

UNIVERSITÉ DE NANTES

FACULTÉ DE MÉDECINE

Année : 2020

N°

THÈSE

pour le

DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE

(DES DE GENETIQUE MEDICALE)

par

Leïla GHESH

Née le 25 novembre 1992 à Nantes

Présentée et soutenue publiquement le 26 octobre 2020

Des variants "perte de fonction" dans ARHGEF9 sont responsables d'une déficience
intellectuelle dominante liée à l'X

Président : Monsieur le Professeur Stéphane BEZIEAU

Directeur de thèse : Docteur Benjamin COGNE

Remerciements

A tous ceux qui ont participé à la réalisation de ce travail et à mon épanouissement tant professionnel que personnel.

Au Professeur Stéphane BEZIEAU, merci d'avoir accepté de présider mon jury et de m'avoir soutenue et défendue dans mes projets, cela a été pour moi un grand honneur.

A Benjamin mon directeur de thèse, merci pour ton soutien, ta patience et ta pertinence dans ce travail, sans toi celui-ci n'aurait probablement pas vu le jour. Merci également pour tes encouragements qui m'ont beaucoup touchée.

Merci au Professeur Elise Launay d'avoir accepté d'être dans mon jury. J'ai pris beaucoup de plaisir à travailler à vos côtés lors de mon semestre en pédiatrie au cours duquel j'ai énormément appris.

Merci au Professeur Nohert Winer d'avoir accepté sans hésiter de faire partie de mon jury. Je vous remercie également pour votre confiance et votre soutien dans nos différents projets.

A Mathilde, je te remercie de faire partie de mon jury mais également de ta disponibilité au cours de toutes ces années, j'ai toujours pu faire appel à toi qu'importe la situation.

Je remercie de manière générale tout le service de génétique médicale qui m'a vu évoluer et grandir durant ces quatre années intenses mais merveilleuses.

Tout d'abord l'équipe de génétique clinique : Bertrand, Sandra, Marie-Line et Laura G. Merci pour vos enseignements précieux, vous m'avez transmis l'amour de la génétique et du contact aux familles, et tout cela dans la bienveillance. Je remercie plus particulièrement Mathilde et Marie V qui ont toujours été plus que disponibles lorsque j'avais le moindre doute. Merci à Ophélie, Chantal, Suzanne et Geneviève pour leur bonne humeur.

Ensuite, au laboratoire qui est devenu comme une deuxième maison après y avoir passé 5 semestres. Merci à toute l'équipe de cyto qui fut mon soleil lors des tournées de « Bonjour » quotidiennes et de mes petites pauses « mentales » et grignotages (Virginie, Marie-Jo, Aurélie, Elodie, Nadine, Annaïg, Anne-Laure, Aurélien, Cécile, Laura J, Thierry,

Olivier). Et cela même si vous m'avez bannie, même sans cookies je reviendrai, vous voilà prévenus !

Merci également à l'équipe de biologie moléculaire (Gaëlle LT, Patricia, Jessy, Christine, Christine, Marie-Annick, Nadège, Anne, Thierry, Eva, Ingrid, Delphine, Flora, Clémence, Jérôme, Karine) grâce à vous j'ai eu une formation au top, toujours dans la bonne humeur et dans l'entraide. Merci de m'avoir aidée dans mes petits projets, un jour peut être serais-je autonome.

Un grand merci à tous les biologistes (Sébastien S, Gaëlle M, Pierre, Thomas, Sébastien K, Kamran, Minh) qui m'ont coaché et donné les clés me permettant de devenir grande.

Merci à Monsieur Piloquet de m'avoir formée et d'avoir rêvé à la cytogénétique de demain avec moi.

Merci à tous ceux qui ont partagé un bureau avec moi et qui m'ont supportée : Mathilde encore, Chrystelle, Wallid, Elise, Pierre, Cindy, Marie K.

Mention spéciale pour toi Claire, dès mon premier semestre tu m'as accordé ta confiance, grâce à toi j'ai pu devenir de plus en plus autonome en toute sécurité. J'ai adoré collaborer avec toi tant en clinique que sur nos projets qui m'ont tant stimulées. De plus tu as toujours été un soutien sans faille dans les moments difficiles.

A Madeleine et Marie DM, j'ai pu découvrir une spécialité merveilleuse à vos côtés et ce de la meilleure des façons. Merci pour votre accompagnement et de m'avoir transmis la passion de la fœtopathologie.

A toute l'équipe de l'HDJ ado : Chrystelle, Samuel, Stéphanie, Valérie et les autres. Je ne vous oublie pas, mon arrivée n'était à la base pas prévue mais allez savoir pourquoi, même 3 ans plus tard je reviens encore et toujours. A Marion ma super ex-co-interne de l'HDJ qui même des années après ne m'oublie pas.

A mes très chers co-internes qui sont aussi mes amis. Si Jessica et Wallid m'ont accueilli dans la famille, celle-ci s'est bien vite agrandie : Solène, Elise, Adeline, Silvestre, Servane, Martin, Xavier. A nos midis, nos cohabitations lors des cours, séminaires, à nos trop rares verres, à nos décompensations. Merci pour votre folie et votre bonne humeur. Et si ce travail est abouti c'est aussi grâce à Adeline et Wallid qui m'ont bien sauvé la mise.

Merci à mes amis, à Marc et Julien avec qui nous avons survécus à nos années médecines ensemble (et aux autres qui sont loin maintenant). A Camille et Alexia, votre amitié est un don précieux que je chéris depuis de nombreuses années. Je suis comblée d'avoir grandi et d'être devenue mère à vos côtés.

A mon filleul, j'ai nommé le beau Raphaël, j'espère être toujours un modèle pour toi et te voir grandir. A ton frère Gabriel, et à Augustin et Léon que je veux voir grandir également. Alexandre, à nos vacances tous ensemble, vivement que l'on recommence.

Et bien évidemment je tiens à remercier ma famille pour qui ça n'a pas toujours été facile. Merci de m'avoir supportée et épaulée au cours de toutes ces années qui ont pu parfois me rendre un peu folle. Merci à mes grands-parents qui m'ont tant donné et à qui je dois tant. A mes frères et sœur que j'aime tendrement, vous pourrez toujours compter sur moi. A mes beaux-parents qui se dévouent corps et âme à leur petit-fils et qui sont un pilier pour nous. A mes oncles, tantes, cousins, cousines (Tonton Eric, tonton Antoine, Evaëlle, Maïlys, Anaïs, Tatie Marie-Jeanne, Céline, Nanie, Jean-Marie et tous les autres) merci de faire en sorte que cette famille soit ce qu'elle est, merveilleuse !

Et pour finir, merci aux deux hommes de ma vie. Maxime pour tout l'amour et la confiance indéfectible que tu me portes. Tu es un père merveilleux. Merci à toi Nathaël, tu es ma plus belle réussite. Tu illumines ma vie (et mes nuits !), restes tel que tu es, tu es parfait.

Tables des matières

Remerciements	1
Tables des matières	2
Liste des abréviations	3
Liste des figures	4
I. Introduction	5
1. Généralité sur la déficience intellectuelle	5
2. La déficience intellectuelle liée au chromosome X (DILX)	6
3. Difficultés et errance diagnostiques	8
4. Evolution des technologies diagnostiques en génétiques humaines	8
5. Le séquençage haut-débit.....	11
6. Stratégie diagnostique actuelle (figure 7)	13
7. <i>ARHGEF9</i>	16
II. Article.....	18
III. Discussion	36
1. Interprétation des variants synonymes :.....	36
2. Interprétation des variants localisées sur le chromosome X	38
3. Prise en charge médico-sociale :.....	39
IV. Perspectives	42
Bibliographie	44

Liste des abréviations

AAIDD *American Association on Intellectual and Developmental Disabilities*

ACPA analyse comparative sur puce à ADN

ACMG-AMP *American College of Medical Genetics & the Association for Molecular Pathology*

AD autosomique dominant

ADN Acide DésoxyriboNucléique

APA *American Psychiatric Association*

AR autosomique récessif

ARN Acide RiboNucléique

ARNm ARN messenger

Cb collybistine

CGH-array Comparative Genomic Hybridization array

CIM Classification internationale des maladies

CNV *Copy Number Variation*

DI déficience intellectuelle

DILX déficience intellectuelle liée à l’X

DLX dominant lié à l’X

DPI diagnostic préimplantatoire

DPN diagnostic prénatal

DSM-5

FISH *Fluorescence in situ hybridization*

NGS *Next Generation Sequencing*

OMS Organisation mondiale de la santé

PFMG 2025 Plan France Médecine Génomique 2025

PNMR plan national maladies rares

RLX récessif lié à l’X

SHD séquençage haut débit

Liste des figures

Figure 1 : Etiologies des déficiences intellectuelles.....	7
Figure 2 : Evolution des technologies pour le diagnostic des causes de DI.....	9
Figure 3 : Les différentes étapes du séquençage.....	10
Figure 4 : Diminution du coût du séquençage.	12
Figure 5 : Identification croissante des gènes responsables de DI.....	13
Figure 6 : Stratégie diagnostique dans la DI.....	15
Figure 7 : Représentation schématique des interactions de la collybistine dans la formation de la géphyrine et du recrutement des récepteurs GABAergiques et glycinergiques dans la membrane postsynaptique.....	17

I. Introduction

1. Généralité sur la déficience intellectuelle

Trois autorités proposent une définition de la déficience intellectuelle (DI) : l'Organisation mondiale de la santé (OMS) dans sa Classification internationale des maladies (CIM-11), l'American Association on Intellectual and Developmental Disabilities (AAIDD) dans la 11^e édition de son manuel de définitions (2010) et l'American Psychiatric Association (APA) avec le DSM-5 (2013) avec en commun les trois critères suivants (1) : 1) la présence 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 ; 2) la présence **de déficits du fonctionnement adaptatif** dans les habiletés conceptuelles, sociales et pratiques apprises qui permettent de fonctionner dans la vie quotidienne ; 3) l'apparition de ces déficits intellectuels et de ces limitations adaptatives au cours de la **période développementale**. La DI peut entraîner des déficits sociaux, on parle alors de « handicap intellectuel ».

En France, la DI représente un enjeu majeur de santé publique et touche environ 1% de la population. La DI est extrêmement hétérogène tant sur le plan clinique que sur le plan étiologique (2). Cette hétérogénéité représente un frein dans l'établissement du diagnostic étiologique, hors l'identification de l'étiologie est primordiale car cela permet aux familles de mettre un nom sur la maladie, de préciser le pronostic, de mettre en place un suivi médical approprié, des mesures de prévention, de préciser le conseil génétique (risque de récurrence pour les prochaines grossesses), d'avoir accès au diagnostic prénatal (DPN) ou au diagnostic préimplantatoire (DPI), d'accéder aux protocoles thérapeutiques (possible uniquement pour un petit nombre de pathologies), ou encore de pouvoir se rapprocher d'associations de patients.

Si une DI est majoritairement la conséquence d'anomalies survenues au cours développement cérébral, les causes possibles sont multiples, et beaucoup ne sont pas encore identifiées, ce qui entrave le processus diagnostic. Ainsi, le diagnostic étiologique reste indéterminé dans près de 40% des cas. Néanmoins, les causes identifiées peuvent être classées en deux groupes principaux : les causes environnementales (toxiques, infectieuses, périnatales, maternelles, traumatiques...) dans 15-20% des cas et les causes génétiques dans 80% des cas (2) (figure 1).

Il est estimé qu'environ 1/3 des 20 000 gènes sont exprimés au niveau du cerveau, expliquant l'extrême hétérogénéité génétique des DI avec tous les modes de transmission, mendéliens (autosomique dominant (AD), récessif (AR) ou lié au chromosome X (LX)) et non mendéliens (sans altération de la séquence nucléotidique) décrits. Parmi les causes génétiques de DI, les anomalies chromosomiques sont fréquentes. Elles sont observées dans 0,7 % des naissances vivantes, et concernent environ 15 % des DI (3,4). Il s'agit d'anomalies de nombre de chromosomes (aneuploïdie), ou d'anomalies de structures (délétions, duplications, dérivés de translocations ou d'inversions). La trisomie 21 ou syndrome de Down est la première maladie génétique à avoir été identifiée (5), c'est également la première cause de DI avec une prévalence d'environ 1 sur 2000 naissances vivantes en France.

Les autres causes génétiques de DI sont essentiellement monogéniques, c'est à dire secondaire à l'altération d'un gène, avec actuellement plus de 1000 gènes impliqués. Dans les pays occidentaux, où il y a un faible taux de consanguinité, environ 80% des cas sont expliqués par des variants *de novo*, c'est-à-dire des variants non hérités, apparues de façon accidentelle chez l'individu atteint. Les pathologies à transmission récessive n'expliquent quant à elles que 3 à 4% des cas.

2. La déficience intellectuelle liée au chromosome X (DILX)

La plus grande prévalence de la DI chez les hommes (ratio hommes/femmes de 1,3/1) et l'observation de grandes familles avec uniquement des garçons atteints ont suggéré une contribution importante des anomalies génétiques du chromosome X aux étiologies des DI, avant même l'avènement de l'ère moléculaire. La possibilité de réaliser des analyses de liaison dans ces grandes familles dans lesquelles la DI ségrégeait selon un mode de transmission lié au chromosome X ont conduit à analyser les gènes présents sur le chromosome X à partir des années 1990 (6). Aujourd'hui, une centaine de gènes ont été identifiés comme responsables de DILX, et expliquent entre 10 et 12 % des causes de DI chez l'homme (7).

