*, *HNRNPR*, et un variant de signification indéterminé chez 14% d'entre eux dont 3 patients présentant potentiellement plusieurs maladies génétiques.

Ces résultats confirment l'intérêt de la réanalyse d'exome et du partage des données pour la découverte de nouveaux gènes et nouveaux mécanismes mutationnels. De nouvelles technologies comme le séquençage de génome, le RNA-seq et les analyses épigénétiques laissent entrevoir un apport supplémentaire au bénéfice des patients pour réduire les situations encore trop nombreuses d'impasses diagnostiques.

Mots-clés : séquençage haut débit, exome, anomalies du développement, malformation, déficience intellectuelle

SUMMARY

Developmental disorders (DD) include malformative disorders and neurodevelopmental disorders such as intellectual disability (ID) and autism spectrum disorders (ASD). These clinical features can be isolated or combined in a given patient, they affect around 3% of worldwide population. These disorders are responsible for major morbidity and mortality, for hospitalization especially in pediatric departments, for disabilities and, therefore, this represents a public health priority. Since 2004, in France, several national plans were conducted and extended, the last one being the 3rd National Plan for Rare Disorders (PNMR 2018-2022) which include 10 main goals, among others increasing diagnosis yield, decrease diagnosis wandering, improving healthcare journey, access to innovation and boost research.

Diagnosis yield in patients suffering from DD depends on knowledge increase and technology improvement which evolve in tight connections. Gathering cohorts of patients affected by similar disorders allowed to search for a shared molecular disorder while identification of genetic variants having similar predicted consequences allowed to describe phenotypic spectrum of syndromes which had not been previously clinically identified and to decipher some heterogeneous disorders. Diagnosis yield of DD could be assessed around 5-15% with chromosomal analyses (especially microarrays), 10-25% according to the number of genes within panels, and until 40% by means of exome sequencing which targets gene coding regions. Despite these improvements, genetic cause remains unknown in many patients, in part due to undetectable mutational mechanisms with used technologies, but also because of interpretation limits in sequencing data although pathogenic variants were detected.

This phD aims at going on the research effort of new mutational spectrum of known disease-causing genes and identification of new ones, after non-contributive diagnosis analyses. This work consisted in exome reanalysis either in a recognizable cohort of ID associated with marfanoid features, or in a cohort of patients with non-syndromic ID. Otherwise, from a cohort of unexplained sporadic DD patients after solo exome reanalysis, we sequenced exome of both parents to interpret in a trio-based strategy. Every time a candidate gene was suspected, it was searched for recurrence in our in-house exome database and through international datasharing platforms to strengthen the association with the disorder and describe the phenotypic spectrum ; functional analyses were conducted when necessary to argue for the pathogenicity.

This strategy allowed to identify several candidate genes which involvement in the disorders could be confirmed in a large proportion of them and led to scientific publications, either after works entirely conducted within our lab (*DLG4* truncating variants identified in 3 patients from the marfanoid ID cohort), or by means of international collaborations: new mutational mechanism and new inheritance associated with *KCNQ3* confirmed by functional analyses (ID, seizures, consanguinity), trio-exome reanalysis in a cohort of 70 unexplained sporadic DD patients allowing to identify a strong candidate gene in 26% of them (including *KCNMA1*, *FBXO11*, *HNRNPR*), and variants of unknown significance in 14%, with 3/70 patients probably affected by multiple disorders.

These results confirm the great interest of exome reanalysis and datasharing to improve new pathogenic variants and new disease-causing genes. New technologies such as whole-genome, RNA sequencing, epigenetic analyses are promising for future discoveries, all this for patient care improvement and to decrease the still too numerous unsolved situations.

Keywords: high throughput sequencing, exome, developmental disorders, malformations, intellectual disability.

Titre : L'approche basée sur le génotype déterminé par séquençage haut-débit en première intention et le partage international des données pour identifier de nouveaux gènes et nouveaux syndromes responsables d'anomalies du développement

Mots clés : séquençage haut débit, exome, anomalies du développement, malformation, déficience intellectuelle

Résumé : Les anomalies du développement (AD) regroupent les syndromes malformatifs et les pathologies neurodéveloppementales comme la déficience intellectuelle (DI) et les troubles du spectre autistique. Ces manifestations peuvent être isolées ou non, et touchent environ 3% de la population. Il s'agit donc d'un problème de santé publique. Une cause génétique est fortement suspectée dans environ 80% des cas mais le diagnostic est posé dans 30 à 40% des cas. Le taux diagnostique chez les patients atteints de maladies rares est tributaire de l'amélioration des connaissances cliniques et des innovations technologiques, qui évoluent en connexion permanente.

