THÈSE

Présentée par: **Claire REDIN**

Pour l'obtention du grade de: **Docteure de l'Université de Strasbourg**

Discipline: Sciences du Vivant

Spécialité: Aspects Moléculaires et Cellulaires de la Biologie

---

NGS-based approaches for the diagnosis of intellectual disability and other genetically heterogeneous developmental disorders

Bardet-Biedl Syndrome & related ciliopathies, leukodystrophies

---

THÈSE dirigée par:

Mr MANDEL Jean-Louis Professeur, Collège de France, Paris

co-encadrée par :

Mr MULLER Jean Maitre de conférences, Praticien Hospitalier, Strasbourg

Soutenue publiquement le 2 Mai 2014

MEMBRES DU JURY:

RAPPORTEURS:

Mr CHELLY Jamel Professeur, Praticien Hospitalier, Institut Cochin, Paris

Mr ROPERS Hans-Hilger Professeur, Max-Planck-Institute for Molecular Genetics, Berlin

EXAMINATEURS:

Mr VERLOES Alain Professeur, Praticien Hospitalier, Robert Debré, Paris

Mme PUJOL Aurora Chargée de recherches, Praticien Hospitalier, IDIBELL, Barcelona

« La véritable science enseigne, par-dessus tout, à douter et à être ignorant.»

Miguel de Unamuno

« La science consiste à passer d'un étonnement à un autre.»

Aristote

Remerciements

Je tiens tout d'abord à remercier vivement mes rapporteurs, les Professeurs Jamel Chelly et Hans-Hilger Ropers pour avoir accepté de lire et juger mes travaux de thèse, contenus dans un manuscrit qui n'est - au final - pas aussi court que je ne l'aurais voulu… A Jamel particulièrement, un grand merci pour ton soutien à toute épreuve, notamment pour mon après-thèse. A Hans-Hilger, c'est un véritable honneur que vous ayez accepté de faire partie de ce jury, honneur auquel je n'aurais jamais songé lors de notre première rencontre il y a quatre ans à la conférence MR-NET d'Erlangen…

Je remercie également mes examinateurs, les Professeurs Alain Verloes et Aurora Pujol pour avoir également accepté de juger mes travaux. Vous n'avez pas la lourde tâche de lire en détail l'intégralité de ce dense manuscrit mais j'espère néanmoins que vous y trouverez des résultats intéressants, Alain sur les ciliopathies, Aurora sur les leucodystrophies. A Alain, c'est un très grand honneur pour moi que vous ayez consenti à cette tâche, vous qui faisiez partie des personnes pressenties pour la succession à la présidence de l'ASHG (j'avais bien sûr voté en votre faveur…). Aurora, c'était un véritable plaisir de collaborer avec toi au cours de cette dernière année sur le projet leucodystrophies, et j'espère vivement que nous serons amenées à collaborer à nouveau dans le futur (suite du projet leucos à Boston ? ;)).

Je tiens ensuite à remercier mes deux co-directeurs de thèse (ou co-encadrants… ;)), le Professeur Jean-Louis Mandel et le Docteur Jean Muller, pour ces quelques quatre années passées à vos côtés qui ont été un véritable plaisir. A Jean-Lou, que je ne remercierai jamais assez, pour tout ce que tu m'as apporté sur le plan scientifique, puits sans fond de connaissance, en génétique certes mais pas que ! Et pour tout le reste, parce que c'est tellement agréable de partager autre chose que de la science, et je garderai longtemps en tête ces moments de partage musicaux et artistiques. A Jean, je tiens à te remercier, juste une fois…. ;) Pour tout ce que tu m'as appris en bioinfo, pour tous ces bons moments passés ensemble, et pour toutes ces discussions animées de par nos deux caractères finalement assez trempés…

Au reste de notre (petite) équipe 'Mandel': Amélie, Angélique, Nicolas, et nos interims (Grace, Inès, Sébastien, …). C'était un véritable bonheur de travailler avec vous, j'espère que mon équipe d'accueil sera aussi chouette ! Amélie, tu m'as tellement appris en génétique, j'espère un jour avoir ton lot de connaissances… C'était un réel plaisir de travailler à tes côtés sur le projet DI. Petite mention spéciale aussi sur 'notre' étude de réévaluation des gènes de XLID dont l'idée nous est venue - rappelle-toi - une fin d'après-midi en terrasse, autour d'un verre de vin (alsacien bien sûr). A Angélique, qui m'a rejoint en milieu de parcours, sur un projet tellement difficile… Je te

souhaite beaucoup de réussite dans ces prochaines années de thèse, je suis certaine que tout va se débloquer connaissant ta rigueur et ta persévérance (avec le ChIP-seq ? ;)). Merci pour ta bonne humeur quotidienne et pour tous ces bons moments, merci d'avoir joué le rôle de maître de stage avec moi sur beaucoup de choses, et surtout merci de nous avoir entretenu gourmettement ces deux dernières années (et merci à Laurent, bien sûr !), nos pantalons s'en souviendront longtemps… :)

A notre frère adoptif de l'équipe Mandel, Ricardos, je te remercie pour ta gentillesse et ton inébranlable bonne humeur, et pour tous ces moments de fous rires passés ensemble notamment lors de notre première vraie rencontre, à Erlangen en 2010. J'espère que depuis, tu as pu retourner en Allemagne avec ta Modus ? ;) A Eric aussi bien sûr, notre second membre adoptif de l'équipe Mandel, qui s'occupe si bien de notre soutien psychologique en nous ramenant des chocolats à Pâques, des glaces l'été, et des fruits à toutes saisons.

Je tiens ensuite à remercier tous les autres membres du département de Neurogénétique et Médecine Translationnelle de l'IGBMC et du laboratoire de Bioinformatique et Génomique Intégrative, pour nos discussions, scientifiques et autres, et pour cette générale bonne entente qui règne au sein de ce département et des équipes. A Olga bien sûr car depuis ton départ les journées ne sont plus les mêmes, à Karim sans qui nos journées dénuées de quelques phrases sexistes ne seraient pas pareilles ;), à Brahim pour ta contribution à nos dodécalitres de café matinaux quotidiens, à Fabrice et Hubert pour nous obliger à aller prendre le soleil le midi de temps en temps, et à beaucoup d'autres que j'oublie.

Je tiens également à remercier tous les membres de la plateforme de séquençage de l'institut avec lesquels j'ai collaboré tout au long de ces quatre années de thèse, et qui est un peu comme ma famille d'accueil. A Bernard, le 'chef', pour ta patience à toute épreuve, nos discussions scientifiques mais aussi sportives, et surtout ton incroyable bonne humeur. A Muriel, Serge, et Claire la petite dernière, j'ai adoré travaillé avec vous… Parfois, dans les moments difficiles, j'avais juste envie de descendre et faire des manips à vos côtés tant votre bonne humeur est contagieuse. A Stéphanie et Michaël, nos bioinformaticiens hors-pairs, que nous avons tellement mis à contribution pour nos différents projets… Stéphanie, pour ta bonne humeur, et pour toutes ces heures passées avec toi à essayer de comprendre les données de séquençage de certains patients avec réarrangements complexes! Merci pour ton aide et pour tout ton travail, sans qui cette thèse n'aurait probablement pas pu être totalement ce qu'elle est.

Un grand merci à Mims, d'une efficacité et d'une patience exemplaires pour toutes nos requêtes quelles qu'elles soient au niveau administratif. Et merci pour ton invariable bonne humeur ! Un énorme merci également à Maïté, pour être tellement disponible dès que l'on a besoin d'un article introuvable ou d'une impression infaisable. Merci de nous faciliter à tous la tâche au quotidien.

Je tiens ensuite à remercier tous mes amis, car '*l'homme est un animal social*' d'après Aristote, mais je crois surtout moi plus que tout autre.

Tout d'abord au sein de l'institut, avec la team des 'na', Ana, Dana, Helena, Heena, Lena, Serena sans qui ma première année de thèse à Strasbourg aurait été beaucoup plus difficile ; au groupe des 'russes' (même s'ils ne le sont pas tous) qui m'est devenu essentiel ces dernières années: Anna, Iskander, Irina, Karima, Katarzyna, Vladimir pour tous ces bons moments passés avec vous, notamment un certain week-end de Juillet qui restera longtemps dans mes souvenirs ; à mes compagnons d'OpenLab Morgan, Morgane, Patrick, Léa et Anaïs pour toutes ces journées passées en dehors de Strasbourg à initier les lycéens aux techniques de laboratoire, j'ai la tête pleine de souvenirs avec vous; à Jérôme, pour nos discussions interminables et enjouées détendues pendant les pauses cafés ; et enfin le meilleur pour la fin : à Amélie et Vincent, pour avoir partagé tant de soirées avec vous, tant de discussions aussi, merci d'avoir été là. Vous êtes certainement les Strasbourgeois qui vont le plus me manquer outre Atlantique (trouverons-nous de nouveaux partenaires de jeux aussi assidus ? ;)).

Ensuite, aux autre amis de plus longue date, qui m'ont permis de me changer les idées en dehors de Strasbourg: à Alex, Amandine, Angélique, Aurélie, Clément, Fred, Kro, Laurianne, Perrine et Rémi, parce que vous comptez tous énormément pour moi.

A Nicolas, qui m'accompagne depuis maintenant quelques années et j'espère pour longtemps encore. Pour ton soutien incomparable et le bonheur au quotidien.

Enfin pour terminer, je tiens à remercier ma famille, également pour leur soutien et tous ces bons moments passés et à venir. A mes parents, mes sœurs Sarah et Mathilde, mes grands parents et à tous les autres, parce que j'ai une famille formidable.

Résumé de la thèse

L'intégralité du patrimoine génétique de chaque individu – ou génome – est stockée dans le noyau de chacune des cellules somatiques au sein de l'acide désoxyribonucléique (ADN). Une infime partie de l'ADN (seulement 1,5%) code pour des gènes qui seront alors transcrits en ARN messager (ARNm) puis décodés et traduits en protéines, responsables du phénotype 'moléculaire' de la cellule. Ce dernier conditionne le phénotype cellulaire à l'échelle de l'organe, qui à son tour déterminera le phénotype macroscopique d'un individu. La séquence de cet ADN est extrêmement conservée d'un individu à l'autre (à 99,9%) car soumise à une forte pression de sélection, mais rapportée à la taille conséquente du génome humain cela représente plusieurs millions de variations inter-individus d'impact variable sur le phénotype macroscopique unique de chacun.

La plupart de ces variations inter-individus est retrouvée dans l'ADN non codant (introns, microARNs, lincRNAs…), ayant des conséquences moindres (*a minima* indirectes) au niveau protéique, ou tout du moins difficilement interprétables à ce jour. Celles se trouvant dans les séquences codantes peuvent avoir divers effets plus ou moins dramatiques sur la traduction d'un gène en protéine, et le maintien de ses fonctions: production d'une protéine identique (changement silencieux), modification d'un ou plusieurs acides-aminés (faux-sens, insertions/délétions conservant le cadre de lecture), production d'une protéine tronquée (non-sens, variations d'épissage, petites insertions/délétions décalant le cadre de lecture, larges délétions/duplications/réarrangements d'un fragment d'ADN), ou bien absence de production de protéine (dégradation de l'ARNm portant un codon stop prématuré). Certaines de ces affections sont tolérées par l'organisme et sont donc d'ordre bénin pour l'individu (ex: expression de certains caractères phénotypiques, daltonisme…), d'autres entrainent un dysfonctionnement de la cellule et sont à l'origine de maladies génétiques plus ou moins sévères chez les individus.

Les maladies héréditaires mendéliennes (ou monogéniques) sont causées par une mutation (héritée ou apparue *de novo* dans une lignée germinale parentale) localisée dans un seul gène. A l'inverse, les maladies polygéniques et multifactorielles sont provoquées par la présence chez un même individu de mutations agissant en synergie dispersées dans plusieurs gènes, et par la contribution de certains facteurs environnementaux. On distingue les maladies monogéniques à transmission dominante (un seul des allèles parentaux porte une mutation), de celles à transmission récessive (les deux allèles du même gène portent une mutation). De même, pour les gènes situés sur les chromosomes sexuels, on distingue les maladies monogéniques à transmission dominante liée à l'X (presqu'autant de femmes atteintes que d'hommes), de celles à transmission récessive liée à l'X (plus d'hommes atteints que de femmes, car ils n'ont qu'un seul allèle pour les gènes du X). Enfin, il existe un dernier mode de transmission également appelé transmission maternelle, où la mutation contenue dans l'ADN mitochondrial est transmise à l'embryon exclusivement par l'ovocyte

maternel. Pour les maladies monogéniques dites 'simples', chez les individus exprimant un phénotype macroscopique particulier, les mutations causales seront toujours retrouvées dans un seul et même gène (ex : drépanocytose, mutations retrouvées exclusivement dans le gène *HBB*). Pour les maladies monogéniques génétiquement hétérogènes, chez des individus présentant un même phénotype clinique, les mutations causales pourront être retrouvées dans un gène parmi un ensemble (plus ou moins grand) de gènes décrits comme impliqués dans la maladie (ex: dystrophie myotonique de type 2, mutations retrouvées soit dans le gène *DMPK*, soit dans le gène *ZNF9* selon les patients).

Problématique

Cette hétérogénéité au niveau moléculaire rend la recherche exhaustive de mutations souvent très compliquée et la plupart du temps impossible en raison des coûts élevés des tests génétiques et du temps important de leur réalisation avec les stratégies dites 'classiques' (ex: séquençage Sanger). Ainsi, une très grande proportion de patients atteints de maladies génétiquement hétérogènes demeurent ainsi non diagnostiqués sur le plan moléculaire. Or l'obtention d'un diagnostic étiologique précis et précoce est primordial pour les familles afin d'être en mesure d'anticiper et de proposer une prise en charge et un traitement adaptés notamment chez le jeune enfant, et à plus long terme de développer des approches thérapeutiques personnalisées (thérapies gène/mutation spécifiques ?).

Mon projet principal de thèse a consisté donc - au travers de pathologies neurologiques de complexité croissante (nombre grandissant de gènes impliqués) - à établir et valider de nouvelles méthodes à visée diagnostique par les technologies de séquençage à haut-débit ciblé, afin de répondre à ce besoin.

Partie I – Mise au point du séquençage ciblé à haut débit sur une pathologie 'peu' complexe : le syndrome de Bardet-Biedl

Dans un premier temps, cette approche a été initiée et validée sur une pathologie 'peu complexe' (*a minima* moins hétérogène que la déficience intellectuelle): le Syndrome de Bardet-Biedl (BBS). BBS est une maladie monogénique récessive rare (estimée à 1/150'000) et relativement hétérogène aux niveaux clinique et génétique. En effet, à ce jour les mutations responsables de ce syndrome ont été retrouvées distribuées dans 19 gènes différents (*BBS1-BBS19*). Le phénotype clinique d'expression variable selon les individus se traduit par une rétinopathie pigmentaire, une malformation des extrémités (polydactylie), une obésité (souvent morbide), des anomalies rénales (kystes) et génitales (hypogonadisme), et une déficience cognitive légère à modérée. Des études fonctionnelles sur les protéines BBS et leurs mutations pathogènes ont démontré leur implication centrale dans le maintien structural et fonctionnel des cils primaires, et

leur rôle dans le transport intraflagellaire confirmant l'appartenance de ce syndrome à la famille des ciliopathies. Plusieurs autres affections du même ordre appartenant à cette même famille engendrent un tableau clinique similaire chez les patients (syndrome d'Alström, de Joubert, Néphronopthise,…), compliquant considérablement la mise en place d'un diagnostic différentiel précis. En 2010, plus de 60% des patients avec ce syndrome restait sans diagnostic.

Pour cette étude, j'ai tout d'abord sélectionné 30 gènes d'intérêt, impliqués dans BBS ou d'autres syndromes avec phénotypes chevauchants. Puis, j'ai mis en place un protocole permettant de capturer spécifiquement les exons codants de ces gènes, et de les séquencer en parallèle chez un même patient par séquençage haut-débit. Cette approche permet également de séquencer de nombreux patients simultanément (jusqu'à 96 sur un même séquenceur Illumina, GAIIx).

Cette étude préliminaire sur 52 patients a permis d'identifier la mutation causale dans près de **70%** d'entre eux (*Redin C. et al*., *J. Med. Genet*., *2012*). De manière surprenante, nous avons identifié chez 12% des patients de larges délétions d'un ou plusieurs exons (à l'état hétérozygote ou homozygote) dans un gène BBS, à l'origine du phénotype observé. De tels évènements à l'état hétérozygote n'étaient auparavant pas détectables par les méthodes de séquençage Sanger classique, et étaient donc largement sous-estimés.

Parmi les gènes ciblés figure le gène *ALMS1* impliqué dans le syndrome d'Alström (ALMS), qui de par sa grande taille n'était auparavant testé dans aucun laboratoire de diagnostic français, car trop coûteux. Grâce à notre approche, les patients présentant un phénotype évocateur de ce syndrome (similaire au syndrome de Bardet-Biedl, mais associant également une surdité) ont à présent la possibilité d'obtenir un diagnostic moléculaire. En particulier, notre étude souligne que le diagnostic différentiel entre BBS et ALMS est difficile, de par la présence de quelques patients initialement diagnostiqués BBS mais chez qui une mutation a finalement été identifiée dans le gène *ALMS1*, et inversement. Ceci est d'une importance majeure, puisque le développement de cardiomyopathies a été largement reporté chez de nombreux patients atteints d'ALMS (Marshall *et al.*, 2007). Un diagnostic précoce chez ces patients est donc très important puisque permettant de surveiller et d'anticiper ces anomalies. Enfin, nous montrons également que pour les patients présentant un tableau clinique complet du syndrome de Bardet-Biedl le taux de diagnostic est supérieur à 81%, suggérant que parmi les 30% de patients 'négatifs' de notre cohorte, de nombreux ne sont peut-être pas de 'vrais' patients BBS.

Cette stratégie a depuis été transférée en routine au laboratoire de diagnostic de Strasbourg, où elle est proposée en test de routine depuis l'automne 2012 (plus de 170 patients supplémentaires ont ainsi été analysés).

Partie II – Implémentation de l'approche de séquençage ciblé à haut débit pour le diagnostic de la déficience intellectuelle

Cette démarche ayant démontré sa pertinence pour le diagnostic d'une pathologie moins complexes (moins de gènes impliqués, expliquant cependant une plus grande proportion de patients), nous l'avons ensuite transposée et adaptée pour le diagnostic de la déficience intellectuelle (DI).

La déficience intellectuelle (DI, anciennement retard mental) constitue un enjeu de santé publique majeur, de par son incidence (environ 2%) et sa complexité, et représente un handicap souvent très lourd pour l'individu concerné et sa famille. Une proportion importante de ces déficiences intellectuelles a une origine génétique.

Les anomalies chromosomiques ou variants structuraux (délétions, duplications, inversions, translocations, etc) sont responsables de la déficience intellectuelle généralement associée à des malformations congénitales dans 10 à 15% des patients. La première cause de DI en France reste ainsi le syndrome de Down (ou trisomie 21). De nombreux autres syndromes délétionnels ou duplicationnels sont associés à des formes de DI à pénétrance partielle (*i.e.* peuvent être retrouvés chez des patients asymptomatiques), tels le 15q11-q13 ou 16p11.2.

En raison du biais de sexe conséquent observé dans la déficience intellectuelle (1.3-1.4 hommes atteints pour 1 femme) et suite à l'identification dans les années 60 de grandes familles avec ségrégation clairement liée à l'X, la recherche de gènes responsables s'est principalement concentrée - et ce jusqu'à récemment - sur les gènes du chromosome X, et donc responsables de déficience intellectuelle liée à l'X (XLDI, anciennement XLMR). A ce jour, et grâce aux efforts de larges consortiums internationaux, une centaine de gènes de XLDI a été identifiée (Gecz *et al.*, 2009; Ropers, 2010; Lubs *et al.*, 2012).

La recherche de gènes autosomiques associés à la déficience intellectuelle récessive s'est développée de façon plus récente. La méthode d'identification la plus évidente étant la recherche de régions d'homozygotie dans des familles consanguines couplée à des analyses de liaison, et suivies du séquençage systématique des gènes candidats. Cette approche a permis l'identification de >100 gènes de déficience intellectuelle autosomique récessive (*PRSS12*, *CC2D1A*, *TUSC3*, *GRIK2, CACNA1G, KDM5A*…) (Kaufman *et al.*, 2010; Ropers, 2010; Kuss *et al.*, 2011; Najmabadi *et al.*, 2011; Musante and Ropers, 2014)

En raison de l'impact de la déficience intellectuelle sur l'aptitude à la reproduction, l'hypothèse d'un mode de transmission autosomal dominant familial avait longuement été délaissée, puisque supposé très rare pour des mutations pleinement pénétrantes. L'identification de syndromes associés à des microdélétions a alors prouvé que des mutations *de novo* pouvaient être une cause majeure dans les cas à ségrégation apparente "dominante". Ceci a depuis été largement confirmé par l'identification de nombreuses mutations ponctuelles tronquantes *de novo* reportées dans des cas

simplex de déficience intellectuelle (exomes des parents et du patient) (Vissers *et al.*, 2010; Hamdan *et al.*, 2011; de Ligt *et al.*, 2012; Rauch *et al.*, 2012)

Bien que de nombreux gènes impliqués dans des formes monogéniques de déficience intellectuelle aient à ce jour été identifiés, le diagnostic de la cause génétique reste difficile. Pour les formes clairement syndromiques (Coffin-Lowry, Alpha-thalassémie,…), le séquençage Sanger du gène correspondant est proposé, tandis que pour les formes non syndromiques l'offre se réduit au test de l'X fragile et à la détection de CNVs pathogènes par CGH-array. Une vaste majorité (70%) des patients reste ainsi sans diagnostic moléculaire. L'explosion des technologies de séquençage à haut débit de 'nouvelle-génération' permet de pallier à ce manque, et de développer des stratégies de séquençage systématique et parallèle de tous les gènes de déficience intellectuelle connus.

Nous avons donc sélectionné un grand nombre de gènes décrits comme impliqués dans des formes monogéniques de déficience intellectuelle: une majorité dans des formes non-syndromiques, mais aussi dans quelques formes syndromiques, à condition que la déficience intellectuelle soit une composante majeure et constante du syndrome. Nous avons essayé d'être le plus exhaustif possible pour les gènes de DI liée à l'X (XLDI). Un total de 217 gènes ont ainsi été sélectionnés (99 du chromosome X et 116 des autosomes, impliqués dans des formes dominantes ou récessives). Ce set de gènes a été séquencé à ce jour chez un total de 106 patients adressés par 17 centres nationaux, présentant un retard de développement cognitif sans cause génétique préalablement identifiée (négatifs pour le test de l'X-fragile et l'analyse par CGH).

La majorité des mutations certainement causales identifiées (héritées ou *de novo)* conduisent à des protéines tronquées, et affectent des gènes d'XLDI (16/26) Ceci s'explique principalement par l'enrichissement majeur de notre cohorte en hommes, et d'autre part par notre exhaustivité des gènes de DI situés sur le chromosome X. Nous avons toutefois identifié également quelques mutations tronquantes *de novo* dans des gènes associés à des formes autosomiques dominantes de DI (*RAI1*, *TCF4*, *DYRK1A*). De façon surprenante, ces 1ers résultats ne montrent aucune corrélation entre le degré de sévérité de la DI chez les patients et la détection d'une mutation causale: une proportion similaire de mutations est retrouvée chez les patients à DI légère par rapport à d'autres avec DI profonde. Cette approche nous a aussi permis de confirmer certains gènes pour lesquels l'implication dans la déficience intellectuelle n'avait jamais été répliquée (nouvelle mutation identifiée dans *MAOA* chez une famille avec un syndrome de Brunner, 20 ans après la 1ère étude (Brunner *et al.*, 1993); nouvelle mutation identifiée dans *NLGN3*, etc…). Nous avons également identifié des mutations dans des gènes syndromiques chez des patients avec un phénotype 'atypique', permettant ainsi de réviser et d'élargir les spectres phénotypiques associés à certains gènes (un jeune garçon sans phénotype musculaire avec une mutation tronquante dans *DMD*, un autre patient avec une mutation dans *TCF4* sans la dysmorphie faciale typique du syndrome de Pitt-Hopkins, etc…).

Enfin, l'identification chez certains jeunes patients de mutations dans les gènes tels *TCF4*, *SLC2A1* et *MECP2* peut avoir des conséquences importantes sur la prise en charge médicale et la qualité de vie future de ces patients. En effet, les individus atteints du syndrome de Pitt-Hopkins développent des épisodes d'hyperventilations nocturnes pouvant être fatales vers l'âge de 10-12 ans. Le patient avec une mutation tronquante retrouvée dans *TCF4* étant beaucoup plus jeune, ce diagnostic génétique permettra d'anticiper et de surveiller cette manifestation. Pour le second patient avec mutation dans *SLC2A1* (déficience en GLUT1), étant également assez jeune (7 ans), la mise en place d'un régime cétogène pourrait permettre d'améliorer (ou tout du moins de limiter l'aggravation de) sa condition. Pour le patient avec mutation dans *MECP2* (impliqué dans le syndrome de Rett), suite au report de nombreux cas similaires avec dérégulations cardio-respiratoires (en plus des difficultés neurodéveloppementales) pouvant provoquer une mort subite (Weese-Mayer *et al.*, 2006), cette condition pourra être étroitement contrôlée.

En parallèle de ce travail, nous avons conduit une ré-analyse exhaustive et systématique sur les mutations causales décrites dans les gènes d'XLDI. La présence de certaines de ces mutations à des fréquences conséquentes dans des populations 'contrôles' nous a conduit à émettre de sérieuses réserves quant à l'implication d'une douzaine de gènes publiés comme étant responsables - lorsque mutés - de déficience intellectuelle liée à l'X. Nous proposons ainsi une stratégie d'évaluation lors de l'identification d'une mutation candidate (*Piton* A., Redin* C., Mandel JL., AJHG, 2013*).

Partie III – Implémentation de l'approche de séquençage ciblé à haut débit pour le diagnostic des leucodystrophies

Au vu de l'efficacité de cette approche pour le diagnostic du Syndrome de Bardet-Biedl et autres ciliopathies associées et la déficience intellectuelle, nous l'avons testée pour le diagnostic des leucodystrophies, ou anomalies de la substance blanche.

Le terme leucodystrophies (LDs) regroupe un ensemble vaste et hétérogène de maladies monogéniques résultant d'anomalies du processus de myélinisation (formation et/ou maintien) dans le système nerveux central, provoqué par un dysfonctionnement des cellules gliales (notamment les oligodendrocytes). A ce jour, des mutations dans environ 50 gènes ont été reportées et identifiées comme à l'origine de différentes formes de LDs.

Dans certains cas, les données du patient (analyses IRM, traits phénotypiques, marqueurs biochimiques...) permettent d'orienter le clinicien vers un gène candidat, conduisant ainsi à un rendu diagnostic pour environ 35% d'entre eux. Cependant, des formes incomplètes ou atypiques de LDs peuvent rendre cette approche peu fructueuse. De plus, certaines formes cliniques démontrent une grande hétérogénéité génétique: pour des patients exhibant des manifestations cliniques très similaires, les mutations causales peuvent être retrouvées dans des gènes différents. Cette

hétérogénéité au niveau moléculaire rend la recherche exhaustive de mutations impossible en termes de coût et de temps avec les stratégies de séquençage dites 'classiques'.

Nous avons donc développé une approche de screening moléculaire par séquençage à haut débit ciblé, basé sur la capture et le séquençage des exons codants de 68 gènes impliqués dans des formes monogéniques de LDs et leucoencéphalopathies. Cette approche permet à la fois la détection de mutations ponctuelles, de petites insertions/délétions, mais également de délétions ou duplications d'exons ou d'un gène entier aussi bien à l'état homozygote/hémizygote qu'hétérozygote (comme démontré dans la partie précédente). Cette capacité de détection des CNVs est d'importance majeure, notamment dans la maladie de Pelizaeus-Merzbacher pour laquelle des duplications, triplications ou délétions entières de *PLP1* sont fréquemment retrouvées.

Une preuve de concept a été effectuée sur 15 patients avec mutations connues, permettant de retrouver 96% des allèles mutés. L'ADN de 120 autres patients avec LDs d'origine génétique non déterminée a été séquencé via cette même approche, permettant d'identifier la mutation causale chez 20 patients (dans les gènes *ALDH3A2, ABCD1, EIF2B2, EIF2B5, GFAP, POLR3A, POLR3B, POMGNT1, PLP1, RNASEH2B, RNASEH2C* et *SLC16A2*). Des mutations potentiellement causales nécessitant des analyses complémentaires ont été identifiées chez dix patients. Enfin, chez cinq autres patients, nous avons identifié à l'état hétérozygote une mutation pathogénique (tronquante ou d'épissage) dans des gènes associés à des formes récessives de LD (*ARSA*, *GBE1*, *PEX1, SAMHD1*) sans qu'une deuxième mutation n'ait pu être détectée. Cette observation surprenante nécessite des investigations complémentaires et la confrontation avec les données cliniques de chaque patient. Il parait en effet peu probable que la totalité de ces mutations manquantes se situent dans des régions non codantes et soient donc non détectées par notre stratégie de capture d'exons. Parmi les hypothèses restantes on peut évoquer l'oligogénisme, ou encore la présence d'un variant hypomorphe sur le 2ème allèle dont la fréquence relativement élevée l'aurait fait paraître non pathogénique.

Avec un taux de mutations détectées de 17 à 25%, la stratégie par capture d'exons apparait timidement efficace pour le diagnostic des leucodystrophies. Une analyse approfondie de la clinique des patients négatifs comme l'âge de début de la maladie, la sévérité, le type de leucodystrophie (hypomyélinisante, démyelinisante,…), le nombre d'investigations génétiques et métaboliques effectuées sur les patients avant l'inclusion, etc… permettra d'établir si pour un profil particulier de patients, le séquençage d'un panel de gènes en particulier semble plus approprié que pour d'autres.

Conclusions & Perspectives

Grâce aux travaux menés au cours de ma thèse, les proportions conséquentes de mutations identifiées chez les patients pour chacune des pathologies cibles confirment la nécessité de l'implémentation en routine de cette approche ciblée comme nouvel outil de diagnostic pour les

maladies génétiquement hétérogènes dans les laboratoires de diagnostic génétique en France. Son efficacité pour le diagnostic des leucodystrophies reste cependant insuffisante et nécessite très certainement une réévaluation des gènes cibles à inclure dans la capture, ce qui permettrait d'augmenter le taux de mutations détectées chez les patients 'tout-venants'.

Les résultats obtenus pour la déficience intellectuelle sont très prometteurs, avec 25% de résultats positifs (contre 16% pour l'X-fragile et la CGH-array cumulés). Le séquençage de 230 patients supplémentaires prévu prochainement permettra de mieux évaluer la contribution de chaque gène dans cette pathologie (d'autant plus important pour les gènes récemment identifiés), de réévaluer le spectre clinique associés à certains syndromes et enfin de préciser l'efficacité de cette méthode.

A l'échelle du patient, le rendu diagnostic permet dans certains cas de proposer des traitements pré ou post-symptomatiques, d'anticiper des symptômes d'apparition supposée tardive, et/ou tout simplement de poser un nom sur l'origine de la pathologie stoppant ainsi une errance diagnostique souvent compliquée pour le patient. Ces résultats permettent également d'envisager un conseil génétique pour la famille du patient en précisant le risque de récurrence ou non de la pathologie, de proposer un diagnostic préimplantatoire ou un diagnostic prénatal. Enfin, l'identification de l'anomalie moléculaire est un pré-requis obligatoire qui permettra au patient de bénéficier à l'avenir de la thérapie génique ciblée et/ou d'essais thérapeutiques.

Table of contents

List of Tables

List of Figures

List of appendices

List of abbreviations

ADHD: Attention Deficit and Hyperactivity Disorders
ALMS = Alström syndrome
ARID: Autosomal recessive Intellectual Disability
ASDs : Autism Spectrum Disorders
ASS: Acceptor Splice Site
BBS: Bardet-Biedl Syndrome
cDNA: complementary DNA
CGH: Comparative Genomics Hybridization
CILD = primary CILliary Dyskinesia
CNS: Central Nervous System
CNVs: Copy Number Variants
ddNTPs: dideoxyribonucleotides
DNA: DeoxyriboNucleic Acid
dNTPs: deoxyribonucleotides
DSM-V: Diagnostic and Statistical Manual of Mental Disorders
DSS: Donor Splice Site
ENCODE project: ENCyclopedia Of DNA Elements
EVS: Exome Variant Server
FISH: Fluorescence In Situ Hybridization
FPR: False-Positive Rate
Gb: Giga base pairs (=1x10^9 bp)
gDNA: genomic DNA
GOLD: Genome OnLine Database
GWAS: GenomeWide Association Studies
HGMD: Human Gene Mutation Database
HGVS: Human Genome Variation Society
Hh = Hedgehog
hr: hour
HSF: HumanSplicingFinder
HTS: High Throughput Sequencing
IBD: Identical By Descent
ICD-10 : International Classification of Diseases, version 2010
ID: Intellectual Disability
IEMs: Inborn Errors of Metabolism
IFT = Intraflagellar transport
IGV: Integrative Genome Viewer
indel: insertion/deletion
IQ: Intelligence quotient
JBTS = Joubert syndrome
kb: kilobase pairs (=1x10^3 bp)
LCA = Leber Congenital Amaurosis
LINEs: Long Interspersed Nuclear Elements
LOD: logarithm of odds
Mb: Megabase pairs (=1x10^6 bp)
MES: MaxEntScan
MIP: Molecular Inversion Probes
MKKS = McKusick Kaufmann syndrome
MKS = Meckel syndrome
MRI: MagneticResonance Imaging
mRNA: messenger RNA
mtDNA: mitochondrial DNA
MTOC = Microtubule organizing center
NGS: Next-Generation Sequencing
NPHP = Nephronophthisis
nt: nucleotide
PCP: Planar Cell Polarity
PCR: Polymerase Chain reaction
PGM: Personal Genome Machine
PKD = Polycystic Kidney Disease
PKU: Phenylketonuria
PNS: Peripheral Nervous System
qPCR: quantitative PCR
RNA: RiboNucleic Acid

RT-PCR: Reverse Transcriptase PCR
Shh = Sonic Hedgehog
SINEs: Short Interspersed Nuclear Elements
SLNS = Senior Loken syndrome
SNP: Single Nucleotide Polymorphism
SNV: Single Nucleotide Variant
SVs: Structural Variants
US $: American dollar
UTR: UnTranslated Region
WES: Whole-Exome Sequencing
WGS: Whole-Genome Sequencing
XLID: X-linked Intellectual Disability

General Introduction

I- From DNA alteration to disease: the basis of genetics

1. The Human genome

The entire genetic heritage of each individual – otherwise referred as human genome – is stored mostly in the nucleus of each somatic cell within chemical molecules of deoxyribonucleic acid (gDNA, for genomic DNA). A tiny fraction is contained in the mitochondria (mtDNA, for mitochondrial DNA). These molecules of DNA store all the genetic information necessary to control all the steps of embryogenesis, development, homeostasis, metabolism, etc. Only a small proportion of genomic DNA (about 1.5%) encodes genes, which correspond to the units of genetic information. The human genome is assessed to contain about 21,000 genes (for a comparison, the mitochondrial chromosome contains only 37 genes; reviewed in (Frazer, 2012), linearly positioned along the chromosomes at precise loci. Genetic disorders are the consequences of alterations encoded in this molecule of DNA, and more precisely in its sequence. Our understanding of genetic disorders has tremendously improved with the completion of the Human Genome Project, reporting the entire sequence of a human genome in 2001 (Lander *et al.*, 2001; Venter *et al.*, 2001; International Human Genome Sequencing, 2004). Knowledge of the entire sequence provided researchers with a physical map to annotate genes and other biological elements onto a single reference. This allowed studying the human genome as a continuous entity rather than discretely, gene by gene. The draft of the human genome is continuously improved over the years thanks to patches and through novel releases (*e.g.* March, 2014: GRCh38 release, referred as hg38).

2. Genome organization and structure

All somatic cells of the body contain 23 pairs of chromosomes (22 autosomal, 1 sexual), one from each pair being inherited from one of the two parents. Germ cells only carry 23 chromosomes. Chromosomes are numbered according to their size that often but does not necessarily correlate to their gene contents. Genes are not distributed randomly along the chromosomes, but tend to cluster together. As a result, certain regions of chromosomes are highly gene-rich while other regions are called gene deserts, and certain chromosomes themselves are gene-rich as opposed to others.

Apart from genes, which as mentioned represent a very minor fraction, 5% of the genome – only –contains some regulatory elements that control gene expression, such as noncoding RNAs. The rest of the genome is mostly composed of repetitive sequences (over 50%; including satellite DNA, SINEs and LINES retrotransposons elements), whose role remains poorly understood but would supposedly be related to maintaining chromosomal structures. As for the genes, the repartition of those repetitive structures is not random, and they are clustered together along the chromosomes. Subsequently to the release of the human genome assembly, several projects that aimed at extensively annotating the human genome such as the Encyclopedia of DNA elements

(ENCODE) were launched. Launched in 2003, the ENCODE project aimed to identify all functional elements of the human genome and provide an annotated reference to the scientific community, after a pilot phase of annotating defined regions of the human genome as a training (Consortium *et al.*, 2007; Weinstock, 2007). After years of intense labor, they proposed (among a plethora of other results) an extensive reference set of annotated genes and transcripts (GENCODE; (Derrien *et al.*, 2012; Djebali *et al.*, 2012; Harrow *et al.*, 2012; Howald *et al.*, 2012; Pei *et al.*, 2012).

3. Gene organization

Genes are the units of the genetic information, which are mostly transcribed into mRNAs in the nucleus, translocated to the cytoplasm where they are translated into proteins, the molecules ultimately responsible for the 'molecular' phenotype of the cell. The genetic information is stored in the DNA via the genetic code, which from the sequence of nucleotides determines the sequence of aminoacids in a protein. Linear polypeptide chains originating from mRNA translation are not yet functional. Those premature proteins have to fold to acquire a specific three-dimensional structure, may assemble with other subunits to acquire a quaternary structure, may be chemically modified (*e.g.* addition of phosphate groups) and then are ready to operate. Most genes are capable of producing several associated transcripts by alternative splicing, generating a multitude of proteins. As a result, the 21,000 genes encode as many as a million of different proteins. However, not all genes code for proteins: the final product might be a non-coding RNA.