C'est ainsi que le syndrome de l'X fragile (FXS) (MIM : 300624) a été décrit, tout d'abord par la description d'un arbre généalogique de DILX en 1943 (8), puis par la description d'un chromosome X inhabituel associé à la DI en 1969 (9). Ce n'est qu'en 1991 que le mécanisme moléculaire (expansion de triplets CGG au niveau du promoteur du gène

FMR1) a été identifié (10). Le FXS est la seconde cause de DI (1%) après la trisomie 21 et représente la première cause monogénique et la première cause de DI héritée (11).

Comme nous venons de le voir, les DILX ont été initialement décrites chez des garçons atteints, avec des apparentées féminines conductrices de la maladie mais asymptomatiques. Ceci s'explique par la présence du deuxième chromosome X chez la femme qui permet de compenser l'allèle défaillant. Cependant, un nombre croissant de gènes ont été identifiés comme responsables de DI chez des filles. Le plus fréquent serait *DDX3X* qui, en 2015, a été estimé responsable de 1 à 3% des DI non diagnostiquées chez les filles (12). Dans ces familles, peu de variants sont hérités, ils sont majoritairement *de novo* avec une probable létalité chez les garçons. Cependant, pour beaucoup de ces gènes, des garçons atteints ont été décrits avec des variants hérités de mères asymptomatiques. Les variants sont souvent faux-sens et hypomorphes avec une conservation partielle de la fonction protéique (12,13). Les modes de transmission dominant lié à l'X (DLX) et récessif lié à l'X (RLX) peuvent donc être retrouvés dans les mêmes gènes.

Le diagnostic moléculaire des DILX est primordial car le risque de récurrence dans ces familles est très important. Cela permet donc de donner aux familles un conseil génétique optimal et de leur offrir la possibilité de recourir à un DPN ou à un DPI, ce qui n'est possible que lorsque le variant causal a été identifié.

Complications de la prématurité	5 %
Causes environnementales	13 %
Anomalies chromosomiques	15 %
Maladies métaboliques	8 %
Syndromes reconnaissables	2 %
DI liées au chromosome X	10 %
Autres maladies monogéniques connues	10 %
DI idiopathiques	35-40 %

Figure 1 : Etiologies des déficiences intellectuelles.

Issu de : Inserm. Déficiences intellectuelles. Collection Expertise collective. Montrouge : EDP Sciences, 2016.

3. Difficultés et errance diagnostiques

Comme nous l'avons vu précédemment, le diagnostic de la DI est complexe de par l'extrême hétérogénéité clinique et étiologique, mais également par la complexité du parcours de soin (difficulté à être orienté vers un professionnel et une structure de santé adéquates). Cette difficulté de diagnostic est responsable d'errance diagnostique avec aujourd'hui, seulement une personne atteinte de DI sur deux disposant d'un diagnostic précis. En moyenne le délai diagnostic est de un an et demi, et peut dépasser 5 ans pour plus d'un quart des individus. L'errance diagnostique est définie par le plan national maladies rares (PNMR) 3 comme étant la période allant de l'apparition des premiers symptômes à la date à laquelle un diagnostic précis est posé. Celle-ci peut engendrer : des soins ou examens médicaux inappropriés, une incompréhension des familles, une mauvaise reconnaissance du handicap, des démarches administratives superflues, ainsi qu'une information médicale inadaptée notamment sur le risque de récurrence (conseil génétique). La réduction de cette errance diagnostique est un des objectifs du PNMR3. Un des moyens est de favoriser l'accès des patients aux récentes avancées technologiques du séquençage haut débit (SHD).

4. Evolution des technologies diagnostiques en génétiques humaines

L'évolution des technologies avec l'apparition successive du caryotype, de la FISH (*Fluorescence in situ hybridization*), du séquençage Sanger, de l'ACPA (analyse chromosomique sur puce à ADN) et enfin du séquençage haut débit (SHD) a permis de détecter des événements moléculaires de plus en plus petits, avec une efficacité diagnostique qui ne cesse d'augmenter (figure 2).

En effet, après les premières observations de chromosomes en 1842 par Carl Wilhelm von Nägeli, c'est en 1956, que le nombre exact de chromosomes humains a été établi à 46 par Joe Hin Tjio et Albert Levan (14). Cette technique utilisant un choc hypotonique permettant d'étaler les chromosomes était très peu résolutive et permettait le classement des chromosomes selon leur taille et leur indice centromérique. C'est ainsi que la première maladie génétique, la trisomie 21 a été découverte en 1959 par Jérôme Lejeune et al. (5), rapidement suivi par l'identification d'autres aneuploïdies dans les années 60 (syndrome de

Klinefelter, syndrome de Turner, trisomies 13 et 18 (15–18)). La technique de caryotypage a été améliorée en 1970, lorsque Torbjorn Caspersson et ses collègues ont réussi à colorer les chromosomes pour la première fois permettant une analyse plus fine de la structure des chromosomes. Néanmoins, l'analyse reste limitée par la visualisation des bandes avec une résolution théorique d'une bande chromosomique soit environ 5Mb.

Dans les années 1986-1988 l'utilisation de sondes complémentaires de l'ADN marquées par fluorescence ont permis le développement de l'hybridation in situ en fluorescence (FISH) par les équipes de Pinkel (19,20) et de Cremer (21). Si cette technique de cytogénétique moléculaire permet une augmentation de la résolution à la sous bande chromosomique (environ 150kb), il s'agit d'une analyse ciblée n'autorisant pas l'exploration du génome entier.

En 1977, le séquençage Sanger est décrit par l'équipe de Frederick Sanger(22). Il s'agit également d'une technique ciblée reposant sur le recopiage d'un brin matrice spécifique d'ADN en incorporant des nucléotides marqués en fluorescence (figure 3). Cette technique permet l'identification de variants ponctuels. Elle a été progressivement automatisée permettant une augmentation des capacités de séquençage mais s'est heurtée à des limitations de coûts et de débits. A titre d'exemple, le séquençage du premier génome humain par une collaboration internationale (*Human Genome Project*)(23) et une société privée (Celera Genomics)(24), aura coûté 3 milliards de dollars et nécessité une dizaine d'années en utilisant des séquenceurs répartis dans plusieurs laboratoires à travers le monde.

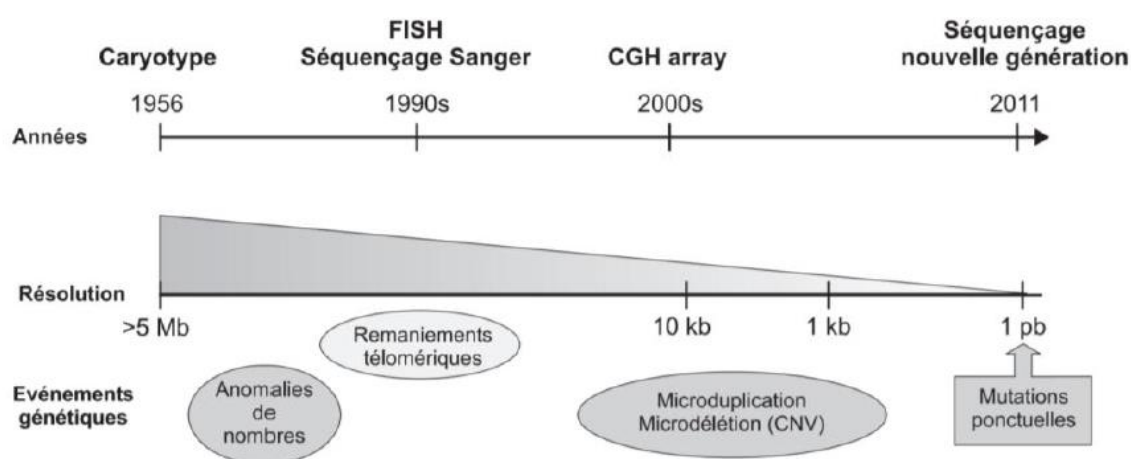
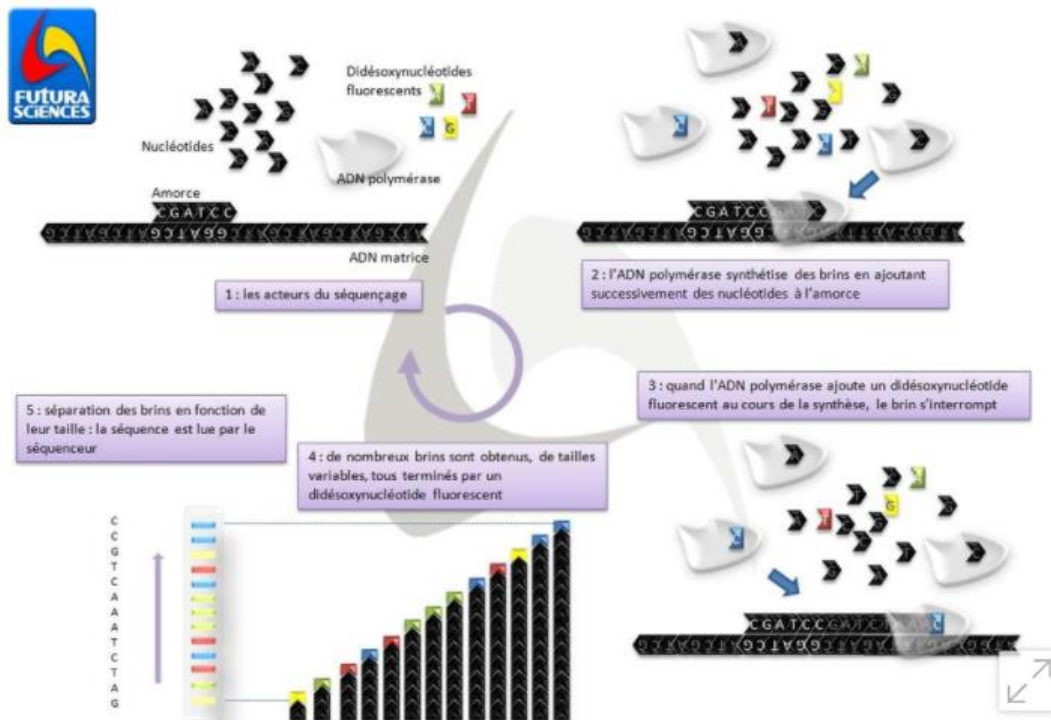


Figure 2 : Evolution des technologies pour le diagnostic des causes de DI.

Issu de : Inserm. Déficiences intellectuelles. Collection Expertise collective. Montrouge : EDP Sciences, 2016.



Les différentes étapes du séquençage. © Claire Peltier/Futura-Sciences

Figure 3 : Les différentes étapes du séquençage.

Issu de <https://www.futura-sciences.com/sante/actualites/genetique-sequençage-adn-nuls-26754/>

Dans les années 2000, c'est l'avènement de la technique d'analyse comparative sur puce à ADN (ACPA ou CGH-array pour Comparative Genomic Hybridization array) qui est une technique de quantification relative pangénomique. Cette technique est dite comparative et compétitive et repose sur l'hybridation compétitive de deux ADN (individu à tester et référence), marqués par des molécules fluorescentes, sur une même puce à ADN et aboutissant à un ratio comparatif de l'intensité des deux fluorescences. La CGH-array permet de mettre en évidence des remaniements chromosomiques déséquilibrés (appelés CNV pour Copy Number Variant) de petite taille avec une résolution de l'ordre du gène (résolution moyenne de 400kb voire jusqu'à quelques dizaines de kb), ce qui a permis de détecter deux fois plus d'anomalies chromosomiques chez les patients avec DI que les techniques de cytogénétiques usuelles.

5. Le séquençage haut-débit

Le séquençage haut débit (SHD) également connu sous le nom de NGS (pour Next Generation Sequencing) est une révolution biotechnologique apparue vers 2004 avec les séquenceurs 2^{ème} génération. le pyroséquençage 454(25). Cette technologie permet de séquencer de grandes quantités d'ADN en des temps record en reposant sur un séquençage massif, parallèle et en temps réel. Il permet d'obtenir la séquence d'un génome complet ou de parties de génome bornées par PCR (séquençage ciblé comme l'exome ou les panels de gènes). Depuis, plusieurs technologies ont été développées mais reposent sur des étapes communes : 1) **préparation des librairies** (ensemble des fragments d'ADN que l'on veut séquencer) par fragmentation aléatoire de l'ADN (soit mécanique soit enzymatique) suivie de la liaison avec des petites séquences spécifiques (adaptateurs). 2) Uniquement si séquençage ciblé, étape de **capture**: sélection des fragments d'ADN désirés par hybridation à des séquences complémentaires et amplification par PCR. Les fragments d'ADN qui ne s'hybrident pas sont éliminés. ; 3) **séquençage** : l'ADN est séquencé en utilisant différentes approches en fonction de la technologie utilisée. À chaque incorporation d'un nucléotide, un signal est détecté : des photons (Roche 454), de la fluorescence (Illumina, SOLiD), une variation du pH (Ion Torrent).

Ensuite, place à la bioinformatique. L'ensemble des données est enregistré dans un fichier Fastq contenant les séquences (appelées « reads ») et leurs scores de qualité (score Phred). Trois grandes étapes sont encore nécessaires avant d'obtenir et de pouvoir interpréter les données finales : 1) **l'alignement** : permettant de localiser des séquences nucléotidiques par rapport au génome de référence ; 2) **l'appel des variants génétiques** : permettant de détecter les variants nucléotidiques par rapport au génome de référence ; 3) **l'annotation des variants** permettant d'obtenir les informations nécessaires à leur interprétation via de nombreux outils (bases de données génotypiques, phénotypiques, outils de prédictions...) (figure 4).