Cette thèse s'inscrit dans l'effort d'identification de nouveaux types de mutations et nouveaux gènes impliqués dans les pathologies du spectre des AD lorsqu'une cause génétique est fortement suspectée mais que les analyses habituelles en procédure diagnostique n'ont pas permis de l'identifier.

Ce travail a donc consisté en la réanalyse de données d'exome une cohorte phénotypiquement dans reconnaissable (DI et syndrome marfanoïde) mais hétérogène génétiquement pour rechercher des mécanismes similaires chez certains patients, et dans une cohorte de patients atteints de DI peu ou pas syndromique. Par ailleurs, à partir d'une cohorte de patients restés négatifs malgré une réanalyse des données

d'exome solo en recherche, un séquençage des parents a été complété et une réinterprétation en trio a été menée. Dans tous les cas, lorsqu'un gène candidat a été identifié, la base de données locale d'exomes était interrogée et le gène partagé sur une plateforme internationale pour tenter d'augmenter le nombre de patients concernés, confirmer son implication et en décrire un spectre phénotypique ; des analyses fonctionnelles prouvant l'effet délétère d'un type de variant génétique ont été menées pour apporter des arguments supplémentaires lorsque cela a été nécessaire.

Cette stratégie a permis d'identifier plusieurs gènes candidats dans le cadre de cette thèse, dont *DLG4* impliqué dans un syndrome DI marfanoïde, et *KCNQ3* (nouveau mécanisme mutationnel et nouveau mode d'hérédité). La réinteprétation en trio dans une série de 70 patients a permis d'identifier un gène fortement candidat chez 26% d'entre eux dont *KCNMA1* et une variation de signification indéterminée chez 14%.

Ces résultats confirment l'intérêt de la réanalyse d'exome et du partage des données pour la découverte de nouveaux gènes et nouveaux mécanismes mutationnels. De nouvelles technologies comme le séquençage de génome, le RNA-seq et les analyses épigénétiques laissent entrevoir un apport supplémentaire au bénéfice des patients pour réduire les situations encore trop nombreuses d'impasses diagnostiques.

Title : The "genotype-first" approach using next-generation sequencing and international data sharing as an efficient strategy to identify new genes and new syndromes responsible for developmental disorders

Keywords : high throughput sequencing, exome, developmental disorders, malformations, intellectual disability

Abstract : Developmental disorders (DD) include malformative disorders and neurodevelopmental disorders such as intellectual disability (ID) with or without autism spectrum disorders (ASD). These clinical features can be isolated or combined in a given patient, they affect around 3% of worldwide population and represent a public health priority. A molecular diagnosis is reached in around 30-40% of patients but a genetic etiology probably account for 80% of cases, the yield depends on knowledge increase and technology improvement which evolve in tight connections.

This phD aims at going on the research effort of new mutational spectrum of known disease-causing genes and identification of new ones, after non-contributive diagnosis analyses. This work consisted in exome reanalysis either in a recognizable cohort of ID associated with marfanoid features, or in a cohort of patients with non-syndromic ID. Otherwise, from a cohort of unexplained sporadic DD patients after solo exome reanalysis, we sequenced exome of both parents to interpret in a trio-based strategy. Every time a candidate gene was suspected, it was searched for recurrence in our in-house exome database and through international datasharing platforms to strengthen the association with the disorder and describe the phenotypic spectrum, functional analyses were conducted when necessary to argue for the pathogenicity. This strategy allowed to identify several candidate genes including *DLG4* involved in marfanoid ID, and *KCNQ3* (mutational spectrum expansion). Trio-exome reanalysis in a cohort of 70 unexplained sporadic DD patients allowed to identify a strong candidate gene in 26% of them (including *KCNMA1*), and variants of unknown significance in 14%.

These results confirm the great interest of exome reanalysis and datasharing to improve new pathogenic variants and new disease-causing genes. New technologies such as whole-genome, RNA sequencing, epigenetic analyses are promising for future discoveries, all this for patient care improvement and to decrease the still too numerous unsolved situations.