A human gene is generally composed of an alternation of exons and introns. Introns are portions of DNA that are initially transcribed into RNA, but are excised from the mature mRNA and are not represented into the final protein product. RNA splicing is a highly efficient process, guided by specific consensus sequences in the primary RNA transcript at the exon/intron junctions: 5' donor splice site consensus sequence: AG| gt, 3' donor splice site: an adenine branchpoint around 20 nucleotides before beginning of the exon, a polypyrimidine tract, and a consensus ag| G (**Figure 1**). At each extremity of the coding sequence (delimited by the methionine start codon and the stop codon), the gene is flanked by 5' and 3' untranslated regions (5'UTR, 3'UTR). A promoter region is located at the 5' end of each gene, which contains sequences regulating the time and place of gene expression. Other regulatory elements are dispersed throughout the gene (enhancers, silencers, etc), and can be located in the introns, at the 5' or 3' end of a gene but also at a considerable distance away. The 3' end of a gene contains a signal sequence of AAUAAA tract, which promotes the termination of the mRNA and addition of the polyadenylation tail.

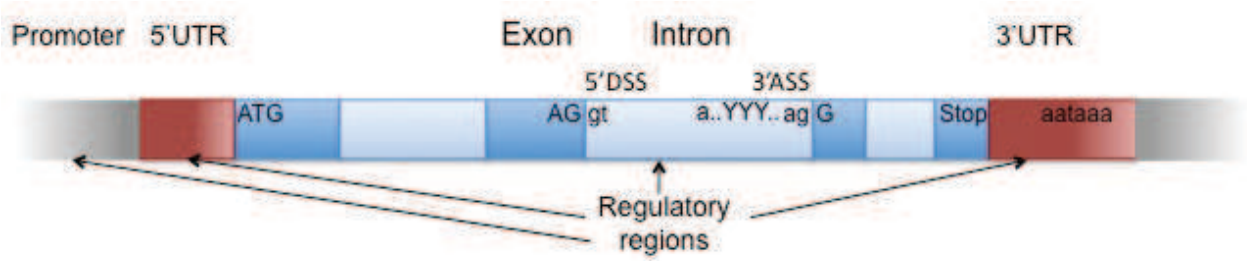


Figure 1: Schematic structure of a human gene.

UTR: untranslated region; DSS: donor splice site; ASS: acceptor splice site; Y: pyrimidines.

The correct expression of a gene and its associated protein production depends on a tightly orchestrated number of events, including gene dosage and structure, transcription, RNA splicing, mRNA stability, translation, structure, function and degradation of the protein product. Any genetic variation that would eventually impair one of these steps can have a dramatic effect and may lead to a disease phenotype. Those genetic variations include large structural variations, or punctual variations affecting protein-coding regions but also variations located in any of the essential regulatory elements (promoter, consensus splice sites and branching sites, regulatory sequences, polyA sequence signal, etc).

4. Genetic variations inter-individuals

The human genome is highly conserved (at 99.9%) from one individual to another mostly due to important selective pressure. However, due to the gigantic size of the diploid human genome (about $6x10^9$ base pairs) this represents millions of base pairs differing between individuals, which genetically determine macroscopic differences among individuals. A vast majority of those variation have no or benign effect on one's phenotype (*e.g.* eye color, hair color, etc) while others are directly responsible for causing disease. Due to the minimal fraction of coding sequences, most of those variations are located in non-coding regions (introns, noncoding RNAs, repeat elements…) and are expected to have a minimal impact at the level of protein production, structure and function. Our actual understanding of non-coding regions being limited, no actual models hence no bioinformatics tools allow a precise interpretation of such non-coding variants apart from splicing predictions.

In some disciplines, a mutation is defined as any change in the nucleotide sequence of a genome compared to the reference, devoid of any notion of pathogenicity. However, to avoid confusion and as recommended by HGVS (Human Genome Variation Society) guidelines, I will use the neutral terms 'variation' or 'variant' instead, and refer to 'mutation' for a genetic variation that causes a disease. Genetic variations can be classified into two main categories: the ones involving large fragments of DNA that can encompass up to entire chromosomes, and point variations that affect one or a few base pairs. Large structural variants arise either from impaired

chromosomal segregation during cell division leading to an abnormal number of chromosomes, or spontaneous events during chromatid recombination leading to translocations, inversions, deletions, insertions, etc. Point variations arise mostly from DNA replication errors (frequency rate: $1x10^{-10}$, or less than one new mutation per cell division; (McCulloch and Kunkel, 2008)) or impaired DNA repair, and are either spontaneous or fostered by the exposure to mutagen agents. The different types of variations affecting one or few nucleotides in a gene are detailed below.

Single nucleotide variants (SNVs)

The substitution from a base to another within the coding sequence can lead to a silent/synonymous change (no effect on the protein sequence, which are relatively frequent when the affected base is the third base of a codon), the replacement of one aminoacid by another referred to as missense variations (mostly when affecting the first or second bases of a codon), the replacement of one codon by a stop codon called nonsense variations, or the replacement of the key start or stop codons by another codon. Silent variations are usually benign since they do not affect the protein sequence, but they can be disruptive if affecting splicing (substitutions of nucleotides near/within splice sites or creating strong cryptic splice sites). Missense variations alter the sequence of the protein, and can be either benign or pathogenic depending on the importance of the affected residue with respect to structure or function of the protein. Nonsense variations incorporate a premature stop codon within the sequence, leading to an mRNA carrying a premature stop codon and that is highly vulnerable for nonsense mediated mRNA decay (reviewed in (Nicholson *et al.*, 2010)). Either the protein is not translated at all due to the degradation of the mRNA, or it is translated under a truncated form that can be more or less stable and functional. Conversely, SNVs can destroy a termination codon and allow translation to continue resulting in a longer protein that may also be more or less stable and functional. Lastly, abolition of the start codon generally prevents any kind of translation resulting in the total absence of protein product.

Outside of the coding sequence, SNVs affecting the acceptor or donor splice sites, branching sites, promoter region or the polyadenylation sequence, or even creating competitor splice sites may result in aberrant amounts of mRNA products or altered RNA processing.

Short insertions/deletions (indels)

Gene variations include also the insertions or deletions of a few base pairs. In coding regions, these either conserve the reading frame if they are a multiple of three (in-frame variations), or cause a shift of the reading frame at the point of the insertion/deletion (frameshift variations). Some tools have been recently developed to predict their impact (*e.g.* (Hu and Ng, 2012)). In in-frame variations, the corresponding number of aminoacids will be inserted/deleted within the normal protein sequence, which can have variable impact on the protein structure and function.

Conversely, frameshift variations incorporate a number of abnormal aminoacids before a premature stop codon, which will have the same possible consequences as nonsense variations.

Dynamic mutations

Some variations are a bit more complex, and involve the amplification of a repeated region (generally trinucleotide tracts) during replication in the gametes. This amplification leads to an increased number of repeats (located or not within the coding sequence) that becomes toxic when bypassing a certain threshold. Within the coding sequence, this amplification result in an abnormally long protein, while within intronic or regulatory sequences it leads to altered transcription, mRNA processing or translation.

Polymorphisms vs rare variants

As mentioned, not all variations reported in an individual are necessarily pathogenic. Some are private or rare, others are relatively common in the general population. Variants with a frequency superior to 1% in the general population are generally called polymorphisms, the ones with a lower frequency are referred as rare variants. Polymorphisms can be of any of the previously discussed type, from SNVs, indels to structural variants. Polymorphisms are predominantly innocuous, as their high frequency makes them incompatible with a full penetrant role in rare diseases. They can nonetheless be associated to genetic risk or protective factors.

5. Discriminating benign variants vs disease-causing mutations

The entire challenge in human genetics – and even more with the arrival of the next-generation technologies flooding databases with variations of unknown significance – is to assess with the highest possible level of confidence whether a variant reported in a patient is responsible for the observed clinical features or not (summarized decisional tree in **Figure 2**).

First, the more the variants are supposed to affect the protein sequence, the more deleterious they shall be presumed. Frameshift, nonsense, start loss, and splice site variants along with large structural variants are the most likely to have an effect on the phenotype since *a priori* directly impacting on RNA processing and protein production.

Then, many additional keys can be gathered from *in silico* tools giving information on the conservation of the affected nucleotides but also of the affected aminoacid residue (if located within a coding sequence), the putative impact on the resulting protein structure (*e.g.* whether it is located within a protein domain, binding sites, etc), the putative effect on RNA splicing (abolition/creation

of splice sites), the pattern of gene expression, … All those are precious information that help predicting the functional impact of a variant (detailed in **General methods**, **III**).

Once a variant seems deleterious enough, co-segregation studies allow verifying whether the candidate variant segregates with the disease status in a family. The study of large cohorts of controls from similar geographical origin also ensures that the detected candidate is not a common variant in the population of concern. This should be computed carefully in relation to the disease of interest (*e.g.* prevalence, mode of inheritance, proposed penetrance, etc; see (Piton *et al.*, 2013)). Once the variation becomes a stronger candidate, functional analyses of the variation can be performed at any protein/cellular/organism level to demonstrate the functional impact of the variation.

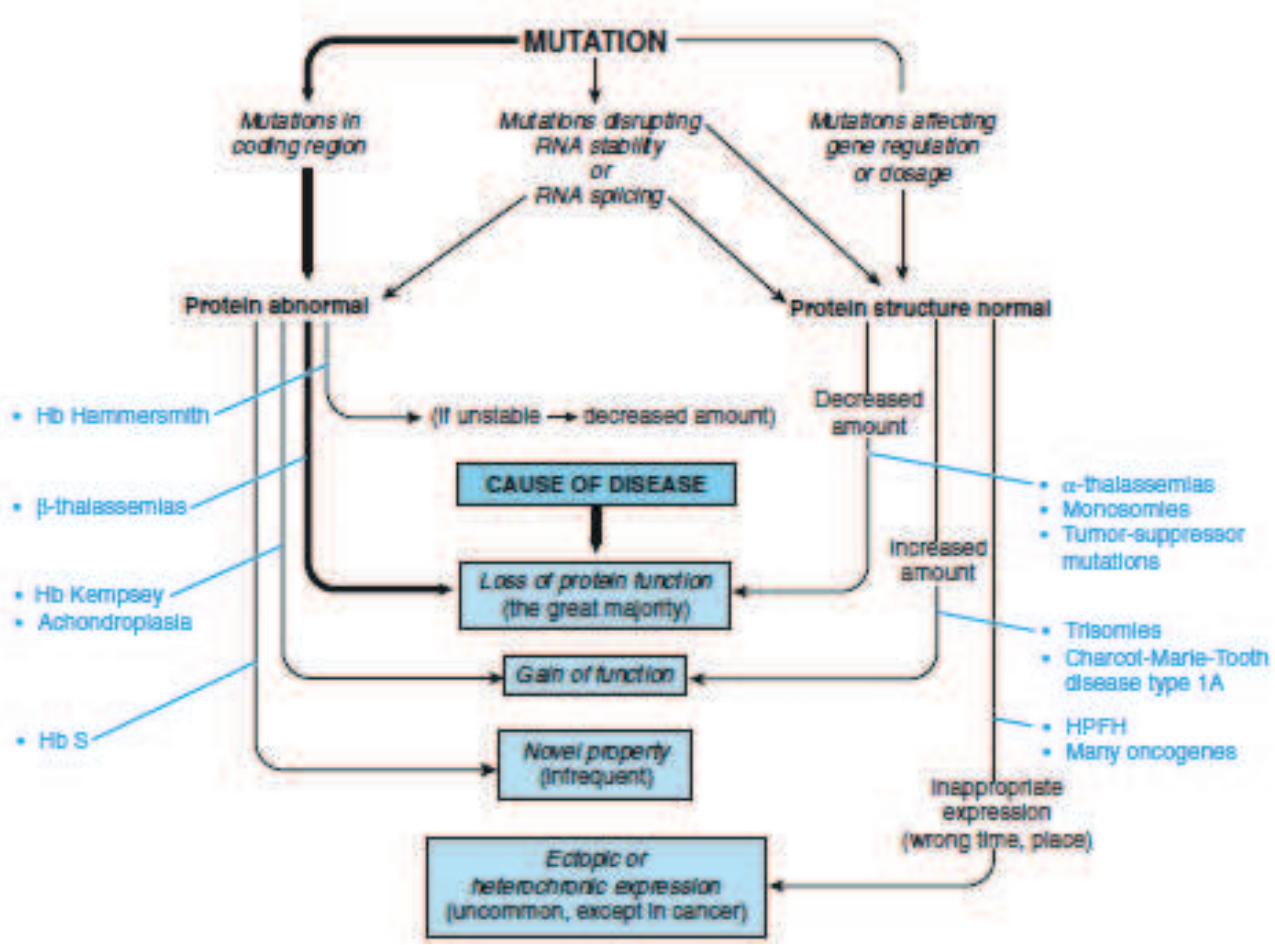


Figure 2 (Thompson & Thompson, Genetics in Medicine): Decisional tree to illustrate the possible mechanisms for a variant to cause a disease.

II- Classification of genetic disorders

The contribution of genetic factors in disease can be large or small, from causing the disease or modifying the course of the disease, to only predisposing to disease. Disorders caused entirely or partially by genetics factors are generally classified into several types: chromosomal disorders, monogenic or single gene disorders, and polygenic or multifactorial disorders.

1. Chromosomal diseases

In chromosomal diseases, structural defects from large chromosomal fragments to whole chromosomes cause the disease. The phenotype originates from an excess or deficiency in genes contained in the chromosomal abnormality. A single gene may be the driver or several genes can each contribute partially to the phenotype. Chromosomal diseases are rather common, with an assessed incidence of 0.7% in newborns (**Table 1**).

The most common type of chromosomal abnormalities is aneuploidy, which consists in either an extra copy or the absence of a chromosome. Trisomies are generally more viable than monosomies, since trisomy 13, 18, 21, X and Y are compatible with life as opposed to the single monosomy X, aka Turner syndrome. The only autosomal trisomies that are not lethal are the ones involving the three autosomes with the lowest number of genes. The most frequent chromosomal abnormality is the trisomy 21 or Down syndrome that has an assessed incidence of 0.1% (Weijerman and de Winter, 2010). The proposed mechanism for aneuploidies is a non-disjunction of chromosomal pairs during meiosis. If occurring during a mitotic division in the zygote, non-disjunction of a pair of chromosomes leads to mosaic aneuploidies. Aneuploidies are universally associated to physical and cognitive manifestations.

Abnormal Karyotype	**1st trimester abortuses**	**Fetuses of mother > 35 years**	**Live births**
Total incidence	1/2	1/50	1/160
Numerical abnormalities	96%	85%	60%
Structural abnormalities	4%	15%	40%
Balanced	0%	10%	30%
Unbalanced	4%	5%	10%

Table 1 (adapted from Thompson & Thompson, Genetics in Medicine): Incidence in fetuses and newborns of numerical and structural chromosomal abnormalities.

Other chromosomal abnormalities are structural defects involving chromosomal fragments that have been isolated from chromosome breakage, and which are eventually reconstituted in an abnormal combination. Chromosomal breakage occurs spontaneously at low frequency but can also be fostered by breaking reagents. The recombination between highly similar sequences (*e.g.*

pseudogenes, LINEs elements, low copy repeats, etc) or mispaired chromosomes may also provoke such fragility and lead to gene duplications or deletions. Like in numerical chromosomal abnormalities, structural chromosomal rearrangements can be mosaic.

Structural rearrangements can be either balanced (*i.e.* the final set of chromosomes has all chromosomal material, although displaced) or imbalanced (*i.e.* the final set of chromosomes has additional or missing chromosomal material). Unbalanced rearrangements include mostly large deletions (generally reflecting haploinsufficiency), duplications, unbalanced translocations, unbalanced inversions, and more rarely the formation of marker chromosomes (small supernumerary chromosomes), isochromosomes (chromosomes with a duplicated arm while the other is missing) or dicentric chromosomes (two arms of two separate chromosomes fused together). Unbalanced rearrangements are more likely to result in an abnormal phenotype since disrupting the normal balance of gene expression. During my thesis, I was involved in unraveling the molecular mechanism provoking an unbalanced-rearrangement responsible for globozoospermia ((Koscinski *et al.*, 2011; Elinati *et al.*, 2012); see **Appendices 1 & 2**).

Balanced rearrangements include paracentric and pericentric inversions and translocations. Carriers of balanced rearrangements generally do not have phenotypic manifestations since the full genetic material – although displaced – is present, unless the breakpoints disrupt the coding sequence of a particular gene.

Many copy number variants (CNVs) are rare private CNVs and one challenge is to evaluate their contribution in a particular phenotype. Indeed, although it is widely accepted that large structural variants are mostly disease causing, this is not the case for smaller CNVs that have a relatively high frequency within the general population (incidence of 1:8 for de novo microdeletions, and 1:50 for de novo microduplications; (Guilmatre *et al.*, 2009)). The penetrance of some CNVs also appears to be highly variable (*e.g.* can be inherited from 'healthy' parents) and may lead to different phenotypic outcomes complicating the discrimination between risk factors and benign CNVs. One possibility to assess the contribution of a CNV in a phenotype is to compare their frequency in a huge cohort of controls, or to find the driver gene (or the disrupted gene in translocations) and identify additional patients with a similar phenotype carrying CNVs or point mutations affecting the same gene.

2. Monogenic disorders

Monogenic disorders are caused by mutations (inherited, or that appeared *de novo* in a parental germ line) located within a single gene, and are often called Mendelian disorders as their mode of inheritance follows Mendel's laws. Although individually rare, they represent – as a group – a considerable proportion of all pediatric and adult disorders. As seen in the previous section, the mutation can be any of a single nucleotide variant (SNV), a small indel, or a submicroscopic CNV

affecting a single gene (*e.g.* a single exon deletion). Monogenic disorders can be inherited as dominant (*i.e.* one of both parental allele carry the mutation) or recessive (*i.e.* both parental alleles carry a mutation) and affect genes located on the autosomes, or sex chromosomes (usually the X-chromosome). Similarly to chromosomal abnormalities, mutations reported in monogenic disorders can be mosaic.

Dominant disorders – especially if severe enough that showing a high impact on fitness – can be caused by *de novo* mutations. Inherited mutations are reported either when the disorder does not have an impact on fitness, or when associated to incomplete penetrance or variable expressivity. Recessive disorders are frequent among populations of weakly mixed genetic background or of high level of consanguinity as they present an increased proportion of homozygosity regions, which may contain a disease-causing allele. However, consanguinity is not the most common explanation for autosomal recessive traits: most cases originate from mating of unrelated people, each one carrying a disease-causing allele (especially if one recurrent disease-causing allele is frequent in the general population).

For the genes located on the sex chromosomes, monogenic disorders can also be either X-linked (or Y-linked) dominant (*i.e.* supposedly as many affected females than affected males), or recessive (*i.e.* more affected males than affected females, who are more protected because of their extra copy of the X-chromosome; *e.g.* hemophilia A). The affected status of females in X-linked disorders is complicated by the patterns of X-inactivation. Indeed, females harbor two populations of cells, one in which one of the X-chromosome is active, the other in which the other X-chromosome is active, harmonizing the expression of most X-linked genes in both males and females. Depending on the pattern of X-inactivation (which can be random or not), two females with the same genotype can express different phenotypes depending on the proportion of cells with the active X-chromosome carrying the disease-causing allele.

Finally, a last mode of inheritance is called ‘maternal transmission’, where the mutation is located within maternal mtDNA and transmitted to the embryo by the maternal oocyte.

Inheritance patterns of monogenic disorders can be complicated according to parental genomic imprinting, leading to the expression or not of the disease-causing allele depending on whether it was maternally or paternally inherited (*e.g.* pseudohypoparathyroidism and related disorders). Another class of monogenic disorders giving rise to atypical inheritance patterns are those involving dynamic/repeat expansions disease-causing mutations, with the characteristic of the genetic anticipation phenomenon (*e.g.* Fragile X syndrome, Huntington disease).

Monogenic disorders can be rather simple and homogeneous, where for patients harboring a similar macroscopic phenotype the mutation is always located in a same gene. Conversely, some monogenic disorders are characterized by a vast genetic heterogeneity. In individuals with similar clinical phenotype, causative mutations can be found in one gene within a more-or-less large subset of genes described as implicated in the disease. Depending on the disorder, this heterogeneity may

be more or less consequent (*e.g.* Bardet Biedl Syndrome: 19 genes reported, intellectual disability: over 300 genes reported).

3. Polygenic/oligogenic and multifactorial disorders

Contrary to monogenic disorders, polygenic diseases are provoked by the synergistic effect of mutations dispersed in several independent genes. Several variations in different genes, instead of one single mutation, are responsible for the phenotype. The oligogenism model gathers different scenarios: the 'synergistic' model, where several variants participate together in the phenotype and have a synergistic effect (*i.e.* carriers of single variants are asymptomatic), and the 'modifier' model, where one variant is the major contributor to the phenotype while the remaining variants acts as modifiers (*i.e.* carriers of the contributor variant are affected but with milder or incomplete phenotype, carriers of the modifiers may be unaffected).

Only very few disorders were demonstrated as purely polygenic, mostly because the larger the number of involved genes, the closest it gets to nearly impossible to prove. The few disorders demonstrated as such are digenic disorders. For many neurodevelopmental disorders this polygenic hypothesis has been largely proposed, but cannot be strictly demonstrated for now since supposedly involving a very large number of genes. The simplest example of multigenic disorders is the digenic *retinitis pigmentosa*. Patients presenting with this disorder carry two heterozygous mutations in two different genes (*PRPH2* and *ROM1*) encoding for two photoreceptor membrane proteins. Carriers of the single heterozygous alleles do not have a phenotype (Kajiwara *et al.*, 1994).

Multifactorial (or 'complex') disorders originate from a complex combination of both genetic and environmental factors, and are thought to be responsible for a majority of common disorders. The contribution of genetic factors is assessed through the study of concordance in monozygotic twins or by assessing the recurrence risk in relatives of affected individuals. The environmental part is often highlighted by the non-full concordance of disease status between monozygotic twins, or the higher recurrence risks in siblings reared together as opposed to siblings reared apart. The recurrence of such disorders in some families (or familial aggregation) is then explained by the shared genetic background combined with a shared environment, increasing the chances to get the exact same gene-environment interactions leading to the disease. Environmental factors are not necessarily negative, and may either accelerate, worsen or conversely protect against a disease phenotype.

Many common disorders such as diabetes, congenital heart defects, Alzheimer's disease, autism spectrum disorders, and psychiatric disorders appear to result from the interaction of several variants in independent genes together producing or predisposing to a phenotype in concert with environmental factors. Their pattern of inheritance does not follow Mendelian's laws.

III- Gene discovery and diagnostic methods

1. Gene mapping methods

The methods used for discovering novel genes responsible for genetic disorders have been jostled with the emergence of NGS technologies, as illustrated by the exponential increase of gene discovery ever since (**Figure 11**). A quick overview of the different methods that were used prior to NGS is given, most of which are still commonly used alone or in combination with NGS to accelerate the searches.

Linkage analysis

One of the standard strategies used to identify causative genes is linkage analysis. Linkage is the terminology used for the tendency of several alleles that are located close together on the same chromosome to be transmitted together through meiosis without recombination event. The linkage analysis approach is based on the study of a family (as large and informative as possible), with several affected and healthy members. It explicitly uses the disease status of the family members to follow the inheritance of the disease over generations and see whether the disease locus appears linked with specific polymorphic genetic markers when transmitted to the next generation. Estimation of linkage and distance between the disease locus and polymorphic markers is made through assessing the recombination fraction Θ between them (Θ=0 if both are linked, Θ=0.5=50% if unlinked). A best estimate of the recombination frequency Θ_{max} is determined, which reflects the estimated linkage and distance of the disease locus with genetic markers according to observations (*i.e.* the family). A statistic measure, the LOD (logarithm of odds ratio) score is computed to determine whether the hypothesis of linkage with a distance of Θ is likely, and with which strength. A LOD score above 3 is usually considered as linkage evidence (*i.e.* hypothesis of linkage 1000 times more probable than the hypothesis of independence).

$$\text{LOD score} = \log_{10} \frac{\text{Likelihood of the data if loci are linked at a particular } \theta}{\text{Likelihood of the data if loci are unlinked } (\theta = 0.50)}$$

Linkage evidence of genetic markers with the disease locus suggests the gene of interest may map closely to their genomic positions. A genome scan (*i.e.* linkage analysis performed on genetic markers distributed along the genome) usually gives one or several genomic intervals with high likelihood of containing the causative gene, among many others (see **Figure 3** for illustration). During my PhD, I preformed linkage analyses to identify candidate regions in three families with pyruvate transport defects that partially contributed to the discovery of the human mitochondrial pyruvate transporter ((Bricker *et al.*, 2012); **see Appendice 1**). The main drawbacks of this

approach is that 1/ it needs to be absolutely certain of the disease status for each family member (which can be a difficult task for late-onset disorders or disorders of incomplete penetrance for instance), 2/ the disease status may segregate with some polymorphic markers just by chance leading to erroneous conclusions, 3/ the output intervals may contain several dozens of genes if not many more, considerably complicating the search for the causative gene.

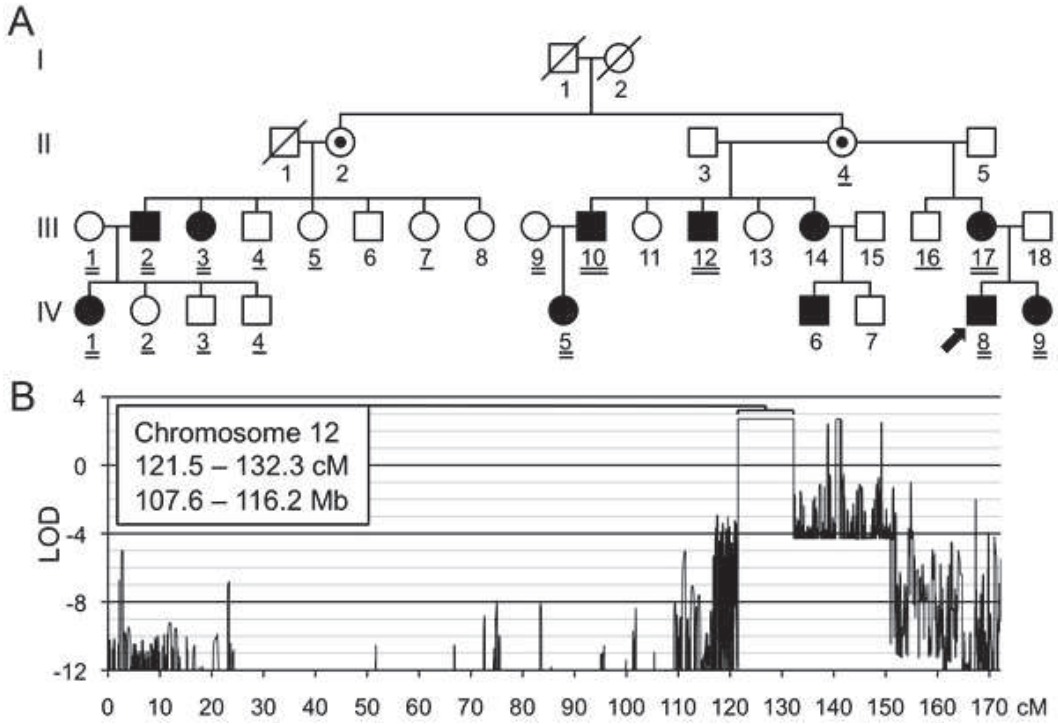


Figure 3 (from (Bowen *et al.*, 2011)): Example of linkage analysis showing linkage evidence between the disease locus responsible for metachondromatosis and genetic markers located on chromosome 12.

A. Pedigree of the family used for the analysis, individuals whose DNA was used for the analysis are underlined. **B.** LOD score values on Chromosome 12. 1 cM (centimorgan) represents the genetic length over which an average of one crossing over occurs per 100 meioses (*i.e.* 1% of recombination rate in meioses).

Identity by descent and homozygosity mapping

Identity by descent analysis is a family-based method, which surveys the genome of affected and unaffected members of a family through genetic markers. This genome survey is performed by genotyping simultaneously from several thousand up to millions of single nucleotide polymorphisms (SNPs) genomewide using SNP microarrays in several family members, both affected and unaffected. The resolution of such arrays has tremendously improved, from the initial 1.5k SNPs GeneChip HuSNP Mapping array (Mei *et al.*, 2000) to the current 6 millions of SNPs from Affymetrix (Genome-Wide SNP Array 6.0). Identical by descent (IBD) regions are those where contiguous SNPs share the same genotype across different family members. Mapping a disease gene then consists in highlighting IBD regions common between affected members while

excluding the ones shared between unaffected and affected family members. The approach can then be repeated on independent families to further narrow down the candidate region.

Homozygosity mapping is a particular subcategory of IBD studies used when looking at suspected autosomal recessive disorders, and mostly in consanguineous families or population of inbreeding and weakly mixed genetic background. It assumes that the causative mutation is homozygous and that both copies segregate within the same IBD haplotype. Indeed, the rarer the disease, the higher the probability that it originates from consanguinity or inbreeding in those families. Since chromosomal regions are usually transmitted as a whole, all markers located in close proximity to the disease allele will also be homozygous. The search then consists on screening shared regions of homozygosity between all affected members, which are not shared with unaffected ones. This presumes that all affected members have the exact same disease and caused by the same disease allele, which can be erroneous in very large highly consanguineous families where two different mutations may arise concomitantly in different branches of the family (*e.g.* (Laurier *et al.*, 2006)).

Association studies

A fully different approach is the association study, which is not based on family study but rather on population study. This approach consists on identifying specific alleles that are associated with a disease in a large population of affected and controls. Since looking at alleles with major to relatively low effect, it requires studying large populations of both affected and controls to reach statistical significance. This method is particularly used for multifactorial and complex disorders. It is based on the reported frequency of alleles in a population of affected individuals as compared to a population of controls, which if increased or decreased is considered as a disease association. It was initially used only on a limited set of alleles, located for instance in genes encoding for proteins from a particular pathway relevant for a specific disease (candidate gene association studies). Association studies are now performed genome-wide (GWAS, for genome-wide association studies), probing simultaneously several millions of alleles without necessary prior knowledge regarding the pathophysiology of a disease. The main drawback of association studies is that when an allele is associated with an increased relative risk for a disease, this does not definitely prove that the allele or the gene is effectively involved in the pathogenesis (*e.g.* it could be due to another allele in linkage disequilibrium). Also, for more complex disorders most GWAS failed to identify any risk loci, presumably due to extensive genetically and phenotypically heterogeneity (*e.g.* diabetes; (Lohmueller *et al.*, 2013)).

2. Gene identification approaches

Positional cloning

Positional cloning is performed once the disease locus has been narrowed to a candidate region, for instance via linkage analysis or homozygosity mapping. Such generic method is usually applied when nothing is known about the function of the gene of interest. To reduce the efforts, the candidate region has to be narrowed as far as possible. Indeed, it was historically a very effort-intensive task, involving initial screening of successive genomic DNA libraries to identify clones containing overlapping DNA fragments spanning the candidate region. Once those were identified, several methods could be applied to extract the position and the structure of genes within the candidate region (*e.g.* cross-species hybridization, exon trapping, cDNA hybridization). As an illustration, the Duchenne/Becker muscular dystrophy gene was mapped following intense positional cloning efforts (Hammonds, 1987; Hoffman *et al.*, 1987). Since the publication of the human genome, those methods became obsolete and have been replaced by bioinformatic analyses allowing to directly pinpoint genes located within the region. Candidate genes are then screened in patients for deleterious mutations using Sanger sequencing. However, these candidate regions may contain up to several hundred genes and successive sequencing of all genes until the identification of the causative mutations can take years. The development of next-generation sequencing (see **section IV**) has considerably facilitated this step.

Functional candidates and comparative genomic approaches

Conversely, other genes were identified independently of their position but through their molecular function or their involvement in a particular metabolic pathway that was relevant to the disorder of concern. Candidate genes could be identified through functional studies (*e.g.* Fanconi anemia type C), or by homology to other species in which the protein encoded by the gene was known (*e.g.* Hemophilia A). As an illustration, *PLP1* was mapped by using cDNA of bovine proteolipid protein combined with Southern blot to fish and map both human and murine PLP genes (Willard and Riordan, 1985). The location of murine *plp1* was highly suggestive of its implication in the phenotype of two myelin-deficient mutant mice, and corresponding human *PLP1* that was located in conserved synteny blocks as compared to the murine gene was finally shown to be implicated in the analogous human Pelizaeus-Merzbacher disease (Willard and Riordan, 1985; Mattei *et al.*, 1986; Cremers *et al.*, 1987).

This functional approach became even more common with the emergence of the bioinformatics era, where many genes have been uncovered by large-scale genomic analyses either looking at a set of candidate genes sharing a similar function, pathway, or expression pattern, or interrogating the human genome looking for recognizable sequence counterparts of a particular

gene with known function in other species. The Bardet-Biedl gene *BBS7* for instance, was discovered by a comparative genomics approach, through a motif-based sequence similarity comparison using *BBS2* as a probe in human and zebrafish genomes (Badano *et al.*, 2003a).

3. Diagnostic tools

Several approaches can be used to detect the presence in patients of a chromosomal abnormality that may be responsible for a chromosomal disorder, or of point variations that may be responsible for monogenic disorders when the gene has been previously identified, which may eventually lead to a molecular diagnosis. Additional biochemical and imaging approaches are also generally useful in directing toward a specific disorder, but are not detailed here. Obtaining a definitive molecular diagnosis is essential for establishing recurrence risks and providing genetic counseling in the family. Moreover, such diagnosis may have a direct impact for the medical prognosis of a patient and even can indicate specific therapeutic options.

Karyotype and FISH analyses

Both karyotype and FISH allow detecting gross chromosomal abnormalities (larger than 10-Mb). The karyotype is the 'minimal' cytogenetic analysis, giving the total number of chromosomes, sex chromosomes and macroscopic chromosomal abnormalities (chromosomes of abnormal size, abnormally located centromeres, etc; example in **Figure 4**).

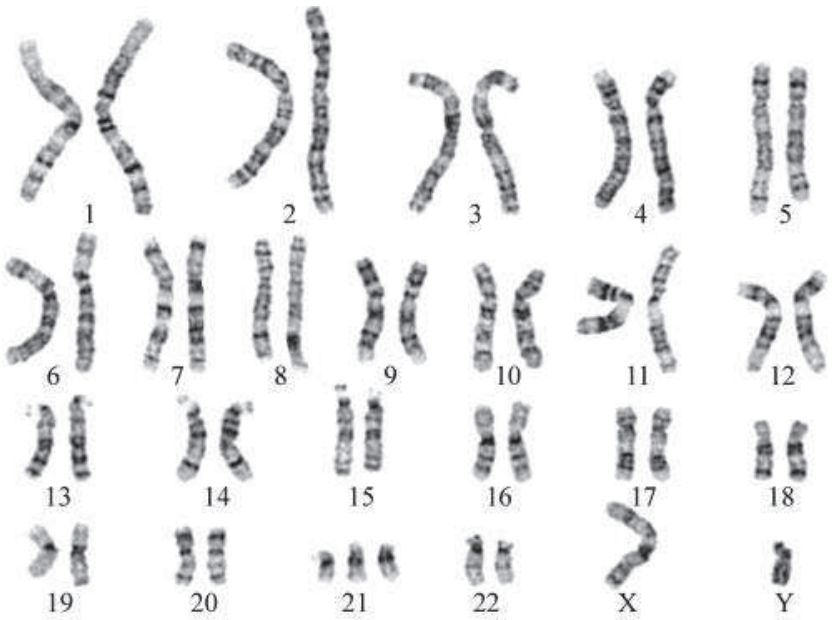
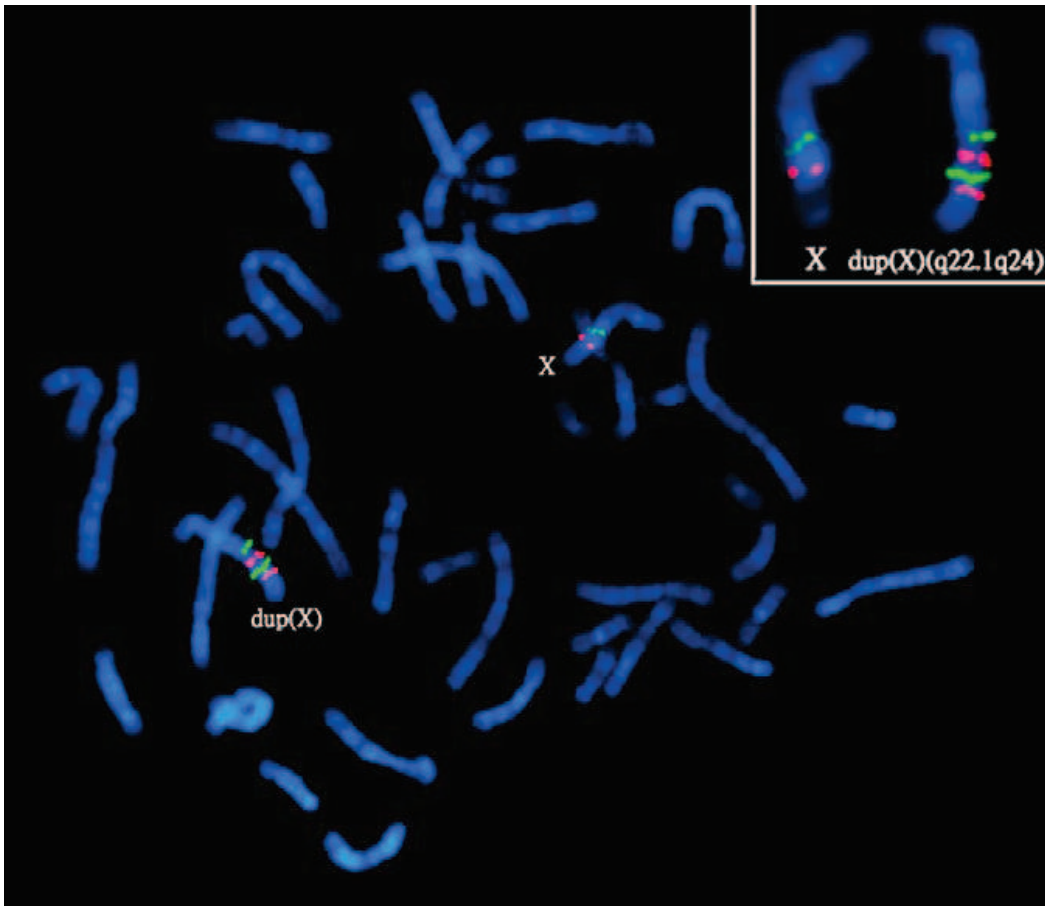


Figure 4 (from (Williamson, 2006)): Karyotype analysis demonstrating an extra-copy of chromosome 21 in a male with Down syndrome.