A l'ère du SHD, l'étape limite n'est plus le séquençage mais la gestion des données. Les exomes génèrent environ 50 000 variants et les génomes entre 3 et 4 millions. L'interprétation constitue donc un véritable défi et repose sur un faisceau d'arguments : pertinence clinique du gène, impact moléculaire de la variation (prédictions théoriques sur le gène, l'ARN messager et la protéine), fréquence de la variation en population générale, présence de la variation chez d'autres patients, ségrégation familiale (recherche de la variation chez les apparentés sains ou atteints), études fonctionnelles (pour évaluer l'impact de la

variation sur l'ARNm, la protéine)... Pour harmoniser l'interprétation des variants, des recommandations ont été établies et approuvées par la communauté, en particulier celles de la société américaine de génétique (ACMG-AMP pour American College of Medical Genetics & the Association for Molecular Pathology). Ces recommandations précisent les éléments pertinents à prendre en compte, ainsi que leur pondération dans la décision et aboutissent à une classification en 5 classes : Classe 1 : Variant bénin ; Classe 2 : Variant probablement bénin ; Classe 3 : Variant de signification inconnue ; Classe 4 : Variant probablement pathogène ; Classe 5 : Variant pathogène (26,27).

Le SHD par sa possibilité d'explorer l'ensemble du génome permet de répondre à l'hétérogénéité extrême des malades rares. Il a permis de diminuer drastiquement la durée et le coût du séquençage, avec un coût divisé par plus d'un million en une quinzaine d'années. Actuellement, le séquençage d'un génome nécessite trois jours et à un coût réactif avoisinant les 1000 dollars (figure 5). Par ailleurs, son arrivée a accéléré la découverte de nouveaux gènes avec chaque année, plus de 60 nouvelles maladies génétiques décrites (figure 6).



Figure 4 : Diminution du coût du séquençage.

Issu de : <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>

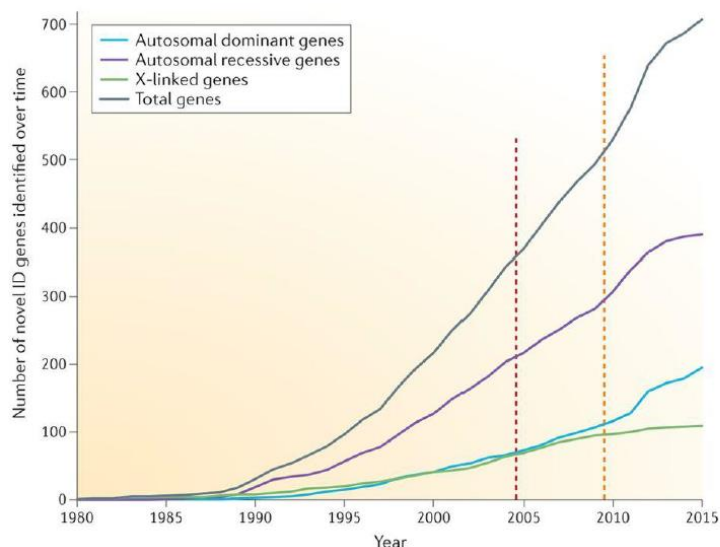


Figure 5 : Identification croissante des gènes responsables de DI.

Tiré de Vissers *et al.* 2015.

6. Stratégie diagnostique actuelle (figure 7)

La reconnaissance initiale d'une anomalie du développement ou d'une DI chez un individu est le plus souvent faite par un médecin généraliste, un pédiatre ou même les parents de l'individu atteint. L'individu est alors orienté en consultation de génétique qui a plusieurs rôles : rencontrer les familles pour comprendre leurs histoires, évaluer cliniquement les individus pour rechercher des signes associés/comorbidités, étayer les hypothèses diagnostiques, proposer des explorations complémentaires si nécessaires, poser un diagnostic, assurer un suivi adapté, prodiguer un conseil génétique...

En pratique, la consultation de génétique médicale se déroule en plusieurs temps : 1) Un **interrogatoire** complet et rigoureux centré sur l'individu et sa famille. L'enquête familiale est un élément essentiel du bilan. Celle-ci permet parfois d'évoquer un mode de transmission spécifique et donc de restreindre le champ des hypothèses diagnostiques. L'interrogatoire se poursuit par l'histoire personnelle de l'individu : grossesse (prise de toxiques, alcool, agents infectieux...), accouchement, période néonatale (hypotonie, difficultés alimentaires...), enfance (étapes du développement psychomoteur, épilepsie...). 2) Un **examen clinique** exhaustif comprenant un examen général complet en insistant sur l'examen neurologique et l'examen morphologique notamment du visage, des extrémités, des organes génitaux externes... 3) Des **examens complémentaires orientés** : de biologie, d'imagerie et enfin, le cas échéant, des examens génétiques.

Le généticien clinicien ne travaille pas seul mais en lien étroit avec de nombreux interlocuteurs (autres généticiens cliniciens, pédiatres, neuropédiatres, pédo-psychiatres, généticiens biologistes...) permettant ainsi des « regards croisés ». Dans certaines situations, la DI est associée à un syndrome reconnaissable permettant la réalisation d'analyses ciblées (FISH, séquençage Sanger d'un gène, panel de quelques gènes...). Cependant, dans la majorité des cas, l'hétérogénéité clinique et la diversité des causes génétiques ne permettent pas l'orientation vers un syndrome génétique particulier.

Si le caryotype a été une technique précieuse qui a permis historiquement l'identification des premières maladies génétiques, il a été démontré que l'ACPA pouvait détecter une majorité des anomalies identifiées par le caryotype mais également des remaniements cryptiques, à l'exception des remaniements structuraux équilibrés (translocations et inversions). L'ACPA est aujourd'hui l'examen de choix en routine.

Par ailleurs, nous avons vu précédemment que le FXS était la seconde cause de DI et la première cause de DI héréditaire. Le diagnostic s'effectue par des méthodes de PCR ciblées qui quantifient la taille des répétitions CGG et le statut de méthylation de l'allèle. Ces deux analyses sont simples et accessibles en routine. Ainsi, la recherche de FXS est aujourd'hui, avec l'ACPA, réalisée en première intention chez les individus présentant une DI sans orientation clinique.

Le SHD par sa possibilité d'explorer l'ensemble du génome permet de répondre à l'hétérogénéité extrême des DI. Son arrivée a accéléré la découverte de nouveaux gènes impliqués dans les DI avec de façon générale assez peu de récurrence, justifiant l'adoption de l'exome dans la démarche diagnostique. Cette technique est aujourd'hui largement utilisée en routine dans le diagnostic de la DI, et arrive de plus en plus tôt dans la stratégie diagnostique, permettant de réduire l'errance diagnostique. Cette stratégie notamment utilisée au CHU de Nantes permet de poser un diagnostic étiologique pour un tiers des familles environ. Néanmoins, la démarche diagnostique est encore très dépendante des possibilités locales où l'exome n'est pas toujours accessible. Si la plupart des gènes ne sont responsables que d'une très faible proportion de cas, certains sont plus fréquents. Ainsi, des panels de gènes impliqués dans la DI existent, notamment le panel DI 44 comportant les 44 gènes majeurs permettant un diagnostic dans environ 10% des cas.

En l'absence de diagnostic étiologique (environ 50 %), un suivi médical, une réévaluation clinique et une ré-analyse des données de séquençage en fonction des connaissances actuelles sont réalisés et depuis quelques années, il existe la possibilité de

poursuivre les explorations par le séquençage du génome en recherche. Le Plan France Médecine Génomique 2025 (PFMG 2025), porté par l'Etat, vise à intégrer la médecine génomique dans le système de soins. Il doit permettre la généralisation de l'accès au séquençage pangénomique ainsi qu'une prise en charge diagnostique et thérapeutique plus personnalisée. Aujourd'hui, deux plateformes de séquençage génomique à très haut débit sont opérationnelles et permettent le séquençage du génome entier des patients dans le cadre du soin.

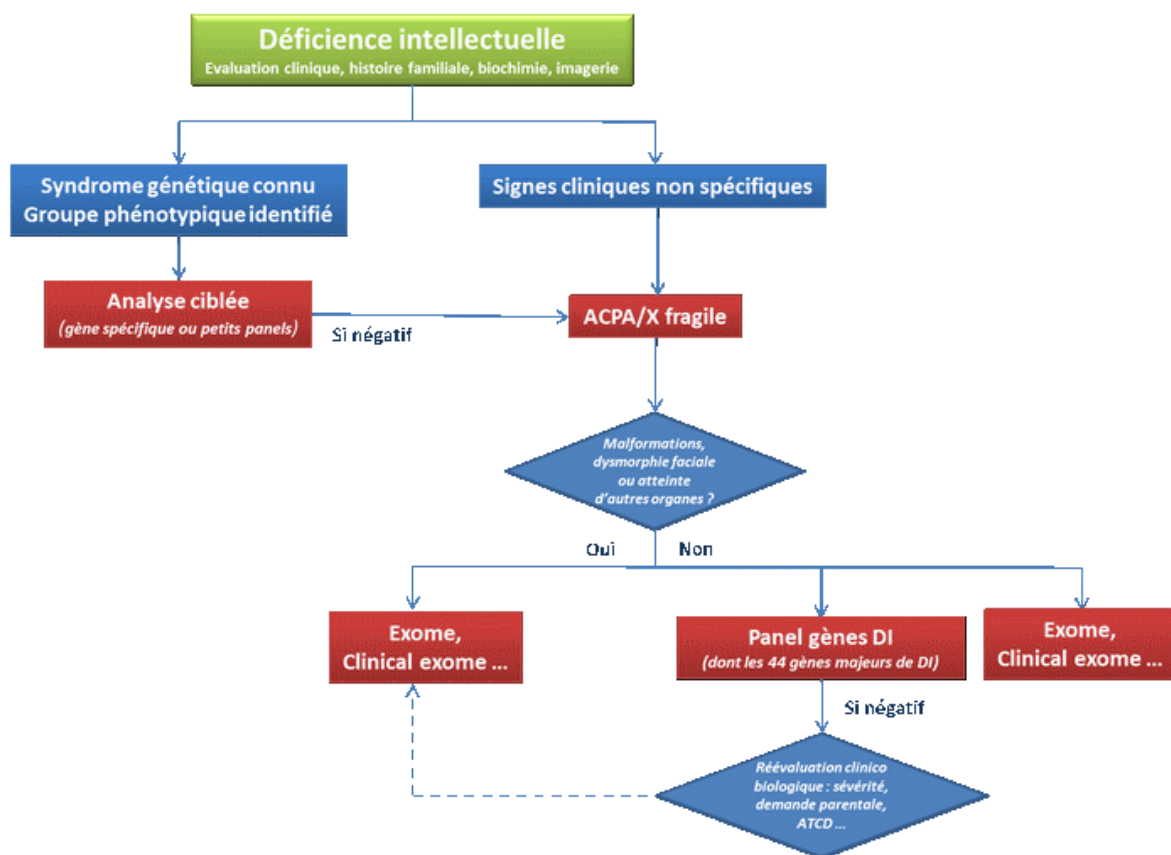


Figure 6 : Stratégie diagnostique dans la DI.

Issu de : <https://www.defiscience.fr/diagnostic/diagnostic-etilogique/>

7. *ARHGEF9*

ARHGEF9 est un gène localisé en Xq11.1 qui comporte dix exons et qui code une protéine de 516 acides aminés, la collybistine (Cb), une Rho-like GTPase spécifique du cerveau.

Avec la géphyrine, la Cb permet le recrutement de récepteurs GABAergiques et glycinergiques au niveau post-synaptique. En effet ces neurorécepteurs s'ancrent dans la membrane plasmique en reposant sur un réseau sous membranaire constitué de géphyrine. La Cb possède trois domaines fonctionnels : 1) un **domaine** C-terminal **PH** (pour *Pleckstrin homology domain*) permettant la liaison à la membrane plasmique via le phosphatidylinositol-3-phosphate ou PI3P ; 2) un **domaine** **DH** (pour *Dbl-homology domain*) ou **RhoGEF** (pour *Guanine nucleotide exchange factors*), permettant d'une part la liaison avec la géphyrine et d'autre part de catalyser les échanges GDP-GTP de *Cdc42* (petite RhoGTPase impliquée dans la régulation du cycle cellulaire) ; 3) un **domaine** N-terminal **SH3** (pour SRC Homology 3 Domain) permettant la liaison à la neurologine 2 (28,29) (figure 9). Ainsi, la Cb a un rôle essentiel dans la synaptogénèse inhibitrice et dans la neurogenèse.

Chez l'homme, des variants génétiques pathogènes dans *ARHGEF9* ont été associées à une DILX initialement décrite chez des garçons. Le tableau clinique est très variable allant d'une DI modérée sans épilepsie à l'encéphalopathie épileptique. Par ailleurs, le spectre phénotypique associe : une DI modérée à sévère, un retard de développement, une épilepsie, un trouble du spectre autistique (TSA) et des traits dysmorphiques. À ce jour, 25 variants pathogènes ont été rapportés dans 26 familles correspondant à 26 hommes et 10 femmes atteints (28–43). Des variants ponctuels et des délétions ont été identifiées chez les hommes alors que seuls de grands remaniements chromosomiques ont été rapportés chez les femmes atteintes (40,43). Parmi les dix femmes atteintes, les cinq premières rapportées présentaient un biais d'inactivation du chromosome X (ICX) (40). Par ailleurs, les mères hétérozygotes des hommes atteints ont été décrites comme asymptomatiques. L'hypothèse avancée était alors celle d'un mode de transmission récessif lié à l'X (RLX) qui affecte les femmes lorsque l'ICX est biaisée (43).

Dans l'article qui suit, nous rapportons trois nouveaux variants (« perte de fonction ») dans *ARHGEF9* (NM_015185.2): un variant synonyme *de novo* affectant l'épissage (c.1056G>A, p.(Lys352=)) chez une fille atteinte de DI avec épilepsie, un variant non-sens chez une autre fille atteinte de DI isolée (c.865C>T, p.(Arg289*)) hérité de son père asymptomatique en mosaïque (24%), et un variant non-sens *de novo* chez un garçon

(c.899G>A; p.(Trp300*)) atteint d'encéphalopathie épileptique. Nous suggérons donc que les variants faux-sens sont responsables d'une DI RLX affectant les garçons, tandis que les variants « perte de fonction » sont responsables d'une DI DLX affectant les garçons et les filles avec ou sans biais d'ICX.

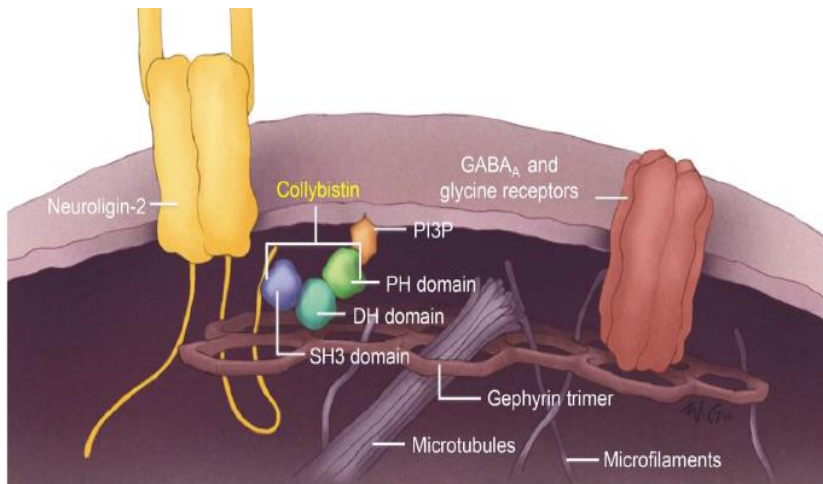


Figure 7 : Représentation schématique des interactions de la collybistine dans la formation de la géphyrine et du recrutement des récepteurs GABAergiques et glycinergiques dans la membrane postsynaptique.

Tiré de Alber *et al.* 2017.

L'article qui suit a été soumis dans Human mutation et est en cours de review.

II. Article

Loss-of-function variants in *ARHGEF9* cause an X-linked intellectual disability dominant disorder.

Leïla Ghesh¹, Thomas Besnard^{1,2}, Mathilde Nizon^{1,2}, Eva Trochu¹, Gaëlle Landeau-Trottier¹, Flora Breheret¹, Christel Thauvin-Robinet^{3,4,5,6}, Ange-Line Bruel^{3,4,6}, Paul Kuentz^{3,4}, Christine Coubes⁷, Cyril Mignot^{8,9}, Boris Keren⁹, Stéphane Bézieau^{1,2}, Benjamin Cogné^{1,2}.

¹Centre Hospitalier Universitaire de Nantes, Service de Génétique Médicale, 9 quai Moncoussu, 44093, Nantes, France.

²Université de Nantes, CNRS, INSERM, l'institut du thorax, 44000, Nantes, France.

³FHU TRANSLAD, Centre Hospitalier Universitaire Dijon-Bourgogne et Université de Bourgogne-Franche Comté, Dijon, France.

⁴Inserm UMR 1231 GAD « Génétique des Anomalies du Développement », Université de Bourgogne, Dijon, France.

⁵Centre de Génétique et Centre de Référence Déficience Intellectuelle de causes rares, Hôpital d'Enfants, Centre Hospitalier Universitaire Dijon-Bourgogne, Dijon, France.

⁶UF Innovation en diagnostic génomique des maladies rares, CHU Dijon, Dijon, France.

⁷Département de Génétique Médicale, Maladies Rares et Médecine Personnalisée, Hôpital Arnaud de Villeneuve, CHU de Montpellier, Montpellier, France.

⁸INSERM, U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, Paris, France.

⁹Service de Génétique clinique et Médicale, CHU Paris, Hôpital Pitié-Salpêtrière, Paris, France

Correspondence to :

Dr Benjamin Cogné, Service de Génétique Médicale, CHU Nantes, benjamin.cogne@chu-nantes.fr, 0033240084020

ABSTRACT

ARHGEF9 defects lead to an X-linked intellectual disability (XLID) disorder related to inhibitory synaptic dysfunction. This condition is more frequent in males with a few affected females reported. Up-to now, sequence variants and gross deletions have been identified in males, while only chromosomal aberrations have been reported in affected females who showed a skewed pattern of X-chromosome inactivation (XCI), suggesting an X-linked recessive (XLR) disorder.

We report three novel loss-of-function (LoF) variants in *ARHGEF9*: a *de novo* synonymous variant affecting splicing (NM_015185.2: c.1056G>A, p.(Lys352=)) in one female; a nonsense variant in another female (c.865C>T, p.(Arg289*)) and present as a mosaic in her father, and a *de novo* nonsense variant in a boy (c.899G>A; p.(Trp300*)). Both females showed a random XCI. Thus, we suggest that missense variants are responsible for an XLR disorder affecting males and that LoF variants, occurring *de novo*, may be responsible for an X-Linked dominant (XLD) disorder affecting males and females.

KEYWORDS: *ARHGEF9*, X-linked intellectual disability, female, loss-of-function, splice-site variant

MAIN TEXT

Intellectual disability (ID) is a developmental disorder characterized by significant limitations in both intellectual and adaptive functioning, in conceptual, social, and practical domains, with onset in the developmental period. It affects approximately 1% to 3% of the population(1), with an excess of males (male/female ratio of 1.3:1). Ten to 12% of ID in males are attributable to X-linked intellectual disability (XLID)(2). However, a growing proportion of ID in females is explained by *de novo* variants in genes located on the chromosome X such as *DDX3X* (3).

ARHGEF9 is located in Xq11.1 and consists of ten exons, spanning a genomic region of about 150kb. This gene encodes collybistin (Cb), a brain specific Rho-like GTPase. Together with gephyrin, Cb allows receptor recruitment in GABAergic and glycinergic synapses. Thus, Cb has a critical role in inhibitory synaptogenesis and in neurogenesis(4,5). In humans, *ARHGEF9* defects have been associated with moderate to severe ID, developmental delay (DD), epilepsy, autism spectrum disorder (ASD) and dysmorphic features. To date, 25

pathogenic variants have been reported in 26 families corresponding to 26 affected males and ten affected females(4–18). Point variants and gross deletions have been identified in males while only chromosomal aberrations have been reported in affected females(15,19). Among the ten affected females, the first five reported showed a skewed pattern of X-chromosome inactivation (XCI)(15). Furthermore, heterozygous mothers of affected males were reported as unaffected. It was thus concluded that pathogenic variants in *ARHGEF9* are responsible for an X-linked recessive (XLR) disorder, affecting females when the XCI is skewed(19).

We report three loss-of-function (LoF) variants identified by exome sequencing (ES) in two unrelated females presenting with moderate ID and in one severely affected male. All three individuals are not related to each other and were assembled by results of next-generation sequencing (NGS) using ClinVar (February 2019). Informed consent was obtained from the parents of the probands for molecular genetic analysis, and publication of the clinical data. Clinical and genetic data were provided by individual physician coauthors in accordance with local research and ethics requirements.

Female individual 1 was the fourth child, born to a non-consanguineous couple with no relevant familial history. She was born at term after an uneventful pregnancy with normal growth parameters. Hypotonia was observed in the neonatal period. At age four, she was referred to specialized genetic consultation because of developmental delay associated to speech delay. At the time of the study at 9.5 years of age, she presented with moderate ID and could say a few short sentences. She had attention disorder associated to shyness. At 8.5 years she experienced her first seizure, consisting in a hemifacial spasm occurring five times per month and becoming generalized tonic-clonic seizure (GTCS) once in a month. Leviteracetam was inefficient and seizures were partially controlled by valproate. Physical examination was without particularity. She had normal brain MRI and EEG showed posterior slow waves.

Female individual 2 was previously reported in a large cohort of 216 individuals with neurodevelopmental disorders screened by medical exome (20). She was the second child born to healthy but related parents. Her older brother was affected by DD associated to pervasive developmental disorders related to a *de novo* 1p36 deletion. Pregnancy and neonatal period were uneventful. At age ten, she was referred to specialized genetic consultation because of speech delay. At that time she presented with moderate ID, could say only a few words and was joyful. Physical examination was without particularity. She had normal brain MRI and EEG was normal apart from posterior slow waves.

Male individual 3 was born at 35 weeks of gestation to unrelated and healthy parents, after an uncomplicated twin pregnancy achieved by assisted reproduction. He was referred to specialized genetic consultation at six years because of epileptic encephalopathy. He showed early developmental delay with walking achievement after five years of age associated to nonverbal severe ID and severe epilepsy, his first myoclonic seizure occurred at 16 months, refractory to any monotherapy or combination of clonazepam, valproate, clobazam, leviteracetam and zonisamide. Physical examination showed axial hypotonia associated to light spastic hypertonia of limbs and neck. Brain MRI identified liquid space enlargement of the anterior hemisphere.

The three probands were negative for routine Fragile X syndrome testing and array comparative genomic hybridization analysis. ES was consequently performed in all three individuals as no specific clinical diagnosis was suggested, either on the proband only (P1) or in trio (P2 and P2). ES were performed following protocols described in Supplemental Methods. Variants were classified according to the American College of Medical Genetics (ACMG) guidelines. All candidate pathogenic variants were confirmed by Sanger sequencing (more details in supplemental file).

In female individual 1, an heterozygous synonymous *ARHGEF9* variant (c.1056G>A, p.(Lys352=)) was identified. This variant was absent from gnomAD. It was located on the last nucleotide of exon 7, and was predicted to affect the donor splice-site of exon 7. Sanger sequencing of the proband and her parents showed that the variant occurred *de novo*. In patient 2 and 3, LoF variant were identified. In female individual 2, medical exome identified an heterozygous variant (c.865C>T, p.(Arg289*)) in *ARHGEF9*. Familial segregation showed that her father carried it in a mosaic state (24% of reads). This variant was not present in the gnomAD database and it was predicted to induce a premature termination codon (PTC) in exon 6/10, thus probably leading to mRNA degradation through nonsense-mediated decay (NMD). Individual 2 was previously reported by Chérot et al(20) in a large cohort of 216 individuals with neurodevelopmental disorders screened by medical exome. The c.865C>T p.(Arg289*) variant identified has also been reported as pathogenic twice in ClinVar, in two unrelated individuals. This recurrent variant is located at a hypermutable CpG site, probably explaining its higher frequency. In male individual 3, ES identified an hemizygous variant (: c.899G>A; p.(Trp300*)) in *ARHGEF9*. Familial segregation showed the *de novo* origin of the variant. This variant was not present in the gnomAD database and as the previous variant, it was predicted to be pathogenic through the activation of NMD.

To assess the predicted impact of the synonymous c.1056G>A variant on splicing, blood samples from PAXgene Blood RNA tubes were collected from female individual 1 and her mother. Sanger sequencing of the proband's cDNA (performed following protocol described in Supplemental Methods) showed exon 7 skipping for the transcript from the allele carrying the variant, resulting in the in-frame deletion of the well conserved DH-PH linker domain p.(Gly309_Lys352del) (figure 1). A pathogenic missense variant in the DH-PH linker domain (c.950C>G; p.(Ser317Trp)) has already been associated to severe ID and moderate epilepsy in a male patient(13), and two intragenic deletions encompassing this domain have been reported in two females with moderate ID(19). The loss of a major functional domain thus probably leads to Cb LoF. We identified two variants inducing a PTC in exon 6 and predicted to lead to haploinsufficiency of *ARHGEF9* through NMD. Thus, all three variants reported here, were expected to result in Cb LoF, even if no functional study were performed.

We report here three novel predicted LoF variants within *ARHGEF9* in two unrelated females and in one male. The first female individual 1 exhibited moderate ID with moderate epilepsy; while the second female individual 2 exhibited moderate ID without seizure. The male individual 3 was severely affected by epileptic encephalopathy. To our knowledge, this is the first report of affected females with sequence variants in *ARHGEF9* and random XCI pattern.

Defects in *ARHGEF9* lead to an XLID disorder with high clinical variability ranging from moderate ID without epilepsy to epileptic encephalopathy. In affected individuals of both sexes, ID and speech delay are constant, furthermore, the phenotypic spectrum may include inconsistently, epilepsy, DD, ASD and dysmorphic features of varying clinical severity. Up-to now, 25 pathogenic variants have been reported in 26 families corresponding to 26 affected males and ten affected females (4–16). In males, 15 pathogenic variants have been identified including five altering the open reading frame (three deletions, one nonsense and one affecting a splice-site) and ten missense variants (table II, figure 2). An attempt of genotype-epilepsy phenotype correlation in males showed that the presence and severity of epilepsy was potentially related to the location and the molecular effect of the variant. 4/5 LoF variants were associated with severe ID, and among them the three deletions occurred *de novo*. The variant resulting in p.(Gln2*) was maternally inherited but potentially escapes NMD and the c.1300+2T>C variant was associated with moderate ID and was maternally inherited (review by Wang et al.(15) and Alber et al.(13), table II).

In affected females, ten *de novo* chromosomal abnormalities disrupting *ARHGEF9* have been identified (three deletions, two intragenic deletions, two translocations and one paracentric inversion). They all presented with a less severe phenotype than males, comprising constant moderate-to-severe ID and speech delay, and inconsistently moderate epilepsy and ASD (review by Aarabi et al.(19)). Interestingly, no sequence variant was identified in affected women, but only in apparently healthy mothers of affected males, which, combined with the fact that the first five affected females presented a skewed XCI pattern, led to consider this defect as an XLR disorder.