FISH (fluorescence in situ hybridization) analysis is a cytogenetic technique that allows detecting finer chromosomal defects such as deletions, duplications or translocations (example in

Figure 5). Fluorescent probes complementary of a DNA or RNA sequence of interest are immobilized within corresponding parts of fixed metaphase or prometaphase chromosomes that have been denatured *in situ*. The fluorescent signal is then detected by microscopy to locate the



specific mapping on chromosomes. By using sophisticated image-processing tools, a cocktail of probes (with a characteristic color ratio) complementary of entire chromosomes are mixed together. This enables to map precisely and detect large deletions, duplications, and translocations and is referred to as 'whole chromosome painting' or spectral karyotyping (SKY). FISH resolution can go as low as one megabase when using the more extended prometaphase chromosomes.

Figure 5 (from (Chen *et al.*, 2011)): FISH analysis using two probes of the X-chromosome (RP11-34P3 in Xq24 encompassing *PLP1* in green, RP11-22P2 in Xq22 encompassing *PGRMC1* in red) illustrating a tandem duplication on the X-chromosome that disrupts the NXF gene cluster in a woman.

Array CGH-analysis

Array-CGH (comparative genomic hybridization) analysis is another form of genomewide painting to detect unbalanced structural variants but at much higher resolution (as small as 10-kb up to several Megabases). DNA probes spanning the entire genome are immobilized, and labeled DNA of interest is hybridized to such probes where abnormal fixation revealed by a fluorescent signature will reveal potential deletions, duplications, translocations, etc (example in **Figure 6**). In standard

CGH, two DNA samples are compared one to another, by labeling each one with a particular fluorophore and simultaneously hybridizing them to DNA probes. An abnormal fluorescent signal at a particular probe will reflect unbalanced molecular ratio between the two samples and hence a putative deletion/duplication.

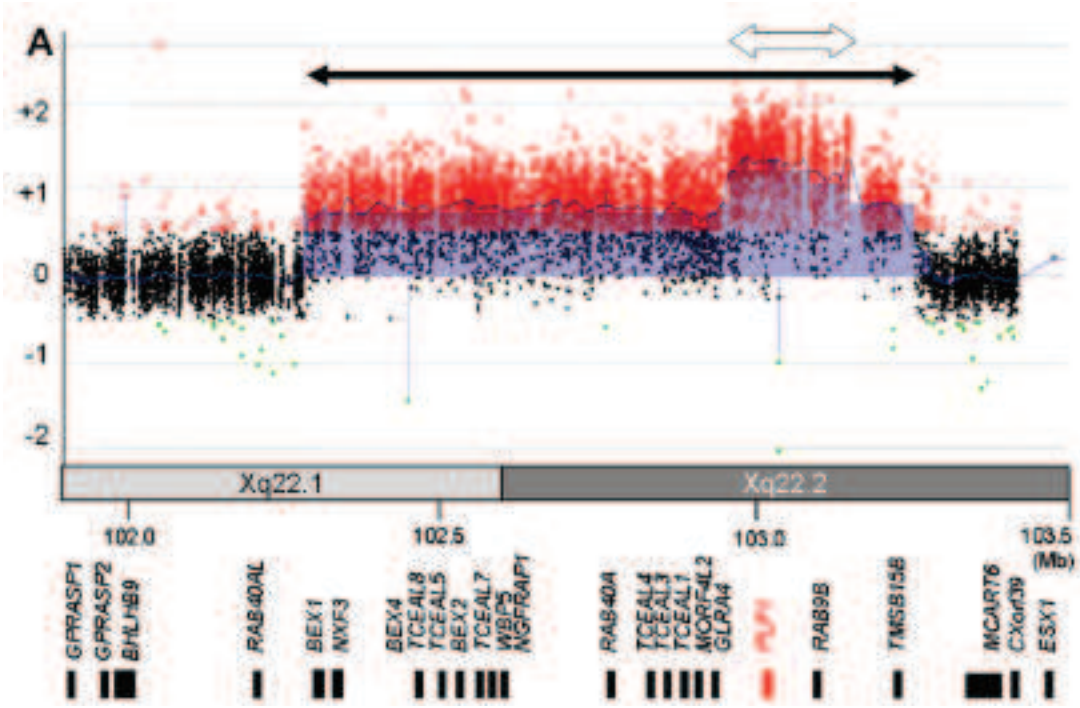


Figure 6 (from (Shimojima *et al.*, 2012)): Array-CGH analysis with abnormal patient/control signal ratio illustrating a duplication encompassing several genes of the X-chromosome (black arrow), and a triplication of *PLP1* (white arrow) in a female with Pelizaeus-Merzbacher disease.

Sanger sequencing

Sanger sequencing is generally used for the detection of point mutations and small indels and is the universal method for DNA sequence analysis. It relies on using dideoxynucleotides that lack a 3'-OH group on the desoxyribose, preventing the attachment of the next base by the DNA polymerase during synthesis. The synthesis is thus prematurely terminated, hence the generic term of 'termination method'.

In Sanger sequencing, the DNA fragment to be sequenced is first amplified by PCR. They are then used as templates for DNA synthesis, and the four fluorescently (or radioactively) labeled dideoxynucleotides are added along with an excess mixture of the four normal nucleotides. Depending on whether a normal or a dideoxyribonucleotide is incorporated, the synthesis either continues or is terminated. DNA fragments are subsequently denatured and separated by electrophoretic migration, which by recovering the fluorescent (or radioactive) signal from the terminating nucleotide allows determining the nucleotide sequence.

Initially, the sequence was determined manually from the pattern of the four reactions run in parallel (with radiolabeled ddATP, ddGTP, ddCTP or ddTTP only) after revelation by autoradiography or UV light. The development of fluorescently labeled ddNTPs, each one with a different dye, allowed the running of all four reactions in one single sequencing reaction and the detection of signals with optical tools. Finally, the latest (and revolutionary) improvement of the method was the one of capillary electrophoresis using different matrices (from polyacrylamide gel to other polymers) to replace the labor intensive of slab gel electrophoresis ((Cohen *et al.*, 1988; Ruiz-Martinez *et al.*, 1993); reviewed in (Karger and Guttman, 2009)). The initially single capillary sequencing was upgraded to multicapillary sequencing, which was the ancestor of the one used now in routine in diagnostic laboratories (Huang *et al.*, 1992).

One major drawback of Sanger sequencing is the impossible detection of large heterozygous deletions or duplications such as exon deletions/duplications. Allele drop out can also occur, preferentially amplifying one out of the two alleles. Finally, even though the automation of the method has tremendously improved both the sequencing capacities and the throughput, Sanger sequencing remains expensive and of low throughput since interrogating one single locus at the time. It remains however the method of choice to sequence long DNA fragments. Of note, the first human genome was sequenced using this method for a total estimated cost of 2.7 billion US $ in a 13-year effort (Lander *et al.*, 2001; Venter *et al.*, 2001; Collins *et al.*, 2003).

IV. Next-generation sequencing technologies: a 2-in-1 tool?

For about two decades, the Sanger sequencing approach remained the universal gold standard for DNA sequencing, especially after its full automation in multicapillary sequencers ('first-generation' sequencing technology). However, as discussed previously, it was not suited for diagnosis purposes in many configurations due to its prohibitive costs and still low throughput. Different massive parallel sequencing methods were subsequently developed to overcome such disadvantages. All next-generation sequencers are based on the same principle, which is live neosynthesis of DNA fragments allowing the detection of the sequence as it gets synthesized by diverse detection methods, getting rid of the time consuming migration step of Sanger sequencing. Also, this neosynthesis is performed in parallel in millions of independent synthesis reactions, allowing an ultra-high throughput of generated sequences.

1. General definitions

Before moving forward, some generic terms universally used when referring to NGS need to be defined. High throughput sequencers generate several millions of independent sequences, of variable length. The term 'reads' or 'sequencing reads' are used to refer to generated sequences during a sequencing run. NGS data is generally visualized on a genome viewer (see **Figure 7**).

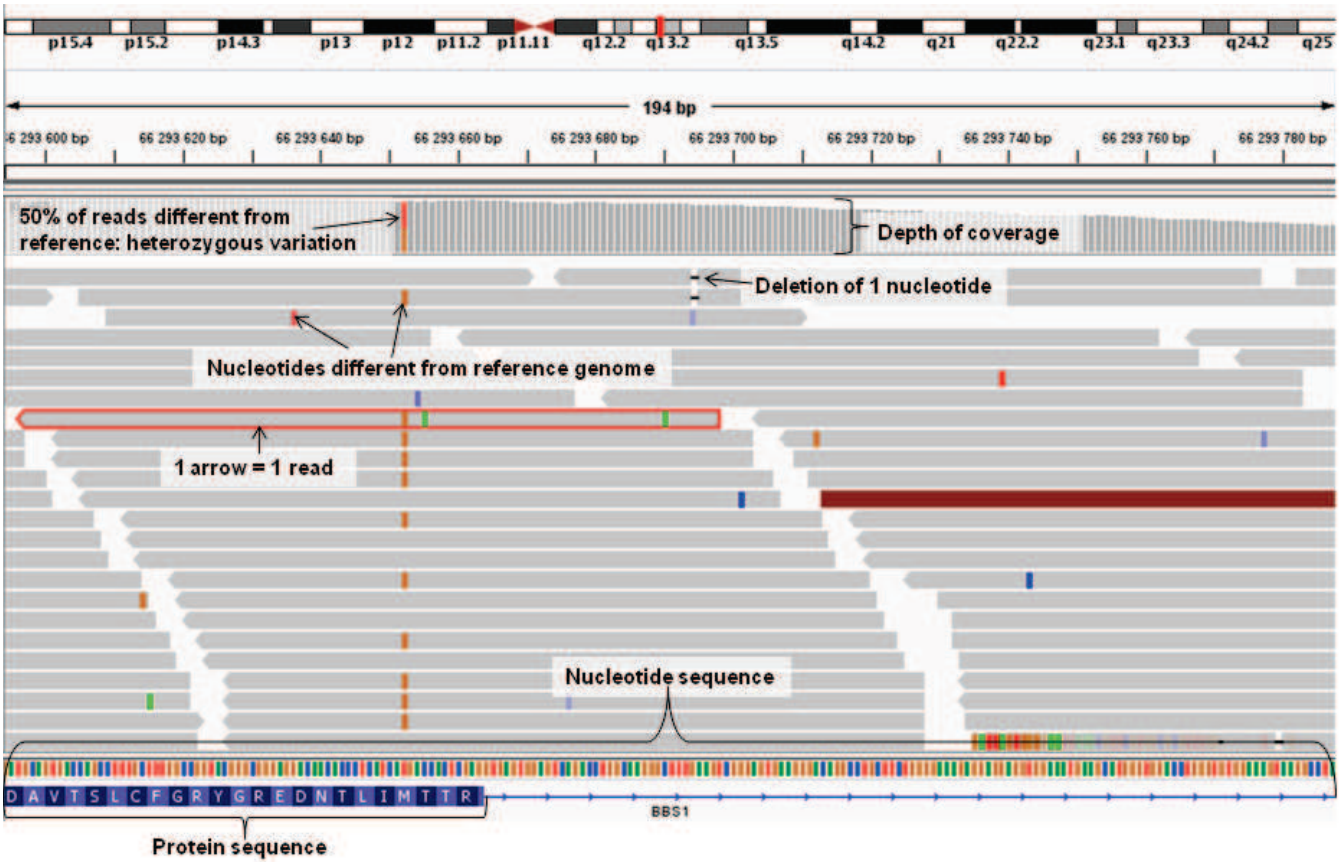


Figure 7: Visualization of high throughput sequencing data zoomed on *BBS1* for one BBS patient, using integrative genome viewer (IGV; http://www.broadinstitute.org/igv/).

The ‘depth of coverage’ or ‘sequencing depth’ corresponds to the number of generated sequencing reads at a particular genomic position. It can be computed either when integrating the total number of generated reads, or after filtering (*i.e.* after discarding duplicate reads, multimapped reads and reads of low quality).

‘Multiplexing’ or ‘pooling’ several samples is often performed since the power of sequencers enables a huge excess of reads per sample that are not useful for most applications. Several samples can be mixed together on a single sequencing lane prior to sequencing. To do so, a short index or barcode of several oligonucleotides, specific for one particular sample is added during the library preparation at each DNA fragment extremity for every sample (see **Figure 8**). At the end of the sequencing run, the sequences originating from a particular sample are retrieved by recovering the ones containing its specific barcode.

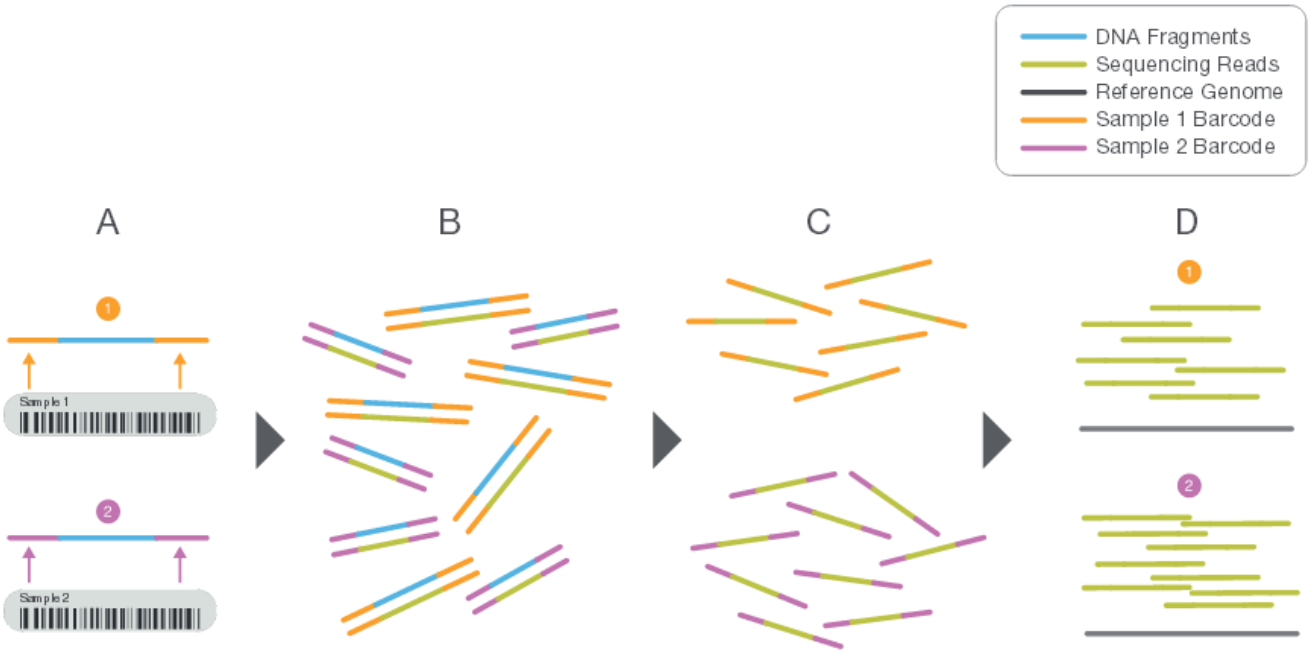


Figure 8 (from Illumina sequencing introduction): Schematic overview of sample multiplexing prior to high throughput sequencing.

A: specific barcodes are ligated to the extremity of fragments from two samples; B: libraries of both samples are mixed together prior to sequencing; C: barcode sequences are used to de-multiplex the reads; D: each set of reads is independently aligned on the reference genome.

2. Next-generation technologies, or the 2nd generation of sequencers

All sequencers of the 2nd generation mainly differ in the template preparation and underlying chemistry of the neosynthesis while imaging and data analysis remain similar (for a general summary, see **Table 2**).

The template preparation requires an amplification step to bypass the imaging detection threshold, which is usually either an emulsion PCR or a solid phase amplification (reviewed in

(Metzker, 2010)). The amplification step needs to remain minimal not to induce too much bias in the amplification product (*e.g.* AT-rich and GC-rich sequences may be underrepresented). The neosynthesis sequencing step itself is then based either upon cyclic reversible termination, pyrosequencing, or successive ligations. Since the 2000s and the first NGS machines, several improved sequencing machines have replaced previous ones, further improving throughput capacities, speed, accuracy and costs.

The first high-throughput sequencer commercially available (in 2005) was the GS20 sequencer from 454 Life Sciences. It combined single-molecule emulsion PCR for gDNA amplification and pyrosequencing (*i.e*. detection of pyrophosphatase release when a nucleotide is incorporated during DNA synthesis). The first completed genome using high-throughput sequencing used this technology and provided the genome assembly of *Mycoplasma genitalia* in a single run (Margulies *et al.*, 2005). The 454 company was later acquired by Roche that developed an improved version of the sequencing machine, the GS FLX, which was launched two years later.

In the meanwhile, another company Solexa, developed its own technology using clonal bridge amplification and sequencing by synthesis using reversible dye terminators. Through the use of the termination method, their technology is somewhat the successor of Sanger sequencing. The Genome Analyzer from Solexa was launched in 2006. The company was then bought by Illumina, which drastically improved the technology over time as reflected by successive versions of sequencers (Genome Analyzer IIX, HiSeq2000/2500 and MiSeq).

The third sequencing machine of this 'second generation' based on ligation technology was developed by Applied Biosystems and was commercially available through the SOLiD system in 2007 (Shendure *et al.*, 2005). Similarly to the 454 system, clonal amplification is performed through emulsion PCR, and sequencing is performed by successively ligating octamers containing two probe-specific nucleotides. The final sequence is deduced by deconvolution of the color code. Each nucleotide is probed twice with this technology that supposedly gives highly accurate results.

3. The 'newest' generations: 3rd and 4th generations of sequencers

All those 'second generation' sequencing methodologies are still limited in terms of the length of sequences they generate (mostly 30-40 bp; see **Table 2**) and the costs, due to the inevitable image acquisition, processing and analysis. The sequencers of 'third generation' and onwards were developed to circumvent some of those drawbacks.

The development of ion torrent sequencing platforms (Ion PGM and Ion proton, later acquired by Life Technologies) was a real revolution in terms of costs. The detecting method during neosynthesis differs from the other sequencers, avoiding the intermediary of light to translate genetic information into digital data and hence getting rid of the costly optic tools. It is based on the sensing of hydrogen ions that are released when a base is incorporated during DNA polymerization.

When two identical bases are incorporated the voltage doubles, but the linear response is accurate up to 6 bases and the detection of longer homopolymers remains challenging. The Ion PGM became the first sequencer largely affordable for most laboratories. Even though the first platform did not allow much throughput (not enough for a human exome for instance) and was more suited for targeted sequencing, its successor Ion Proton now does.

Sequencing machine	**Amplification and sequencing technologies**	**1 run technical performances**	**Read length**	**Strengths and weaknesses**
Capillary sequencer	Sanger sequencing (dye terminators)	86-kb 1hr	900-bp	Low throughput (max. 96 sequences)
Roche				
GS Junior	Emulsion PCR, pyrosequencing	0.04 Gb 10 hrs	36-100-bp	Long reads, suited for de novo genome assembly/repeated regions and metagenomics High cost High error rates in homopolymers >3-4nt
GS FLX Titanium XLR70		1 billion reads 0.5 Gb 10 hrs	450-600-bp	
GS FLX Titanium XL+ (initially 454 Life Sciences)		1 billion reads 0.7 Gb 23 hrs	700 bp- 1 kb	
Illumina				
Genome Analyzer (formerly Solexa)	Solid-phase bridge amplification, reversible dye terminators, sequencing by synthesis	40 million reads 1-3 Gb 3-6 days	36-bp	Most widely used platform Short reads AT-rich regions underrepresented
Genome Analyzer II		80-100 million reads 2-15 Gb 1-6 days	25-50-bp	
HiSeq2000/2500		Up to 6 billion PE reads 100 - 600 Gb (dual flow-cells) 2-11 days	36-100 bp	
MiSeq		12-50 million reads 25-300 bp 4-65 hrs		
Applied Biosystems/Life Technologies				

SOLiD	Emulsion PCR, sequencing by successive ligation reactions	4 Gb 6 days	35-bp	Encoding system provides inherent error correction, lowest error rate Long runs AT-rich regions underrepresented
SOLiD 3		1 billion reads 30-60 Gb 3-14 days	35-50-bp	
SOLiD 4		1.4 billion reads 80-100 Gb 3-16 days	35-50 bp	
Ion Proton	Semiconductor sequencing	60-80 million reads 10 Gb 2-4 hrs	Up to 200-bp	Fast runs Low-cost sequencer and consumables
Ion PGM		Up to 5.5 million reads Up to 2 Gb 2-7 hrs	200-400 bp	
Helicos BioSciences				
HeliScope	Single molecule sequencing, reversible dye terminator	1 billion reads 21-35 Gb 8 days	35	No amplification bias No GC/AT bias High error rates (3-5%)
Oxford Nanopore Technologies				
GridIon & MinIon systems	Single molecule sequencing, exonuclease sequencing, nanopore sensing	NA NA NA	Theoretically infinite	High accuracy No amplification bias Theoretically infinite length of sequencing reads No errors associated to homopolymers

Table 2: Comparison and evolution of higher throughput DNA sequencing methods showing a large improvement between the first platforms and the newest, in terms of sequencing capacities, throughput, speed and accuracy.

Other highly promising sequencers are those based on 'single molecule sequencing', bypassing the amplification step. The HeliScope from Helicos BioSciences technically belong to this category, even though it was a pioneer in the field and was developed along with the second-generation sequencers (Braslavsky *et al.*, 2003). Like the Illumina technology it also uses the principle of dye-termination, but with much higher resolution of imaging analysis. Terminator nucleotides are added successively in a predetermined order to pause the process, in order to internally reflect the emitted fluorescence that is then captured as an image. Terminator groups are then cleaved before adding the next nucleotide.

The Nanopore sequencing technique is still under development and is proposed by Oxford Nanopore Technologies (Clarke *et al.*, 2009). It is one of the rare sequencer that does not use the sequencing by neosynthesis principle, but rather by excision. The approach relies on the principle that when a nucleotide passes through a nanopore it generates an electric current due to conductions of ions through the pore, which can be sensed and is characteristic of the nucleotide base (and even of its methylation state). Nanopore Sequencing Technologies couples this system with exonuclease digestion, forcing the translocation of each base through the nanopore.

At the end of this second 'vague' of sequencers, some companies developed 'benchtop' sequencers of lower throughput that became affordable for many laboratories (*e.g.* ion PGM from Life technologies, MiSeq from Illumina) and are particularly suited for targeted gene sequencing, small genome sequencing, metagenomics, etc.

4. NGS applications in human genetics

The development of 'all generations' sequencing technologies revolutionized the field of genomics, at both research and diagnosis levels, particularly in rare Mendelian disorders as illustrated in **Table 3**. It also marked the start of the international 1,000 genomes and ENCODE projects, energized by the cut costs of full genome sequencing using NGS platforms. NGS are also widely used in other fields such as transcriptomics, metagenomics, etc that are not detailed here.

Inheritance pattern	January, 2007	March, 2014
Autosomal	1,851	3,707
X-linked	169	284
Y-linked	2	4
Mitochondrial	26	27
Total	**2,048**	**4,022**

Table 3 (adapted from (Koboldt *et al.*, 2013)): Impact of NGS technologies on unraveling the molecular basis of OMIM-listed disorders

Full genomes

The first applications of such technologies were the decoding of entire genomes, particularly the human one since it took such efforts to assemble by classic Sanger sequencing. One of the first human genome sequenced by massively parallel sequencing was this of James D. Watson released in 2008, sequenced in two months on a 454 at a mean depth of coverage of 7.4X (Wheeler *et al.*, 2008). Many genomes from other species were subsequently assembled *de novo* (principally from bacteria and lower eukaryotes) thanks to those technologies (**Figure 9**).

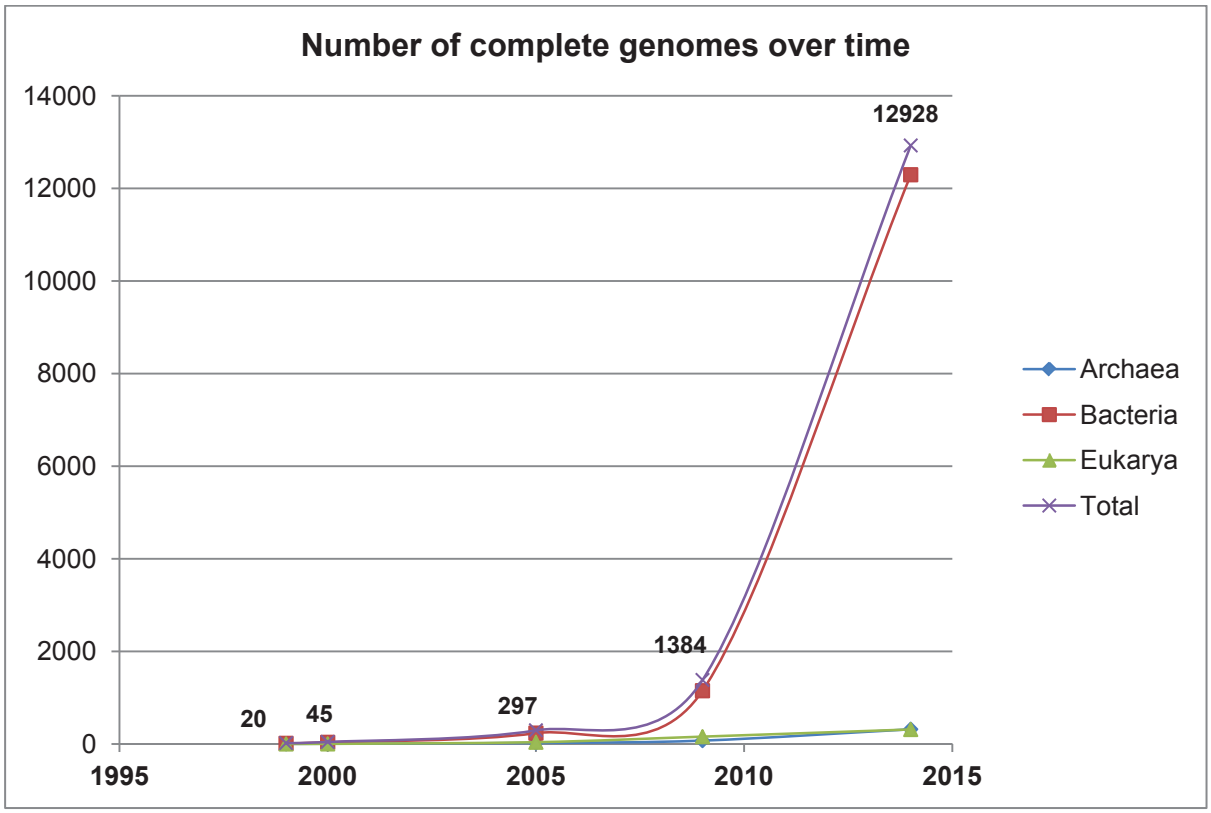


Figure 9: Evolution of complete published genomes, according to the three kingdoms of life

(source: Genome OnLine Database, GOLD; (Kyrpides, 1999; Bernal *et al.*, 2001; Liolios *et al.*, 2006; Liolios *et al.*, 2010)).

Exome and targeted sequencing

Although the first applications of second-generation sequencers in human genetics started with full genomes, it soon became obvious that full-genome sequencing was still prohibitively expensive, unsuited to the power of sequencing machines, not of enough quality (low depth of coverage), and a wealth of mostly variants with neutral or weak impact (or at least unready to interpret) on one's phenotype. Researchers were particularly interested in looking at specific regions of the genome and especially coding regions, since most genetic disorders are caused by alterations disrupting protein-coding sequences (Stenson *et al.*, 2014). Enrichment and amplification of DNA in selective targets became essential prior to proceed to the high-throughput sequencing step. By selectively retrieving specific regions of interest, costs and efforts are drastically reduced (many samples can be sequenced in a single sequencing lane), but also quality of the generated sequences is highly improved (higher depth of coverage).

The different enrichment approaches that have been developed can be classified into three different groups: enrichment by hybridization (NimbleGen and Agilent Technologies respectively), selective circularization methods (molecular inversion probes, MIP), and PCR-based approaches (RainDance and Fluidigm technologies; see **Figure 10** (Albert *et al.*, 2007; Okou *et al.*, 2007; Porreca *et al.*, 2007; Tewhey *et al.*, 2009)).

Hybridization methods consist in annealing fragmented gDNA with probes complementary of the sequences of interest, which will allow to specifically isolate the targeted regions from the rest of the input gDNA. Following several washing steps, unspecific DNA is eliminated and enriched DNA is amplified before sequencing. Nimblegen technologies use this approach, where the single-stranded complementary DNA baits are fixed on an array (**Figure 10, 1a**). Conversely, Agilent SureSelect Target Enrichment System uses in solution RNA baits, which are longer than those from Nimblegen and biotin-labeled so that the duplexes RNA baits/gDNA can be captured using streptavidin-coated beads (**Figure 10, 1b**). RNA baits are eventually digested.

The selective circularization technique is a different process, using oligonucleotides containing flanking regions of the target at each extremity and a universal core for all probes. The probe hybridizes to single-stranded fragmented gDNA and DNA polymerase is used to copy and circularize the targeted sequences (molecular inversion probes, MIP; **Figure 10, 2a**). Alternatively, the HaloPlex system from Agilent Technologies uses restriction enzymes to fragment the DNA, probe extremities are complementary of the restriction pattern, and the hybridized targeted DNA is replicated by circularization (**Figure 10, 2b)**.

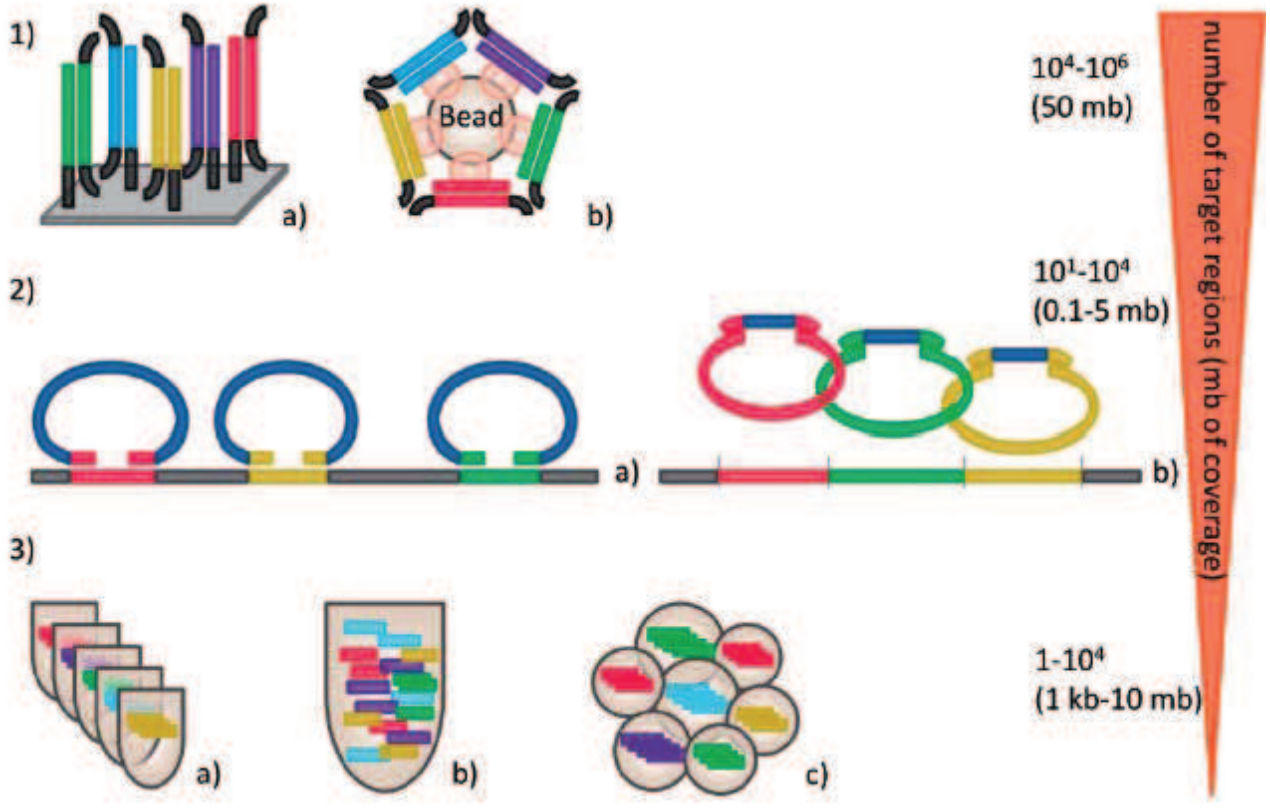


Figure 10 (from (Mertes *et al.*, 2011)): Overview of the different targeted enrichment techniques.

The third group of enrichment methods is based on PCR amplification of the target sequences either by multiplex PCR (multiple reactions in one tube) or multiple simplex PCRs reactions (multiple tubes with single reactions). Fluidigm technologies use the multiplex PCR

method and allow up to 480 PCR amplicons in one container (**Figure 10, 3b**). RainDance technologies use microfluidic devices to generate microdroplets containing primers and others containing fragmented gDNA and PCR reagents, which are fused together as a single droplet in which one specific PCR reaction takes place (**Figure 10, 3c**).

Any of those above methods can be used to target particular regions of interest. However, the different methods differ in enrichment specificity (proportion of sequences in-targets versus sequences off-targets), coverage homogeneity across all targeted regions, coverage homogeneity across different samples (*i.e*. reproducibility of the method), enrichment capacity (maximum size of total target regions), and associated time and costs of sample preparation. The method of enrichment has to be carefully adapted to the overall finality of a project.

The terminology of 'exome' was first proposed in 2009, following a study reporting the exome sequencing of 12 humans (Ng *et al.*, 2009), including 4 with Freeman-Sheldon syndrome and with known molecular diagnosis. This was the first proof of concept of using exome sequencing in a few unrelated affected individuals to unravel the genetic cause of a Mendelian disorder. A year later, a first study confirmed this proof of concept by uncovering the genetic cause of Miller syndrome, which was confirmed concomitantly by full genome sequencing (Ng *et al.*, 2010; Roach *et al.*, 2010). That was approximately the time I started my PhD. Since then, mutations in presumably a few hundred genes have been identified as the primary cause of genetic disorders using that same exome sequencing approach.

In parallel to exome sequencing, many teams developed projects of smaller-scale, first targeting and screening genes located in candidate regions for novel gene discovery (*e.g.* in deafness or Clericuzio-type poikiloderma with neutropenia (Rehman *et al.*, 2010; Volpi *et al.*, 2010)), then targeting and screening a set of genes implicated in disorders of a same family in large cohorts of patients (*e.g.* (Brownstein *et al.*, 2011; Simpson *et al.*, 2011)). When I started my PhD targeted sequencing projects had been exclusively published for the first approach*, i.e*. positional cloning.

NGS technologies have been widely used ever since for both molecular and clinical diagnosis (mainly through targeted sequencing) and novel gene discovery (mainly trough exome sequencing), becoming a two-in-one tool.

V- Problematic, aims of the project

In the early 2010s, as seen previously next generation sequencing technologies were promising a new era of massive and 'affordable' generating and analyzing DNA sequences. The first exome allowing to uncover the genetic cause of a Mendelian disorder reported in 2010 opened the way to many more (see **Figure 11**; (Ng *et al.*, 2010)).

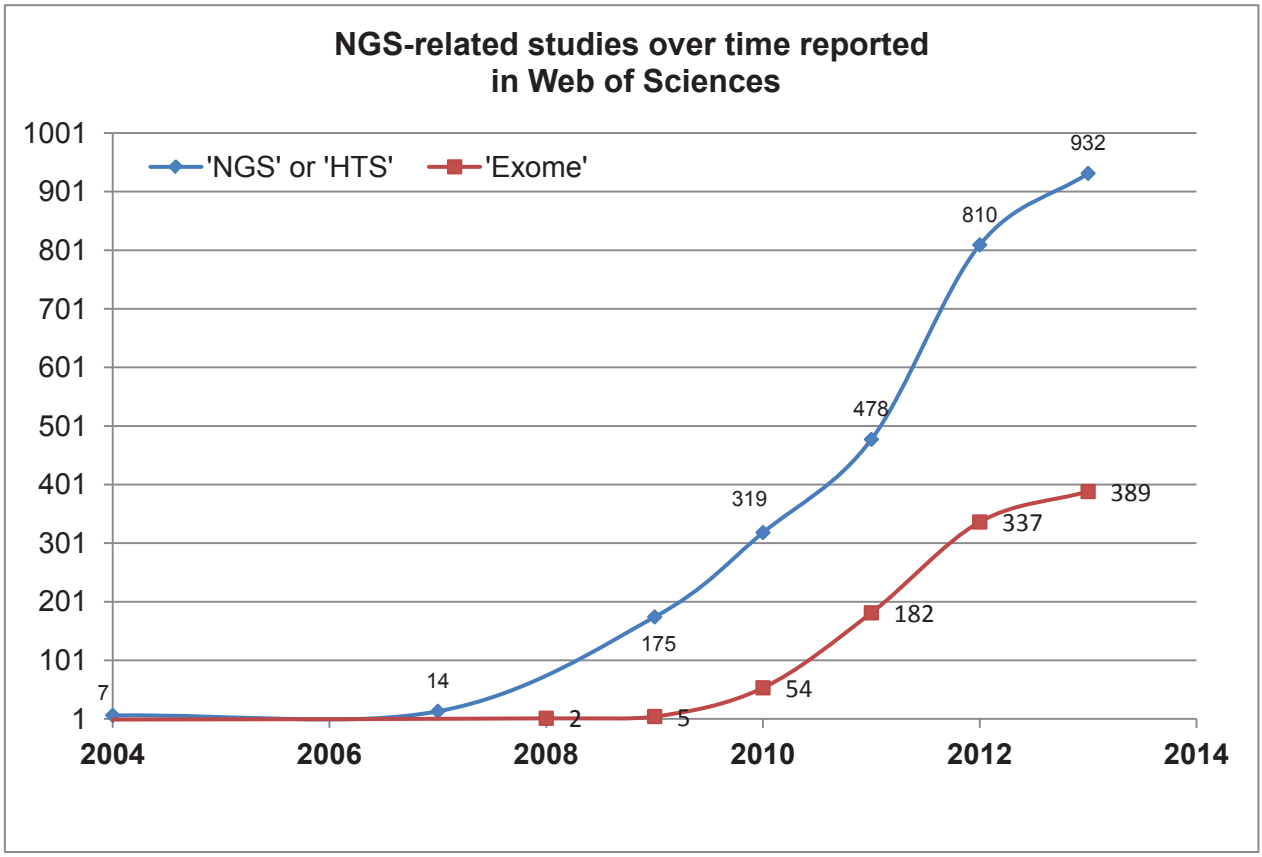


Figure 11: Evolution of the number of studies using next-generation sequencing and exome sequencing applications over time.