No sequence or structural variants resulting in LoF are listed in the gnomAD database, making *ARHGEF9* extremely LoF intolerant ($pLi = 1$). Moreover, only five nonsense variants are documented in the ClinVar database, including two potentially escaping NMD (the previously reported p.(Gln2*)(5) and the unreported p.(Trp505*)), and one splice-site, all of them classified as pathogenic or likely pathogenic. Among the 25 pathogenic variants described in the literature, in males, we observe a majority of missense variants (10/15) mostly inherited (2/10 occurred *de novo*), the other five were predicted LoF (5/15), three of which occurred *de novo*. All pathogenic variants identified in affected females (10/10) were chromosomal abnormalities and were non inherited events. Therefore, LoF variants in *ARHGEF9* are very rare events occurring *de novo*, only in affected individuals (except for the c.1300+2T>C variant (10), and for the variant resulting in p.(Gln2*) which probably escapes NMD, both of them were maternally inherited.), including females, in whom LoF variants are the only disease-causing events reported to date. Thus, even if functional data will be required to support it, we suggest that missense variants are responsible for a XLR disorder affecting males and females with skewed XCI and that LoF variants may be responsible for a XLD disorder affecting males and females with or without XCI and mainly caused by *de novo* variants.

In conclusion we report three novel LoF variants in *ARHGEF9*, including a synonymous variant, in two unrelated females presenting with moderate ID, and in one severely affected male. Synonymous variants can easily be missed by ES, as they are often filtered out during analysis. This report highlights the possible deleterious effect of such variants and the necessity to annotate them with splicing prediction score during exome analysis. Pathogenic variants in *ARHGEF9* were so far linked to an XLR disorder. We provide here evidences that LoF variants are probably responsible for a XLD condition, which is important to consider for future exome and genome analysis.

Individual	1	2	3
ARHGEF9 variant			
cDNA (NM_015185.2)	c.1056G>A p.(Lys352=)/Exon 7	c.865C>T	c.899G>A
Protein (NP_056000.1)	skipping	p.(Arg289*)	p.(Trp300*)
Genomic DNA (NC_000023.11)	g.63665886C>T	g.63674097G>A	g.63674063C>T
Exon	7	6	6
Zygosity	Heterozygous	Heterozygous Paternally (mosaic 24% of reads)	Hemizygous
Inheritance	<i>De novo</i>		<i>De novo</i>
X inactivation	Random	Random	N/A
Method of mutation detection	ES	ME	ES
Clinical Informations			
Gender	Female	Female	Male
Presenting age and symptoms	2y/DD	2y/Speech delay	14mo/Epileptic encephalopathy
Age at last investigation	9 ,5 y	10y	6y
Birth			
Gestational week	36	39	assisted reproduction 35
Birth weight (grams/SDS)	2660 (-1SD)	3200g (M)	2140 (-1SD)
Birth length (cm/SDS)	45 cm (-1,5SD)	50 cm (M)	45.5 (M/-1SD)
Birth head circumference (cm/SDS)	33 cm (M)	35 cm (M)	32 (M/-1SD)
Growth			
Growth failure yes/no	-	-	-
Height at age last investigation (cm/SDS)	134 cm (M)	140cm (+1SD)	112 (M/-1SD)
Weight at age last investigation (kg/SDS)	28kg (M)	30kg (M)	17 (-1/-2SD)
Head circumference at age last investigation (cm/SDS)	54,5 cm (+1,5SD)	56 cm (+2,5SD)	52,5(M/+1SD)

Central nervous system			
Degree of DD or ID	Moderate	Moderate	Severe
Age of walking	23 mo	16 mo	>5y
Age of first words	3y	2y	N/A
Speech abilities	Short sentences	Few words	No language
Hypotonia/hypertonia	Neonatal hypotonia	-	axial hypotonia++, light spastic hypertonia limbs and neck
Brain imaging done, age	Yes	Yes	Yes/2y
Anomalies in brain imaging	-	-	liquid space enlargement (anterior hemispheric)
Behavioural anomalies	Difficulty concentrating, shyness	Joyful	Tears, cry
Epilepsy			
Seizures with age of onset,	+/(8,5 y)	-	+/(14mo)
Seizures types	Hemifacial spasm//GTCS	N/A	Myoclonic
Seizures frequency	5 times per month	N/A	NR
EEG anomalies	Posterior slow waves	Posterior slow waves	interictal
Antiepileptic drugs	Valproate after leviteracetam	N/A	Clonazepam after valproate, clobazam, leviteracetam and zonisamide
Seizures outcome	Partially controled	N/A	Refractory
Dysmorphic features	-	-	-
Other features	-	-	-

Table I. Genetic and phenotypic features of the three individuals with *ARHGEF9* loss-of-function variants reported here.

DD: Developmental delay; ES: exome sequencing; GTCS: generalized tonic-clonic seizures; ID: intellectual disability; M: mean; ME: medical exome; mo: months, N/A: not applicable; NR: not reported; SD: standard deviation; y: years; +: abnormal finding positively identified; -: abnormal finding not present.

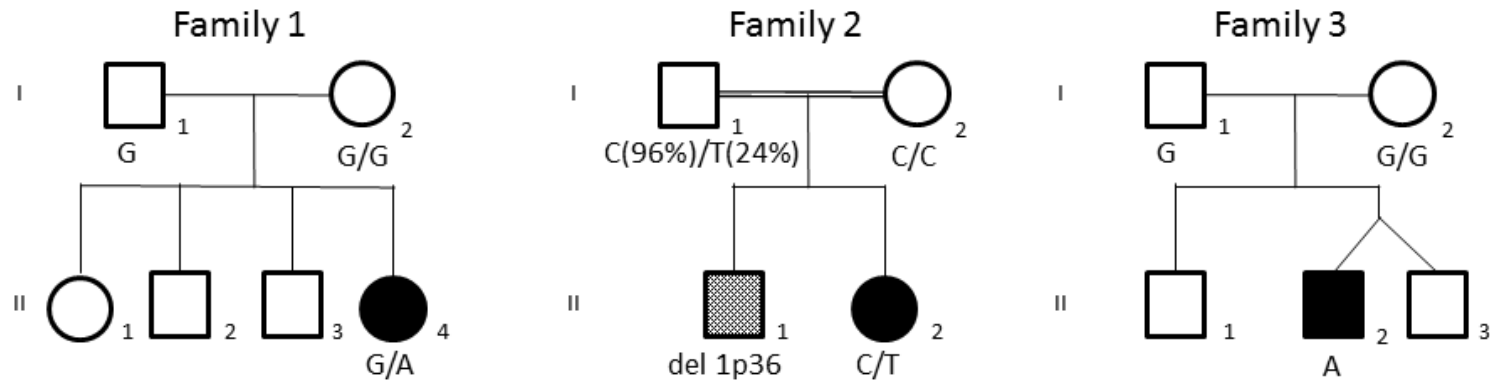
Sex	Variant type	Variant (recurrence)	Inheritance	Exon (10)	X Inactivation	ID	Epilepsy	Reference
M	Missense	c.164G>C/p.G55A	NR	2	NA	Severe	EE	[4]
M		c.311G>A/p.R104Q	De novo	3	NA	Severe	EE	[13]
M		c.530T>C/p.L177P	De novo	4	NA	Severe	Partially controlled	[13]
M		c.869G>A/p.R290H	NR	6	NA	Moderate	+	[13]
M		c.868C>T/p.R290C (4)	Maternally	6	NA	Severe	EE	[15]
M		c.950C>G/p.S317W (2)	Maternally	7	NA	Severe	Partially controlled	[13]
M		c.967G>A/p.G323R (4)	Maternally	7	NA	Moderate	Controlled	[14]
M		c.1012C>T/p.R338W (3)	Maternally	7	NA	Moderate	-	[12]
M		c.1067G>A/p.R356Q (3)	Maternally	8	NA	Moderate	-	[13]
M		c.1198G>A/p.E400K	Maternally	8	NA	Moderate	-	[10]
M	Nonsense	c.4C>T/p.Q2*	Maternally	1	NA	Severe	EE	[5]
M		c.899G>A/p.W300*	De novo	6	NA	Severe	EE	This study
M	Splicing	c.1300+2T>C	Maternally	-	NA	Moderate	+	[10]
M	Gross deletion	ChrX:61848414-63138698	De novo	-	NA	Severe	Controlled	[8]
M		ChrX:62321746-63058548	De novo	-	NA	Severe	Controlled	[5]
M		ChrX:62856174-63072861	De novo	-	NA	Severe	Controlled	[17]
F		ChrX:62970571-63052696	De novo	-	Random	Moderate	-	[18]
F		ChrX:62838630-62865334	De novo	-	Skewed	Moderate	-	[13]
F		ChrX:62854862-62862403	De novo	-	Skewed	Moderate	Controlled	[13]
F		NR	De novo	-	Random	Severe	Controlled	[11]
F		NR	De novo	-	Random	Severe	-	[11]
F	Complex rearrangement	46,X,inv(X)(q11.1;q27.3)	De novo	-	Skewed	Moderate	-	[6]

F		46,X,t(X;18)(q11.1;q11.21)	De novo	-	Skewed	Severe	+	[7]
F		46,X,t(X;20)(q12;P13)	De novo	-	Skewed	Severe	Controlled	[13]
F	Intragenic deletion	ChrX:62876206-62900076	De novo	-	Random	Moderate	-	[19]
F		ChrX:62872539-62929093	De novo	-	Random	Moderate	-	[19]
F	Splicing	c.1056G>A/E7 skipping	De novo	7	Random	Moderate	Partially controlled	This study
F	Nonsense	c.865C>T/p.R289*	Paternaly (mosaic 24% of reads)	6	Random	Moderate	-	This study

Table II. Genetic and phenotypic features of individuals with *ARHGEF9* pathogenic variants reported to date.

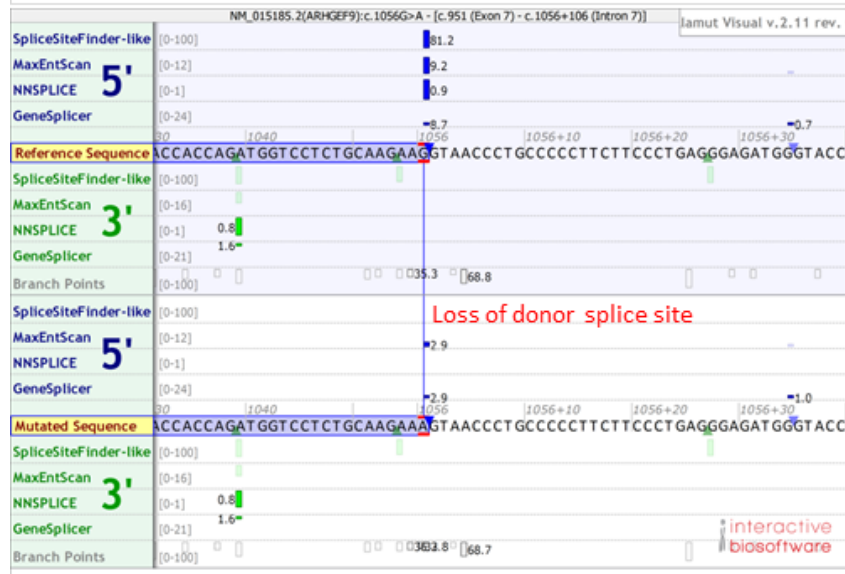
ID: intellectual disability; EE: epileptic encephalopathy; NA: not applicable; NR: not reported; +: abnormal finding positively identified but without precision; -: abnormal finding not present. Adapted from Wang et al. (15)

A

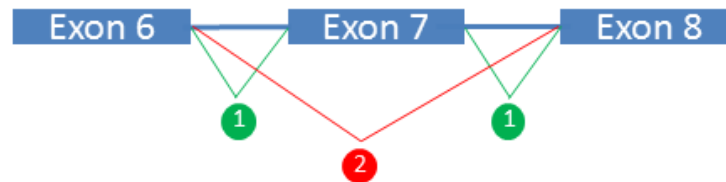


B

NM_015185.2(ARHGEF9):c.1056G>A (Exon 7) (Family 1)



genomic



transcript

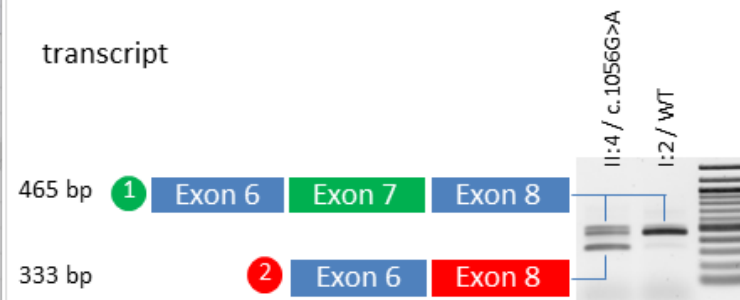


Figure 1.

A. Pedigree of the three families.

B. Predicted effect of the synonymous c.1056G>A variant on splicing. According to prediction programs Splice Site Finder-like, MaxEntScan, NNSplice, and Gene Splicer compiled by Alamut Visual v2.11 (Interactive Biosoftware, Rouen, France), the synonymous c.1056G>A variant induces a loss of canonical donor splice site (see left panel). The most likely predicted consequence is a skipping of in-frame *ARHGEF9* exon 7 (132 bp), which was confirmed in vitro by transcript analysis (see rights panels) ; PCR products were visualized on 1.5% agarose gel and Sanger sequencing were performed (data not shown). Two different transcripts were identified. Transcript 1 corresponds to the physiological splicing of the exon 7. The allele carrying the c.1056G>A variant leads to transcript 2, corresponding to the skipping of the exon 7. DNA variants are described according to the nomenclature established by the Human Genome Variation Society. Nomenclature HGVS V2.0 is defined according to *ARHGEF9* mRNA reference sequence NM_015185.2.

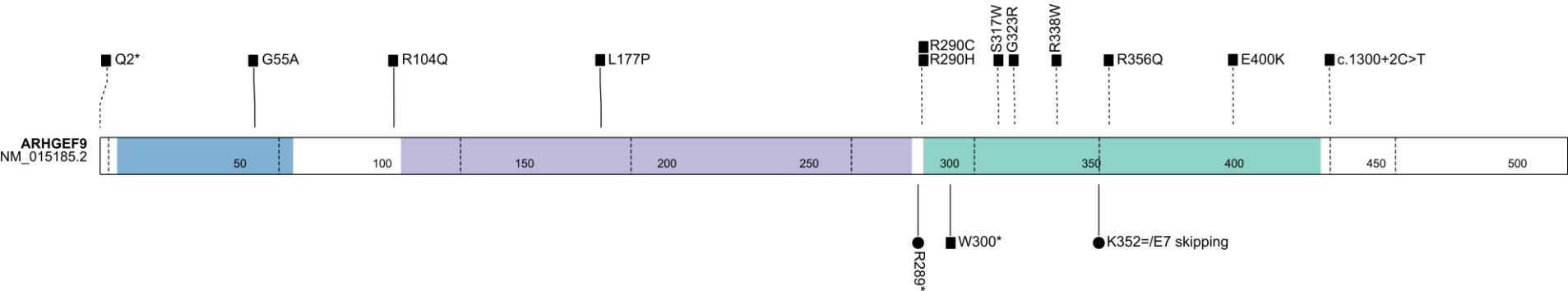


Figure 2. Newly and previously reported sequence variants in *ARHGEF9* (NM_015185.2) represented on Collybistin protein.

Above the protein are the previously reported variants while below the protein are the sequence variants identified in the current report. Squares represent affected male and circles represent affected females. Dotted variants are inherited from an asymptomatic mother; the other variants occurred *de novo*. Adapted from ProteinPaint (21)

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

WEB RESOURCES

DECIPHER, <https://decipher.sanger.ac.uk/>

ClinVar, <https://www.ncbi.nlm.nih.gov/clinvar/>

Ensembl, <https://www.ensembl.org/index.html>

gnomAD, <https://gnomad.broadinstitute.org/>

OMIM, <https://www.omim.org/>

UniProt, <https://www.uniprot.org/uniprot/>

Proteinpaint, <https://proteinpaint.stjude.org/>

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

1. Harris JC. Intellectual disability: Understanding its development, causes, classification, evaluation, and treatment. New York, NY, US: Oxford University Press; 2006. x, 429. (Intellectual disability: Understanding its development, causes, classification, evaluation, and treatment).
2. Ropers H-H, Hamel BCJ. X-linked mental retardation. *Nat Rev Genet*. 2005 Jan;6(1):46–57.
3. Nicola P, Blackburn PR, Rasmussen KJ, Bertsch NL, Klee EW, Hasadsri L, et al. De novo *DDX3X* missense variants in males appear viable and contribute to syndromic intellectual disability. *Am J Med Genet A*. 2019 Apr;179(4):570–8.
4. Harvey K, Duguid IC, Alldred MJ, Beatty SE, Ward H, Keep NH, et al. The GDP-GTP exchange factor collybistin: an essential determinant of neuronal gephyrin clustering. *J Neurosci Off J Soc Neurosci*. 2004 Jun 23;24(25):5816–26.
5. Shimojima K, Sugawara M, Shichiji M, Mukaida S, Takayama R, Imai K, et al. Loss-of-function mutation of collybistin is responsible for X-linked mental retardation associated with epilepsy. *J Hum Genet*. 2011 Aug;56(8):561–5.
6. Marco EJ, Abidi FE, Bristow J, Dean WB, Cotter P, Jeremy RJ, et al. *ARHGEF9* disruption in a female patient is associated with X linked mental retardation and sensory hyperarousal. *BMJ Case Rep*. 2009;2009.
7. Kalscheuer VM, Musante L, Fang C, Hoffmann K, Fuchs C, Carta E, et al. A balanced chromosomal translocation disrupting *ARHGEF9* is associated with epilepsy, anxiety, aggression, and mental retardation. *Hum Mutat*. 2009 Jan;30(1):61–8.
8. Lesca G, Till M, Labalme A, Vallee D, Hugononq C, Philip N, et al. De novo Xq11.11 microdeletion including *ARHGEF9* in a boy with mental retardation, epilepsy, macrosomia, and dysmorphic features. *Am J Med Genet A*. 2011 Jul;155A(7):1706–11.
9. Lemke JR, Riesch E, Scheurenbrand T, Schubach M, Wilhelm C, Steiner I, et al. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. *Epilepsia*. 2012 Aug;53(8):1387–98.
10. de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med*. 2012 Nov 15;367(20):1921–9.
11. Holman SK, Morgan T, Baujat G, Cormier-Daire V, Cho T-J, Lees M, et al. Osteopathia striata congenita with cranial sclerosis and intellectual disability due to contiguous gene deletions involving the *WTX* locus. *Clin Genet*. 2013 Mar;83(3):251–6.
12. Long P, May MM, James VM, Grannò S, Johnson JP, Tarpey P, et al. Missense Mutation R338W in *ARHGEF9* in a Family with X-linked Intellectual Disability with Variable Macrocephaly and Macro-Orchidism. *Front Mol Neurosci*. 2015;8:83.

13. Alber M, Kalscheuer VM, Marco E, Sherr E, Lesca G, Till M, et al. ARHGEF9 disease: Phenotype clarification and genotype-phenotype correlation. *Neurol Genet.* 2017 Jun;3(3):e148.
14. Klein KM, Pendziwiat M, Eilam A, Gilad R, Blatt I, Rosenow F, et al. The phenotypic spectrum of ARHGEF9 includes intellectual disability, focal epilepsy and febrile seizures. *J Neurol.* 2017 Jul;264(7):1421–5.
15. Wang J-Y, Zhou P, Wang J, Tang B, Su T, Liu X-R, et al. ARHGEF9 mutations in epileptic encephalopathy/intellectual disability: toward understanding the mechanism underlying phenotypic variation. *Neurogenetics.* 2018 Jan;19(1):9–16.
16. Chiou T-T, Long P, Schumann-Gillett A, Kanamarlapudi V, Haas SA, Harvey K, et al. Mutation p.R356Q in the Collybistin Phosphoinositide Binding Site Is Associated With Mild Intellectual Disability. *Front Mol Neurosci.* 2019;12:60.
17. Machado COF, Griesi-Oliveira K, Rosenberg C, Kok F, Martins S, Passos-Bueno MR, et al. Collybistin binds and inhibits mTORC1 signaling: a potential novel mechanism contributing to intellectual disability and autism. *Eur J Hum Genet EJHG.* 2016 Jan;24(1):59–65.
18. Bhat G, LaGrave D, Millson A, Herriges J, Lamb AN, Matalon R. Xq11.1-11.2 deletion involving ARHGEF9 in a girl with autism spectrum disorder. *Eur J Med Genet.* 2016 Sep;59(9):470–3.
19. Aarabi M, Kessler E, Madan-Khetarpal S, Surti U, Bellissimo D, Rajkovic A, et al. Autism spectrum disorder in females with ARHGEF9 alterations and a random pattern of X chromosome inactivation. *Eur J Med Genet.* 2018 Jul 23;
20. Chérot E, Keren B, Dubourg C, Carré W, Fradin M, Lavillaureix A, et al. Using medical exome sequencing to identify the causes of neurodevelopmental disorders: Experience of 2 clinical units and 216 patients. *Clin Genet.* 2018;93(3):567–76.
21. Zhou X, Edmonson MN, Wilkinson MR, Patel A, Wu G, Liu Y, et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat Genet.* 2016 Jan;48(1):4–6.

SUPPLEMENTARY FILES

MATERIAL AND METHODS

Exome sequencing strategy

For the first family, ES of the proband (P1) was performed with the Clinical Research Exome kit V2 (Agilent, Santa Clara, CA, USA) on Illumina HiSeq2500 (paired-end sequencing, 2×75 bp). A mean read depth of 87X was obtained. Reads were aligned to the human reference genome sequence hg19 using bwa (v0.7.3). Bioinformatics analyses were then performed according to the best practices of GATK (v3.4). Variants were annotated using ANNOVAR and filtered using in-house scripts to remove variants with an allele frequency above 0.5% in 1000 genome, gnomAD and in a local database of 489 exomes. The possible functional impact of amino-acid changes was predicted by SIFT (Sorting Intolerant from Tolerant), PolyPhen-2 hvar and CADD score (Combined Annotation Dependent Depletion). The effect on splicing of the variants localized in splicing consensus regions (-3 to $+8$ at the 5' *splice* site and -12 to $+2$ at the 3' *splice* site) was predicted with dbcsnv. The AlamutVisual software (Interactive biosoftware) was used to study retained variants.

For the second family trio WES was performed on a NextSeq 500 Sequencing System (Illumina, San Diego, CA), with a 2x 150bp high output sequencing kit after a 12-plex enrichment with SeqCap EZ MedExome kit (Roche, Basel, Switzerland), according to manufacturer's specifications. Sequence quality was assessed with FastQC 0.11.5, then the reads were mapped using BWA-MEM (version 0.7.13), sorted and indexed in a bam file (samtools 1.4.1), duplicates were flagged (sambamba 0.6.6), coverage was calculated (picard-tools 2.10.10). Variant calling was done with GATK 3.7 Haplotype Caller. Coverage for these samples was $>93\%$ at a 20x depth threshold. Variants were then annotated with SnpEff 4.3, dbNSFP 2.9.3, gnomAD, ClinVar, HGMD, OMIM, ACC genes (sup data) and an internal database. Filtering was performed with criteria based on the consequence on the protein (stop gain, splice site, frameshift indel, in frame indel, missense), frequency in gnomAD (threshold of 3 allele counts for de novo variants and 3 homozygous/hemizygous for homozygous, hemizygous and compound heterozygous variants). CNVs were called with CNVKit (PMID 27100738) and Excavator2 (PMID 27507884).

For the last family DNA was extracted from fetal frozen tissues and parental peripheral whole-blood samples, using the QIAamp DNA Mini Kit (Qiagen) following standard procedures. Exome capture and sequencing were performed at Integragen SA from 1 µg of genomic DNA per individual using the SureSelect Clinical Research Exome kit (Agilent). The resulting libraries were sequenced on a HiSeq 2000 (Illumina) according to the manufacturer's recommendations for paired-end 76 bp reads. Briefly, reads from whole exome sequencing were aligned to the human genome reference sequence (GRCh37/hg19 assembly) using Burrows-Wheeler aligner (BWA; version 0.7.15). Duplicate reads were marked using Picard MarkDuplicates (version 2.4.1) [<http://broadinstitute.github.io/picard/>] and aligned read were then processed using GATK BaseRecalibrator and PrintReads (Genome Analysis Toolkit; version 3.8) to recalibrate base quality scores, according to GATK Best Practices recommendations. Quality control was performed on all BAM files by calculating depth of coverage onto RefSeq database (release 2018-11-11) with GATK DepthOfCoverage. SNPs and indels were identified from BAM files using GATK HaplotypeCaller. SNPs present in cluster (more than three SNPs in a window size of 10 bases) were flagged in filter field using GATK VariantFiltration. All variants identified were annotated using SnpEff (version 4.3) to predict the effects of variants on genes. Additionally, SnpSift (version 4.3) was used to annotated variants with allele frequencies from the Exome Aggregation Consortium (ExAC; release 0.3.1) [<http://exac.broadinstitute.org/>], the Exome Sequencing Project (ESP) and the Genome Aggregation Database (gnomAD; release 2.0.2) (<http://gnomad.broadinstitute.org/>), and with some prediction scores (CADD, Polyphen-2, GERP and Grantham) based on dbnsfp database (release 2.9). Finally, an in-house python script was used to annotate variants with ClinVar Clinical significance (release 2019-01-08), OMIM disease-causing genes (Online Mendelian Inheritance in Man; release 2019-01-14) (<https://omim.org/>) and gnomAD constraint metrics (pLI and missense Z score). Variants were then filtered according to the following criteria: filter field is PASS; only variants covered with at least five reads and three alternative base supporting reads were reported; only protein altering variants in genes were selected based on SnpEff annotation; common variants present at a frequency above 1% in ExAC, ESP and gnomAD were excluded and variants present in more than five individuals of 100 in-house control exomes and 94 cohort exomes were excluded.

RT PCR cDNA

To assess the predicted impact of the synonymous c.1056G>A variant on splicing, blood samples from PAXgene Blood RNA tubes were collected from female individual 1 and her mother. Total RNA was isolated with the NucleoSpin RNA Plus Kit (Macherey-Nagel) according to the manufacturer's instructions. The concentration of total RNA was measured on a NanoDrop spectrophotometer (NanoDrop Technologies). Reverse transcription was performed with the ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs). PCRs were performed on cDNA with the HotStarTaq DNA Polymerase (Qiagen) according to the PCR amplicons (PCR conditions are available upon request). PCR products were analyzed on 1% agarose gel electrophoresis and Sanger sequenced.

III. Discussion

Dans cet article, nous avons rapporté trois nouveaux variants « perte de fonction » dans *ARHGEF9*, incluant un variant synonyme affectant l'épissage. Ces variants ont été identifiés par exome chez deux filles présentant une DI modérée plus ou moins une épilepsie et chez un garçon présentant une encéphalopathie épileptique. Nous suggérons donc que les variants faux-sens sont responsables d'une DI RLX affectant les garçons, tandis que les variants « perte de fonction » sont responsables d'une DI DLX affectant les garçons et les filles avec ou sans biais d'ICX.