In this context, when I started my PhD in 2010, the goal was to make the most of those high-throughput sequencing technologies to develop an approach that would incorporate them in routine, in diagnostic settings. Specifically, in the context of highly genetically heterogeneous disorders the need was real for patients that had very few chances of otherwise getting molecular diagnostic. Such heterogeneity had been the topic of intense gene discovery is those disorders but was mostly a burden for developing efficient diagnostic strategies for the patients, by limiting considerably the diagnostic offer. A majority of patients was left without molecular diagnosis. Indeed, in highly genetically heterogeneous disorders the systematic screening of all coding exons from all involved genes by Sanger sequencing is simply prohibitive and would be a herculean task.

Through the study of three different disorders (or group of disorders) of increasing degree of complexity, we aimed at developing an alternative approach based on targeted high-throughput sequencing, which allows to screen simultaneously – and in a large number of patients – all coding regions from a set of genes by the technique of DNA enrichment. The three disorders of focus were

Bardet-Biedl syndrome, leukodystrophies and intellectual disability. All three harbor variable degree of genetic heterogeneity, proportion of patients that are diagnosed with mutations in known genes (*i.e.* correlated to the number of genes that are still left to uncover), and prevalence of the disease among the population (see **Table 4**).

As a proof of concept, the approach was first to be initiated and optimized on Bardet-Biedl syndrome, since the diagnostic laboratory of Strasbourg University Hospital along with the laboratory of Pr Dollfus is one of the reference centers for this disorder. Bardet-Biedl syndrome is a rare autosomal recessive disorder, which belongs to the family of ciliopathies and is relatively genetically heterogeneous: mutations have been reported in 19 different genes in patients. About 60% of the patients remain without molecular diagnosis when applying the current diagnostic method.

Once the approach would be optimized, the goal was to translate it on an even more genetically heterogeneous disorder: intellectual disability. Intellectual disability is a common neurodevelopmental disorder, highly genetically heterogeneous with over 300 genes that have been reported so far. A large majority of patients remain without molecular diagnosis subsequently to the current diagnostic strategy including array-CGH, fragile-X and a few single gene testing.

Lastly, we also aimed at applying a similar approach on leukodystrophies, which may be considered as a subclass of syndromic ID of mostly severe forms. About 50 genes have been implicated in leukodystrophies so far, and over 65% of the patients remain without molecular diagnosis.

Pathology	Incidence	# proposed genes	Estimated contribution in patients
Bardet-Biedl Syndrome	1/100’000	19	70-80%
Leukodystrophies	1/7’600	~50	?
Intellectual disability	1-3/100	>300	?

Table 4: Studied pathologies and associated properties revealing and increasing degree of complexity

The first aim of developing such approach was primarily directed towards improving the medical care of patients when applicable and proposing genetic counseling to families. Apart from that, we also expected our studies to give us more insights into the genetics of each of those three disorders by providing unbiased means to assess the contribution of each individual gene in the disease and maybe highlight some recurrent genes, and redefining the clinical spectrum associated with mutations in some particular genes.

GLOBAL MATERIAL & METHODS

In this section, I detail the general approach that we developed and implemented on three groups of disorders of interest: Bardet-Biedl syndrome and related ciliopathies (see **Results – Part I**), intellectual disability (see **Results – Part II**) and leukodystrophies (see **Results – Part III**).

I- Targeted sequencing

As overviewed in the general introduction, next-generation or high-throughput sequencing (NGS, HTS) seemed to be the ideal diagnostic tool to tackle the genetic heterogeneity in the monogenic forms of above-mentioned disorders. We chose to develop a strategy based on targeted sequencing of specific genes of interest. Indeed, although exome sequencing appears more exhaustive and applicable in a more universal manner whatever the disorder of interest, it remained significantly expensive when we originally started (in 2010) and would have allowed a much lower throughput of patients screened.

1. Gene selection and design

Every targeted sequencing project starts with the (complex) task of selecting the genes to include in the study. This step depends on the disorder of concern, but also mainly on the limitations related to the total size of targeted regions, restricted both by manufacturing price thresholds and by the power of the sequencing machine that will define the maximal number of patients to multiplex in a single sequencing lane. Once the list of genes of interest is complete, the next question to answer is what portion of the genes do we want to target: full genes, regulatory regions, all exons (including 5'UTR and 3'UTR), protein-coding exons only,… This much depends on the question to be answered: if one expects to face with large deletions and wants to map the breakpoints, then full genes would be useful. If mutations were already reported in the 5'UTR or 3'UTR regions then it could be of high value to add such regions as well. Otherwise, protein-coding regions (including intron/exon junctions) can be sufficient since most variants responsible for Mendelian disorders are known to disrupt protein-coding sequences.

The genomic positions of each coding exon (including alternative ones) were retrieved for each gene of interest through the UCSC table browser. Once a file containing all genomic positions in a bed format was generated, it was processed through a script to automatically add 20-bp within the introns for each region. Regions that were in close proximity (<150-pb away) were fused in a single region, and those that were of a size <150-bp were extended on both sides to reach the minimal size of 150-bp. Once this processed file was definitive, it was submitted to the web-service provided by Agilent (previously eArray https://earray.chem.agilent.com/earray/, now SureDesign https://earray.chem.agilent.com/suredesign/). The tiling frequency used was of 5x (*i.e.* number of independent probes covering each nucleotide of a targeted region; a probe being considered as an

orphan if it is >100-bp away from any neighboring probe). For the 'intellectual disability' design, we manually designed 120-mer baits for the regions that were left with no baits (mostly due to multimapping). Finally, for each design since there is a well-known burden of GC-content impacting on the hybridization efficiency the number of probe replicates was customized according to their GC-content (max. replicates for the group %GC>70, then lower replicates for the group 70>%GC>65, then lower replicates for the group 65>%GC>60, and then minimum replicates for the rest). The number of probe replicates was also increased for the group of orphan baits, 1st exon baits, and the one of custom baits for initially uncovered regions.

2. DNA samples quality control

Upon DNA samples arrival (extracted either from peripheral blood using QIASymphony from Qiagen® or saliva extracts using Oragene kits from DNAgenotek®), their integrity was checked on a 1%-agarose gel by electrophoresis. Quantification and further quality checking were performed using Nanodrop® ($A_{260/280}$ ratios > 1.8; $A_{260/230}$ ratios > 1.7, concentration >30ng/µL).

3. Library preparation

DNA samples (3-4 µg) were randomly fragmented to achieve a 200-400-bp distribution of the fragments using a Covaris E220 (duty cycle: 10%, intensity: 5, cycles per burst: 200, time: 300s). Sonication profiles were checked on a DNA-1000 chip using a Bioanalyzer (Agilent Technologies). Sequencing adapters were ligated on 500 ng of sheared DNA using the SPRIworks Fragment Library System I (Beckman Coulter; further size selection: 200-400-bp when using the GAIIX sequencer, 400-600-bp when using the HiSeq2500). Libraries were amplified by PCR (#cycles: 4 to 6) and quality was checked on a Bioanalyzer. The enrichment step was performed on individual DNAs using the in-solution SureSelect Target Enrichment System XT (Agilent) on 500 ng of DNA-prepped library. Multiplexing adaptors (SureSelect Target Enrichment System XT kit, Agilent) were added when performing the last PCR amplification step (#cycles: 12 to 14).

For the leukodystrophies and BBS projects, we added the multiplexing adapters along with the sequencing adapters directly when using the SPRIworks Fragment Library System I (Beckman Coulter). Even though the enrichment efficiency and general depth of coverage in targeted regions were lowered than compared to the post-capture multiplexing protocol (*i.e.* more reads located off-target), generated sequencing data was still passing our quality criteria. This allowed pooling two DNA samples within equimolar amount prior to the enrichment step, increasing by two-fold the number of samples that could theoretically be processed with this protocol (*e.g.* for one enrichment kit of 96 reactions, 192 could be processed with our custom protocol). Conversely, for the intellectual disability project both methods were compared again but the pre-capture multiplexing

protocol was proven to diminish dramatically the capture efficiency due to the larger size of target (1 Mb for the ID panel versus 188 kb for BBS and 281 kb for the leukodystrophies), leading to unsatisfying depth of coverage in some targeted regions.

4. Sequencing

8 pM of final mixed libraries were hybridized per one sequencing lane of the flowcell, which was then introduced in the cBot platform to perform cluster amplification. Sequencing runs of 72-bp single-read were then performed on a GAIIx (Illumina, until 2011) or of 110-bp paired-end on a HiSeq2500 (Illumina, 2011-2014).

II- Bioinformatic pipeline

The bioinformatic analysis can be divided into two major steps: the first one that allows going from the image acquisition and analysis to a list of variants detected in the sample when compared to a reference genome, and the subsequent step that takes this list of variants as an input, and allows annotating extensively those variants in order to facilitate the prioritization and ranking of candidate mutations.

1. Raw bioinformatic pipeline

This pipeline was developed by bioinformaticians of the IGBMC sequencing platform, mainly Stéphanie Le Gras, and Michaël Dumas.

Image analysis and base calling are performed using the pipeline provided by Illumina: Real-Time Analysis (RTA) v1.9. Barcode sequences and low-quality called-bases are trimmed from the reads, and low-quality reads are discarded. The remaining sequences are aligned on the human reference genome hg19/GRCh37 using BWA v.0.5.9 (Burrows-Wheeler 10/01/12 Aligner). Reads that mapped at more than one location and duplicate reads (same start and end positions, same strand) that could reflect PCR-duplicates are removed. Reads are then locally re-aligned, to improve the alignment quality (especially for indels).

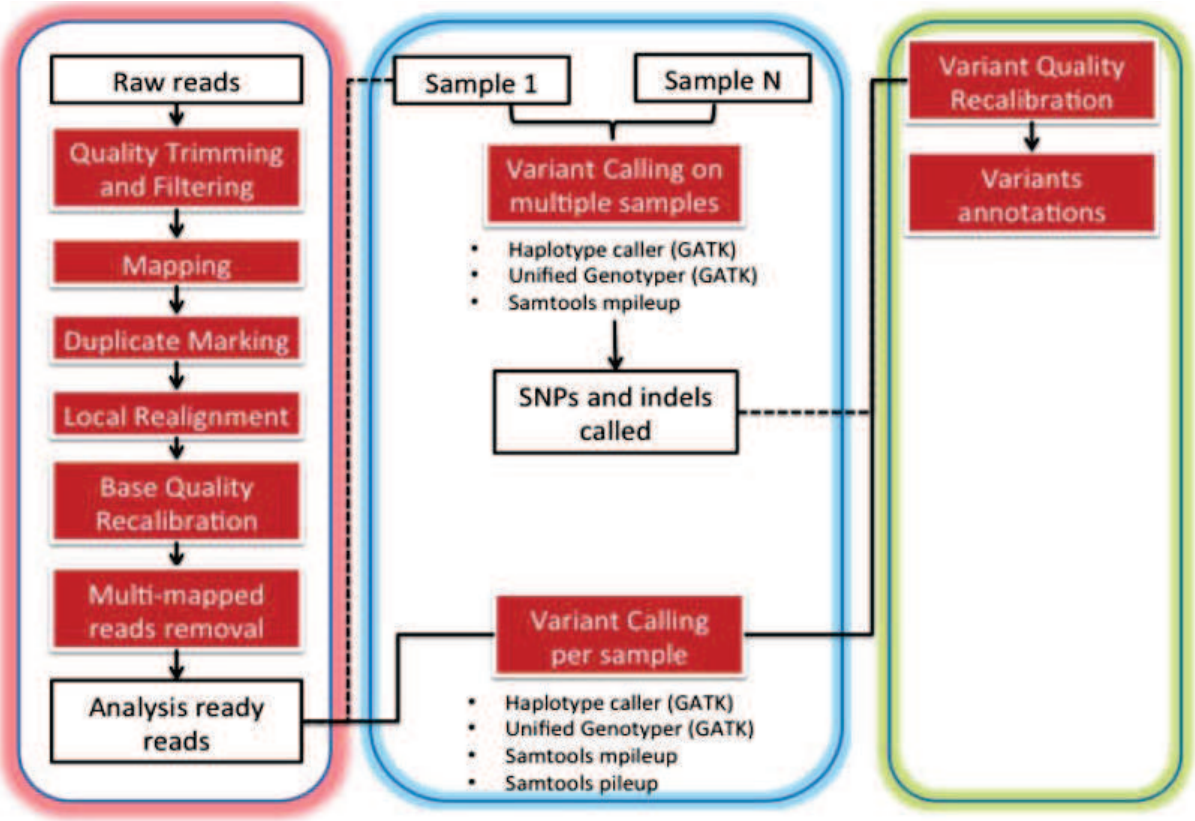


Figure 12: Improved version v2 of the bioinformatic pipeline for read mapping and variant calling (released in August, 2013) .

GATK: Genome Analysis Toolkit; SNPs: single nucleotide polymorphisms, indels: small insertion/deletions.

In the initial version of the pipeline, variant calling was performed using a single variant caller: Samtools pileup v0.1.7-4. The pipeline was improved (especially for the indel calling) by using in parallel four different variant callers (see **Figure 12**). An additional step of variant calling on multiple samples was also added, that performs the call on the data across multiple samples, increasing the probability to detect recurrent variants that are of low frequency (*i.e.* <15% of the total reads) in some samples and would be otherwise missed. It is of particular importance to detect sequencing artifacts that are present in multiple samples, or when performing trio-exome sequencing: when screening for *de novo* mutations, most candidate are in fact true positive, present at low-frequency in one parent but that was missed by the pipeline. This strategy allows reducing such false-positive *de novo* calls. The annotation of the detected variants is finally performed using Sequence Variant Analyzer (SVA).

2. Copy number variants/Structural variants detection pipeline

In parallel, a specific pipeline is run to detect larger events spanning from >50-bp to several hundred of kb (*i.e.* spanning several exons): deletions, duplications, translocations, inversions, insertion of repeated elements, etc. To achieve this, we used a depth-of-coverage comparison method. Once all samples processed in the same sequencing lane are normalized, the depth-of-coverage in non-overlapping windows of 20 nucleotides is compared between one sample and eight others selected randomly, using the Bioconductor package DE-seq. The fold change is then computed between the average depth of coverage the eight samples and the depth of coverage of the sample. When taking the log2 ratio of this fold change, hemizygous/homozygous deletions will be characterized by a high values >>1, heterozygous deletions by a value close to 1, heterozygous duplications by values close to -0.6 and triplications or double duplications (4 copies) by a value close to -1. Candidate SVs were recovered when log2ratio was either >0.8 or <-0.5, and if adjusted p-values were <0.1 (Benjamini and Hochberg adjustment procedure for multiple testing; (Benjamini and Hochberg, 1995)).

3. Variant annotation

Once the list of variant is generated, an extensive annotation is performed via the tool VaRank that was developed in-house by Cécile Pizot, Véronique Geoffroy and Jean Muller. This tool uses the Alamut-HT (Alamut-high throughput, Interactive Biosoftware) program to retrieve several pieces of information: putative effect of the variant at the protein level (Sift, Polyphen2, Mutation Taster), putative impact at the DNA level (splice site prediction scores: NNSplice, MaxEntScan, HumanSplicingFinder, other local splice effects), conservation cues (DNA level: phyloP and PhastCons, protein level: Grantham score), frequency of the variant in several generic

databases (dbSNP, 1000 genomes, Exome Variant Server). Importantly Alamut is compliant with the HGVS mutation nomenclature allowing an accurate nomenclature of variations (http://www.hgvs.org/mutnomen/). Lastly, VaRank also computes the frequency of each variant among the cohort of sequenced patients affected with a same disease, and indicates its allelic state in other patients.

4. Bioinformatic tools and databases

Properties of most bioinformatic tools and databases used are summarized in **Table 5**.

Impact at the DNA level

PhyloP and *PhastCons*

Both are methods to determine the level of conservation of a specific nucleotide and are freely provided for the entire human genome in the UCSC browser. PhyloP is based on the multiple alignments of 46 species, considering not only single nucleotide conservation but also the one of neighboring residues (Pollard *et al.*, 2010). The output value is a probability value, reflecting the confidence that the nucleotide belongs to a conserved region. Conversely, PhastCons is based on the conservation of the single nucleotide only, completely ignoring the neighboring effects (Siepel *et al.*, 2005). The output value ranges from -14.1 to 6.4, with positive values reflecting conservation (slower evolution of the site than expected under neutral drift) while negative values reflect fast-evolving sites (acceleration of the evolution: faster than expected).

Impact at the mRNA level: splicing prediction software

MaxEntScan, NNSplice, HumanSplicingFinder

Alamut provides five widely used splicing prediction software: the three detailed in this section plus GeneSplicer and SpliceSiteFinder-like. In VaRank, only three of them are used due to the observation that when performing a principal component analysis, we can observe that GeneSplicer and SpliceSiteFinder-like give highly redundant results to the three others due to similar underlying computational methods (see **Figure 13**). In order not to bias the results, they were not included in VaRank. MaxEntScan is based on the 'maximum entropy distribution' principle, whose parameters were estimated from available known signal sequences (Yeo and Burge, 2004). NNSplice is the improved version of Genie, with only genes with constraint consensus splice sites were considered in the training set (GT' for donor splice sites, AG' for acceptor splice sites; (Reese *et al.*, 1997)). Lastly, HumanSplicingFinder is based on position weight matrices assessing the strength of 5' and 3' splice sites and branch points, and was trained

with a dataset of 83 intronic and 35 exonic variants resulting in splicing defects (Desmet *et al.*, 2009). All three software give a score result reflecting the relative strength of acceptor/donor splice sites.

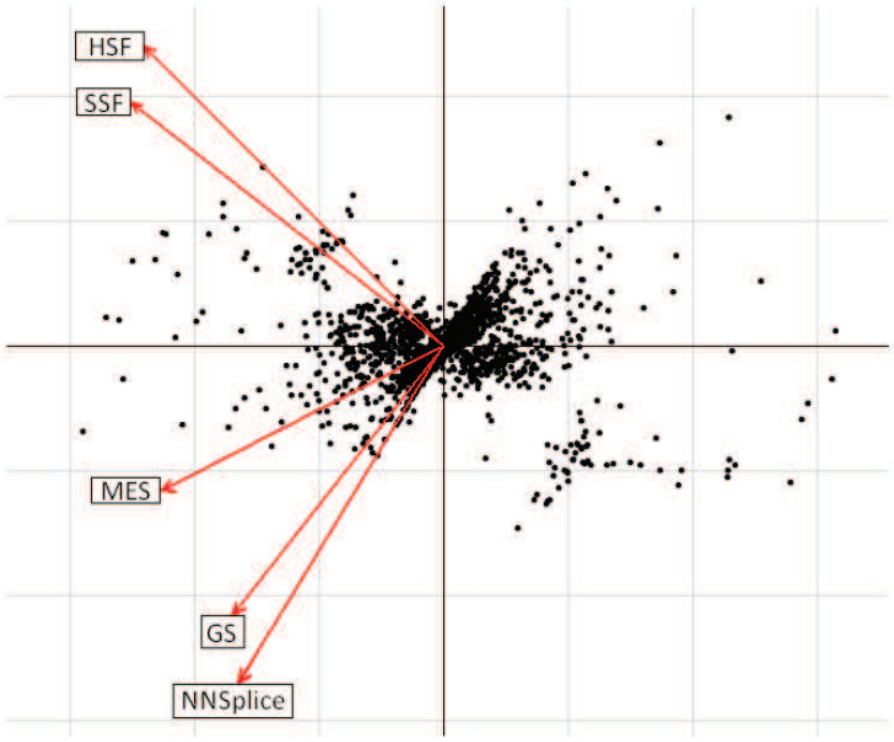


Figure 13: Principal component analysis performed on splicing predictions of 5 splicing prediction software in order to discriminate software with redundant results.

The analysis shows correlated results for Human Splicing Finder (HSF) and SpliceSiteFinder-like (SSF) on one hand (top panel), and for GeneSplicer and NNSplice on the other hand (low panel). MaxEntScan although giving results within the same trend as GS and NNSplice cannot be considered as fully redundant with those. source: Master's Thesis, Cécile Pizot.

Based on literature, we consider splice sites do be putatively affected when there is a decrease of over 10% in the strength of the splice site compared to the wild type in at least two prediction software (Houdayer *et al.*, 2012).

Impact at the protein level

Several attributes are considered to assess the pathogenicity of a missense variant: pairwise comparison of physicochemical characteristics of the wild-type versus the substituting residue, evolutionary constraints at the affected residue, comparison of the substituting residue and the evolutionary tolerated aminoacids in other species, and structural considerations. Depending on the prediction software, one or several of those attributes are used.

Regarding the evolutionary constrains, the pathogenicity can be assessed via two rules that are universally accepted (Tavtigian *et al.*, 2008):

- Missense variations affecting residues that are evolutionarily constrained are often pathogenic, whereas those falling at positions that are not constrained are often neutral or have minimal impact.

- Missense variations that are not observed among the cross-species range of variations at the affected position are often pathogenic, whereas substitutions falling within the cross-species range of variations are often neutral or have minimal impact.

Sift

Sift is a prediction software for the pathogenicity of missense variants (Ng and Henikoff, 2003; Kumar *et al.*, 2009). It is based only on protein alignments of highly similar sequences (>90%), with the following assumption: the more conserved the aminoacid position, the more important its function in the protein. Output score ranges from 0 to 1, variants with output scores below 0.05 are considered deleterious. A median info value is also given in the input (range: [0 ; 4.32]), reflecting the diversity of the sequences used in the alignment (the higher the more closely related the full set of sequences, hence the lower the confidence in the result since all positions will be highly conserved).

PolyPhen-2

PolyPhen-2 is the natural evolution of PolyPhen, predicting the impact of a missense variant on the structure and function of a protein using multiple alignments of homologous sequences, but also functional annotations and structural information when available (Adzhubei *et al.*, 2010). It uses the UniProtKB/Swiss-Prot database to see whether a specific protein feature might be affected by the variation. If a 3D structure is available, it checks whether the substitution site is in spatial contact with one of these critical features. Two different training datasets have been developed (HumDiv and HumVar). The HumDiv model (the one we used) is trained with all damaging alleles causing human Mendelian diseases reported in the UniProtKB database, together with differences between human proteins and their closely related mammalian homologs presumably innocuous. For each variation, PolyPhen-2 calculates a Bayesian probability that the mutation is damaging and gives an associate estimate of false-positive rates (FPR, *i.e.* chances that the mutation is classified as damaging when it is not). The HumDiv model uses the 5%/10% FPR cutoff to classify variants as probably/possibly damaging.

Integrated tools: all at once

Mutation Taster

Mutation Taster is an integrated tool that combines predictions for any alteration at the nucleotide level: nucleotide substitutions but also indels (Schwarz *et al.*, 2010). It uses as training set of over 390,000 known disease mutations from HGMD (Human Gene Mutation Database) and over 6,800,000 innocuous SNPs and indel polymorphisms from the 1,000 Genomes Project to predict the disease potential of any alteration. It is basically split into three Bayesian models: silent alteration ('without_aae'), single amino-acid substitution/insertion/deletion ('simple_aae') or larger amino-acid sequence modification ('complex_aae'). The output score is a probability value, reflecting the confidence level in the following prediction assumption (polymorphism or disease causing).

With the output score, it provides a plethora of information: conservation score based on PhyloP and PhastCons, splice site prediction using NNSplice, aminoacid conservation level using homologues of 10 other species, whether a protein feature is impacted (using the SwissProt protein features), the resulting length of the altered protein with corresponding position of the stop codon, the NMD probability (if the termination codon occurs >50 nucleotides upstream of the final intron/exon junction), the length of the CDS, etc.

Variant databases

dbSNP

dbSNP is one of the oldest polymorphism databases. It is freely available, provided by NCBI and aimed at serving as a central repository for both single nucleotide variations (SNV) and short indel polymorphisms to help discriminating them from pathogenic variations, in human but also in other model organisms (http://www.ncbi.nlm.nih.gov/SNP/). Alamut and as a result VaRank uses one of the most recent dbSNP build 137 (January, 2013). However, a major concern emerged with the latest versions of dbSNP: since anyone can now submit variants uncovered within NGS data, a lot of true potentially mutations (*i.e.* false-positive polymorphisms) lie in those datasets, which was not the case in the former versions. Moreover, the newly added field 'clinical significance' is supposed to help discriminating the real SNPs to the true mutations present in the databases, but is not curated properly. Several true mutations are not annotated as pathogenic while innocuous polymorphisms may be annotated as pathogenic (*e.g.* variant c.1292G>A, p.Arg431His in the autosomal recessive ID gene *PRODH*, whose MAF = 8.16% in exome variant server).

Tool	Use	Range scores	Pathogenicity criteria
Sift	Missense prediction	[0 ; 1]	≤0.05: deleterious
Polyphen2 (HumDiv)	Missense prediction	None	≤5% FPR: probably damaging; 5%<FPR≤10%: possibly damaging, >10% FPR: benign
Mutation Taster	Any nucleotide change prediction	[0 ; 1] (probability value)	Value close to 1 (*i.e.* high confidence on the prediction) along with a 'disease causing' prediction
MaxEntScan	Splice site prediction	[0 ; 12] (strength of splice site)	≥10% decrease
NNSplice	Splice site prediction	[0 ; 1] (strength of splice site)	≥10% decrease
HumanSplicingFinder	Splice site prediction	[0 ; 100] (strength of splice site)	≥10% decrease
PhyloP	Nucleotide conservation	[-14.1 ; 6.4]	≥2
PhastCons	Nucleotide conservation	[0 ; 1] (probability value)	≥0.9
Grantham	Amino-acid biophysical difference for missense	[0 ; 215]	≥100

Table 5: Summary of annotation tool incorporated through VaRank, and associated threshold for considering a variant as potentially pathogenic.

1000 Genomes

The 1,000 genomes project released at the end of 2012 an integrated map of genetic variations from over 1,000 genomes (http://www.1000genomes.org/, phase I release). The ultimate aim is to provide the genome of about 2,500 people from about 25 populations worldwide. During our analysis, the low frequency of candidate mutations in this genomic data is also verified.

Exome Variant Server

Exome variant server (EVS; http://evs.gs.washington.edu/EVS/) is another variant database, provided freely online by the National Health Lung and Blood Institute (NHLBI). It regroups the variant data from 6,503 exomes of individual with European American or African American origin. Since individuals were gathered for cardiac, lung and metabolic phenotypes, they can be considered as controls with respect to cognitive function (especially when considering that they had to sign consent forms prior their enrollment in the study). This repository does not include only allelic

frequencies of associated variants but also provides genotypes of the individuals, which is of particular relevance when studying X-linked disorders. It contains additional information such as coverage data, conservation scores (GERP, PhastCons), predictions for missense variants (PolyPhen-2), whether it was reported as a hit in GWAS, the clinical link if any, etc. Most our results are filtered using those data, requiring a variant to be either absent of with low frequency in such database.

III- Data analysis

Before starting analyzing the data, one needs to set up a hypothesis concerning the type of genetic disorder of concern. In the most simple single-gene disorders, the phenotype is caused by mutations (located on one or both alleles) in a single nuclear or mitochondrial gene. In polygenic disorders, the additive effect of mutations dispersed in several genes is responsible for the phenotype. Such hypothesis is hard to prove. Lastly, in multifactorial disorders, the combination of environmental factors with one or several predisposing genetic risks leads to the disease phenotype. Of note, chromosomal anomalies affecting several genes can fall in either categories, depending if a single gene is the driver of the disease phenotype or not. In our case, we were mainly focusing on mutations with substantial effect with full penetrance, aka monogenic disorders (without however ruling out the polygenic hypothesis, since it seems reasonable for neurodevelopmental disorders, and triallelism has been proposed in Bardet-Biedl syndrome).

1. Variant prioritization

Once this extensively annotated list of variants (SNVs and indels) is generated, it has to be filtered down to highlight the best candidate as responsible for the phenotype of the patient. Our primary filtering criteria is based on the quality of the variant: the corresponding base has to be covered by at least 10 independent sequences/sequencing reads, in which the variant shall be seen in a minimum of 15% of all such sequences. A variant is considered as presumably heterozygous if reported in 15 to 80% of all sequences, homozygous if in over than 80% (such assumption shall not prevent manual checking of the variant to verify the coherence with sequencing data).

Then, variants are filtered out if there are overrepresented (inconsistently with the frequency of the disease) in the Exome Variant Server (*e.g.* for ID, a maximum of 1 heterozygous carrier or hemizygous male for autosomal dominant/X-linked disorders, and 50 carriers for autosomal recessive disorders, even stricter parameters for leukodystrophies or BBS that are less frequent in the population).

Finally, the variants that are too frequent in the cohort of sequenced patients (incompatible with the frequency of the most recurrent mutation in the disease) are also filtered out. Within all remaining variants, the ones that are predicted pathogenic by prediction software (Sift and PolyPhen-2 for missense variants, two out of the three splicing prediction software for putative splice variants) are kept as candidate for further analysis. Among those, the first step is to manually check the quality of the variant (*i.e.* shall not be observed in other patients, see below).

Then, the allelic state of the variant is confronted with the mode of inheritance associated with the affected gene. If a single heterozygous mutation is detected in a gene associated with recessive

mode of inheritance, the entire gene is checked manually in IGV to ensure that no other potentially pathogenic event was missed by both the SNVs/indels and the CNVs/SVs pipelines.

For putative CNVs/SVs, all candidate regions are checked manually to exclude false positive calls, and those that appear as true positives are processed to the validation step if they fit with the associated mode of inheritance of the affected gene. The manual checking allows verifying the allelic frequency of other SNPs in the region to see if they match with the putative event (*e.g.* all SNPs shall appear as homozygous in a region that is proposed to be deleted at the heterozygous state, conversely in a putative duplicated region all SNPs shall have an allelic frequency of about 33%/66%). It also allows checking whether the size of the proposed abnormality is underestimated or not (*e.g.* do the neighboring exons appear affected?).

2. Data visualization

Once interesting candidates are identified, the quality of such variants is always manually checked in a genome browser such as IGV). This essential step is the equivalent of the necessary checking of electropherograms for Sanger sequencing. This gives a general overview of the sequencing quality in the region, and tells whether this is also observed in other patients of the cohort (see **Figure 14**).

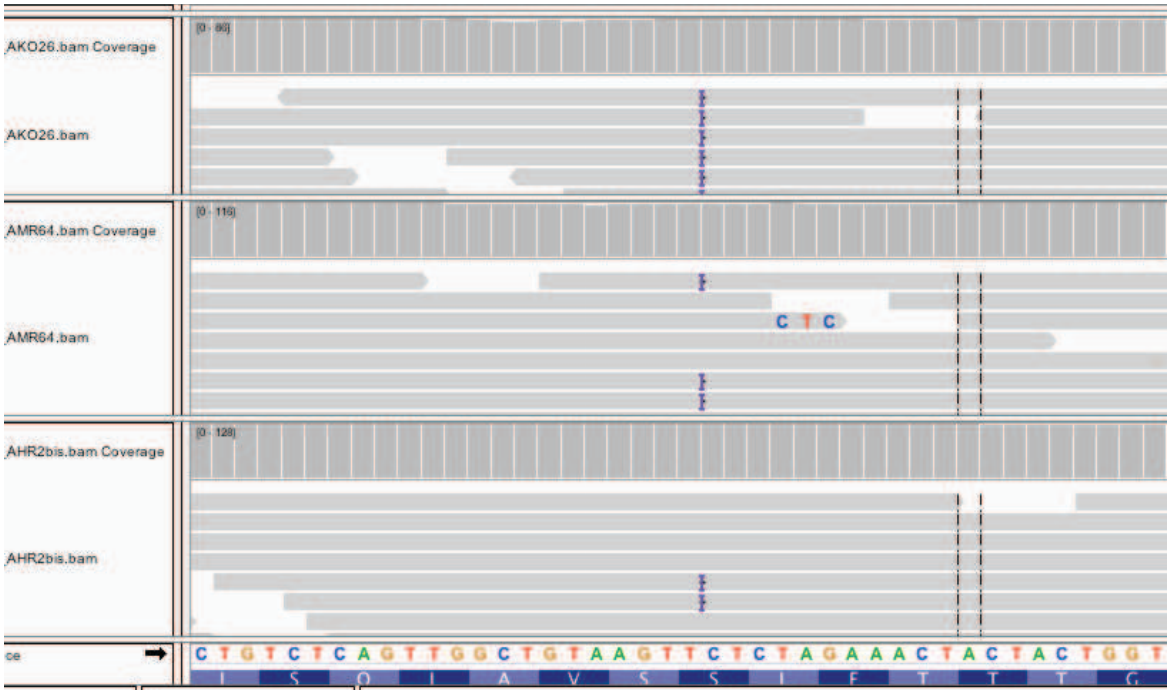


Figure 14: Visualization of sequencing data (zoomed on *ALMS1*, exon #8) in IGV for three different BBS patients.

One grey arrow represents one read of 72-bp. Nucleotide and corresponding aminoacid sequences are indicated for the zoomed region. Purple vertical bars indicate the insertion of several nucleotides. In this case, the CTC insertion (c.1577_1579dup, p.Pro526Dup) is present in three different patients, diminishing its credibility for being pathogenic.

IV- Validation, diagnosis

If both the allelic state of the candidate variant and the associated mode of inheritance coincide, the variant is validated by Sanger sequencing (for SNVs, indels, large homozygous/hemizygous deletions), or qPCR (for large heterozygous deletions/duplications or homozygous duplications). Putative splicing mutations were confirmed by RT-PCR on patients' fibroblasts when available. The clinical data of the patient is screened, and extensive literature search is performed on the affected gene. Subsequent co-segregation analyses were performed on family members as extensively as possible. Finally, the variant was considered as a certainly pathogenic mutation when it co-segregated with the disease status in the family and either: 1- was truncating (nonsense, frameshift, splice site mutation, large deletions/duplications/etc), 2- was a missense that had been already convincingly reported or 3- was a new missense but that we further confirmed by functional studies. All the other missense variants with potentially pathogenic effect according to predictions were transiently considered as potentially pathogenic, waiting to be further confirmed as mutations or excluded (by additional co-segregation analysis of functional studies).

All certainly-pathogenic mutations were then reported to the diagnostic laboratory of Strasbourg University Hospital, to ascertain official diagnosis reports to the clinicians that then relayed the information to the patients and their families.

RESULTS - PART I

TARGETED-SEQUENCING FOR THE DIAGNOSIS OF BARDET-BIEDL SYNDROME

I- Introduction - Cilia, this omnipresent organelle

1. Definition

Cilia were first described in the 19th century – if not earlier – coinciding with the descriptions of eukaryotic cell structure (Zimmerman, 1898). Later observations were facilitated by the development of electron microscopy. However, the utility of cilia was debatable for long, widely thought to be just an evolutionary vestige. Cilium arose early in the eukaryotic evolution, and was lost from several groups leaving it a dispensable organelle for some species (Cavalier-Smith, 2002) while necessary protuberance with essential motile and sensory functions in others. Cilia are found in specific cell types in invertebrates (mostly sensory neurons, (Wheatley *et al.*, 1996)) while almost ubiquitously present in vertebrate cells. It is undistinguishable in structure and composition from the flagellum, abundant in sperm and unicellular eukaryotes whose role is primarily cell motility. In epithelial cells, motile cilia act in symphony to orchestrate extracellular fluid movement (Sleigh, 1974). Primary cilia usually project as single immotile organelles from nearly all cell surfaces (Wheatley *et al.*, 1996), whereas motile cilia are most frequently concentrated in clusters on multi-ciliated cells (*e.g.* trachea cells). The role of primary – or non-motile – cilia is however more obscure, but it would act as a sensor for physiological, chemical and mechanical cues (light, odor and fluid flow in vertebrates, osmolarity changes, chemo-attractants/repellants and sound in invertebrates; (Evans *et al.*, 2006)). Cilia overall appear to play a crucial role in regulating vertebrate development and homeostasis.

2. Structure

Cilia are essentially composed of an association of microtubules forming the cytoskeletal structure called ‘axoneme’, and surrounded by a plasma membrane extending from the cell surface into the extracellular space but with a distinct repertoire of membrane and channel proteins. The axoneme is a highly structured backbone, composed of 9 microtubule doublets circling a central core containing (motile cilia, possess dynein arms) or not (primary cilia, lack dynein arms) a tenth microtubule doublet. There exist a few exceptions of non-motile 9+2 (kinocilium of hair cells) or motile 9+0 (nodal cilia in the embryo) cilia in human (see **Figure 15**; (Fliegauf *et al.*, 2007)). Ciliary motility is accomplished by dynein motor activity (motor complexes which produce the force needed for bending) transiting from one microtubule doublet to another. The cilium emerges from and is anchored to the basal body, containing 9 microtubule triplets and serves as a nucleation site for the growth of the axoneme microtubules. This microtubule-organization center (MTOC) derives from the older pair of centrioles. The transition zone is essentially composed of transition fibers, and marks the junction of the basal body and the axoneme. It is proposed to function as a

diffusion barrier, regulating the entry of molecules towards or onwards the cilium. Most intraflagellar transport (IFT) and cargo proteins have been shown to accumulate in this zone.