Le variant c.1056G>A décrit dans l'article est un variant synonyme (changement nucléotidique n'entraînant pas de changement d'acide aminé) au niveau de la dernière base de l'exon 7 du gène *ARHGEF9*. Ce variant est prédit comme abolissant le site donneur d'épissage de cet exon par les outils de prédictions *in silico*. Afin de confirmer cet effet possible sur l'épissage, nous avons réalisé une étude d'ARN par séquençage Sanger de l'ADN complémentaire de la patiente 1 et nous avons pu confirmer un saut de l'exon 7 et une délétion en phase d'un domaine fonctionnel majeur.

1. Interprétation des variants synonymes :

Le séquençage de l'exome d'un individu génère environ 50 000 variants par rapport au génome de référence, ce qui représente un véritable défi d'interprétation. Une série de filtres permet de réduire le nombre de variants pour permettre une analyse biologique. Parmi ces filtres, les quelques 10 000 variants synonymes générés par exome(44) sont souvent considérés comme ayant un impact moléculaire minime. Les outils de prédictions *in silico* tels que SIFT(45), PolyPhen(46) ou encore CADD(47) sont des aides diagnostiques mais sont peu efficaces pour interpréter ces variants, car ils se basent essentiellement sur la conservation et l'écart physico-chimique des acides aminés. Ainsi, les variants synonymes sont souvent exclus d'emblée dans le processus d'interprétation.

Cependant, des variants synonymes ont été impliqués dans de nombreuses pathologies, notamment dans la déficience intellectuelle. En fonction de leur localisation, ces variants peuvent avoir un effet délétère sur de nombreux processus indispensables à la survie cellulaire. Un des effets le plus connu et le plus étudié de ces variants est l'altération de

l'épissage de l'ARN. Mueller et al. ont montré que 23% des variants synonymes localisées dans l'exon 7 du gène *SMN1* avaient un effet sur l'épissage(48). Ainsi, les algorithmes capables de prédire l'impact de variants sur l'épissage comme GeneSplicer(49), MaxEntScan, Human Splicing Finder(50) ou SPiCE(51) (dédiés au sites consensus donneur/accepteur) Branchpointer(52) (dédié au site de branchement), ESEFinder(53) (dédié aux séquences régulatrices) ou SPiP(54) (prédiction d'un défaut d'épissage indépendamment de la position du variant et du motif d'épissage impacté) constituent une première approche pour l'interprétation de ces variants. Il est également possible de réaliser des études fonctionnelles afin de confirmer l'éventuel impact délétère sur l'épissage. Toutefois, les pipelines actuels ne permettent pas l'annotation de tous les variants impactant l'épissage.

Les outils de prédiction d'épissage ne sont pas cependant pas suffisants. Les variants synonymes peuvent également impacter la transcription en altérant un site de liaison à un facteur de transcription(55), en altérant un îlot GpC régulant la méthylation de l'ADN et la reconnaissance des exons(56). Par ailleurs, il existe un biais dans l'usage des codons, les codons synonymes n'étant pas utilisés à la même fréquence. Ce biais est corrélé à l'expression des gènes(57), le taux de traduction(58), le repliement co-translationnel et la structure des protéines(59,60). En effet, plus un ARNm est riche en GC (notamment au niveau de la troisième base d'un codon), plus il est stable permettant une traduction plus intense. Ainsi, les gènes fortement exprimés sont riches en GC(61). Ensuite, remplacer un codon par un autre codon synonyme peut ralentir voire stopper la traduction par défaut de l'ARN de transfert (ARNt) correspondant au nouveau codon. Une hypothèse est que la composition des codons dans les régions codantes a coévolué avec l'abondance en ARNt pour atteindre un taux de traduction idéal(62). Des outils de prédictions *in silico* dédiés aux variants synonymes ont été développés en prenant en compte ces différents mécanismes (figure). Cependant, une étude récente visant à évaluer leur performance a montré qu'aucune de ces méthodes ne permet une interprétation fiable des variants synonymes à grande échelle(63). Par ailleurs, tous les mécanismes physiopathologiques en lien avec des variants synonymes ne sont pas élucidés, il reste donc de nombreux progrès à faire quant à la compréhension de l'impact de telles variants sur l'organisme et leur implication en pathologie humaine.

2. Interprétation des variants localisés sur le chromosome X

Le chromosome X a fait l'objet de nombreuses études génétiques et il contient environ 15 % des gènes connus impliqués en DI(64). De nombreuses pathologies LX ont d'abord été décrites avec un mode de transmission unique soit DLX (*DDX3X*), ou RLX (*ARHGEF9*). Cependant, à l'instar de *DDX3X*, pour lequel des garçons atteints ont été décrits avec des variants hérités de mères asymptomatiques, nous rapportons pour la première fois des variants « perte de fonction » dans *ARHGEF9* chez deux filles. Nous suggérons donc que les variants faux-sens sont responsables d'une DI RLX affectant les garçons, tandis que les variants « perte de fonction » sont responsables d'une DI DLX affectant les garçons et les filles. Ainsi, des modes de transmission DLX et RLX peuvent être retrouvés dans les mêmes gènes, ce qui est important à considérer lors d'une analyse génomique.

De façon général, la validation d'un variant candidat repose en partie sur la ségrégation familiale et sur l'absence de ce même variant dans une population contrôle. En effet, dans la classification ACMG, le caractère *de novo* d'une variation est un critère fort dans l'établissement de la pathogénicité de celle-ci. Cependant, de nombreuses variants pathogènes LX sont héritées de mères asymptomatiques. Il est donc parfois difficile de classer un variant LX en classe 4 ou 5 en l'absence de ce critère. Lorsque cela est possible, il est utile de tester les ascendants des mères conductrices afin de voir si ces variants pathogènes sont survenues *de novo* dans les générations précédentes. Par ailleurs, comme ces variants peuvent être hérités de mères conductrices dites asymptomatiques, celles-ci peuvent être retrouvées dans les bases de données de population contrôle. Il est donc important de vérifier le sexe des individus rapportés dans ces bases de données lors de l'interprétation d'un variant localisé sur le chromosome X. Il est également essentiel de procéder à un examen clinique complet des femmes conductrices dites asymptomatiques et de rechercher des signes cliniques minima. Il peut également être utile d'effectuer une étude du biais d'inactivation de l'X chez ces femmes hétérozygotes. Par ailleurs, dans une famille avec plusieurs individus atteints, un variant non causal peut co-ségréger fortuitement avec la pathologie et ainsi fausser l'interprétation. Afin de préciser ce point, il a été proposé de calculer la probabilité (« N ») que la présence du variant d'intérêt (selon le mode de transmission présumé) chez les individus atteints, soit due au hasard plutôt que due à une co-ségrégation(65).

Si l'interprétation des données issues du SHD est complexe, l'interprétation des variants LX est particulièrement délicate. Hors dans le cadre d'une anomalie LX, où le risque

de récurrence est très important, l'identification du variant causal est primordiale pour permettre d'une part un conseil génétique optimal et d'autre part l'accès au DPN ou au DPI dans le cadre d'un projet parental.

3. Prise en charge médico-sociale :

Le diagnostic moléculaire d'une DI est une étape essentielle dans le parcours de soin des familles. En effet, outre le conseil génétique et l'accès au DPN/DPI, cela leur permet de mettre un nom sur la maladie, de préciser le pronostic, de mettre en place un suivi médical approprié, des mesures de prévention, d'accéder aux protocoles thérapeutiques (s'ils existent), et de pouvoir se rapprocher d'associations de patients.

Si le SHD est un outil diagnostique très performant, les médecins et le grand public ne connaissent pas toujours son existence. Dans ce contexte, la France joue un rôle pionnier dans le domaine des maladies rares : elle est le premier pays en Europe à avoir élaboré et mis en œuvre des plans nationaux dédiés. En effet, depuis 2004, à travers trois plans nationaux successifs, la France est structurée en réseau de diagnostic et de prise en charge sur le territoire, pour organiser le parcours de soins des personnes concernées par les maladies rares.

Un des objectifs principaux du PNMR 3 qui est en cours de déploiement et synergique avec le PFMG 2025, est de réduire l'errance diagnostique avec un objectif d'un an. Cet objectif est en partie réalisable en favorisant l'accès au SHD.

En Pays de la Loire, environ 165.000 personnes sont atteintes d'une maladie rare. C'est ainsi qu'a été créée la plate-forme régionale d'information et d'orientation des maladies rares (Prior : <https://www.chu-nantes.fr/plate-forme-regionale-d-information-et-d-orientation-des-maladies-rares-nantes-angers-prior>). Pluridisciplinaire et multiprofessionnelle, cette plateforme s'adresse aux patients et à leurs proches (orientation vers le CRMR/CCMR le plus proche, vers des associations de patients, une aide psychologique, une aide financière, administrative...), aux professionnels qui les accompagnent (information et mise en relation avec d'autres professionnels) et aux associations qui les soutiennent.

Aujourd'hui, il existe des centaines d'associations de patients. L'alliance maladies rares (<https://www.alliance-maladies-rares.org/>) est une association rassemblant plus de 230 associations comme par exemple : l'association Xtraordinaire (<http://www.xtraordinaire.org/index.php>) représentant les familles et patients confrontés à la DILX et l'association Sans Diagnostic et Unique

(<https://www.facebook.com/sansdiagnosticetunique/>) représentant les familles et patients confrontés aux anomalies du développement et syndromes malformatifs sans diagnostic

Une DI peut avoir des conséquences qui limitent l'activité ou réduisent la participation en société, et peut alors entraîner un handicap. Sur un plan pratique, toutes les ressources citées ci-dessus permettent d'orienter les familles vers les nombreux dispositifs visant à améliorer la vie quotidienne des patients et de leurs familles. Une partie de ces dispositifs nécessite une reconnaissance du handicap par la **Maison Départementale des Personnes Handicapées (MDPH)**. Un Plan Personnalisé de Compensation (PPC) contenant des mesures visant à répondre au projet de vie de la personne handicapée est soumis à la **Commission des droits et de l'autonomie des personnes handicapées (CDAPH)** qui évalue le taux d'incapacité de la personne handicapée, attribue la prestation de compensation, reconnaît la qualité de travailleur handicapé, et se prononce sur les dispositifs d'inclusion scolaire.

Par exemple, il existe notamment des aides financières visant à compenser le handicap soumises ou non à conditions de ressources (PCH, AAH, AEEH, AJPP, pension d'invalidité). Il existe également des dispositifs facilitant l'insertion scolaire et professionnelle. En effet, tout enfant de plus de 6 ans doit être scolarisé, que ce soit à l'école, en institution spécialisée, ou à domicile, et selon la loi, chaque enfant handicapé a le droit à un parcours de formation adapté, et peut le cas échéant bénéficier d'un accompagnement par un service d'accueil médico-social (ESMS). Ainsi le PPS (Projet Personnalisé de Scolarisation) inclus dans le PPC permet: 1) **en classe ordinaire** d'avoir l'aide d'un AESH (Accompagnant des Elèves en Situation de Handicap) ; 2) **en milieu ordinaire** d'être scolarisé en ULIS (Unités localisées pour l'inclusion Scolaire) ; 3) d'être scolarisé dans un **établissement spécialisé** (IME pour Instituts Médico-Educatifs, EEAP pour Etablissement ou service pour enfants et adolescents polyhandicapés, ITEP pour Institut thérapeutique, éducatif et pédagogique) ; 4) **à domicile** d'obtenir les services d'un SESSAD (Service d'éducation spéciale et de soins à domicile). Des ESMS existent également pour les adultes de moins de 60 ans et d'autres pour les adultes de plus de 60 ans.

Si tous ces dispositifs visent à améliorer la qualité de vie des patients et des familles confrontées au handicap, les démarches administratives permettant l'ouverture des droits peuvent se révéler ardues. Le dossier MDPH est rempli par le patient ou son représentant légal et vise à représenter le parcours et le quotidien de la personne malade afin d'évaluer ses besoins. Dans ce dossier, le certificat médical est une pièce obligatoire. Cependant, il n'est pas toujours évident de décrire tous les problèmes (de santé ou autres) auxquels sont

confrontés les patients, ce qui a une incidence sur l'obtention des droits. L'établissement d'un diagnostic génétique est un élément objectif qui permet souvent la simplification des dossiers et une meilleure reconnaissance du handicap.

IV. Perspectives

Le PFMG 2025, porté par l'Etat, vise à intégrer la médecine génomique dans le système de soins en permettant la généralisation de l'accès au génome. Ainsi, l'exome acteur incontournable dans la stratégie diagnostique des DI pourrait être supplanté par le génome dans les années à venir. Alors que les exons ne représentent qu'environ 1.5% du génome, le séquençage du génome entier permet d'étudier des régions jusque-là inexplorées avec la mise en évidence de nouveaux mécanismes moléculaires et la détection des variants structuraux. Si sur le plan technique tout est mis en place avec des capacités de séquençage en augmentation exponentielle, le challenge de demain sera dans le stockage, et l'interprétation des données. En effet, nous avons abordé la difficulté d'interprétation de l'exome qui génère environ 50 000 variants. Le génome génère quant à lui entre 3 et 4 millions de variants. Dans cette masse d'informations, identifier la variation responsable du phénotype représente un véritable défi. De plus, en séquençant la totalité du génome, nous augmentons la probabilité de générer des données non sollicitées, sans lien avec le phénotype du patient (données secondaires ou incidentales). Ces résultats peuvent être difficiles à gérer pour le clinicien et peuvent être anxiogènes pour les familles. Les données incidentes font l'objet de nombreux débats éthiques sur la divulgation de ces informations alors que leur impact sur la santé des patients n'est pas suffisamment établi.

Les techniques de séquençage actuelles sont basées sur des fragments courts rendant difficile la détection de variants de structure de taille moyenne ou grande. Des technologies émergentes basées sur l'analyse d'une seule molécule d'ADN de grande taille, permettent de détecter des variants de structure jusqu'ici inaccessibles (Oxford Nanopore, PacBio, Bionano genomics). Cependant, ces technologies présentent des inconvénients propres à chacune et demandent à être améliorées (taux d'erreur important, débit insuffisant, coûts conséquents).

Si le SHD représente une révolution biotechnologique et est un outil précieux dans l'identification des causes de la DI, il ne représente qu'une des nombreuses techniques consistant à appréhender la complexité du vivant. En effet, les approches « omiques » définies comme des techniques à haut débit permettant une analyse simultanée d'un grand nombre de variable comprennent entre autres : la génomique, l'épigénomique, la transcriptomique, la protéomique, la métabolomique... Par ailleurs, au-delà de la représentation linéaire des molécules d'ADN, de nouvelles technologies visent à établir les interactions entre fragments

d'ADN et leur implication en pathologie humaine (capture de conformation des chromosomes ou génome 3D).

Les technologies sont en constante évolution et se développent très rapidement. Cependant, le chemin est encore long avant d'élucider la complexité du fonctionnement du vivant. La génétique médicale est une discipline fascinante, en plein essor, et est promise à de nombreuses révolutions futures.

Bibliographie

1. APA (American Psychiatric Association). Diagnostic and statistical manual of mental disorders. fifth ed. Washington, DC: APA; 2013.
2. Inserm. Déficiences intellectuelles. Collection Expertise collective. Montrouge : EDP Sciences, 2016, XV-1157 p. - <http://hdl.handle.net/10608/6816>.
3. Raymond FL, Tarpey P. The genetics of mental retardation. *Hum Mol Genet.* 15 oct 2006;15 Spec No 2:R110-116.
4. Ellison JW, Rosenfeld JA, Shaffer LG. Genetic basis of intellectual disability. *Annu Rev Med.* 2013;64:441-50.
5. Lejeune J, Gautier M, Turpin R. [Study of somatic chromosomes from 9 mongoloid children]. *Comptes Rendus Hebd Seances Acad Sci.* 16 mars 1959;248(11):1721-2.
6. Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev.* 2002;8(3):117-34.
7. Ropers H-H, Hamel BCJ. X-linked mental retardation. *Nat Rev Genet.* janv 2005;6(1):46-57.
8. Martin JP, Bell J. A PEDIGREE OF MENTAL DEFECT SHOWING SEX-LINKAGE. *J Neurol Psychiatry.* juill 1943;6(3-4):154-7.
9. Lubs HA. A marker X chromosome. *Am J Hum Genet.* mai 1969;21(3):231-44.
10. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, 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.* 31 mai 1991;65(5):905-14.
11. Ciaccio C, Fontana L, Milani D, Tabano S, Miozzo M, Esposito S. Fragile X syndrome: a review of clinical and molecular diagnoses. *Ital J Pediatr.* 19 avr 2017;43(1):39.
12. Snijders Blok L, Madsen E, Juusola J, Gilissen C, Baralle D, Reijnders MRF, et al. Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. *Am J Hum Genet.* 6 août 2015;97(2):343-52.
13. Kellaris G, Khan K, Baig SM, Tsai I-C, Zamora FM, Ruggieri P, et al. A hypomorphic inherited pathogenic variant in DDX3X causes male intellectual disability with additional neurodevelopmental and neurodegenerative features. *Hum Genomics.* 01 2018;12(1):11.
14. Tjio JH, Levan A. The Chromosome Number of Man. *Hereditas.* 1956;42(1-2):1-6.

15. Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature*. 31 janv 1959;183(4657):302–3.
16. Ford CE, Jones KW, Polani PE, De Almeida JC, Briggs JH. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet Lond Engl*. 4 avr 1959;1(7075):711–3.
17. Patau K, Smith DW, Therman E, Inhorn SL, Wagner HP. Multiple congenital anomaly caused by an extra autosome. *Lancet Lond Engl*. 9 avr 1960;1(7128):790–3.
18. Edwards JH, Harnden DG, Cameron AH, Crosse VM, Wolff OH. A new trisomic syndrome. *Lancet Lond Engl*. 9 avr 1960;1(7128):787–90.
19. Pinkel D, Straume T, Gray JW. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc Natl Acad Sci U S A*. mai 1986;83(9):2934–8.
20. Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J, et al. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc Natl Acad Sci U S A*. déc 1988;85(23):9138–42.
21. Cremer T, Landegent J, Brückner A, Scholl HP, Schardin M, Hager HD, et al. Detection of chromosome aberrations in the human interphase nucleus by visualization of specific target DNAs with radioactive and non-radioactive in situ hybridization techniques: diagnosis of trisomy 18 with probe L1.84. *Hum Genet*. déc 1986;74(4):346–52.
22. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. déc 1977;74(12):5463–7.
23. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 15 févr 2001;409(6822):860–921.
24. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 16 févr 2001;291(5507):1304–51.
25. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 15 sept 2005;437(7057):376–80.
26. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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 Off J Am Coll Med Genet*. mai 2015;17(5):405–24.
27. Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet*. 02 2016;98(6):1067–76.

28. Alber M, Kalscheuer VM, Marco E, Sherr E, Lesca G, Till M, et al. ARHGEF9 disease: Phenotype clarification and genotype-phenotype correlation. *Neurol Genet.* juin 2017;3(3):e148.
29. Chiou T-T, Long P, Schumann-Gillett A, Kanamarlapudi V, Haas SA, Harvey K, et al. Mutation p.R356Q in the Collybistin Phosphoinositide Binding Site Is Associated With Mild Intellectual Disability. *Front Mol Neurosci.* 2019;12:60.
30. Harvey K, Duguid IC, Alldred MJ, Beatty SE, Ward H, Keep NH, et al. The GDP-GTP exchange factor collybistin: an essential determinant of neuronal gephyrin clustering. *J Neurosci Off J Soc Neurosci.* 23 juin 2004;24(25):5816–26.
31. Shimojima K, Sugawara M, Shichiji M, Mukaida S, Takayama R, Imai K, et al. Loss-of-function mutation of collybistin is responsible for X-linked mental retardation associated with epilepsy. *J Hum Genet.* août 2011;56(8):561–5.
32. Marco EJ, Abidi FE, Bristow J, Dean WB, Cotter P, Jeremy RJ, et al. ARHGEF9 disruption in a female patient is associated with X linked mental retardation and sensory hyperarousal. *BMJ Case Rep.* 2009;2009.
33. Kalscheuer VM, Musante L, Fang C, Hoffmann K, Fuchs C, Carta E, et al. A balanced chromosomal translocation disrupting ARHGEF9 is associated with epilepsy, anxiety, aggression, and mental retardation. *Hum Mutat.* janv 2009;30(1):61–8.
34. Lesca G, Till M, Labalme A, Vallee D, Hugononq C, Philip N, et al. De novo Xq11.11 microdeletion including ARHGEF9 in a boy with mental retardation, epilepsy, macrosomia, and dysmorphic features. *Am J Med Genet A.* juill 2011;155A(7):1706–11.
35. Lemke JR, Riesch E, Scheurenbrand T, Schubach M, Wilhelm C, Steiner I, et al. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. *Epilepsia.* août 2012;53(8):1387–98.
36. de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med.* 15 nov 2012;367(20):1921–9.
37. Holman SK, Morgan T, Baujat G, Cormier-Daire V, Cho T-J, Lees M, et al. Osteopathia striata congenita with cranial sclerosis and intellectual disability due to contiguous gene deletions involving the WTX locus. *Clin Genet.* mars 2013;83(3):251–6.
38. Long P, May MM, James VM, Grannò S, Johnson JP, Tarpey P, et al. Missense Mutation R338W in ARHGEF9 in a Family with X-linked Intellectual Disability with Variable Macrocephaly and Macro-Orchidism. *Front Mol Neurosci.* 2015;8:83.
39. Klein KM, Pendziwiat M, Eilam A, Gilad R, Blatt I, Rosenow F, et al. The phenotypic spectrum of ARHGEF9 includes intellectual disability, focal epilepsy and febrile seizures. *J Neurol.* juill 2017;264(7):1421–5.

40. Wang J-Y, Zhou P, Wang J, Tang B, Su T, Liu X-R, et al. ARHGEF9 mutations in epileptic encephalopathy/intellectual disability: toward understanding the mechanism underlying phenotypic variation. *Neurogenetics*. janv 2018;19(1):9–16.
41. Machado COF, Griesi-Oliveira K, Rosenberg C, Kok F, Martins S, Passos-Bueno MR, et al. Collybistin binds and inhibits mTORC1 signaling: a potential novel mechanism contributing to intellectual disability and autism. *Eur J Hum Genet EJHG*. janv 2016;24(1):59–65.
42. Bhat G, LaGrave D, Millson A, Herriges J, Lamb AN, Matalon R. Xq11.1-11.2 deletion involving ARHGEF9 in a girl with autism spectrum disorder. *Eur J Med Genet*. sept 2016;59(9):470–3.
43. Aarabi M, Kessler E, Madan-Khetarpal S, Surti U, Bellissimo D, Rajkovic A, et al. Autism spectrum disorder in females with ARHGEF9 alterations and a random pattern of X chromosome inactivation. *Eur J Med Genet*. 23 juill 2018;
44. Shen H, Li J, Zhang J, Xu C, Jiang Y, Wu Z, et al. Comprehensive characterization of human genome variation by high coverage whole-genome sequencing of forty four Caucasians. *PloS One*. 2013;8(4):e59494.
45. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 1 juill 2003;31(13):3812–4.
46. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. janv 2013;Chapter 7:Unit7.20.
47. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 08 2019;47(D1):D886–94.
48. Mueller WF, Larsen LSZ, Garibaldi A, Hatfield GW, Hertel KJ. The Silent Sway of Splicing by Synonymous Substitutions. *J Biol Chem*. 13 nov 2015;290(46):27700–11.
49. Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res*. 1 mars 2001;29(5):1185–90.
50. Desmet F-O, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. mai 2009;37(9):e67.
51. Leman R, Gaildrat P, Le Gac G, Ka C, Fichou Y, Audrezet M-P, 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*. 06 2018;46(15):7913–23.
52. Signal B, Gloss BS, Dinger ME, Mercer TR. Machine learning annotation of human branchpoints. *Bioinforma Oxf Engl*. 15 2018;34(6):920–7.

53. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 1 juill 2003;31(13):3568–71.
54. Raphaël Leman. Développement d'outils biostatistiques et bioinformatiques de prédiction et d'analyse des défauts de l'épissage : application aux gènes de prédisposition aux cancers du sein et de l'ovaire. Médecine humaine et pathologie. Normandie Université, 2019. Français. NNT : 2019NORMC418. tel-02454489.
55. Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Reynolds A, et al. Exonic transcription factor binding directs codon choice and affects protein evolution. *Science.* 13 déc 2013;342(6164):1367–72.
56. Gelfman S, Cohen N, Yearim A, Ast G. DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res.* mai 2013;23(5):789–99.
57. Komar AA. The Yin and Yang of codon usage. *Hum Mol Genet.* 01 2016;25(R2):R77–85.
58. Sørensen MA, Kurland CG, Pedersen S. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol.* 20 mai 1989;207(2):365–77.
59. Zhou T, Weems M, Wilke CO. Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol Biol Evol.* juill 2009;26(7):1571–80.
60. Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol.* févr 2013;20(2):237–43.
61. Duan J, Shi J, Ge X, Dölken L, Moy W, He D, et al. Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. *Sci Rep.* 2013;3:1318.
62. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet.* janv 2011;12(1):32–42.
63. Zeng Z, Bromberg Y. Predicting Functional Effects of Synonymous Variants: A Systematic Review and Perspectives. *Front Genet.* 2019;10:914.
64. Neri G, Schwartz CE, Lubs HA, Stevenson RE. X-linked intellectual disability update 2017. *Am J Med Genet A.* 2018;176(6):1375–88.
65. Jarvik GP, Browning BL. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *Am J Hum Genet.* 02 2016;98(6):1077–81.

NOM : GHESH

PRENOM : Leïla

Titre de Thèse : Des variants "perte de fonction" dans *ARHGEF9* sont responsables d'une déficience intellectuelle dominante liée à l'X

RESUME

Le gène *ARHGEF9* a été impliqué chez des patients atteints de déficience intellectuelle (DI) liée à l'X. La plupart des individus rapportés sont des garçons porteurs majoritairement de variations faux-sens hérités de mères dites asymptomatiques. Une dizaine de filles ont été décrites, toutes porteuses de remaniements chromosomiques de novo avec environ la moitié d'entre elles qui présentaient un biais d'inactivation de du chromosome X. L'hypothèse avancée était alors celle d'un mode de transmission récessif lié à l'X.

Nous rapportons trois nouveaux variants « perte de fonction », non hérités, dont un variant synonyme affectant l'épissage, chez deux filles et un garçon atteints de DI. Nous suggérons donc que les variants faux-sens sont responsables d'une DI récessive liée à l'X, tandis que les variants « perte de fonction » sont responsables d'une DI dominante liée à l'X affectant les garçons et les filles avec ou sans biais d'inactivation de l'X.

MOTS-CLES

ARHGEF9, déficience intellectuelle liée à l'X, filles, variant synonyme, variant « perte de fonction »