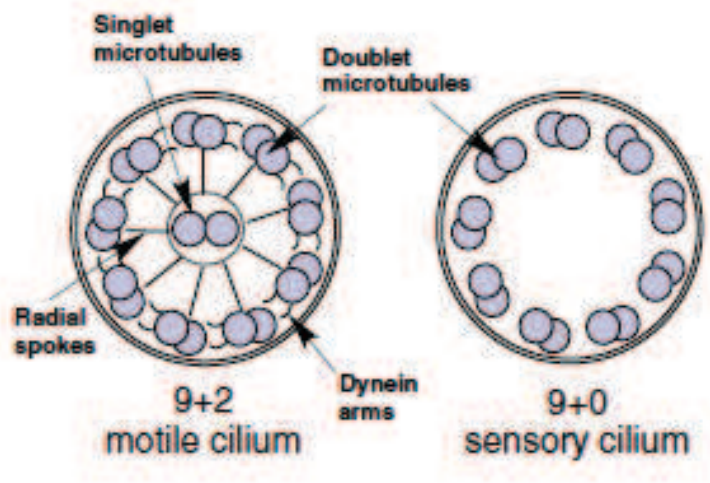


Figure 15 (from (Bisgrove and Yost, 2006)): 9+0 and 9+2 axoneme structure.

Cross-section of the 9+2 and 9+0 cilium. Cilia are broadly divided into 2 types based on the presence or absence of a central pair of microtubule singlets in the axoneme. Inner and outer dynein arms, which are usually associated to 9+2 cilia, can be present in either type of cilium and ensure ciliary motility.

3. Ciliary biogenesis

Ciliogenesis is synchronized with the cell cycle and occurs primarily during the G0-G1 phase. Cilia assembly begins with the nucleation and growth of the axonemes from the template that are basal bodies, which are formed from modified centrioles. Their assembly is thus highly influenced by the maturation and orientation of the basal body (Ishikawa and Marshall, 2011). Cilia growth starts emerging from the transition zone, but further elongation is restricted to the distal extremity of the cilia involving the recruitment of building blocks (proteins and metabolites) mediated through the bidirectional transit of two IFT complexes (IFT-B and IFT-A) to and from the cilia (Ishikawa and Marshall, 2011). Some protein complexes act as molecular motors for the transportation of those IFT macromolecular rafts, such as kinesin-2 involved in anterograde movement (from the basal body to the tip of the cilium) as opposed to cytoplasmic dynein-2 involved in retrograde transport (from the tip to the basal body, summarized in **Figure 16**; (Rosenbaum and Witman, 2002)). The axoneme elongation and the ciliary membrane development occur concurrently independently from the cell body, and this process is called compartmentalized ciliogenesis (Avidor-Reiss *et al.*, 2004). Primary cilia are dynamically assembled and disassembled throughout cell cycle.

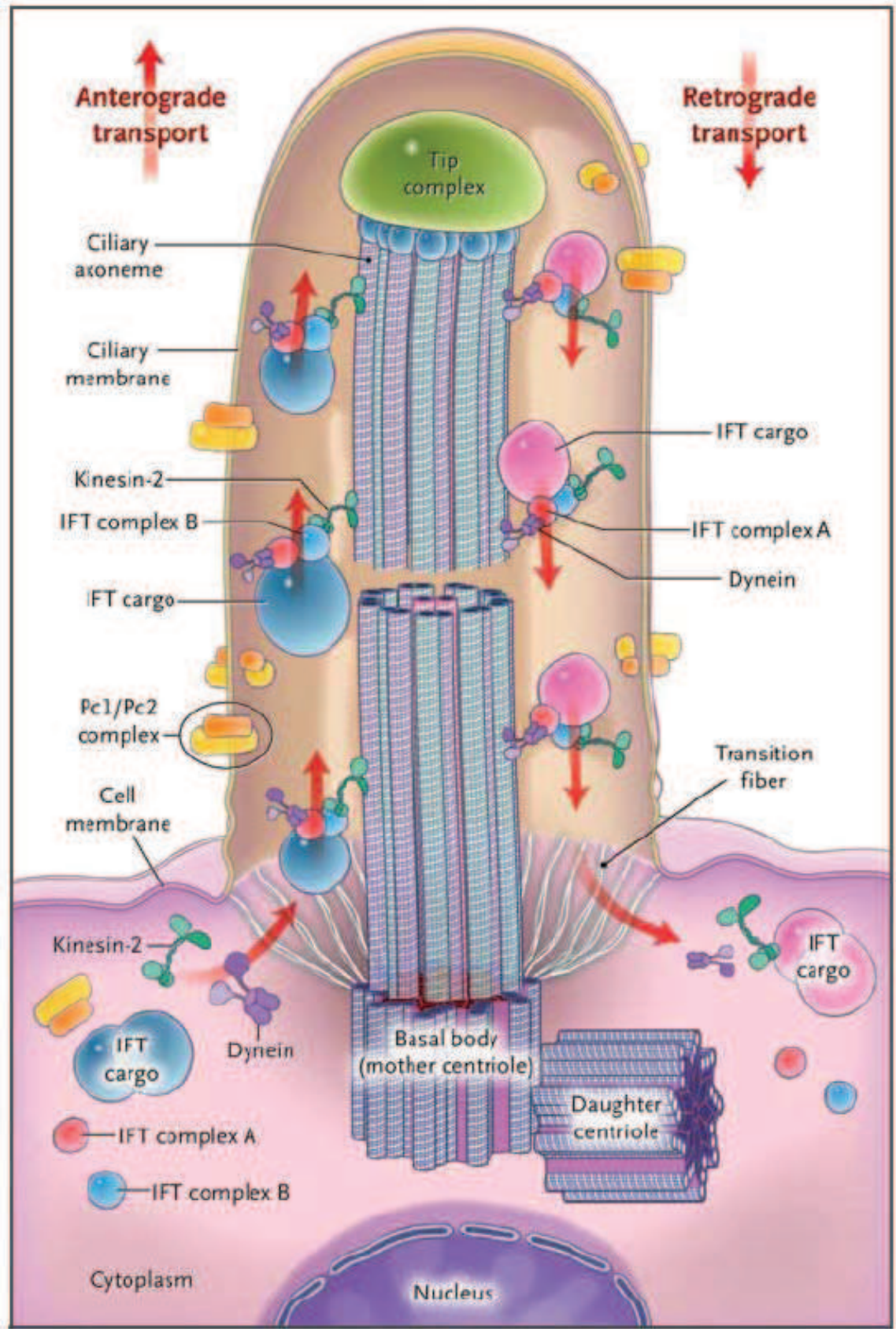


Figure 16 (from (Hildebrandt *et al.*, 2011)): Ciliary structure and components.

The cilium is a hair-like protuberance at the cell surface of nearly all vertebrate cells, which consists of a microtubule-based axoneme covered by a specialized plasma membrane. It is assembled from the basal body, or mother centrioles. Transition fibers act as a filter for molecules passing into or out of the cilium.

Axonemal and membrane components are transported in raft macromolecular particles (complexes A and B) by means of IFT along the axoneme toward the tip of the cilium by kinesin-2. Retrograde transport occurs by means of the motor protein cytoplasmic dynein.

4. Cilia: preponderant modulators of key signaling pathways

Shh

Primary cilia act as a signaling and transducing platform ensuring the regulation of both upstream and downstream elements of the Hedgehog (Hh) cascade, a key regulator of proper embryonic development. The main targets of Hh signaling are the glioma (GLI) transcription factors GLI1-3. In the absence of Hh, patched-1 sequesters smoothened which is predominantly on intracellular vesicles, and GLI3 is constantly cleaved into the GLI3 repressor form (GLI3R) at the ciliary tip (reviewed in (Hui and Angers, 2011)). GLI3R is then transported via dynein down the axoneme to the nucleus, binding Hh regulatory elements and blocking the transcription of downstream targets (Haycraft *et al.*, 2005). In the presence of Sonic Hedgehog (Shh, the best-known ligand for the Hh pathway), smoothened is released by patched-1, and is transported via IFT cargo at the tip of the cilium. There, it interacts with several GLI proteins preventing its proteolysis into GLI3R and promoting the formation of a GLI transcription activator (GLI3A). The active GLI3A form is subsequently transported to the nucleus down the axoneme where it activates Hh-dependent transcription.

Wnt

Primary cilia also participate in the regulation of both canonical and non-canonical pathways, respectively implicated in general regulation of gene transcription and planar cell polarity (PCP) pathway regulating cell cytoskeleton. Some proteins located at the transition zone such as the jouberin (implicated in Joubert Syndrome) were shown to prevent the phosphorylation-dependent degradation of β-catenin and mediate its translocation to the nucleus upon WNT binding, activating transcription of WNT target genes. Such stabilization is mediated by the binding of WNT to its frizzled receptor, causing the activation of cytoplasmic Disheveled that in turn inhibits the complex promoting β-catenin degradation. Other ciliary proteins are involved in the sequestration of both β-catenin and jouberin at the tip of the cilium allowing to modulate the Wnt pathway responsiveness (Lancaster *et al.*, 2011). In PCP, in the presence of fluid flow, intracellular calcium release through the cilium promotes the expression of inversin (NPHP2), which acts as a molecular switch between canonical and non-canonical Wnt pathways. Wnt binds to frizzled, recruiting and activating Disheveled at the plasma membrane, resulting in cytoskeleton regulation and remodeling through RhoA and Rho-kinase activation (Habas *et al.*, 2003).

5. Implication in disease: 'ciliopathies'

Ciliopathies regroup disorders associated with genetic mutations occurring in genes encoding for ciliary-related proteins, resulting in abnormal function or structure of the cilia. All those disorders affect a protein of the ciliary proteome, (over 1,000 polypeptides identified to date; (Gherman *et al.*, 2006; Inglis *et al.*, 2006)) and include polycystic kidney disease (PKD), nephronophthisis (NPHP), Alström Syndrome (ALMS), Joubert syndrome (JBTS), Meckel syndrome (MKS), Senior-Loken syndrome (SLSN) or Leber congenital amaurosis (LCA). Due to the ubiquitous presence of cilia in vertebrate (and hence human) cells (see **Figure 17**), ciliary dysfunction may manifest via a large range of phenotypes, from organ-specific to broadly pleiotropic disorders. Such features generally include retinal degeneration, renal anomalies and cerebral defects, and less consistently diabetes, liver diseases, obesity and skeletal dysplasias (**Table 6**).

Ciliary Function	Disease phenotype
Nodal flow	Heterotaxy (Nonaka *et al.*, 1998; Yost, 2003)
Photoreception	Retinal Degeneration (Li *et al.*, 1996; Liu *et al.*, 2004)
Odorant reception	Anosmia (Kulaga *et al.*, 2004)
Mechanosensation	Polycystic kidney disease (Nauli *et al.*, 2003)
GLI repressor formation	Polydactyly, neural patterning defects (Liu *et al.*, 2005)
GLI activator formation	Neural patterning defects (Huangfu *et al.*, 2003; Liu *et al.*, 2004; Corbit *et al.*, 2005)
Convergent extension	Neural tube closure defects (Ross *et al.*, 2005; Simons *et al.*, 2005; Kyttala *et al.*, 2006; Park *et al.*, 2006)

Table 6 (from (Singla and Reiter, 2006)): Defects in ciliary functions cause several human diseases.

Defects in motile cilia

The first cilia-related disorder to be reported was associated to defects of motile cilia, through the study of infertile individuals with immotile sperm whose flagellum exhibited absence of dynein-arms (Afzelius, 1976; Eliasson *et al.*, 1977). This condition belonged to the large family of primary ciliary dyskinesia (CILD), a highly genetically heterogeneous group of disorders (CILD 1-28 to date) characterized in patients by respiratory defects, infertility, anosmia and uneven *situs inversus*. Rare manifestations include hydrocephalus, retinal degeneration and cystic kidney disease. Clinical features in CILD perfectly reflect the distribution of motile ciliated cells throughout affected organs and tissues. Ciliary defects in the respiratory tract result in impaired mucociliary transport manifesting as repetitive respiratory infections. Male infertility is caused by loss of ciliary movement in the efferent duct or loss of sperm flagellar motility leading to failure of immature sperm to reach the vas deferens. *Situs inversus* or heterotaxy results from absence of nodal ciliary laminar leftward flow during embryogenesis required for normal visceral left/right

patterning, hence leaving random chances for definitive (normal or reversed) symmetry (El Zein *et al.*, 2003). Lastly, hydrocephalus was proposed to result from defective laminar flow of cerebrospinal fluid in the brain ventricles and through the cerebral aqueduct, owing to loss of ependymal cilia motility (Eley *et al.*, 2005).

Defects in immotile cilia

Contrary to disorders affecting primarily motile cilia, defects in non-motile (or sensory or primary) cilia lead to a wider range of manifestations: cystic kidney diseases, liver fibrosis, anosmia, retinal degeneration, obesity, diabetes, skeletal anomalies, hearing loss, cognitive defects and hypogonadism.

The underlying mechanisms linking cystic kidney diseases (from polycystic disease, renal medullary cystic diseases, to cystic renal dysplasias) to primary cilia are still poorly understood. Some proteins defective in such disorders are located at renal primary cilia and were shown to mediate planar cell polarity (PCP) ensuring planar organization of epithelium and correct mitotic spindle orientation and hence normal diameter of tubular structures such as renal ducts by aligning mitotic orientation of proliferating cells (Masyuk *et al.*, 2008). One hypothesis is that loss of PCP signaling would lead to renal cystogenesis.

Cholangiocyte cilia regulate bile formation via mechano, osmo and chemosensory cues. Structural defects of such cilia and/or of their transduction signaling result in decrease intracellular calcium and increased cAMP, causing cholangiocyte hyperproliferation, abnormal cell matrix interactions and altered fluid secretion/absorption that eventually lead to hepatic cystogenesis.

Anosmia can be explained by the trapping of some olfactory proteins in dendrites and cell bodies of olfactory receptor neurons, a thinner olfactory ciliary layer and unstable microtubules, resulting in at most shortened axonemal cilia preventing efficient transduction signal of sensed olfactory stimuli (Kulaga *et al.*, 2004).

Retinal dystrophy is caused by degeneration of retinal photoreceptors. Maintenance of photoreceptors integrity depends on continuous IFT, ensuring the transit of essential molecules (arrestin, transducin, opsin) from the inner segment via the connecting cilium to the outer segment, containing the membrane stacks itself containing the photo-pigment. Reduced anterograde transport of rhodopsin across connecting cilium of photoreceptors link to rhodopsin accumulation in the cell body, engendering cell death (Nishimura *et al.*, 2004). In ciliopathies, several photoreceptor structural defects have been reported, from absence of outer segments, disorganized outer segments to simple photoreceptor degeneration without obvious abnormalities in their morphology.

Leptin is a protein expressed in adipocytes, proportionally to fat mass, acting in the CNS to promote weight loss by reducing food intake and increasing energy expenditure. Surprisingly,

hyperleptinemia was reported in most patients with Bardet-Biedl syndrome (Rahmouni *et al.*, 2008), but with associated leptin resistance caused by the inability of leptin to activate the downstream intracellular machinery associated with its receptor (Bates *et al.*, 2003).

It was shown that disregulated Hh pathway (particularly increased Hh signaling through overexpression of GLI2) leads to impaired non-ciliated b-cells function and massive insulin secretion resulting in glucose intolerance in transgenic mice (Landsman *et al.*, 2011). Altered Hh pathway is most probably at the origin of the diabetes reported in some patients, but when concomitant might as well be the direct outcome of obesity, suggested to be the primary cause of type 2 diabetes.

Hedgehog pathway is one of the key regulatory networks involved in animal development (Jiang and Hui, 2008). In vertebrates, in the case of embryonic limb bud development, Shh is expressed in the posterior limb bud mesenchyme and plays a preponderant role in anteroposterior patterning of the limbs (Benazet and Zeller, 2009). Loss-of-function mutations to SHH result in digit loss, while mutations in GLI3 cause polydactyly (McGlinn and Tabin, 2006). Defective ciliary proteins result in Shh cascade response (especially impaired processing of GLI3), directly impacting on mammalian development (Zhang *et al.*, 2012).

Hearing require integrity of sensory hair cells of the inner ear that carry multiple actin-based stereocilia and one single tubulin-based kinocilium. Auditory stimuli are thought to be transduced via mechanically gated ion channels converting vibrations into electrical signals by depolarization of the hair cell (Kim *et al.*, 2003). In affected patients, the defective ciliary proteins may prevent such depolarization.

Some manifestations include malformed external genitalia such as micropenis, undescended testes and hypospadias in males, a recapitulated phenotype in Shh-null mice (Haraguchi *et al.*, 2001), demonstrating that Shh also acts to pattern genital development. Genital and limb phenotypes seen in some ciliopathies may share a similar etiology caused by aberrant Shh signaling. Reduced fertility is also reported in females with CILD, possibly due to a defective transport of the egg traditionally ensured by a complex interaction between ovarian duct ciliated epithelium, muscle contractions and flow of tubal secretions.

As cilia play a crucial role in cell-cycle regulation that is responsible for the coordination of cancer-related signaling molecules (GLI proteins), altered ciliary functions may lead to repressed or stimulated tumorigenesis depending on the context (Han *et al.*, 2009; Wong *et al.*, 2009).

Cilia dysfunction and neurodevelopmental disorders

Neurological manifestations are often observed in patients affected with some types of ciliopathies, suggesting an important role of cilia in brain functioning and/or development. First reports of

primary neuronal cilia dated back to 50 years ago. The role of primary cilia in SHH signaling appears to be one of the most plausible explanations for cerebral manifestations, due to its preponderant involvement in neuronal specification and cerebellar growth and patterning. Mice presenting with defective primary cilia in cells of the developing CNS exhibit decreased SHH signaling coupled with cerebellar hypoplasia and defects in foliation (Chizhikov *et al.*, 2007; Spassky *et al.*, 2008)

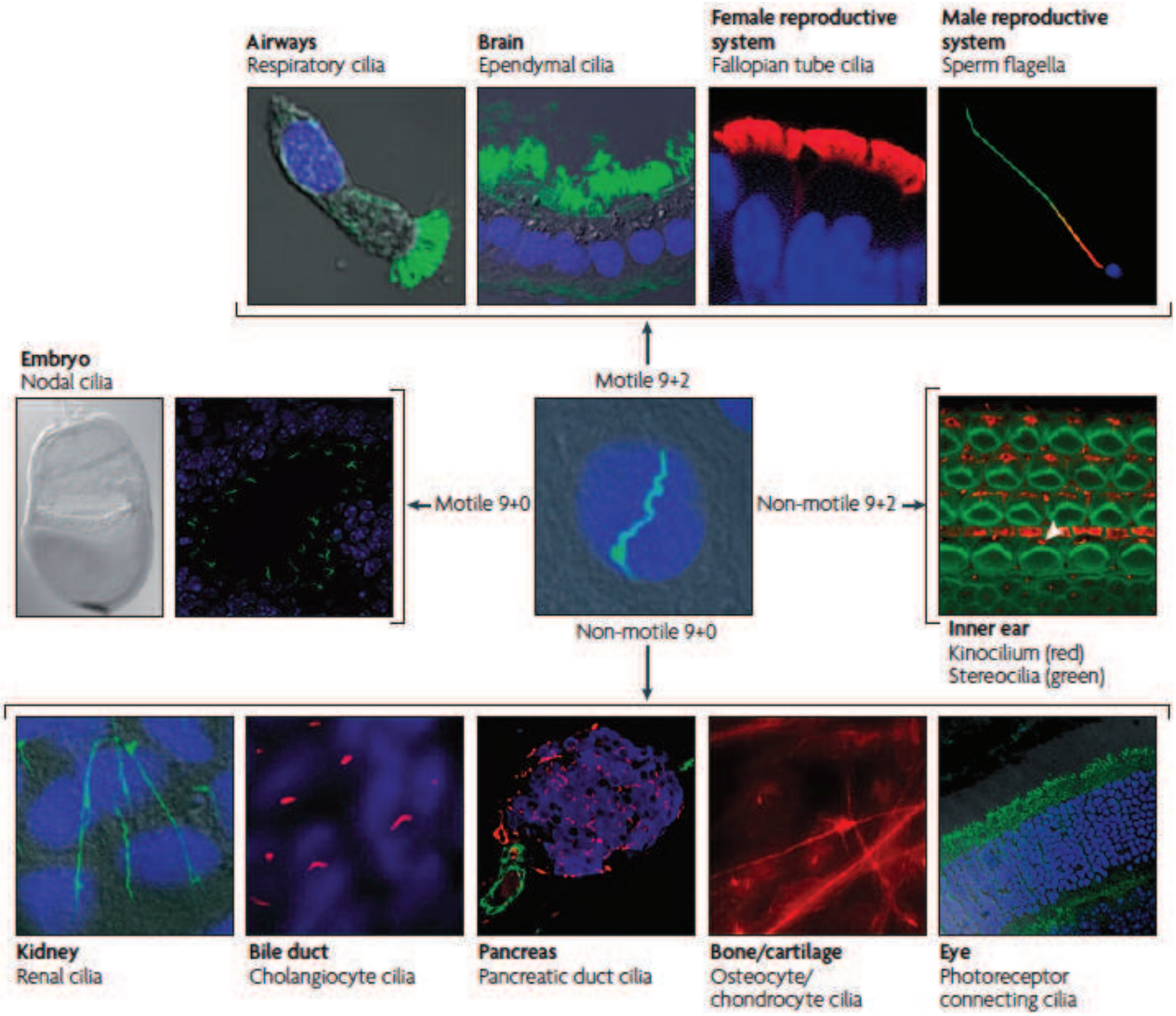


Figure 17 (from (Fliegauf *et al.*, 2007)): Ciliary dysfunction in human disease: why are those organs/tissues affected?

Distribution of all types of ciliated cells within vertebrate organs and tissues. Motile monocilia (9+0 axoneme) are found at the embryonic node, multiple motile cilia (9+2 axoneme) are located on respiratory epithelial cells, brain ependymal cells and epithelial cells lining the fallopian tubes. The sperm flagellum (co-stained with antibodies against the dynein heavy chain in red) represents a single elongated motile cilium (9+2). Non-motile monocilia (9+0) extend from the surface of most quiescent cells of the body (*e.g.* monocilia of the tubular epithelia of the kidney, of the bile duct and pancreatic ducts, etc). Non-motile 9+2 cilia are found in the inner ear. Axonemes were stained in red or green using the cilia-specific α−tubulin marker. Nuclei were stained using Hoechst or DAPI.

Primary cilia also participate in the regulation of the canonical Wnt signaling cascade that plays a major role in neurodevelopment (Salinas, 2012). Deregulation of the Wnt pathway preventing the β-catenin translocation to the nucleus in mice was shown to be associated with cerebellar hypoplasia and midline fusion defects, resuming the phenotype observed in Joubert patients. Finally, primary cilia play also a role in adult neurogenesis (for instance in synapse formation in adult-born hippocampal neurons (Kumamoto *et al.*, 2012) proposing that they are required for the correct integration of such adult-born neurons into existing brain circuitry, to eventually become undistinguishable from neurons born during embryogenesis.

II- Bardet-Biedl Syndrome

1. Origins

Bardet-Biedl Syndrome was initially described in 1866, by John Laurence and Robert Moon subsequently to the observation of one family whose affected members presented with retinal dystrophy, obesity, intellectual disability (formerly called 'mental retardation), hypogonadism and spastic paraplegia (Laurence and Moon, 1995). This condition was then considered as identical as the one reported in the 1920's by Dr Georges Bardet (Bardet, 1920) and Artur Biedl (Biedl, 1922) with an additional polydactyly, and was re-attributed the name of Laurence-Biedl or Laurence-Moon-Biedl Syndrome (Solis-Cohen, 1925). However, the later observation of an extremely rare co-occurrence of spastic paraplegia and obesity in Laurence-Biedl Syndrome led to a final scission of both entities, leaving as Bardet-Biedl Syndrome the most frequent manifestation.

2. Etiology

Bardet-Biedl Syndrome (BBS) is a rare autosomal recessive genetic disorder, whose incidence may vary depending on the genetic background of the population of concern. Indeed, it is very rare in populations with highly mixed genetic background and low levels of consanguinity, such as in northern Europe where its incidence is estimated at 1:150,000 births. It is however much more frequent in less mixed population, such as either regions of higher rates of consanguinity (Bedouins: from Kuwait 1:12,000 births), or geographically constraint populations with low migration patterns (Newfoundland: 1:15,000 births).

3. Physiopathology

BBS proteins have been early shown to locate within primary (non-motile) cilia, mainly in basal bodies or centrosomes and are involved either in ciliary structure or function. The first protein bringing such evidence was TTC8/BBS8, since among the affected patients carrying mutations in this gene one proband presented with *situs inversus* (Ansley *et al.*, 2003). Also, it was demonstrated to localize in centrioles. BBS proteins can be classified in 2 subcategories: the first one comprises BBS1, BBS2, BBS4, BBS5, BBS7, BBS8 and BBS9 that assemble together to form the BBSome (Nachury *et al.*, 2007), the second includes BBS6, BBS10 and BBS12 that share high similarity and belong to family of chaperonins (Seo *et al.*, 2010). The BBS chaperonin complex mediates the formation of the BBSome (Seo *et al.*, 2010), which in turn ensures trafficking along the axoneme of the cilium and is required for ciliogenesis (Nachury *et al.*, 2007). One particular study demonstrated

that BBS7 and BBS8 both mediated the association of two IFT cargo proteins (kinesin and OSM-3) to ensure the correct rate and function of global IFT transportation (Ou *et al.*, 2005). The latest BBS protein identified to date is one IFT protein, IFT27 (Aldahmesh *et al.*, 2014).

4. Genetic heterogeneity

The first gene identified as responsible for BBS when mutated was MKKS/*BBS6* in 2000 (Katsanis *et al.*, 2000; Slavotinek *et al.*, 2000) a few month after being shown to be responsible for McKusick-Kaufman Syndrome, a more severe ciliopathy (Stone *et al.*, 2000). Most of the first genes were identified following a combination of linkage analyses, homozygosity mapping and haplotype reconstructions in large consanguineous families (see **Table 7**), or by positional cloning. With the rapid expansion and evolution of bioinformatics, others were identified either by comparative genomics (selecting candidate genes according to homology to other BBS proteins, or link to ciliary function), or by direct Sanger sequencing following the identification of their implication in other ciliopathies. The very last BBS genes recently identified were proposed subsequently to exome sequencing. The identification of up to 19 BBS genes now fully demonstrates the extensive genetic heterogeneity of this disorder.

The contribution of each locus in the pathology is highly variable, the most recurrent mutated genes being *BBS10*, *BBS1*, *BBS2* and *BBS12* (**Table 7**). Mutations in *BBS1* or *BBS10* are reported in almost 50% of the patients. Interestingly, most genes recently identified as implicated in BBS (*BBS11*, *BBS13-BBS19*) have only be found in one or a few BBS families suggesting there are minor contributors to the disease. The allelic frequency of some recurrent mutations is also variable across populations: the most recurrent mutations among European patients are located in *BBS1* (c.1169T>G, p.Met390Arg) and *BBS10* (c.271dup, p.Cys91fsLeu*95), yet the one in *BBS1* is not observed in families of Arabic origin suggesting it arose later in the human diaspora (Mykytyn *et al.*, 2003; Katsanis, 2004; Stoetzel *et al.*, 2006). Altogether, mutations in known BBS genes account for only 80% of the patients, suggesting there are some other genes to uncover. Even though the mode of inheritance commonly proposed is autosomal recessive, a few studies suggested an oligogenic mode of transmission, with the report of a potential 3rd pathogenic allele in a second BBS gene in a few families (Katsanis *et al.*, 2001; Katsanis *et al.*, 2002; Badano *et al.*, 2003b; Beales *et al.*, 2003; Katsanis, 2004; Badano *et al.*, 2006; Zaghloul *et al.*, 2010). However, in spite of the emergence of NGS, no further supporting evidence was reported over the last decade (Hichri *et al.*, 2005; Muller *et al.*, 2010).

Gene	Gene size bp (CDS)	Contribution in BBS patients	Method of identification	References
BBS1	22,966 (1,782)	22%	Linkage candidate, CG	(Mykytyn *et al.*, 2002)
BBS2	35,750 (2,166)	7%	Linkage candidate, positional cloning	(Nishimura *et al.*, 2001)
BBS3/ ARL6	33,779 (561)	1%	Comparative genomics, PC	(Chiang *et al.*, 2004; Fan *et al.*, 2004)
BBS4	52,292 (1,560)	5%	Positional cloning	(Mykytyn *et al.*, 2001)
BBS5	27,160 (1,026)	3%	Comparative genomics, PC	(Li *et al.*, 2004)
BBS6/ MKKS	29,034 (1,713)	4%	Positional cloning, Linkage candidate, ciliopathy gene	(Katsanis *et al.*, 2000; Slavotinek *et al.*, 2000; Stone *et al.*, 2000)
BBS7	46,008 (2,148)	2%	Comparative genomics	(Badano *et al.*, 2003a)
BBS8/ TTC8	53,358 (1,518)	2%	Positional cloning, CG	(Ansley *et al.*, 2003)
BBS9/ PTHB1	479,529 (2,559)	2%	Linkage candidate, comparative genomics	(Nishimura *et al.*, 2005)
BBS10	3,957 (2,172)	21%	Linkage candidate	(Stoetzel *et al.*, 2006)
BBS11/ TRIM32	13,999 (1,962)	<0.5%	Linkage candidate	(Chiang *et al.*, 2006)
BBS12	12,242 (2,133)	8%	Linkage candidate, comparative genomics	(Stoetzel *et al.*, 2007)
BBS13/ MKS1	14,170 (1,680)	<1%	Linkage candidate, ciliopathy gene	(Kyttala *et al.*, 2006; Leitch *et al.*, 2008)
BBS14/ CEP290	93,204 (7,440)	<0.5%	Positional cloning, ciliopathy gene	(den Hollander *et al.*, 2006; Sayer *et al.*, 2006; Valente *et al.*, 2006; Leitch *et al.*, 2008)
BBS15/ WDPCP	467,317 (2,469)	<0.5%	Molecular Function	(Kim *et al.*, 2010)
BBS16/ SDCCAG8	224,161 (12,510)	1%	Ciliopathy gene, Exome sequencing,	(Otto *et al.*, 2010; Schaefer *et al.*, 2011)
BBS17/ LZTFL1	15,816 (900)	<1%	Exome sequencing, molecular function	(Marion *et al.*, 2012)
BBS18/ BBIP1	20,637 (312)	<0.5%	Exome sequencing, molecular function	(Scheidecker *et al.*, 2014)
BBS19/ IFT27	1,104 (561)	<0.5%	Exome sequencing, linkage candidate	(Aldahmesh *et al.*, 2014)

Table 7: BBS genes and associated method used for their identification

Clinical manifestations

BBS belongs to the family of ciliopathies, and hence is a multisystemic disorder affecting multiple organs in patients. The main affected ciliated organs are retina and kidney, but others such as limbs, brain and heart may also be concerned. There is a large inter and intrafamilial phenotypic variability in patients, and that is why, early on some critical differential diagnostic criteria had to be drawn in order to discriminate BBS patients from other with related syndromic ciliopathies. Such criteria, originally proposed by Beales in 1999 are still used as an international reference for the diagnosis of BBS (Beales *et al.*, 1999). Subsequently to the study of 109 BBS patients, some features were shown to be more (retinal dystrophy, polydactyly) or less (renal anomalies, cognitive impairment) consistently present in patients, leading to the establishment of diagnostic rules to follow: presence in patient of either 4 major or 3 major and 2 minor features (see below) in order to be BBS diagnosed (**Table 8**).

Major Features	**Minor Features**
Rod-cone dystrophy	Strabismus/cataracts/astigmatism
Polydactyly	Brachydactyly/syndactyly
Obesity	
Learning disability	Speech disorder/delay
	Developmental delay
Hypogonadism (in males)	
Renal anomalies	Polyuria/polydipsia (nephrogenic diabetes insipidus)
	Diabetes mellitus
	Hepatic fibrosis
	Mild spasticity (especially lower limbs)
	Ataxia/poor coordination/imbalance
	Dental crowding/hypodontia/small roots/high arched palate
	Left ventricular hypertrophy/congenital heart disease

Table 8: Beales' criteria for the diagnosis of Bardet-Biedl Syndrome. From (Beales *et al.*, 1999).

The expansion of our understanding of the disorder has led to the addition of previously unmentioned traits but that are consistent with ciliary defects *(e.g.* anosmia, hearing defects) while questioning others such as intellectual disability that is now generally considered as mild cognitive impairment or learning disabilities.

5. Absence of genotype to phenotype correlations

Such genetic heterogeneity is also reported in most other ciliopathies (Alström syndrome being the exception, with only one gene reported so far), and a consequent genetic overlap is observed between such disorders (**Figure 18**). Many genes have been shown as implicated in two or more different ciliopathy disorders. One of the best illustrations is CEP290, centrosomal ciliary protein, first reported as defective in patients with Joubert and Senior-Loken syndromes (Sayer *et al.*, 2006; Valente *et al.*, 2006). A few months later, mutations in the encoding gene are reported in patients with LCA (den Hollander *et al.*, 2006), then in patients with MKS (Baala *et al.*, 2007) and eventually in patients with BBS (Leitch *et al.*, 2008). However, no precise genotype-to-phenotype correlations could be drawn (Coppieters *et al.*, 2010). Other examples show that hypomorphic mutations in *NPHP3* cause NPHP, whereas loss-of-function mutations result in the much more severe phenotype of MKS.

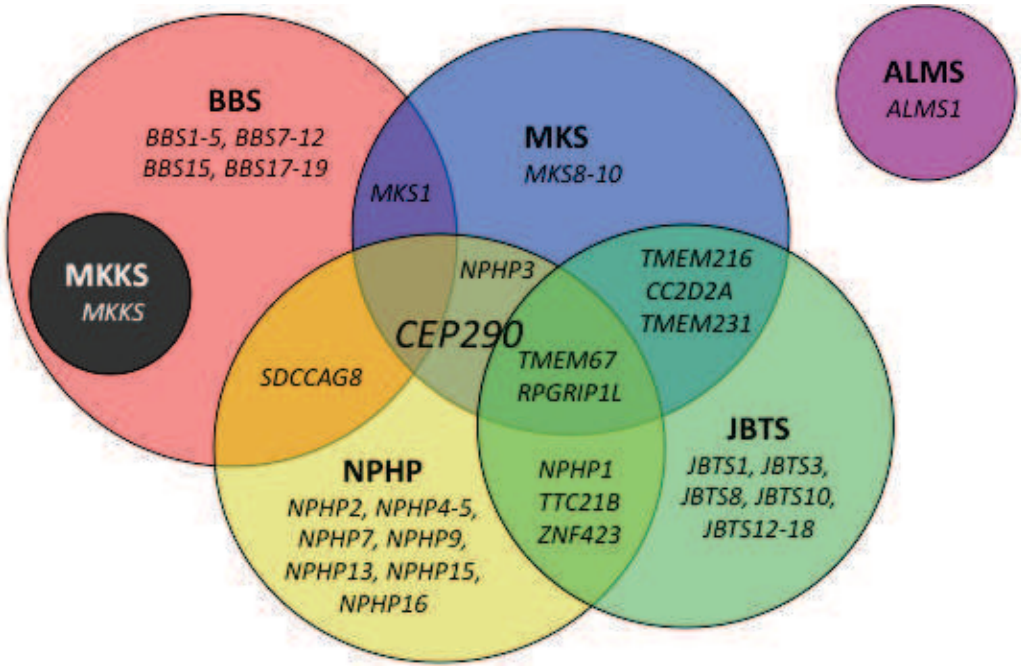


Figure 18: Distribution of ciliopathy genes among six major ciliopathy disorders, demonstrating extensive genetic overlap between such conditions.

ALMS: Alström syndrome, BBS: Bardet-Biedl Syndrome, JBTS: Joubert syndrome, MKS: Meckel-Grubler syndrome, MKKS: McKusick-Kaufmann syndrome, NPHP: nephronophthisis. The CEP290 is the most central gene, which was found mutated in patients with any of five different ciliopathy disorders (MKS, BBS, JBTS, NPHP, but also in SLSN, not represented in this graph).

In BBS itself, although extensively genetically heterogeneous no major genotype to phenotype correlation emerged that would allow to guide the molecular analysis towards one particular gene, despite extensive studies on the topic. One study proposed the association of milder severe retinal impairment in *BBS1* patients (Daniels *et al.*, 2012). Patients carrying mutations in the chaperonin BBS genes (*BBS6, BBS10, BBS12*) appear to develop more severe renal disease (Imhoff *et al.*, 2011). We also recently reported two proposed genotype-phenotype correlations: *BBS16*

mutations were found in patients with specifically severe early-onset renal disease and absence of polydactyly ((Schaefer *et al.*, 2011), confirmed by (Billingsley *et al.*, 2012)), and *BBS17* mutations were found in patients with atypical and specific mesoaxial insertional polydactyly in patients (Schaefer *et al.*, 2013). However, only two families with *BBS17* mutations have been reported so far limiting the significance of this correlation, so the identification of additional families is needed to confirm this indication. Finally, no clear correlations were established between the nature of the mutation itself (missense versus null mutations) and the severity of the phenotype in BBS patients (Daniels *et al.*, 2012).

6. Phenotypic overlap and differential diagnosis

BBS, as a ciliopathy and one of the most pleiotropic one, shares also a number of clinical characteristics and can be confounded with other syndromes in patients (**Figure 19**). The most overlapping syndromes are ALMS, JBTS, and MKS syndromes. ALMS (whose associated clinical traits are retinal dystrophy, obesity and diabetes) can be discriminated from BBS by the additional report of sensorineural hearing loss and cardiomyopathy in patients while absence of polydactyly and learning disabilities. JBTS clinical features include retinal dystrophy, renal defects and cognitive impairment, but additional specific cerebellar defects ('molar tooth sign'), respiratory abnormalities that are not reported in BBS patients, and no polydactyly

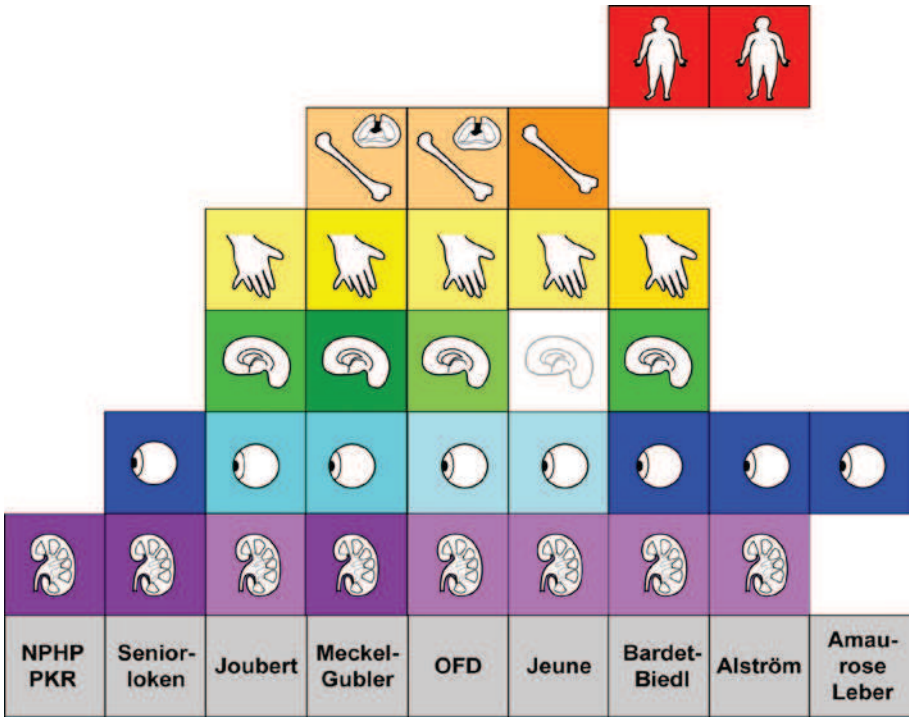


Figure 19 (from (Mockel *et al.*, 2011)): Main disorders belonging to the family of ciliopathy, highlighting the extensive phenotypic overlap in patients, from multiple to isolated-organ dysfunction.

Lastly, MKS (which also largely genetically overlaps with BBS locus) is a much more severe syndrome, lethal in the perinatal period, characterized by cystic kidney disease and polydactyly but additional severe CNS defects (occipital encephalocele and/or hydrocephalus), hepatic fibrosis, and cleft palate. Other ciliopathy affect specifically one particular organ and therefore are less prone to be misdiagnosed with BBS, such as NPHP or PKD (essentially cystic kidney diseases), or LCA (retinal dystrophy pathology).

7. Routine diagnostic offer for BBS patients

At the time this work was initiated, the diagnostic strategy for BBS in France (the diagnostic laboratory of Strasbourg University Hospital being the reference center for this disease) was restricted to direct sequencing of two PCR amplicons containing the most recurrent BBS mutations in *BBS1* (c.1169T>G, p.Met390Arg) and *BBS10* (c.271dup, p.Cys91fsLeu*95), plus screening of the entire *BBS12* coding sequence (see **Figure 20**).

If a single heterozygous mutation was detected, then the entire coding sequence of the gene was screened in search of the second allele. If no mutations were detected and the parents of the proband were related (*i.e.* consanguineous conception), SNPs analysis followed by homozygosity mapping was proposed on proband and another relative, to identify putative BBS genes located in regions of homozygosity shared between affected individuals. If parents were non-consanguineous, or if no known BBS gene fell within such homozygosity regions, the molecular analysis was stopped. Over 60% of all BBS patients were left without molecular diagnosis with this strategy, while it is the first step for patients' relatives towards genetic counseling. In other countries, an array was developed to screen for previously reported mutations in various BBS genes (Pereiro *et al.*, 2011).

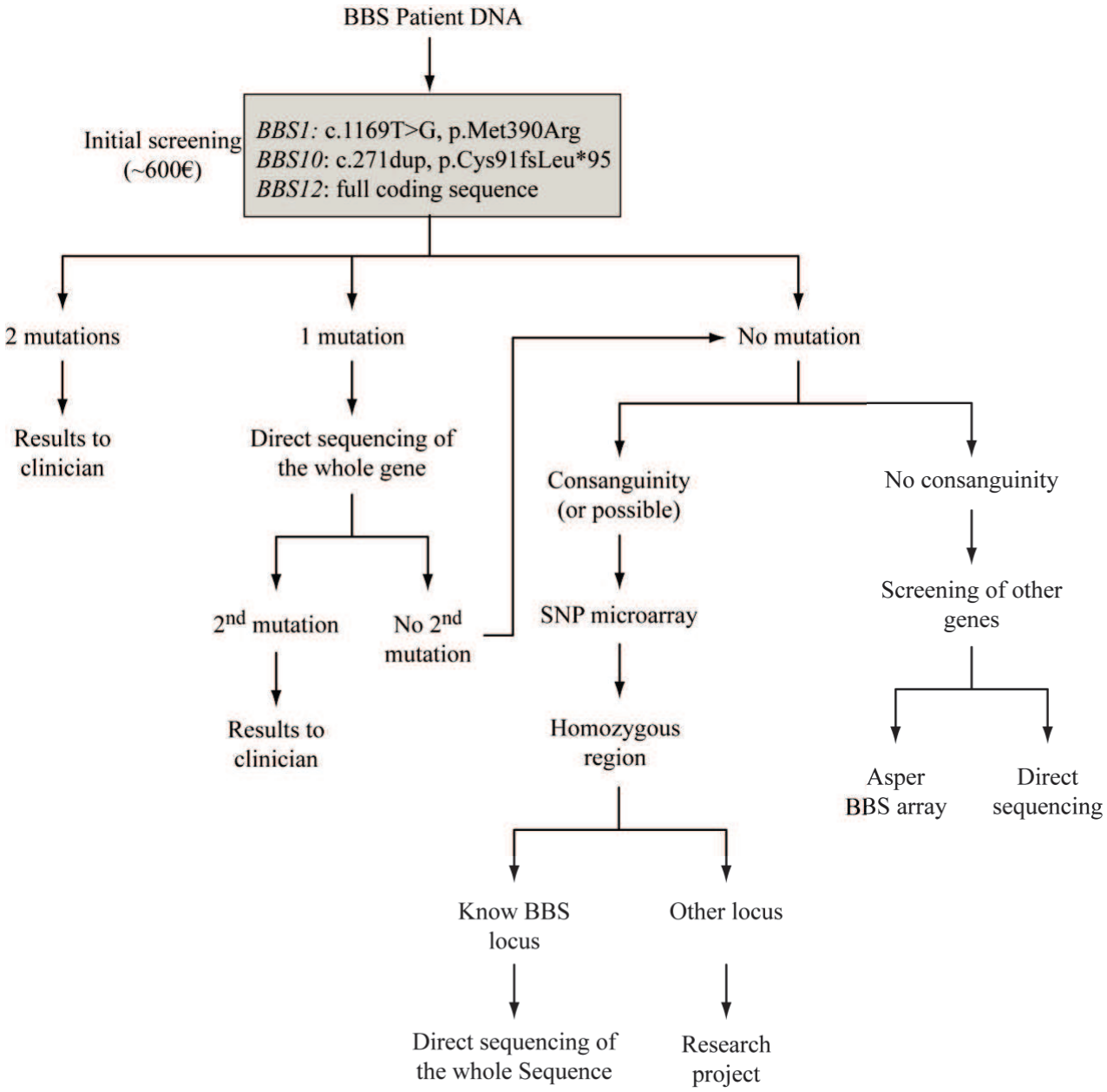


Figure 20: Decisional tree for the molecular diagnosis of BBS in Strasbourg until 2012

III- Results

Paper I: Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström syndromes.

Redin C*, Le Gras S, Mhamdi O, Geoffroy V, Stoetzel C, Vincent MC, Chiurazzi P, Lacombe D, Ouertani I, Petit F, Till M, Verloes A, Jost B, Chaabouni HB, Dollfus H, Mandel JL, Muller J. J Med Genet. 2012 Aug.

Aim of the study: Development of an alternative method for the diagnosis of Bardet-Biedl and Alström syndromes using NGS technologies. Proof of concept of the method on 14 patients with known mutations (positive controls), application on 38 additional patients without previous molecular diagnosis.

Contribution: Major. Performed the design of the capture, quality controls of DNA samples, preparation of sequencing libraries, analyzed the processed data, wrote the paper.

JMG Online First, published on July 7, 2012 as 10.1136/jmedgenet-2012-100875

ORIGINAL ARTICLE

Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström Syndromes

Claire Redin,[1] Stéphanie Le Gras,[2] Oussema Mhamdi,[3] Véronique Geoffroy,[4] Corinne Stoetzel,[5] Marie-Claire Vincent,[6] Pietro Chiurazzi,[7] Didier Lacombe,[8] Ines Ouertani,[3] Florence Petit,[9] Marianne Till,[10] Alain Verloes,[11] Bernard Jost,[2] Habiba Bouhamed Chaabouni,[3] Helene Dollfus,[5,12] Jean-Louis Mandel,[1,6,13] Jean Muller[1,6]

► Additional materials are published online only. To view these files please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2012-100875/content/early/recent).

For numbered affiliations see end of article.

Correspondence to
Professor Jean-Louis Mandel, Department of Neurogenetics & Translational medicine, IGBMC, 1 rue Laurent Fries, Illkirch cedex 67404, France; jlmandel@igbmc.fr

Received 7 March 2012
Revised 15 May 2012
Accepted 28 May 2012

ABSTRACT

Background Bardet-Biedl syndrome (BBS) is a pleiotropic recessive disorder that belongs to the rapidly growing family of ciliopathies. It shares phenotypic traits with other ciliopathies, such as Alström syndrome (ALMS), nephronophthisis (NPHP) or Joubert syndrome. BBS mutations have been detected in 16 different genes (*BBS1-BBS16*) without clear genotype-to-phenotype correlation. This extensive genetic heterogeneity is a major concern for molecular diagnosis and genetic counselling. While various strategies have been recently proposed to optimise mutation detection, they either fail to detect mutations in a majority of patients or are time consuming and costly.

Method We tested a targeted exon-capture strategy coupled with multiplexing and high-throughput sequencing on 52 patients: 14 with known mutations as *proof-of-principle* and 38 with no previously detected mutation. Thirty genes were targeted in total including the 16 BBS genes, the 12 known NPHP genes, the single ALMS gene *ALMS1* and the proposed modifier *CCDC28B*.

Results This strategy allowed the reliable detection of causative mutations (including homozygous/heterozygous exon deletions) in 68% of BBS patients without previous molecular diagnosis and in all *proof-of-principle* samples. Three probands carried homozygous truncating mutations in *ALMS1* confirming the major phenotypic overlap between both disorders. The efficiency of detecting mutations in patients was positively correlated with their compliance with the classical BBS phenotype (mutations were identified in 81% of 'classical' BBS patients) suggesting that only a few true BBS genes remain to be identified. We illustrate some interpretation problems encountered due to the multiplicity of identified variants.

Conclusion This strategy is highly efficient and cost effective for diseases with high genetic heterogeneity, and guarantees a quality of coverage in coding sequences of target genes suited for diagnosis purposes.

INTRODUCTION

Bardet-Biedl syndrome (BBS; OMIM# 209900) is a pleiotropic recessive disorder with high non-allelic genetic heterogeneity. Its incidence varies from an estimated 1:160 000 in northern Europe to 1:13 500–17 000 in Bedouins and Newfoundlanders, respectively.[1] BBS belongs to the large and growing family of ciliopathies and, therefore, shares phenotypic traits with Joubert (JBTS), Alström (ALMS) and Meckel (MKS) syndromes.[1 2] Differential clinical diagnosis may thus be difficult, especially in young probands who do not yet show some later onset-specific manifestations.[3 4] In particular, recent reports highlight a significant clinical overlap between BBS and ALMS.[3 5]

The main phenotypic features of BBS comprise retinal dystrophy, polydactyly, obesity, mild developmental delay, polycystic kidneys and hypogenitalism. Other minor features can also be observed in patients, such as cardiac abnormalities, other digit or eye anomalies, diabetes, hypertension, hearing defects, anosmia.[6 7] Up to now, mutations have been detected in 16 different genes (*BBS1-BBS16*), but no clear genotype-to-phenotype correlation could be observed, besides the suggested exception of *BBS16*.[8]

Alström syndrome (OMIM #203800) was reported to be much less prevalent than BBS, with an estimated incidence of 1:1 000 000. Its phenotypic features overlap with those of BBS in early infancy and include: cone-rod dystrophy, obesity, type 2 diabetes mellitus, hearing loss but also hypertriglyceridemia, dilated cardiomyopathy, and progressive pulmonary, hepatic, or renal dysfunction.[9] To date, only one gene (*ALMS1*) has been identified, but recent reports showed some families with suggestive ALMS-carrying mutations in BBS genes.[3 5] The large size of *ALMS1* coding sequence appears to have impaired widespread diagnostic testing of ALMS.

Exhaustive conventional Sanger sequencing for BBS diagnosis is prohibitively expensive because of the large number of genes involved, and so also for ALMS due to the large size of *ALMS1* coding sequence (12 kb, 24 exons; table 1). Alternative

Table 1 Genes included in the targeted enrichment strategy and their associated disorders

Official gene symbol	BBS#	NPHP#	Other disease-related symbols	# Exons in consensus	# Total exons in all isoforms	Size of coding exons (bp)	Gene size (bp)	Targeted region size
BBS1*	BBS1	–	–	17	17	1782	22 966	23 285*
BBS2	BBS2	–	–	17	17	2166	35 750	3801
ARL6	BBS3	–	RP55	8	11	561	33 779	2123
BBS4*	BBS4	–	LCA	16	18	1560	52 292	52 611*
BBS5	BBS5	–	–	12	17	1026	27 160	3997
MKKS	BBS6	–	–	6	7	1713	29 034	3326
BBS7	BBS7	–	–	19	20	2148	46 008	5308
TTC8	BBS8	–	RP51	15	16	1518	53 358	3137
BBS9	BBS9	–	–	23	24	2559	476 529	5196
BBS10	BBS10	–	–	2	2	2172	3957	3941
TRIM32	BBS11	–	LGMD2H; STM	2	2	1962	13 999	4077
BBS12	BBS12	–	–	2	3	2133	12 242	3829
MKS1	BBS13	–	MKS1	18	20	1680	14 170	3745
CEP290	BBS14	NPHP6	MKS4; JBTS5; LCA10; SLSN6	54	55	7440	93 204	10 510
WDPCP	BBS15	–	–	18	19	2469	467 317	4405
SDCCAG8	BBS16	NPHP10	SLSN7	18	21	2141	244 087	3599
ALMS1		–	ALMS; LCA	23	24	12 510	224 161	13 682
NPHP1		NPHP1	JBTS4; SLSN1	20	22	2202	81 726	3264
INVS		NPHP2	–	17	19	3198	201 916	4103
NPHP3		NPHP3	MKS7; RHPD	27	27	3993	41 823	5328
NPHP4		NPHP4	SLSN4	30	30	4281	129 662	5693
IQCB1		NPHP5	SLSN5	15	15	1797	65 317	2585
GLIS2		NPHP7	–	8	8	1575	7374	2175
RPGRIP1L		NPHP8	MKS5; JBTS7; CS	27	27	3708	103 954	5243
NEK8		NPHP9	–	15	15	2079	13 953	3096
TMEM67		NPHP11	MKS3; JBTS6; CS	28	30	2745	64 389	4797
TTC21B		NPHP12	JBTS11; ATD4	29	29	3951	79 894	5414
TMEM216	–	–	MKS2; JBTS2	5	5	447	6504	1795
AHI1	–	–	JBTS3	27	29	3591	213 794	5175
CCDC28B	–	–	–	6	6	603	4790	1079
			Total	483	545	79 781	2 865 109	200 319

*Sequence of the entire gene (coding/non-coding exons, introns) was targeted.
ALMS, Alström syndrome; ATD, asphyxiating thoracic dystrophy; BBS, Bardet-Biedl syndrome; CS, COACH syndrome; JBTS, Joubert syndrome; LCA, Leber congenital amaurosis; LGMD, limb-girdle muscular dystrophy; MKKS, McKusick-Kaufman syndrome; MKS, Meckel-Gruber syndrome; NPHP, nephronophthisis; RHPD, renal-hepatic-pancreatic dysplasia; RP, retinitis pigmentosa; SLSN, Senior-Loken syndrome; STM, sarcotubular myopathy.

cost-conscious strategies have been proposed for BBS diagnosis, such as: initial screening of recurrent mutations and frequently mutated genes (*BBS1*, *BBS10*, *BBS12*) combined with homozygosity mapping for consanguineous families[10 11]; or primer extension arrays to test a series of known BBS mutations.[5] Another approach recently proposed is the pooling of patients' DNAs with subsequent PCR-amplification and massive parallel resequencing of *BBS1-12* coding exons, followed by heteroduplex screening to identify the mutation carrier.[12] Such a method presents some limitations as it will miss exon deletions and may not be suited for diagnostic purposes. Considering the clinical overlap with other ciliopathies, another approach would be to test, systematically and simultaneously, all corresponding genes for such overlapping syndromes, which would be particularly relevant for patients with atypical or incomplete clinical phenotypes. We describe here the results of such an approach, based on a targeted exon capture of 30 genes coupled to next-generation sequencing (NGS).

SUBJECTS AND METHODS

Detailed protocols are available in Supplementary Methods.

Subjects

DNA samples from 52 unrelated patients were collected. Most patients had been addressed to the diagnostic laboratory, or to the National Reference Center for rare ophtalmogenetic diseases in Strasbourg. Eleven DNA samples stemmed from Tunisian patients included in an independent BBS epidemiology study.

The *proof-of-principle* cohort included 14 non-Tunisian patients with a confirmed BBS molecular diagnosis (identified prior to this study by Sanger sequencing). Twenty-six out of the 38 patients without known mutations, and recruited in Strasbourg, had been initially screened for *BBS1* and *BBS10* recurrent mutations, plus the entire coding sequence of *BBS12*.

For all patients, a written consent for genetic testing was obtained, either from adult probands or from the legal representative in case of minors.

Library preparation, targeted capture and sequencing

DNA samples were prepared and controlled following standard procedures.

The capture design was performed with eArray following the manufacturer's instructions (Agilent).

DNAs (3 μg) were sheared mechanically using Covaris E220 (duty cycle: 10%; intensity: 5; cycles per burst: 200; time: 300 s).

For the *proof-of-principle* experiment, sequencing adaptors were added on 500 ng of sheared DNA using the SPRIworks Fragment Library System I (Beckman Coulter). After amplification and quality assessments, targeted capture was performed on individual samples using the *in-solution* SureSelect Target Enrichment System (Agilent) on 500 ng of DNA-prepped library. Additional

steps of washing, purification and elution were performed, and multiplexing adaptors (TruSeq Illumina DNA indexes) were added by PCR during the post-capture amplification step.

For all following experiments, multiplexing adapters were added simultaneously to sequencing adapters using the SPRI-works system. Equimolar amounts of two tagged libraries were then pooled prior to the capture reaction. All other following steps prior to sequencing remained identical. A 72-bp single-read sequencing was performed on a Genome Analyser IIx (GAIIx, Illumina).

Bioinformatic pipeline

Read mapping and variant calling were performed following standard procedures. Variant filtering was performed using VaRank, an in-house software which collects variant-specific information to rank them according to their predicted pathogenicity (figure 1, Supplementary Methods).

Copy-number variation (CNV) detection method

CNVs were identified using a depth-of-coverage method.[13 14] For each patient, read counts in non-overlapping windows of 20 nucleotides were computed, normalised and then compared randomly with eight other samples from the same experiment (considered as replicates) using the Bioconductor package DEseq (initially designed for RNA-seq data).[15] Candidate regions for CNVs were retrieved when log2 ratios (controls/sample) were either ≥ 0.84 (fold change >1.8, potential deletion) or ≤ -0.51 (fold change <0.7, potential duplication), and if p values adjusted for multiple testing (Benjamini and Hochberg procedure)[16] were smaller than 0.1.

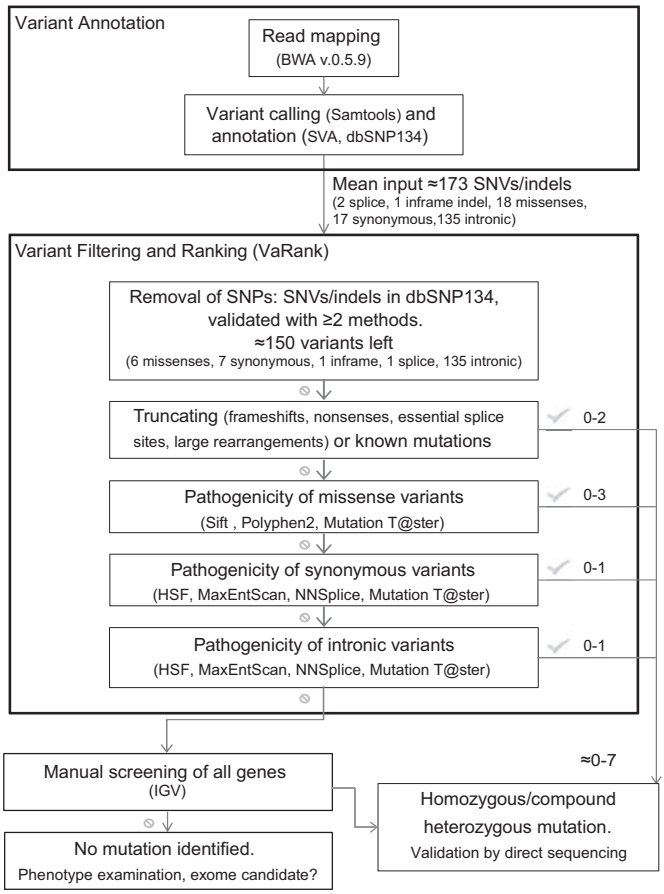


Figure 1 Global flowchart of the bioinformatic pipeline implemented for mutation detection. Software acronyms: BWA, Burrows-Wheeler Aligner; SVA, Sequence Variant Analyser; SIFT; Polyphen2; HSF, Human Splicing Finder; MaxEntScan, Maximum Entropy Scanning; NNSplice; Mutation Taster; IGV, Integrative Genome Viewer.

Statistical methods

Confidence intervals were computed for proportion estimates and indicated in brackets. Fisher's exact test was computed to compare distributions of small populations. Subsequent p value is given at $\alpha=0.05$.

RESULTS AND DISCUSSION

Targeted regions: design strategy

Our primary goal was to develop an efficient mutation-screening strategy for the diagnosis of patients with phenotypes evocative of BBS, or of clinically overlapping ciliopathies. We chose a target enrichment approach coupled with NGS in order to focus the sequencing on genomic regions of interest. We targeted all exons (including 5′ and 3′ UTRs) of the 16 known BBS genes (table 1). Because of the known clinical overlap, we also included coding exons of *ALMS1*, and of all 12 known nephronophthisis genes (*NPHP1-12*), since retinal degeneration can often be observed in this kidney-specific disease.[9 17] Coding sequences of *AHI1/JBTS3*, *TMEM216/MKS2/JBTS2*, and of the proposed BBS-modifier *CCDC28B/MGC1203*, were also targeted.[18] Because some of these genes are associated with multiple phenotypes, our design includes 6 MKS, 7 JBTS and 4 Senior-Loken syndrome (SLSN) genes (see table 1).

With this first design, we wanted to investigate whether including intronic sequences could favour both, the detection and sizing of exon deletions. We therefore included baits-targeting intronic sequences of *BBS1* and *BBS4*. This choice was dictated by two observations: an apparent excess of patients heterozygous for the BBS1-recurrent mutation M390R with no second mutation detected,[11] and multiple reports of *BBS4* exon deletions in patients.[4 11 19] A maximal threshold of 200 kb for cumulated targeted regions was set because of the manufacturer's pricing limits.

Presence of repeated sequences precluded bait tiling in 19.7% of initially targeted regions. This concerned, almost exclusively, introns of *BBS1* and *BBS4*, besides a small number of 3′UTRs, and only 128 bp of protein coding regions (within first exons of *ALMS1* and *NPHP3*; table S1).

Proof-of-principle and technical results

In our *proof-of-principle* experiment, we selected 16 DNA samples, of which 14 were with known BBS mutations. In this first trial, after barcoding the target-enriched libraries, we sequenced pools of four or eight libraries per lane of a GAIIx (see Supplementary Methods). This *proof-of-principle* analysis was carried blind, that is, without knowledge of implicated BBS genes and their associated mutations. A constellation of all mutation types (missenses, nonsenses, splice mutations, large deletions and complex rearrangements) at different allelic dosage was tested (figure S1). All 14 previously identified mutations, including two heterozygous *BBS1*-deletions (figure 2A), were detected in their correct heterozygous/homozygous state (table S2). In particular, in patient AKE12, we could detect an abnormal local drop of coverage in *BBS12* due to a rare mutation type (insertion of an Alu sequence, figure S1A) although the exact nature of the mutation could only be determined by Sanger sequencing. A

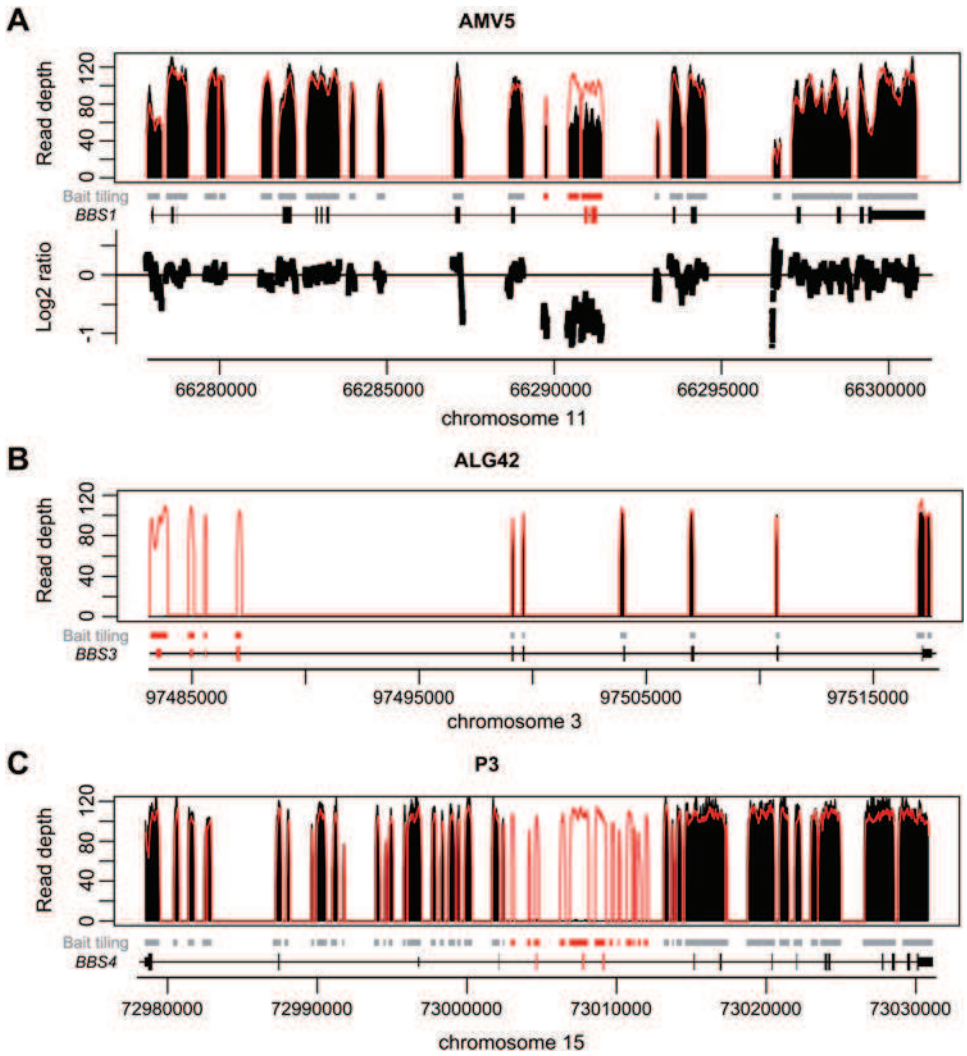


Figure 2 Detection of large deletions in three patients using a depth-of-coverage method. Black peaks: normalized depth of coverage from patients' DNA samples. Empty peaks: normalized mean depth of coverage across samples from the same sequencing lane. Grey squares: bait-covered regions. Black peaks: normalized depth of coverage from patients' DNA samples. Empty peaks: normalized mean depth of coverage across samples from the same sequencing lane. Grey squares: bait-covered regions. Highlighted squares: deleted regions. Gene representation: black squares: exons, dashed lines: introns. Genomic positions are given according to the human reference genome hg19/ GRCh37. (A) Heterozygous deletion of BBS1 (exon #10, 11) in AMV5 patient. Corresponding Log2 ratios between both depths of coverage (normalised mean and AMV5 patient) further highlight the presence of the deletion. (B) Homozygous deletion of *BBS3* (exon #1, 2a, 2b, 3) in ALG42 patient. (C) Homozygous deletion of *BBS4* (exon #4, 5, 6) in P3 patient. (A and C): targeting intronic sequences allows restricting the deletion breakpoints. (B) and C): Log_2 ratios between both depths of coverage (normalised mean and corresponding patients) could also allow detecting both deletions but are not shown (supplementary figure S2).

similar drop in coverage was observed for a second patient, AHX91, with another complex mutation detected previously by Sanger sequencing (insertion/inversion in *BBS5*).

In this first experiment, we almost systematically reached the maximal theoretical coverage of 144x illustrated by a mean coverage of 127±4x after removal of duplicate reads (table 2). Due to this global saturating coverage when considering unique reads, we used all reads, including duplicates, when applying our depth-based method for the detection of CNVs.

These promising depth-of-coverage results (table 2, table S3) encouraged us to further increase the number of pooled samples. In the next experiments we used a single capture reaction for two barcoded libraries, allowing both cost and bench-time savings, and pooled 12 libraries per sequencing lane (maximum number of barcodes proposed at that time by Illumina).

This new protocol was performed on a second cohort of 36 patients with unknown mutations. Sequencing resulted in a mean coverage of 78±17× (283±153x before discarding duplicate reads) with 91.4±6.4% of targeted regions being covered more than 40x (table 2). This relative drop of coverage appears to be a consequence of a lower capture efficiency that might be due to: (1) an input amount of individual library reduced by half, due to the pre-capture pooling and (2) the addition of barcodes before capture, leading to less efficient blocking and unspecific hybridisation. The resulting coverage still guarantees a reliable detection of variants and of their homozygous/heterozygous state.

A small proportion of targeted regions was weakly covered in some patients (ie, depth <10x after duplicates filtering), with very few of them in a systematic way in other patients (table S4). This only concerned 0.63±0.68% of protein coding regions, and mostly included intronic GC-rich sequences (GC content: 68.3±5% vs 40.2±10% across all targeted regions), or some first exons (tables S3 and S4).

Table 2 Sequencing statistics of both coverage (in captured regions) and capture efficiency

	Mean coverage (×)		Targeted regions with 5× coverage	Targeted regions with 40× coverage	Targeted regions with 80× coverage	Reads in targeted regions	
	Before filters	After filters				Before filters	After filters
Pool of 4*	2208±416	130±3	100±0%	99.9±0%	98.3±0.6%	76±2%	35±5%
Pool of 8*	1024±151	125±3	99.9±0.3%	99.7±0.3%	96.9±0.9%	74±5%	40±6%
Pool of 12†	283±153	78±17	99.6±0.4%	91.4±6.4%	48.7±28%	25±9%	14±4%

*First set of experiment. The capture-enrichment step was done individually, on untagged DNA libraries.
†Second set of experiment. The capture-enrichment step was realised on equimolar pools of 2 barcoded DNA libraries.

Variant filtering: importance of databases and frequency data

In targeted regions, we detected, on average, 170 variants (Single Nucleotide Variants (SNVs) and indels) per patient. All were systematically analysed for putative effect on protein structure and splice sites using VaRank (figure 1, Supplementary Methods). About 130 of these variants were recorded in dbSNP134 (table S5), but only 20 were validated with at least two independent methods and, therefore, filtered out. Indeed, in the context of a rare recessive disorder, some true mutations can be present at very low frequency in a heterozygous state in controls.

Potential pathogenicity of the remaining 150 variants was assessed using bioinformatic tools and considering their allele frequency in a European-American population, as reported in the Exome Variant Server database (EVSdb). This yielded from zero to six interesting variants per sample, among which were obvious truncating or known mutations in some patients.

The new 'clinical significance' field introduced in dbSNP134 has to be considered with caution since established mutations can now be reported in the database but are not systematically flagged as pathogenic (example: rs179363897, p.R138H) mutation in *BBS5*). Conversely, we detected some false-positive annotations: rs4784677 (p.N70S) in *BBS2*—initially reported as a third allele according to the triallelic hypothesis—[20] is flagged as pathogenic, but is too frequent to be a fully penetrant mutation (0.77% in EVSdb). Filters have to be carefully adapted to the disorder of interest, and to the constantly evolving updates of databases.

Detection of exon deletions

One advantage of NGS-based strategies, as opposed to Sanger sequencing, is the opportunity to detect—in addition to SNVs and small indels—CNVs affecting one or more exons (figure 2 and S2). In the *proof-of-principle* experiment, two heterozygous deletions could be detected in *BBS1*. Among the unknown samples, two homozygous deletions in *BBS3/ARL6* and *BBS4* were identified. To our knowledge, we provide here the first report of large deletions in *BBS1* and *BBS3/ARL6*, while several deletions affecting *BBS4* have been previously observed.[4 11 19] Since we also targeted intronic sequences of *BBS1* and *BBS4*, we were able to narrow the boundaries of subsequent detected deletions (figure 2). For patient AMV5, by using coordinates of affected exons, the estimated size of *BBS1* deletion would be between 466 and 4707 bp, while with our design, we could restrict it to 1862–3841 bp (figure 2A, table S2). In patient P3, we could similarly reduce the assessed size of the *BBS4* deletion from 4626–12 975 bp based on exon positions down to 9376–10 469 bp (figure 2C, table 3A). Lastly, since the *BBS3/ARL6* deletion in patient ALG42 encompasses the first three exons of the gene (figure 2B), we tested whether it may extend and affect *EPHA6* located upstream, encoding an ephrin receptor. Direct PCR testing excluded such extended deletion (data not shown).

Thanks to this method, in the six patients in whom we detected a single heterozygous potentially pathogenic mutation, we can ascertain that no heterozygous deletion is present in *trans*, or at least none encompassing exonic sequences.

Distribution of detected BBS mutations in the 38 unknown patients

Of the 38 samples with unknown BBS mutations (36+2 from the *proof-of-principle* experiment) we detected clearly pathogenic biallelic mutations in 26 cases (68.4%; table 3A). To our knowledge, seventeen of these mutations have not been reported previously, indicating that the BBS mutation spectrum is far from being saturated in spite of numerous BBS mutation reports. Homozygous mutations were found in 88% (23/26) of mutated patients, coherent with the large number of consanguineous probands included in the cohort (75%; 25/33). In two patients of consanguineous origin, the BBS mutation was located outside the homozygous regions detected by prior SNP array analysis and would, therefore, have been missed using a homozygosity mapping strategy (patients AGL23, AKX44; table 3A).

Among the remaining 12 patients with no biallelic mutations identified (table 3B), one patient (AHR2) had a heterozygous clearly pathogenic splicing mutation in *BBS3*. Five patients had heterozygous missenses predicted to be damaging by SIFT, Polyphen2 and/or Mutation T@ster (AIY87, AIX45, AMO77, AMA28, AHL86) with the latter two carrying such variants in two different genes. One consanguineous Melanesian patient, AKE98, presented with classical BBS-inclusion features, including polydactyly. He carried two homozygous variants which initially appeared as potentially pathogenic: a distal frameshift in *INVS/NPHP2* predicted to add 14 amino-acids to the C-terminus of the longer protein isoform, and a non-reported missense P2679L affecting a conserved residue in ALMS1 (figure S3). Subsequent segregation analysis ruled out their implication in the disease since both variants were heterozygous in a similarly affected brother.

In five patients, no potentially pathogenic variant could be identified in any of the 30 targeted genes. These patients are thus candidate for exome sequencing which might either help in identifying novel genes or in reconsidering the clinical diagnosis.

Mutation load in BBS and other targeted genes: importance of *ALMS1*

The mutation load among BBS genes in our cohort appears consistent with previous reports.[7 11] Observed occurrences for *BBS1* (7/38, 18.4%) and *BBS2* mutations (5/38, 13.2%), the most frequently mutated genes in our study, are similar to the respective reported figures of 16.9% and 12%.[7] Considering frequently mutated genes, our study was strongly biased against *BBS1*, *BBS10* and *BBS12*, since two-thirds of the patients' DNAs were previously tested negative for *BBS1* and *BBS10* recurrent mutations, plus all *BBS12* protein coding sequence. This explains the total absence of *BBS12* mutations in our cohort, and the relatively low contribution of *BBS10* (5/38, 13.2%) as compared with the literature (≥20%).[7 11] The contribution of other BBS genes was low, with frequently only one proband involved.

We did not find any mutation in the 'new' BBS genes (*BBS13–16*) suggesting that, cumulatively, they have a small contribution to the total mutation load. *BBS13/MKS1* was indeed shown to be mostly implicated in MKS since only one BBS patient was reported with two heterozygous mutations p.[C492W]; [F371del] others carried only heterozygous missenses, sometimes in addition to homozygous truncating *BBS1* mutations.[21] Likewise, for *BBS14/CEP290*, a homozygous truncating mutation (p.E1903*) was found in a single BBS patient,[21] while other mutations are much more often implicated in Joubert, Senior Loken, Leber Congenital Amaurosis or Meckel syndromes. Similar observations can be made for *BBS15/WDPCP* and *BBS16/SDCCAG8*.[22 23] Like in all other studies of BBS cohorts, no mutation was identified in *BBS11/TRIM32* raising the question of its real implication in BBS: only one homozygous missense mutation was described in a single

Table 3 Identified mutations and other potentially pathogenic variants in the 38 patients with previously unknown genotype. A) Patients with two clearly pathogenic variants in one gene; B) Patients with a single or no clear pathogenic variant in one gene. Mutations are described according to the latest nomenclature conventions described in HGVS

Patient#	Principal mutations (p.)	Principal mutations (c.)	Sanger validation/ segregation	Mutation, as predicted by (among SIFT, PPhen2 & Mutation T@ster)	Additional potentially pathogenic variants	Frequency of additional variants in EVSdb	Geographic origin	Inbred	BBS inclusion criteria
A)									
BBS1									
ALO47	**p.[R160Q]; [R160Q]**	c.[479G→A]; [479G→A]	SV	PP, MT	–		Turkey	Yes (BBS1, BBS4)	No
P1	p.? ; ?	c.[1473+4A→G]; [1473+4A→G]	SV + S	MT (Ex14 splice site altered]	–		Tunisia	Yes*	Yes
P9	p.? ; ?	c.[1110G→A]; [1110G→A]	SV + S	MT (Ex11 splice site altered]	AHI1/JBTS3: p.[R830W]; [=]	2.83%	Tunisia	Yes*	No
AMK19	**p.? ; [M390R]**	c.[951+1G→A]; [1169T→G]	SV + S	MT (Ex10 splice site altered]; S	TTC21B/NPHP12: p.[R713I]; (=), BBS7: p.[D412G]; [=]	NF, 0.23%	NA	NA	No
P12	p.[A14Lfs*28] ; [A14Lfs*28]	c.[39del]; [39del]	SV + S	MT	–		Tunisia	Yes*	Yes
P11	**p.[R146*]; [R146*]**	c.[436C→T]; [436C→T]	SV + S	MT	CCDC28B: Ex2 splice site altered (F110F]; [=]	2.07%	Tunisia	Yes*	Yes
AIO57	p.[E224K]; [E224K]	c.[670G→A]; [670G→A]	SV	S, PP	NPHP3: p.[R1074H]; (=), ALMS1: p.[S2102L]; [=]	0.01%, 3.07%	Morocco	Yes (BBS1, BBS3, NPHP5, JBTS2)	No
BBS2									
P2	p.? ; ?	c.[345+5G→A]; [345+5G→A]	SV + S	(Ex2 splice site altered]	–		Tunisia	Yes*	Yes
AGL23	p.[H665Tfs*11]; [H665Tfs*11]	c.[1992del]; [1992del]	SV + S	MT	TTC21B/NPHP12: p.[R616C]; [=]	0.31%	India	Yes (ALMS1, NPHP1, NPHP3, NPHP6, NPHP8]	No
P7	**p.[R189*]; [R189*]**	c.[565C→T]; [565C→T]	SV + S	MT	NPHP4: p.[R674H]; [R674H]	NF	Tunisia	Yes*	Yes
AKX44	**p.[R272*]; [R272*]**	c.[814C→T]; [814C→T]	SV	MT	SDCCAG8/BBS16: p.[Q505E]; [=]	NF	Algeria	Yes (BBS7, BBS12, NPHP8)	Yes
ALG76	p.[L209P]; [L209P]	c.[626T→C]; [626T→C]	SV + S	S, PP, MT	–		Turkey	Yes (BBS2, NPHP8)	Yes
BBS3/ARL6									
ALG42	p.[del Ex1-3]; [del Ex1-3]	c.[(?_−30)_(123+?))del; [(?_−30)_(123+?)]del	SV + S	–	–		Turkey	Yes (ALMS1, BBS3, BBS4, BBS9, BBS14, NPHP2, NPHP3)	Yes
BBS4									
P3	**p.[del Ex4-5-6] ; [del Ex4-5-6]**	c.[(157−?)_(405+?)del]; [(157−?)_(405+?)del] *(g.[24029..24285]_ (33661..34498))del]; g.[24029..24285]_ (33661..34498))del]*	SV + S	–	BBS5: p.[N184S]; [=]	0.57%	Tunisia	Yes*	Yes
BBS5									
P13	p.[L50R]; [L50R]	c.[149T→G]; [149T→G]	SV + S	S, PP, MT	–		Tunisia	Yes*	Yes
ALG5	p.[R138H]; [R138H]	c.[413G→A]; [413G→A]	SV + S	S, PP, MT	–		France	No	No
BBS6/MKKS									
P10	p.? ; ?	c.[1272+1G→A]; [1272+1G→A]	SV + S	MT (Ex5 splice site altered)	NPHP4: p.[R959Q]; [=]	NF	Tunisia	Yes*	Yes
BBS8/TTC8									
P14	p.? ; ?	c.[329+1G→A]; [329+1G→A]	SV + S	MT (Ex4 splice site altered]	AHI1/JBTS3: p.[R830W]; [=]	2.83%	Tunisia	Yes*	Yes

Continued

Table 3 Continued

Patient#	Principal mutations (p.)	Principal mutations (c.)	Sanger validation/ segregation	Mutation, as predicted by (among SIFT, PPhen2 & Mutation T@ster)	Additional potentially pathogenic variants	Frequency of additional variants in EVSdb	Geographic origin	Inbred	BBS inclusion criteria
BBS10									
P8	p.[S396Lfs*6]; [S396Lfs*6]	c.[1171_1181dupGCATTTATACC]; [1171_1181dupGCATTTATACC]	SV + S	MT			Tunisia	Yes*	Yes
JSL	**p.[V707*]**; [R95S]	c.[2119–2120delGT]; [285A→T]	SV + S	MT; S, PP, MT			NA	NA	NA
BBH64	p.[T483Nfs*8]; [?]	c.[1448–1452delCTCAA]; [?]	S	MT	CCDC28B: Ex2 splice site altered (F110F) ; (=]	2.07%	NA	No	No
AMR64	**p.[L414S]; [L414S]**	c.[1241T→C]; [1241T→C]	SV	S, PP	CCDC28B: Ex2 splice site altered (F110F) ; (=]	2.07%	NA	NA	Yes
AKR68	**p.[L414S]; [L414S]**	c.[1241T→C]; [1241T→C]	SV	S, PP	CEP290/BBS14: p.[K1870Nfs*4]; [=]	NF	NA	Yes (BBS10, JBTS3)	No
ALMS1									
AIA84	p.[E1114Rfs*9); (E1114Rfs*9]	c.[3340del]; [3340del]	SV	MT			Turkey	Yes (ALMS1, BBS14)	Yes
AKO26	**p.[R3629*]; [R3629*]**	c.[10885C→T]; [10885C→T]	SV + S	MT	WDPCP/BBS15: p.[V329M]; (=), AHI1/JBTS3: p.[R548H]; [=]	0.13%, 1.65%	France	Yes (ALMS1, BBS12)	Yes
ALB64	p.[S577*]; [S577*]	c.[1730C→G]; [1730C→G]	SV + S	MT	IQCB1/NPHP5: p.[E481K]; [=]	0.17%	Portugal	Yes (BBS7, BBS12, BBS14, ALMS1)	No

Patient#	Gene	Potential mutations (p.)	Potential mutations (c.)	Frequency in EVSdb	Mutation, as predicted by (among SIFT, PPhen2 & mutation T@ster)	Geographic origin	Consanguinity	BBS inclusion criteria
B)								
AHR2	***BBS3/ARL6***	p.[D179N]; [=]	c.[535G→A]; [=]	NF	MT (Ex8 splice site altered]	Reunion	Yes. BBS3, JBTS3	No
AIY87	***BBS9***	p.[P641S]; [=]	c.[1921C→T]; [=]	NF	S, PP, MT	NA	Yes: BBS5, ALMS1, NPHP12, NPHP5, BBS14	No
AKE98	***ALMS1***	p.[P2679L]; [P2679L]	c.[8036C→T]; [8036C→T]	NF	S, P	Melanesia	Yes: BBS12, BBS9, CEP290, NPHP1, MKS1	Yes
	NPHP4	p.[V1228G]; [=]	c.[3683T→G]; [=]	0.10%	S, P			
	NPHP3	p.[A519G]; [=]	c.[1556C→G]; [=]	NF	P, MT			
	INVS/NPHP2	p.[N1061Kfs*20]; [N1061Kfs*20]	c.[3176_3177insA]; [3176_3177insA]	NF	–			
AIX45	***TMEM67/ NPHP11***	p.[Y486C]; [=]	c.[1457A→G]; [=]	0.01%	P, MT	NA	Yes	Yes
AMA28	***MKKS/BBS6***	**p.[A242S]; [=]**	**c.[724G→T]; [=]**	0.58%	S, P	France	No	Yes
	NPHP1	p.[E569G]; [=]	c.[1706A→G]; [=]	NF	S, P, MT			
AMO77	***TTC21B/ NPHP12***	p.[R863W]; [=]	c.[2587C→T]; [=]	0.23%	S, P, MT	NA	NA	Yes
AHL86	***AHI1/JBTS3***	p.[R830W]; [=]	c.[2488C→T]; [=]	2.83%	S, P, MT	NA	NA	No
	INVS/NPHP2	p.[T122P]; [=]	c.[364A→C]; [=]	NF	S, P, MT			
ALA60						Caucasian	No	Yes
AKT96						France	No	No
AMT10						France	No	Yes
AIF95						NA	No	No
AKK32						France	No	No

*Patients from consanguineous families, but not genotyped on affymetrix SNP arrays. **In bold**: previously reported mutations in other studies. Mutation prediction software equivalents: possibly/probably damaging (polyphen2), damaging (SIFT), disease causing (mutation taster). Ex, exon; S, SIFT; PP, PolyPhen2; MT, mutation T@ster.

(?): The exact nature of the second heterogeneous mutation could not be identified by high-throughput sequencing. An abnormal loss of coverage is observed at the very end of *BBS10*. The exact nature of this apparently complex mutation is still under investigation by direct sequencing, but appears to involve an alu insertion coupled with a duplication/inversion. Amino acid conservation of non-reported missense mutations is shown in supplementary figure S4. SV: Sanger validation; S: segregation validated.

Consanguinity was documented by clinicians, and in most cases, BBS patients from consanguineous families were genotyped on Affymetrix 250k SNP arrays. When so, BBS and other targeted genes located within homozygous regions are thus indicated in ( ).

consanguineous family, while several other mutations were identified in recessive forms of limb girdle muscular dystrophy.[24]

One noteworthy result is the finding of homozygous truncating *ALMS1* mutations in 3/38 patients (AIA84, AKO26, ALB64; 7.9%). In particular, the nonsense found in AKO26 patient p.R3629* seems to be a recurrent *ALMS1* mutation, since already reported in five other ALMS patients.[25–27] The phenotypic overlap between BBS and ALMS seems to be larger than previously thought, as recently suggested with examples of Alström patients with mutations in BBS genes,[5] and the reverse situation, such as in our study, of *ALMS1* mutations in patient with suspected BBS.[3]

Lastly, no clearly pathogenic mutation was found in any NPHP or JBTS genes in the cohort.

Correlation between mutation detection efficiency and clinical phenotype

Comparison of clinical phenotype between patients with two clearly pathogenic mutated alleles (n=26) and those with either a single possible pathogenic variant or no suspicious variant detected (n=12) showed a clear correlation between the number of major BBS clinical features and the probability of detecting two BBS mutated alleles in patients (figure 3). Biallelic mutations were detected in 81% (CI (60% to 92%)) of patients meeting BBS inclusion criteria. In particular, in Tunisian patients recruited upon strict clinical criteria, mutations were found in 11/11 cases and in seven different BBS genes, ruling out a potential founder effect. On the contrary, for some of the 12 patients without clear mutations, BBS was only one suspected diagnosis among others. Furthermore, our initial selection of patients without recurrent mutations in *BBS1* or *BBS10*, and without any mutation in *BBS12* may have enriched our cohort in patients with non-typical BBS phenotypes. The current widely quoted estimation that known BBS genes account for only 70–75% of the total mutation load in BBS patients may thus be underestimated if considering only patients with strictly defined BBS phenotype.

The distribution of BBS inclusion features appears different between patients with two BBS mutations, two *ALMS1* mutations or no biallelic mutation identified (table 4). Patients with no detected mutation presented with significantly less polydactyly, a major BBS clinical sign: only 25% versus 70% in patients with detected BBS mutations (p=0.029*). The other clinical features seem to follow the trend of classical BBS patients.

Regarding *ALMS1*-mutated probands, 2/3 had been sent for suspected BBS (Prader-Willi or ALMS were also considered for AIA84 and AKO26, respectively) and satisfied BBS diagnostic criteria; the last one (ALB64) was addressed for syndromic retinal dystrophy (table 4).[6] AIA84 presents a classical BBS with retinal dystrophy, obesity, cognitive defects, hypogonadism and brachydactyly. AKO26 presents an atypical BBS with the same features along with abnormal severe deafness, specific for ALMS. Lastly, ALB64 presents a typical ALMS with severe deafness and retinal dystrophy. None of them presented with polydactyly. As previously suggested,[3] both, the absence of polydactyly and the prevalence of deafness in *ALMS1*-mutated patients, are keys for genotype-phenotype discrimination between ALMS and BBS mutated patients.

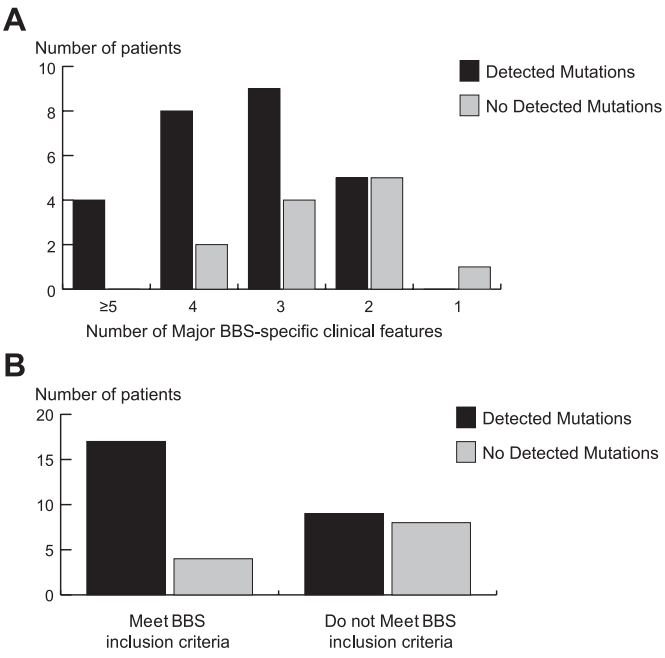


Figure 3 Compliance with classical BBS phenotype is positively correlated to the efficiency to detect principal mutations in BBS genes. (A) The number of BBS diagnostic major inclusion criteria[6] in patients is correlated to an efficient detection of BBS mutations. (B) Efficiency of detecting mutation in patients fulfilling BBS the phenotypic inclusion criteria or not. BBS inclusion criteria presenting with three major features plus at least two minors, or presenting with four major features and more.[6] Primary criteria include: rod-cone dystrophy, polydactyly, obesity, learning disabilities, hypogonadism and renal anomalies. Secondary features comprise speech delay, other eye anomalies, brachydactyly or syndactyly, ataxia, diabetes, developmental delay, dental anomalies, cardiac anomalies and hepatic fibrosis.[6]

Assessment of oligogenism in BBS

The presence and potential effect of triallelism or oligogenism in BBS has been widely discussed and appears controversial ([18 20 28 29] vs[30–32]). In our approach, the simultaneous sequencing of all 16 BBS genes, and of 14 other genes involved in overlapping ciliopathies, allows the systematic detection of most additional potentially pathogenic variants in those genes and, consequently, an unbiased assessment of oligogenism.

Out of the 52 patients analysed, we found only one heterozygous truncating mutation (p.K1870Nfs*4) as a third allele in *BBS14/CEP290* in patient AKR68 who carries a pathogenic missense mutation in *BBS10*. Such a frequency is in fact in the range of what to expect by chance. We previously calculated the probability of carrying a true BBS-pathogenic mutation to be about 1:50.[31] Since we also included in our design other ciliopathy genes, the probability to carry a pathogenic mutation in one of the 30 genes is rather between 1:20 and 1:30 (calculation based on each disease incidence and reported contributions of targeted genes in the mutation load). Potentially, pathogenic heterozygous missenses (not previously reported in patients or in the EVSdb) were also found in eight patients (three of the *proof-of-principle*, table S2; five of the 'unknown' cohort, table 3). Such variants might act as modifiers, but it is unlikely that they are required for full expression of a classical BBS phenotype. Conversely, in some patients where a single clearly pathogenic mutation was found, variants in other genes of the same pathway (especially those encoding proteins of the same complex, such as the BBSome or the BBS-chaperonin complex)[33 34] might contribute to the disease state in a digenic mode of inheritance proposed in few BBS families.[20 35 36]

Potential case of triallelism is illustrated by patient AIZ62, who is compound heterozygous or a nonsense p.E191* and a missense p.A242S in BBS6/MKKS. The pathogenicity of A242S variant has

Table 4 Report of major BBS clinical features in the 38 patients without previously known molecular diagnosis, with or without detected mutations

	Retinitis pigmentosa	Obesity	Polydactyly	Brachydactyly	Hypogonadism	Cystic kidney	Mild intellectual disabilities
2 BBS mutations	95.65% (22/23)	86.96% (20/23)	69.57% (16/23)	43.48% (10/23)	39.13% (9/23)	21.74% (5/23)	73.91% (17/23)
2 ALMS1 mutations	3/3	1/3 (AIA84)	0/3	2/3 (AIA84, AKO26)	2/3 (AIA84, AKO26)	1/3 (ALB64)	2/3 (AIA84, AKO26)
0 or 1 mutation	81.82% (9/11)	81.82% (9/11)	25.00% (3/12)	33.33% (4/12)	41.67% (5/12)	50% (5/10)	81.82% (9/11)

ALMS1 patients: AIA84 was addressed to Strasbourg Diagnostic Laboratory for Bardet–Biedl or Prader-Willi syndromes, AKO26 for suggestive BBS or ALMS with abnormal cognitive defects and ALB64 for a syndromic retinal dystrophy or suggested ALMS. Patients with a proposed Alström syndrome presented with early deafness: at 5 (ALB64) or 6 (AKO26) years of age.

been a subject of discussion.[37–39] Analysis in zebrafish indicated that it affects BBS6/MKKS function and suggested a dominant negative effect.[40] EVSdb allows to infer its frequency at 0.59% (CI (0.43 to 0.80%)), higher than the most frequent BBS mutation, M390R in BBS1 (0.24%, CI (0.15 to 0.39%)). A242S cannot thus be a highly penetrant mutation since it should then be found more frequently in patients than the M390R mutation, which is not the case. In patient AIZ62, a third heterozygous variant was identified in BBS12 (p.Q620R, residue conserved in mammals, but not in more distant vertebrates) thus affecting another subunit of the BBS-chaperonin complex.[34] We suggest that A242S is a hypomorphic allele that may lead to a phenotype when in *trans*, with a complete null mutation, and could be further potentiated by a hypomorphic allele affecting another subunit of the same complex. Segregation analysis in AIZ62 family could not be performed to test this hypothesis.

Lastly, we looked in our cohort for the allelic frequency of the previously proposed BBS-modifier variant c.330C→T in *CCDC28B/MGC1203*,[18] and found a frequency of 3.85% (CI (1.56 to 9.47%)), which is not significantly different (p=0.17) from the 2.07% (CI (1.76 to 2.43%)) observed in EVSdb.

CONCLUSIONS

The extensive non-allelic genetic heterogeneity of Bardet-Biedl syndrome has been a major problem for molecular diagnostic and genetic counselling applications. Various strategies have been proposed in recent years to optimise mutation detection,[5 10–12] but have either low sensitivity or are too time consuming and expensive in diagnostic settings. This problem is shared by other disease entities, such as cardiomyopathies, hearing loss, Usher syndrome or Charcot-Marie tooth neuropathies. Those NGS-based alternative strategies can be divided in whole genome/exome sequencing,[41–43] or targeted sequencing, either by using multiplex PCR,[12] multiple singleplex PCR[44–46] or capture enrichment approaches.

We implemented an *in-solution* targeted capture strategy for the 16 known BBS genes and 14 other genes implicated in ciliopathies that share overlapping clinical features with BBS. We show here that this is very efficient since in a single sequencing lane of an Illumina GAIIx one can simultaneously analyse more than 99.4% of targeted protein coding sequences of these genes in 12 patients, with sufficient coverage to guarantee reliable detection of heterozygous variants, small indels and exon deletions. In a week-long run, one could thus potentially analyse 96 patients. Investigation of 36 patients could be completed in about 3 weeks when including sample preparation and initial bioinformatic analysis. Data analysis and further mutation validation takes a longer time for cases where only variants of uncertain pathogenicity are present. We estimated the overall consumable costs in our settings at about $600. This can be even further decreased by using the latest more powerful sequencers, allowing analysis of larger pools of barcoded samples while keeping a high depth of coverage. Indeed, we recently sequenced a new cohort of 24 patients (using 12 capture reactions) in a single lane of an Illumina HiSeq2000 with even higher coverage than the one obtained in previous experiments with the GAIIx (data not shown). While exome sequencing is clearly more exhaustive, in terms of gene coverage in current implementations, 20–30% of targeted exons are not sufficiently covered for diagnostic accuracy, that is, to ensure low rates of false-positive/false-negative findings and reliable detection of heterozygous mutations or exon deletions. Moreover, the informatics resources needed for exome/genome sequencing data analysis and storage are considerably more important than for targeted sequencing, and can often be a limitation.

This strategy, however, presents some limited pitfalls. Few protein coding regions were not well covered, either because of failure in bait design (presence of repeat motifs) or poor capture efficiency (mostly GC-rich sequences and first exons). Then, our protocol with initial barcoding of libraries followed by capture on pooled samples may be cost effective, but at present, limits the capture efficiency and needs further optimisation, especially if applied to larger gene sets. Finally, like for all targeted exon strategies, deep intronic mutations will be missed. The alternative to targeting entire genes would still miss a high proportion of intronic regions containing repetitive sequences, and would also disproportionately increase the number of rare variants to analyse with splice-prediction bioinformatic tools that are currently not highly reliable. For genes expressed in leucocytes or fibroblasts, another alternative would be selected RNA sequencing (enriched in cognate gene transcripts).

While similar capture strategies have been recently developed for other diseases, most of them included a much smaller cohort, and reported only *proof-of-principle* analysis,[47–49] with the exception of Walsh *et al*.[50] Multiplex PCR approaches may have the potential of covering exons more exhaustively,[46] given that primer design is more flexible than hybridisation bait tiling, but is limited to smaller gene panels than for exon capture. PCR pooling without barcoding has been used for BBS and NPHP,[12 51] a strategy which may be cost effective for analysing large numbers of samples in epidemiological studies, but appears unsuited for diagnosis where a key preoccupation is to limit false-positive/negative rates.

Regarding the specific case of BBS, our study suggests that when strict clinical criteria are complied with, the frequency of detected mutations is higher than the generally quoted 70% figure.[7 11] There may thus be only few strict BBS genes remaining to be identified, especially considering that with most strategies, with the exception of whole-gene sequencing, one will miss deep intronic pathogenic mutations. Additional genes to be discovered may correspond rather to variant BBS-like phenotypes than to strictly defined BBS. Indeed, for patients with *BBS16/SDCCAG8* mutations, genotype-phenotype analysis showed for the first time a clear departure from the typical BBS phenotype with absence of polydactyly and systematic and severe renal manifestations (usually present in only 30% of BBS patients).[7 8] Our finding of *ALMS1* mutations in three patients

Web resources

The URLs for online tools and data presented herein are:
OMIM: http://www.omim/org/
e-Array (Agilent Technologies): https://earray.chem.agilent.com/earray/
UCSC: http://genome.ucsc.edu/
SVA: www.svaproject.org/
dbSNP: http://www.ncbi.nlm.nih.gov/projects/SNP/
Mutation Nomenclature: http://www.hgvs.org/mutnomen/recs.html
Exome Variant Server, NHLBI Exome Sequencing Project (ESP): http://evs.gs.washington.edu/EVS/
Integrative Genomics Viewer (IGV): http://www.broadinstitute.org/igv/
Polyphen 2: http://genetics.bwh.harvard.edu/pph2/
Sift Human Protein: http://sift.jcvi.org/www/SIFT_enst_submit.html
Mutation taster: http://www.mutationtaster.org/

confirms the major clinical overlap with BBS. Finally, we found no evidence for triallelism in our cohort of BBS patients.

Author affiliations
[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR7104, INSERM U964, Université de Strasbourg, Illkirch, France
[2]Microarrays and Sequencing Platform, IGBMC, Illkirch, France
[3]Laboratory of Human genetics, University of Medicine of Tunis, Tunis, Tunisia
[4]Bioinformatics Platform, IGBMC, Illkirch, France
[5]Laboratoire de Génétique Médicale EA3949 Inserm Avenir, Université de Strasbourg, Strasbourg, France
[6]Laboratoire de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
[7]Istituto di Genetica Medica, Universita' Cattolica, Roma, Italy
[8]CHU Bordeaux, University of Bordeaux, Department of Medical Genetics, Bordeaux, France
[9]Service de Génétique Clinique, Hôpital Jeanne de Flandre, CHRU de Lille, Lille, France
[10]Service de Cytogénétique Constitutionnelle, Hospices Civils de Lyon, CBPE, Bron Cedex, France
[11]Department of Genetics, INSERM U676, Assistance Publique Hôpitaux de Paris (AP-HP), Robert Debré University Hospital, Paris, France
[12]Service de Génétique Médicale, Centre de Référence pour les Affections Rares en Génétique Ophtalmologique (CARGO), Hôpitaux Universitaires de Strasbourg, Strasbourg, France
[13]Chaire de Génétique Humaine, Collége de France, Illkirch, France

Acknowledgements We thank Ngoc-Hoan Nguyen for his help in the bioinformatics setup, and Cécile Pizot for the development of VaRank. We warmly thank Géraldine Greff, Anne-Sophie Jaeger, Manuela Antin, Elisabeth Scherrer, Serge Vicaire and Muriel Philipps for their technical assistance. Lastly, we wish to thank all patients and families included in this study, and our clinician colleagues who addressed patients for diagnostic analysis and provided clinical information, in particular: Drs P Abou-Jaoudé, C Baumann, H Flodrops, T Frébourg, D Genevieve, A Goldenberg, B Leheup, P Parent, P Petitjean, C Poitou-Bernert and S Taque.

Contributors CR, JLM and JM designed the study; CR, OM, CS, BJ performed experiments; CR, SLG, VG, JM performed the bioinformatics studies; CR, SLG, OM, VG, CS, JLM and JM collected and analysed data; OM, CS, MCV, PC, DL, IO, FP, MT, AV, HBC and HD provided DNA samples and clinical information; CR, JLM and JM wrote the manuscript; MCV, BJ, HBC, HD, JLM and JM provided technical support, conceptual advice and project coordination.

Funding This work was partially supported by a grant from Agence de Biomédecine to JLM and JM, by funds from APLM and by the Association Française contre les Myopathies (AFM) thanks to its support to the IGBMC sequencing platform.

Competing interests None.

Patient consent For children, consent was signed by the parents.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Zaghloul NA,** Katsanis N. Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy. *J Clin Invest* 2009;**119**:428—37.
2. **Hildebrandt F,** Benzing T, Katsanis N. Ciliopathies. *N Engl J Med* 2011;**364**:1533—43.
3. **Aliferis K,** Helle S, Gyapay G, Duchatelet S, Stoetzel C, Mandel JL, Dollfus H. Differentiating alstrom from Bardet-Biedl syndrome (BBS) using systematic ciliopathy genes sequencing. *Ophthalmic Genet* 2011;**33**:18—22.
4. **Karmous-Benailly H,** Martinovic J, Gubler MC, Sirot Y, Clech L, Ozilou C, Auge J, Brahimi N, Etchevers H, Detrait E, Esculpavit C, Audollent S, Goudefroye G, Gonzales M, Tantau J, Loget P, Joubert M, Gaillard D, Jeanne-Pasquier C, Delezoide AL, Peter MO, Plessis G, Simon-Bouy B, Dollfus H, Le Merrer M, Munnich A, Encha-Razavi F, Vekemans M, Attie-Bitach T. Antenatal presentation of Bardet-Biedl syndrome may mimic Meckel syndrome. *Am J Hum Genet* 2005;**76**:493—504.
5. **Pereiro I,** Hoskins BE, Marshall JD, Collin GB, Naggert JK, Pineiro-Gallego T, Oitmaa E, Katsanis N, Valverde D, Beales PL. Arrayed primer extension technology simplifies mutation detection in Bardet-Biedl and Alstrom syndrome. *Eur J Hum Genet* 2011;**19**:485—8.
6. **Beales PL,** Elcioglu N, Woolf AS, Parker D, Flinter FA. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *J Med Genet* 1999;**36**:437—46.
7. **Deveault C,** Billingsley G, Duncan JL, Bin J, Theal R, Vincent A, Fieggen KJ, Gerth C, Noordeh N, Traboulsi EI, Fishman GA, Chitayat D, Knueppel T, Millan JM, Munier FL, Kennedy D, Jacobson SG, Innes AM, Mitchell GA, Boycott K, Heon E. BBS genotype-phenotype assessment of a multiethnic patient cohort calls for a revision of the disease definition. *Hum Mutat* 2011;**32**:610—19.
8. **Schaefer E,** Zaloszyc A, Lauer J, Durand M, Stutzmann F, Perdomo-Trujillo Y, Redin C, Bennouna Greene V, Toutain A, Perrin L, Gerard M, Caillard S, Bei X, Lewis RA, Christmann D, Letsch J, Kribs M, Mutter C, Muller J, Stoetzel C, Fischbach M, Marion V, Katsanis N, Dollfus H. Mutations in SDCCAG8/NPHP10 cause Bardet-Biedl syndrome and are associated with penetrant renal disease and absent polydactyly. *Mol Syndromology* 2011;**1**:273—81.
9. **Marshall JD,** Maffei P, Beck S, Barrett TG, Paisey RB. Clinical utility gene card for: Alstrom syndrome. *Eur J Hum Genet* 2011;**19**.
10. **Billingsley G,** Deveault C, Heon E. BBS mutational analysis: a strategic approach. *Ophthalmic Genet* 2011;**32**:181—7.
11. **Muller J,** Stoetzel C, Vincent MC, Leitch CC, Laurier V, Danse JM, Helle S, Marion V, Bennouna-Greene V, Vicaire S, Megarbane A, Kaplan J, Drouin-Garraud V, Hamdani M, Sigaudy S, Francannet C, Roume J, Bitoun P, Goldenberg A, Philip N, Odent S, Green J, Cossee M, Davis EE, Katsanis N, Bonneau D, Verloes A, Poch O, Mandel JL, Dollfus H. Identification of 28 novel mutations in the Bardet-Biedl syndrome genes: the burden of private mutations in an extensively heterogeneous disease. *Hum Genet* 2010;**127**:583—93.
12. **Janssen S,** Ramaswami G, Davis EE, Hurd T, Airik R, Kasanuki JM, Van Der Kraak L, Allen SJ, Beales PL, Katsanis N, Otto EA, Hildebrandt F. Mutation analysis in Bardet-Biedl syndrome by DNA pooling and massively parallel resequencing in 105 individuals. *Hum Genet* 2011;**129**:79—90.
13. **Medvedev P,** Stanciu M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. *Nat Methods* 2009;**6**(11 Suppl): S13—20.
14. **Nord AS,** Lee M, King MC, Walsh T. Accurate and exact CNV identification from targeted high-throughput sequence data. *BMC Genomics* 2011;**12**:184.
15. **Anders S,** Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;**11**:R106.
16. **Benjamini Y,** Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodological)* 1995;**57**:289—300.
17. **Simms RJ,** Hynes AM, Eley L, Sayer JA. Nephronophthisis: a genetically diverse ciliopathy. *Int J Nephrol* 2011;**2011**:527137.
18. **Badano JL,** Leitch CC, Ansley SJ, May-Simera H, Lawson S, Lewis RA, Beales PL, Dietz HC, Fisher S, Katsanis N. Dissection of epistasis in oligogenic Bardet-Biedl syndrome. *Nature* 2006;**439**:326—30.
19. **Mykytyn K,** Braun T, Carmi R, Haider NB, Searby CC, Shastri M, Beck G, Wright AF, Iannaccone A, Elbedour K, Riise R, Baldi A, Raas-Rothschild A, Gorman SW, Duhl DM, Jacobson SG, Casavant T, Stone EM, Sheffield VC. Identification of the gene that, when mutated, causes the human obesity syndrome BBS4. *Nat Genet* 2001;**28**:188—91.
20. **Katsanis N,** Ansley SJ, Badano JL, Eichers ER, Lewis RA, Hoskins BE, Scambler PJ, Davidson WS, Beales PL, Lupski JR. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* 2001;**293**:2256—9.
21. **Leitch CC,** Zaghloul NA, Davis EE, Stoetzel C, Diaz-Font A, Rix S, Alfadhel M, Lewis RA, Eyaid W, Banin E, Dollfus H, Beales PL, Badano JL, Katsanis N. Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. *Nat Genet* 2008;**40**:443—8.
22. **Kim SK,** Shindo A, Park TJ, Oh EC, Ghosh S, Gray RS, Lewis RA, Johnson CA, Attie-Bittach T, Katsanis N, Wallingford JB. Planar cell polarity acts through septins to control collective cell movement and ciliogenesis. *Science* 2010;**329**:1337—40.
23. **Otto EA,** Hurd TW, Airik R, Chaki M, Zhou W, Stoetzel C, Patil SB, Levy S, Ghosh AK, Murga-Zamalloa CA, van Reeuwijk J, Letteboer SJ, Sang L, Giles RH, Liu Q, Coene KL, Estrada-Cuzcano A, Collin RW, McLaughlin HM, Held S, Kasanuki JM, Ramaswami G, Conte J, Lopez I, Washburn J, Macdonald J, Hu J, Yamashita Y, Maher ER, Guay-Woodford LM, Neumann HP, Obermuller N, Koenekoop RK, Bergmann C, Bei X, Lewis RA, Katsanis N, Lopes V, Williams DS, Lyons RH, Dang CV, Brito DA, Dias MB, Zhang X, Cavalcoli JD, Nurnberg G, Nurnberg P, Pierce EA, Jackson PK, Antignac C, Saunier S, Roepman R, Dollfus H, Khanna H, Hildebrandt F.

Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat Genet* 2010;**42**:840—50.
24. **Schoser BG,** Frosk P, Engel AG, Klutzny U, Lochmuller H, Wrogemann K. Commonality of TRIM32 mutation in causing sarcotubular myopathy and LGMD2H. *Ann Neurol* 2005;**57**:591—5.
25. **Bond J,** Flintoff K, Higgins J, Scott S, Bennet C, Parsons J, Mannon J, Jafri H, Rashid Y, Barrow M, Trembath R, Woodruff G, Rossa E, Lynch S, Sheilds J, Newbury-Ecob R, Falconer A, Holland P, Cockburn D, Karbani G, Malik S, Ahmed M, Roberts E, Taylor G, Woods CG. The importance of seeking ALMS1 mutations in infants with dilated cardiomyopathy. *J Med Genet* 2005;**42**:e10.
26. **Marshall JD,** Hinman EG, Collin GB, Beck S, Cerqueira R, Maffei P, Milan G, Zhang W, Wilson DI, Hearn T, Tavares P, Vettor R, Veronese C, Martin M, So WV, Nishina PM, Naggert JK. Spectrum of ALMS1 variants and evaluation of genotype-phenotype correlations in Alstrom syndrome. *Hum Mutat* 2007;**28**:1114—23.
27. **Minton JA,** Owen KR, Ricketts CJ, Crabtree N, Shaikh G, Ehtisham S, Porter JR, Carey C, Hodge D, Paisey R, Walker M, Barrett TG. Syndromic obesity and diabetes: changes in body composition with age and mutation analysis of ALMS1 in 12 United Kingdom kindreds with Alstrom syndrome. *J Clin Endocrinol Metab* 2006;**91**:3110—16.
28. **Katsanis N.** The oligogenic properties of Bardet-Biedl syndrome. *Hum Mol Genet* 2004;**1**:R65—71.
29. **Katsanis N,** Eichers ER, Ansley SJ, Lewis RA, Kayserili H, Hoskins BE, Scambler PJ, Beales PL, Lupski JR. BBS4 is a minor contributor to Bardet-Biedl syndrome and may also participate in triallelic inheritance. *Am J Hum Genet* 2002;**71**:22—9.
30. **Hichri H,** Stoetzel C, Laurier V, Caron S, Sigaudy S, Sarda P, Hamel C, Martin-Coignard D, Gilles M, Leheup B, Holder M, Kaplan J, Bitoun P, Lacombe D, Verloes A, Bonneau D, Perrin-Schmitt F, Brandt C, Besancon AF, Mandel JL, Cossee M, Dollfus H. Testing for triallelism: analysis of six BBS genes in a Bardet-Biedl syndrome family cohort. *Eur J Hum Genet* 2005;**13**:607—16.
31. **Laurier V,** Stoetzel C, Muller J, Thibault C, Corbani S, Jalkh N, Salem N, Chouery E, Poch O, Licaire S, Danse JM, Amati-Bonneau P, Bonneau D, Megarbane A, Mandel JL, Dollfus H. Pitfalls of homozygosity mapping: an extended consanguineous Bardet-Biedl syndrome family with two mutant genes (BBS2, BBS10), three mutations, but no triallelism. *Eur J Hum Genet* 2006;**14**:1195—203.
32. **Smaoui N,** Chaabouni M, Sergeev YV, Kallel H, Li S, Mahfoudh N, Maazoul F, Kammoun H, Gandoura N, Bouaziz A, Nouiri E, M'Rad R, Chaabouni H, Hejtmancik JF. Screening of the eight BBS genes in Tunisian families: no evidence of triallelism. *Invest Ophthalmol Vis Sci* 2006;**47**:3487—95.
33. **Nachury MV,** Loktev AV, Zhang Q, Westlake CJ, Peranen J, Merdes A, Slusarski DC, Scheller RH, Bazan JF, Sheffield VC, Jackson PK. A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell* 2007;**129**:1201—13.
34. **Seo S,** Baye LM, Schulz NP, Beck JS, Zhang Q, Slusarski DC, Sheffield VC. BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly. *Proc Natl Acad Sci U S A* 2010;**107**:1488—93.
35. **Badano JL,** Kim JC, Hoskins BE, Lewis RA, Ansley SJ, Cutler DJ, Castellan C, Beales PL, Leroux MR, Katsanis N. Heterozygous mutations in BBS1, BBS2 and BBS6 have a potential epistatic effect on Bardet-Biedl patients with two mutations at a second BBS locus. *Hum Mol Genet* 2003;**12**:1651—9.
36. **Beales PL,** Badano JL, Ross AJ, Ansley SJ, Hoskins BE, Kirsten B, Mein CA, Froguel P, Scambler PJ, Lewis RA, Lupski JR, Katsanis N. Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. *Am J Hum Genet* 2003;**72**:1187—99.
37. **Andersen KL,** Echwald SM, Larsen LH, Hamid YH, Glumer C, Jorgensen T, Borch-Johnsen K, Andersen T, Sorensen TI, Hansen T, Pedersen O. Variation of the McKusick-Kaufman gene and studies of relationships with common forms of obesity. *J Clin Endocrinol Metab* 2005;**90**:225—30.
38. **Slavotinek AM,** Searby C, Al-Gazali L, Hennekam RC, Schrander-Stumpel C, Orcana-Losa M, Pardo-Reoyo S, Cantani A, Kumar D, Capellini Q, Neri G, Zackai E, Biesecker LG. Mutation analysis of the MKKS gene in McKusick-Kaufman syndrome and selected Bardet-Biedl syndrome patients. *Hum Genet* 2002;**110**:561—7.
39. **Stone DL,** Slavotinek A, Bouffard GG, Banerjee-Basu S, Baxevanis AD, Barr M, Biesecker LG. Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome. *Nat Genet* 2000;**25**:79—82.
40. **Zaghloul NA,** Liu Y, Gerdes JM, Gascue C, Oh EC, Leitch CC, Bromberg Y, Binkley J, Leibel RL, Sidow A, Badano JL, Katsanis N. Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. *Proc Natl Acad Sci U S A* 2010;**107**:10602—7.
41. **Choi M,** Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 2009;**106**:19096—101.
42. **Lupski JR,** Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 2010;**362**:1181—91.
43. **Walker RH,** Schulz VP, Tikhonova IR, Mahajan MC, Mane S, Arroyo Muniz M, Gallagher PG. Genetic diagnosis of neuroacanthocytosis disorders using exome sequencing. *Mov Disord* 2011;**27**:539—43.
44. **Hopp K,** Heyer CM, Hommerding CJ, Henke SA, Sundsbak JL, Patel S, Patel P, Consugar MB, Czarnecki PG, Gliem TJ, Torres VE, Rossetti S, Harris PC. B9D1 is revealed as a novel Meckel syndrome (MKS) gene by targeted exon-enriched next-generation sequencing and deletion analysis. *Hum Mol Genet* 2011;**20**:2524—34.
45. **Jones MA,** Bhide S, Chin E, Ng BG, Rhodenizer D, Zhang VW, Sun JJ, Tanner A, Freeze HH, Hegde MR. Targeted polymerase chain reaction-based enrichment and next generation sequencing for diagnostic testing of congenital disorders of glycosylation. *Genet Med* 2011;**13**:921—32.
46. **Mondal K,** Shetty AC, Patel V, Cutler DJ, Zwick ME. Targeted sequencing of the human X chromosome exome. *Genomics* 2011;**98**:260—5.
47. **Brownstein Z,** Friedman LM, Shahin H, Oron-Karni V, Kol N, Rayyan AA, Parzefall T, Lev D, Shalev S, Frydman M, Davidov B, Shohat M, Rahile M, Lieberman S, Levy-Lahad E, Lee MK, Shomron N, King MC, Walsh T, Kanaan M, Avraham KB. Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in middle eastern families. *Genome Biol* 2011;**12**:R89.
48. **Meder B,** Haas J, Keller A, Heid C, Just S, Borries A, Boisguerin V, Scharfenberger-Schmeer M, Stahler P, Beier M, Weichenhan D, Strom TM, Pfeufer A, Korn B, Katus HA, Rottbauer W. Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies. *Circ Cardiovasc Genet* 2011;**4**:110—22.
49. **Shearer AE,** DeLuca AP, Hildebrand MS, Taylor KR, Gurrola J 2nd, Scherer S, Scheetz TE, Smith RJ. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proc Natl Acad Sci U S A* 2010;**107**:21104—9.
50. **Walsh T,** Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, Nord AS, Mandell JB, Swisher EM, King MC. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci U S A* 2010;**107**:12629—33.
51. **Otto EA,** Ramaswami G, Janssen S, Chaki M, Allen SJ, Zhou W, Airik R, Hurd TW, Ghosh AK, Wolf MT, Hoppe B, Neuhaus TJ, Bockenhauer D, Milford DV, Soliman NA, Antignac C, Saunier S, Johnson CA, Hildebrandt F. Mutation analysis of 18 nephronophthisis associated ciliopathy disease genes using a DNA pooling and next generation sequencing strategy. *J Med Genet* 2011;**48**:105—16.

Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström Syndromes

Claire Redin, Stéphanie Le Gras, Oussema Mhamdi, et al.

J Med Genet published online July 7, 2012
doi: 10.1136/jmedgenet-2012-100875

Notes

Paper II: Clinical and genetic characterization of Bardet-Biedl syndrome in Tunisia: defining a strategy for molecular diagnosis.

M'hamdi O*, **Redin C***, Stoetzel C, Ouertani I, Chaabouni M, Maazoul F, M'rad R, Mandel J, Dollfus H, Muller J, Chaabouni H. Clin Genet. 2013 Feb

Aim of the study: Application of the previous method on additional Tunisian patients, to better delineate the genetics of BBS in Tunisia and define a new diagnosis strategy incorporating NGS.

Contribution: Major. Performed the design of the capture, the quality controls of DNA samples, the preparation of sequencing libraries, analyzed the processed data.

CLINICAL GENETICS An International Journal of Genetics, Molecular and Personalized Medicine

Clin Genet 2014: 85: 172–177
Printed in Singapore. All rights reserved

CLINICAL GENETICS
doi: 10.1111/cge.12129

Short Report

Clinical and genetic characterization of Bardet–Biedl syndrome in Tunisia: defining a strategy for molecular diagnosis

M'hamdi O., Redin C., Stoetzel C., Ouertani I., Chaabouni M., Maazoul F., M'rad R., Mandel J.L., Dollfus H., Muller J., Chaabouni H. Clinical and genetic characterization of Bardet–Biedl syndrome in Tunisia: defining a strategy for molecular diagnosis.
Clin Genet 2014: 85: 172–177. © John Wiley & Sons A/S. Published by John Wiley & Sons Ltd, 2013

Bardet–Biedl syndrome (BBS, OMIM 209900) is a rare genetic disorder characterized by obesity, retinitis pigmentosa, post axial polydactyly, cognitive impairment, renal anomalies and hypogonadism. The aim of this study is to provide a comprehensive clinical and molecular analysis of a cohort of 11 Tunisian BBS consanguineous families in order to give insight into clinical and genetic spectrum and the genotype–phenotype correlations. Molecular analysis using combined sequence capture and high-throughput sequencing of 30 ciliopathies genes revealed 11 mutations in 11 studied families. Five mutations were novel and six were previously described. Novel mutations included c.1110G>A and c.39delA (p.G13fs*41) in *BBS1*, c.115+5G>A in *BBS2*, c.1272+1G>A in *BBS6*, c.1181_1182insGCATTTATACC in *BBS10* (p.S396Lfs*6). Described mutations included c.436C>T (p.R146*) and c.1473+4A>G in *BBS1*, c.565C> (p.R189*) in *BBS2*, deletion of exons 4–6 in *BBS4*, c.149T>G (p.L50R) in *BBS5*, and c.459+1G>A in *BBS8*; most frequent mutations were described in *BBS1* (4/11, 37%) and *BBS2* (2/11, 18%) genes. No phenotype–genotype correlation was evidenced. This data expands the mutations profile of *BBS* genes in Tunisia and suggests a divergence of the genetic spectrum comparing Tunisian and other populations.

Conflict of interest

The authors have reported no conflicts of interest.

O. M'hamdi[a,†], C. Redin[b,†], C. Stoetzel[c,†], I. Ouertani[a,d], M. Chaabouni[a,d], F. Maazoul[d], R. M'rad[a,d], J.L. Mandel[b,e,f], H. Dollfus[c,g], J. Muller[b,e] and H. Chaabouni[a,d]

[a] Faculté de médecine de Tunis, Université de Tunis El-Manar, Tunis, Tunisie, [b]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR7104, INSERM U964, Université de Strasbourg, Illkirch, France, [c]Faculté de Médecine, Université Louis Pasteur de Strasbourg, Laboratoire de Génétique Médicale EA3949 (Equipe AVENIR-Inserm), Strasbourg, France, [d]Hôpital Charles Nicolle de Tunis, Service des Maladies Congénitales et Héréditaires, Tunis, Tunisie, [e]Hôpitaux universitaires de Strasbourg, Laboratoire de Diagnostic Génétique, Strasbourg, France, [f]Chaire de Génétique Humaine, Collège de France, Illkirch, France, and [g]Hôpitaux Universitaires de Strasbourg, Service de Génétique Médicale/Centre de Référence pour les Affections Génétiques Ophtalmologiques, Strasbourg, France

[†]These authors have equally contributed.

Key words: Bardet–Biedl syndrome – genotype–phenotype correlation – mutation – next generation sequencing

Corresponding author: Prof Habiba Chaabouni, Laboratoire de Génétique Humaine, Faculté de Médecine de Tunis 17 rue Djebel Lakhdar 1007 Tunis, Tunisia.
Tel.: +216 71 563 709/+216 71 570 756;
fax: +216 71 570 553;
e-mail: Habiba.chaabouni@rns.tn

Received 16 August 2012, revised and accepted for publication 18 February 2013

Bardet–Biedl syndrome (BBS) is a pleiotropic autosomal recessive disorder characterized by a wide spectrum of clinical signs including progressive retinal degeneration, polydactyly, obesity, learning difficulties, renal tract and genital anomalies as well as other less frequent features such as anosmia or Hirschprung disease. Clinical diagnosis is established if at least four major features, or three major and two minor features are present (1). The phenotype spectrum of BBS is variable; some manifestations could appear during childhood. BBS is considered as a rare disorder, its prevalence in Tunisia has been estimated to be 1:156,000 (2), and the current prevalence in North America and Europe ranges from 1:140,000 to 1:160,000 live births. Populations with a high rate of consanguinity or from isolated regions were characterized by a high frequency of BBS such as for Koweit (1:17,000) or Newfoundland (1:13,000) (3, 4). BBS is a genetically heterogeneous disorder. Till date, 17 genes have been described (*BBS1–17*). Seven BBS proteins (BBS1, 2, 4, 5, 7, 8 and 9) form a stable complex BBSome (5, 6). Oligogenic inheritance has been described in some BBS families (7–9). Mutations in *BBS1–17* account for 70–80% of affected BBS families (10).

The spectrum of BBS genes mutations in North Africa and the Arab world remains uncertain as only a few molecular studies were reported in these regions. The aim of this report is to provide a comprehensive clinical and molecular analysis of a cohort of 11 consanguineous Tunisian BBS families, to ascertain the phenotype–genotype correlation, and to delineate a strategy for the clinical implementation of molecular diagnosis of Tunisian BBS patients.

Materials and methods

Patients were recruited by the Department of Congenital and Hereditary Disorders at Charles Nicolle Hospital Tunis, Tunisia. Diagnosis of BBS was made when four of the major features, or three primary and two secondary features were present. All BBS Tunisian patients had thorough clinical evaluation that included personal and family histories, physical traits and dysmorphology examination, anthropometric measurements, ophthalmologic evaluation, psychiatric examination, blood glucose, liver and renal function tests. Peripheral blood was collected in EDTA (ethylene diamine tetraacetic acid) tubes from all the affected and unaffected members who were clinically evaluated. Informed consent was obtained from all patients, their families and control group.

Molecular analysis

All patients included in this study were tested for the two recurrent known mutations M390R in *BBS1* and C91Lfs* in *BBS10* by direct sequencing. No mutations were revealed. All patients were analyzed in the second step using the targeted exon capture strategy coupled with high-throughput sequencing as described by Redin et al. (11). A total of 30 genes were targeted including 12 *NPHP* genes, 16 *BBS* genes, the *ALMS1* gene and the *CCDC28B* gene. Only principal mutations were confirmed by direct sequencing on ABI 3130 (Applied Biosystems, Foster City, CA) for the patients and their parents for segregation analysis. The novel variants not described in the Human Genetics Mutation Database (http://www.hgmd.cf.ac.uk/ac) were verified using the 1000 Genome Database (http://browser.1000genomes.org/index.html) and by sequencing of the control group consisting of 96 unrelated individuals of Tunisian origin.

Bioinformatics tools

Different softwares were used for molecular results analysis: Human Splicing Finder v.2.3. (http://www.umd.be/HSF/) and Splice site prediction (http://fruitfly.org/seq_tools/splice.html) were used for analysis of splicing mutations; polyphen2 (http://genetics.bwh.harvard.edu/pph) and sift (http://sift.jcvi.org/) were used to perform *in silico* prediction of the functional impact of missense mutations.

Results

Clinical characterization of BBS families

A cohort of 11 Tunisian consanguineous families with 15 affected members was included in this study. The patients' ages ranged between 4 and 36 years. All patients studied here were from consanguineous families. The pedigrees in all affected members are consistent with the autosomal recessive mode of inheritance. The phenotype spectrum in our cohort BBS families was broader than the classic package as renal abnormalities have been added as new cardinal feature for the clinical diagnosis of BBS (1). All patients showed signs of retinal dystrophy, obesity was present in 93.33%, intellectual disabilities were noted in 80% and renal anomalies were present in 26.66%. Clinical analysis revealed a low degree of interfamilial variability (Table 1). Except family BB-S-F12, where two affected members had different clinical features, one patient had all principle features of BBS, whereas her sister had renal cysts with mild mental retardation and nystagmus.

Molecular characterization of BBS in Tunisian families

Molecular analysis revealed 11 different mutations in *BBS1*, *BBS2*, *BBS4*, *BBS5*, *BBS6*, *BBS8* and *BBS10* genes, five of them were novel and six were previously described. Identified mutations were summarized in the Table 2. Figure 1 showed the multiple sequences alignment of the missense mutation L50R identified in *BBS5* gene. The following results expand the genetic spectrum of BBS genes in Tunisia.

Table 1. Clinical features of 15 Tunisian patients belonging to 11 families

Patient ID	Gender	Age (years)	Obesity	RP	Polydactyly	Hypogenitalism	Renal anomalies	MR	Dental anomalies	Visual impairment out of RP	Hearing loss	Other features
BBS-F1-A	F	5	+	+	+	+	−	−	+	+	−	Neonatal hypotonia
BBS-F2-A	M	6	+	+	+	+	−	+	−	NA	−	Anemia HTA
BBS-F3-A	F	4.5	+	+	+	+	−	−	+	NA	−	Facial palsy
BBS-F7-A	F	8	+	+	+	−	−	+	−	−	−	−
BBS-F8-A	M	15	+	+	+	+	−	+	+	−	−	−
BBS-F8-B	F	11	+	+	−	+	+Dialysis	+	+	+	−	−
BBS-F9-A	F	21	+	+	−	−	−	+	+	−	−	Diabetes mellitus
BBS-F9-B	F	29	+	+	+	−	−	−	−	+	−	−
BBS-F10-A	F	7	+	+	+	+	−	+	+	−	−	−
BBS-F11-A	F	32	+	+	−	+	NA	+	NA	+	−	−
BBS-F11-B	F	40	+	+	−	−	NA	+	−	+	−	−
BBS-F12-A	F	14	−	−	−	−	PKD + dialysis	+	−	+	−	−
BBS-F12-B	M	4	+	+	+	+	+	+	−	−	−	−
BBS-F13-A	F	11	+	+	+	−	−	+	+	+	−	−
BBS-F14-A	M	4	+	+	+	+	+	−	+	−	−	HTA

F, female; HC, head circumference; HTA, arterial hypertension; M, male; MR, mental retardation; NA, not available; PKD, polycystic kidney disease; RP, retinitis pigmentosa; +, present; −, absent.

Table 2. BBS genes mutations and other variants identified in Tunisian BBS patients

Gene	Family	Patient ID	Principal mutations	Predicted effect	Status in parents		Status in patient	Additional identified alleles	Status in patient	SIFT prediction	POLYPHEN2
					Father	Mother					
BBS1	BBS-F9	BBS-F9-A	c.1110G>A (exon 11)	Splice mutation	Htz	Htz	Hmz	–	–	–	–
		BBS-F9-B	c.1110G>A (exon 11)	Splice mutation	Htz	Htz	Hmz	–	–	–	–
	BBS-F12	BBS-F12-A	c.39delA (exon 1)	p.A14Lfs*28	Htz	Htz	Hmz	*BBS3*: c.-688A>G	Htz	–	–
		BBS-F12-B	c.39delA (exon 1)	p.A14Lfs*28	Htz	Htz	Hmz	*BBS3*: c.-688A>G	Htz	–	–
	BBS-F1	BBS-F1-A	c.1473+4A>G (exon 14)	Splice mutation	Htz	Htz	Hmz	*NPHP1*: p.G154S	Htz	Tolerated	Benign
	BBS-F11	BBS-F11-A	c.448C>T (exon 5)	p.R146*	Htz	Htz	Hmz	–	–	–	–
		BBS-F11-B	c.448C>T (exon 5)	p.R146*	Htz	Htz	Hmz	–	–	–	–
		BBS-F11-C	c.448C>T (exon 5)	p.R146*	Htz	Htz	Hmz	–	–	–	–
BBS2	BBS-F7	BBS-F7-A	c.565C>T (exon 5)	p.R189*	Htz	Htz	Hmz	*MARK3*: p.R674H	Htz	Damaging	Probably damaging
	BBS-F2	BBS-F2-A	c.345+5G>A (exon 2)	Splice mutation	Htz	Htz	Hmz	*BBS7*: c.-1503delC	Htz	–	–
								BBS13: c.*1004delC	Htz	–	–
BBS4	BBS-F3	BBS-F3-A	Deletion Ex4-5-6		Htz	Htz	Hmz	*BBS5*: p.N184S	Htz	Damaging	Probably damaging
BBS5	BBS-F13	BBS-F13-A	c.149T>G (exon 3)	p.L50R	Htz	Htz	Hmz	/	/	Damaging	Probably damaging
								BBS16: p.R464K	Htz	Tolerated	Possibly damaging
								BBS6: c.-3436C>T	Htz	–	–
								BBS15: c.-1012C>T	Htz	–	–
BBS6	BBS-F10	BBS-F10-A	c.1272+1G>A (exon 5)	Splice mutation	Htz	Htz	Hmz	*NPHP4*: p.R959Q	Htz	Damaging	Probably damaging
								BBS9: c.*336A>G	Htz	–	–
BBS8	BBS-F14	BBS-F14-A	c.459+1G>A (exon 4)	Splice mutation	Htz	Htz	Hmz	*AHI1*: p.R830W	Htz	Damaging	Probably damaging
								TMEM216: c.*558G>A	Htz	–	–
BBS10	BBS-F8	BBS-F8-A	c.[1171_1181dupGCATTTATACC] (exon2)	p.S396Lfs*6	Htz	Htz	Hmz	BBS7: c.*729del C	Htz	–	–
								AHI1: c.*27C>T	Htz	–	–

BBS, Bardet–Biedl syndrome; Hmz, homozygous; Htz, heterozygous.

Fig. 1. Multiple sequences alignment covering the pathogenic missense mutation L50R identified in *BBS5* gene in the family BBS-F13. Homo, human; Rattus, rat; Mus, mouse; Bos, cow; Xenopus, *Xenopus tropicalis*; Danio, zebrafish. *, position of the missense mutation.

Heterozygous variants detection and triallelism inheritance

Interestingly others variants were detected in 30 analyzed genes in our BBS patients. For all 11 studied families, novel homozygous and heterozygous variants were identified. The high quality of the high-throughput sequencing technology did uncover an important number of variants in 16 BBS genes and the other 14 ciliopathies analyzed genes. Most variants were heterozygous in coding, non-coding and regulatory regions. The majority were non-coding (100/116 or 86%) with exception of nine UTR variants (*BBS3* c.-688A>G, *BBS6* c.-3436C>T, *BBS7* c.*729del C, c.-1503delC, *BBS9* c.*336A>G, *BBS13* c.*1004delC, *BBS15* c.-1012C>T, *TMEM216* c.*558G>A, *AHI1* c.*27C>T). One interesting heterozygous variant p.R830W in *AHI1* gene was detected in the family BBS-F14 (Table 2).

Discussion

We report the clinical evaluation and molecular analysis of 11 Tunisian consanguineous BBS families using the targeted exon-capture strategy coupled with high-throughput sequencing. Clinical features analysis revealed that most affected patients shared the classical features of BBS. Molecular analysis was able to identify mutations in all families, which is higher than previous reports (12–14). This further shows that upon strict inclusion criteria in the current molecular test is extremely powerful.

In our cohort, *BBS1* and *BBS2* were the most commonly mutated genes, respectively 36% and 18%, respectively. *BBS4*, *5*, *6*, *8* and *10* contributed each one for a single family. These results are partially inconsistent with European reports where *BBS1* and *BBS10* were the most commonly involved genes in BBS patients, respectively 21% and 30% (14, 15).

BBS1 gene mutations remain the principal cause of BBS disease in different populations, whereas we find divergence for other BBS genes. Actual data revealed that *BBS2* and *BBS8* genes are more frequently mutated in Tunisian patients (16, 17), whereas *BBS3* and *BBS4* are frequently mutated in Arab Saudi population (18, 19).

A wider cohort is necessary to confirm the results obtained in our smaller subset of the Tunisian BBS patients in order to better characterize the genetic spectrum of this disorder.

The recognition that BBS is a ciliopathy has extended the understanding of the role of the BBS genes (20) as well as the mechanisms of some observed defects. Several studies showed that BBS proteins play a critical role in the cilium biogenesis and ciliary signaling (20–23). Seven BBS conserved proteins (BBS1, 2, 4, 5, 7, 8 and 9) form a stable complex, the BBSome controlling the vesicular transport to the cilium. *BBS7* and *BBS8* genes played a minor role in mutational spectrum of BBS in multiethnic population (24). Molecular analysis of BBS in Tunisian families were in agreement with previously published studies which suggested that *BBS7* gene were infrequently mutated in different populations (10, 16, 17, 19).

The genotype analysis in Tunisian BBS studied families revealed that mutation spectrum of BBS genes involved *BBS1–BBS10* genes. The current studied cohort was small and was selected on strict clinical criteria. No clear genotype–phenotype correlation could be observed. We will extend in the future the clinical and molecular analysis of all Tunisian patients sharing minimum features of BBS, thus to identify the mutational spectrum in our population.

Two mutations p.R189* (*BBS2* gene) in family BBS-F7 and c.459+1G>A (*BBS8* gene) in family BBS-F14 were previously identified in four Tunisian families from different regions of the country (17). Genotyping analysis showed that these patients shared the same haplotype suggesting the hypothesis of the founder mutation for these two mutations in our population. On the other hand the *BBS4* gene mutation identified in family BBS-F3 was reported in one Algerian BBS family (10), suggesting the hypothesis of the founder mutation in North Africa region.

The principal advantage of the recently used approach in the molecular analysis of Tunisian BBS patients (17) was the simultaneously analysis of all 16 BBS genes and other 14 genes implicated in overlapping ciliopathies, allowing the detection of all principal mutations and triallelism. Interestingly, one heterozygous variant p.R830W in *AHI1* gene was detected in the family BBS-F14. This allele was described as a modifier of retinal degeneration risk in patients with nephronophthisis (25). Clearly this approach showed high efficiency in the mutation detection than other strategies used last year such as screening of all defects BBS genes by direct sequencing and homozygosity mapping using single nucleotide polymorphism (SNP) in consanguineous families. Molecular investigation and data analysis were more quickly achieved than other molecular strategies.

In order to fully explore the mutation spectrum in the Tunisian population and to confirm the founder effect in some of them, it will be necessary to extend the molecular analysis to sporadic cases, and by lowering

the inclusion criteria. Herein we present a specific strategy that could be applied to power mutations detection in Tunisian BBS patients. In first step, we suggest the screening of common mutation in *BBS1*, the most frequently mutated gene and the founder BBS mutations in *BBS2* and *BBS8* genes. In second step, we advocate a targeted exon-capture strategy coupled high-throughput sequencing of 30 ciliopathies genes. Exon capture coupled with high-throughput sequencing strategy will be used in the future for diagnosis purpose of BBS.

Acknowledgements

We sincerely thank the patients and their families for their enthusiastic participation. We thank the members of the IGBMC sequencing platform; we also thank all the Tunisian clinicians who referred the patients. This work was supported by a grant from Agence de Biomedicine France, Association Française contre les myopathies (AFM) and the Ministry of High Education and Scientific Research and Technology, Tunisia.

References

1. Beales PL, Elcioglu N, Woolf AS, Parker D, Flinter FA. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J Med Genet 1999: 36 (6): 437–446.
2. M'hamdi O, Ouertani I, Maazoul F, Chaabouni H. Prevalence of Bardet-Biedl syndrome in Tunisia. J Community Genet 2011: 2 (2): 97–99.
3. Moore SJ, Green JS, Fan Y et al. Clinical and genetic epidemiology of Bardet-Biedl-syndrome in Newfoundland: a 22 year prospective, population-based, cohort study. Am J Med Genet A 2005: 132 (4): 352–360.
4. Farag TI, Teebi AS. High incidence of Bardet-Biedl syndrome among the Bedouin. Clin Genet 1989: 36 (6): 463–464.
5. Nachury MV, Loktev AV, Zhang Q et al. A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. Cell 2007: 129 (6): 1201–1213.
6. Lechtreck KF, Johnson EC, Sakai T et al. The Chlamydomonas reinhardtii BBSome is an IFT cargo required for export of specific signaling proteins from flagella. J Cell Biol 2009: 187 (7): 1117–1132.
7. Katsanis N, Ansley SJ, Badano JL et al. Triallelic Inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science 2001: 293 (5538): 2256–2259.
8. Zaghloul NA, Liu Y, Gerdes JM et al. Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. Proc Natl Acad Sci U S A 2010: 107 (23): 10602–10607.
9. Leitch CC, Zaghloul NA, Davis EE. Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. Nat Genet 2008: 40 (4): 443–448.
10. Muller J, Stoetzel C, Vincent MC et al. Identification of 28 novels mutations in the Bardet-Biedl syndrome genes: the burden of private mutations in an extensively heterogeneous disease. Hum Genet 2010: 127 (5): 583–593.
11. Redin C, Le Gras S, Mhamdi O et al. Targeted High-Throughput Sequencing for Molecular Diagnosis of Genetically Heterogeneous Diseases: Efficient mutation detection in BBS and Alstrom syndromes. J Med Genet 2012: 49 (8): 502–512.
12. Hjortshøj TD, Grønskov K, Philp AR et al. Bardet-Biedl syndrome in Denmark; report of 13 novel sequence variations in six genes. Hum Mutat 2010: 31 (4): 429–436.
13. Harville HM, Held S, Diaz-Font A et al. Identification of 11 novel mutations in 8 BBS genes by high-resolution homozygosity mapping. J Med Genet 2010: 47 (4): 262–267.
14. Janssen S, Ramaswami G, Davis EE et al. Mutations analysis in Bardet-Biedl syndrome by DNA pooling and massively parallel resequencing in 105 individuals. Hum Genet 2011: 129 (1): 79–90.
15. Badano JL, Kim JC, Hoskins BE et al. Heterozygous mutations in BBS1, BBS2 and BBS6 have a potential epistasic effect on Bardet-Biedl patients with two mutations at a second BBS locus. Hum Mol Genet 2003: 12 (14): 1651–1659.
16. Chen J, Smaoui N, Hammer MB et al. Molecular analysis of Bardet-Biedl syndrome families: report of 21 novel mutations in 10 genes. Invest Ophthalmol Vis Sci 2011: 52 (8): 5317–5324.
17. Smaoui N, Chaabouni M, Sergeev YV et al. Screening of the eight BBS genes in Tunisian families: no evidence of triallelism. Invest Ophthalmol Vis Sci 2006: 47 (8): 3487–3495.
18. Abu-Safieh L, Al-Anazi S, Al-Abidi L et al. In search of triallelism in Bardet-Biedl Syndrome. Eur J Hum Genet 2012: 20 (4): 420–427.
19. Abu Safiah L, Aldahmesh MA, Shamseldin H et al. Clinical and molecular characterization of Bardet-Biedl syndrome in consanguineous populations: the power of homozygosity mapping. J Med Genet 2010: 47 (4): 236–241.
20. Ansley SJ, Badano JL, Blacque OE et al. Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. Nature 2003: 425 (6958): 628–633.
21. Berbari NF, Lewis JS, Bishop GA, Askwith CC, Mykytyn K. Bardet-Biedl syndrome proteins are required for the localization of G protein-coupled receptors to primary cilia. Proc Natl Acad Sci U S A 2008: 105 (11): 4242–4246.
22. Blacque OE, Reardon MJ, Li C et al. Loss of C. elegans BBS-7 and BBS-8 protein function results in cilia defects and compromised intraflagellar transport. Genes Dev 2004: 18 (13): 1630–1642.
23. Gerdes JM, Liu Y, Zaghloul NA et al. Disruption of the basal body compromises proteasomal function and perturbs intracellular Wnt response. Nat Genet 2007: 39 (11): 1350–1360.
24. Bin J, Madhavan J, Ferrini W, Mok CA, Billingsley G, Héon E. BBS7 and TTC8 (BBS8) mutations play a minor role in the mutational load of Bardet-Biedl syndrome in a multiethnic population. Hum Mutat 2009: 30 (7): 737–746.
25. Louie CM, Caridi G, Lopes VS et al. AHI1 is required for photoreceptor outer segment development and is a modifier for retinal degeneration in nephronophthisis. Nat Genet 2010: 42 (8): 502–512.

Paper III: Mutations in SDCCAG8/NPHP10 Cause Bardet-Biedl Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly.

Schaefer E*, Zaloszyc A, Lauer J, Durand M, Stutzmann F, Perdomo-Trujillo Y, **Redin C**, Bennouna Greene V, Toutain A, Perrin L, Gérard M, Caillard S, Bei X, Lewis RA, Christmann D, Letsch J, Kribs M, Mutter C, Muller J, Stoetzel C, Fischbach M, Marion V, Katsanis N, Dollfus H.. Mol Syndromol. 2011 Sep.

Aim of the study: Report of 5 BBS families carrying mutations in *SDCCAG8/NPHP10*, proposing it as a novel BBS gene *(BBS16*). Reporting also one of the only hints for genotype-to-phenotype correlations in BBS since strikingly no patients present with polydactyly and present with abnormally severe kidney disease.

Contribution: Minor. Performed the statistical analyses for determining which clinical feature was significantly different in *SDCCAG8*/*BBS16* patients compared to other BBS patients. Wrote the different conclusions accordingly.

Original Article

Molecular Syndromology

Mol Syndromol 2010;1:273–281
DOI: 10.1159/000331268

Accepted: July 23, 2011
by A. Rauch
Published online: September 14, 2011

Mutations in *SDCCAG8/NPHP10* Cause Bardet-Biedl Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly

E. Schaefer[a] A. Zaloszyc[a] J. Lauer[a] M. Durand[a] F. Stutzmann[a] Y. Perdomo-Trujillo[b] C. Redin[c] V. Bennouna Greene[a] A. Toutain[d] L. Perrin[e] M. Gérard[e] S. Caillard[f] X. Bei[g] R.A. Lewis[h] D. Christmann[i] J. Letsch[b] M. Kribs[j] C. Mutter[k] J. Muller[c, l] C. Stoetzel[a] M. Fischbach[m] V. Marion[a] N. Katsanis[g] H. Dollfus[a, b]

[a]Laboratoire de Génétique Médicale EA 3949, Equipe Avenir-Inserm, Faculté de Médecine de Strasbourg, Université de Strasbourg, [b]Centre de Référence pour les Affections Rares en Génétique Ophtalmologique (CARGO), Hôpitaux Universitaires de Strasbourg, Strasbourg, [c]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Département de Neurobiologie et Génétique, Laboratoire de Bioinformatique et Génomique Intégratives, Illkirch-Graffenstaden, [d]Service de Génétique, Hôpitaux Universitaires de Tours, Tours, [e]Département de Génétique, Assistance Publique Hôpitaux de Paris – CHU Robert Debré, Paris, et [f]Service de Néphrologie, Dialyse et Transplantation Rénale, Hôpitaux Universitaires de Strasbourg, Strasbourg, France; [g]Center for Human Disease Modeling, Duke University Medical Center, Durham, N.C., and [h]Department of Ophthalmology, Baylor College of Medicine, Houston, Tex., USA; [i]Service de Radiologie 2, Hôpitaux Universitaires de Strasbourg, Strasbourg, [j]Service de Néphrologie et d'Hémodialyse, Centre Hospitalier Regional, Haguenau, [k]Centre d'Investigation Clinique des Hôpitaux Universitaires de Strasbourg, [l]Laboratoire de Diagnostic Génétique, Nouvel Hôpital Civil, et [m]Service de Pédiatrie 1, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

Key Words

Bardet-Biedl · Ciliopathy · Nephronophthisis · Polydactyly · *SDCCAG8*

Abstract

The ciliopathies are an expanding group of disorders caused by mutations in genes implicated in the biogenesis and function of primary cilia. Bardet-Biedl syndrome (BBS) is a model ciliopathy characterized by progressive retinal degeneration, obesity, polydactyly, cognitive impairment, kidney anomalies and hypogonadism. Mutations in *SDCCAG8 (NPHP10)* were described recently in patients with nephronophthisis and retinal degeneration (Senior-Loken syndrome; SLS). Given the phenotypic and genetic overlap between known ciliopathy genes, we hypothesized that mutations in *SDCCAG8* might also contribute alleles to more severe, multisystemic ciliopathies. We performed genetic and phenotypic analyses of 2 independent BBS cohorts. Subsequent to mutation screening, we made a detailed phenotypic analysis of 5 families mutated for *SDCCAG8* (3 homozygous and 2 compound heterozygous mutations) and conducted statistical analyses across both cohorts to examine possible phenotype-genotype correlations with mutations at this locus. All patients with mutations in *SDCCAG8* fulfilled the diagnostic criteria for BBS (retinal degeneration, obesity, cognitive defects, renal failure, hypogonadism). Interestingly, none of the patients with primary *SDCCAG8* mutations had polydactyly, a frequent but not obligatory BBS feature. In contrast, the same patients displayed early-onset renal failure, obesity, as well as recurrent pulmonary and ENT in-

E.S. and A.Z. contributed equally to this work.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

Accessible online at:
www.karger.com/msy

Prof. Helene Dollfus
Laboratoire de Génétique Médicale EA 3949 Inserm Avenir
Université de Strasbourg, Hôpitaux Universitaires de Strasbourg
11 rue Humann, FR–67085 Strasbourg Cedex (France)
Tel. +33 3 88 12 81 20, E-Mail hdollfus@unistra.fr

fections. Comparison of the phenotypes of these families with our entire BBS cohort indicated that renal impairment and absent polydactyly correlated significantly with causal *SDCCAG8* mutations. Thus, *SDCCAG8* mutations are sufficient to cause BBS in 1–2% of our combined cohorts, and define this gene as the sixteenth BBS locus *(BBS16)*. The absence of polydactyly and the concomitant, apparently fully penetrant association with early kidney failure represents the first significant genotype-phenotype correlation in BBS that potentially represents an indicator for phenotype-driven priority screening and informs specific patient management.

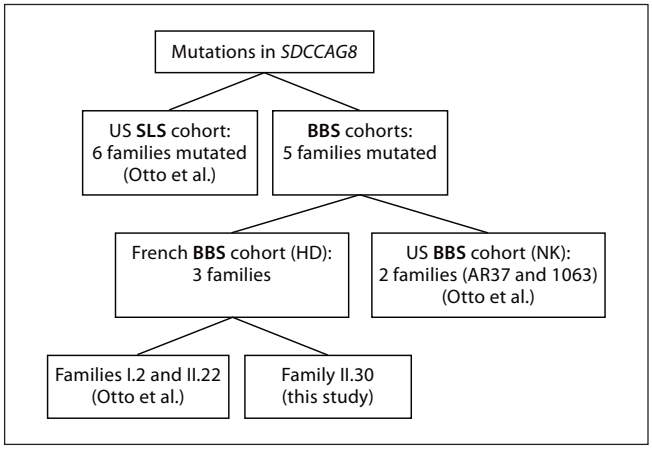


Fig. 1. Representation of families reported with mutations in *SDCCAG8* (H.D.: Helene Dollfus; N.K.: Nicholas Katsanis).

The ciliopathies represent a diverse group of clinically distinct but phenotypically overlapping disorders caused by mutations in genes involved in the biogenesis or maintenance of the primary cilium, a sensory organelle found in most mammalian cells [Gerdes et al., 2009]. The phenotypic range of ciliopathies encompasses both single-organ disorders as well as broad syndromic conditions, and their range of onset likewise varies from prenatally to early childhood. To date, mutations in >30 genes have been reported to be involved in this spectrum of diseases [Badano et al., 2006]. The ciliopathies are also characterized by a high degree of genetic heterogeneity, with little evidence of genotype-phenotype correlations. For example, mutations in 11 genes have been shown to be genetically necessary to cause nephronophthisis (NPHP, or, when present in conjunction with retinal degeneration, Senior-Loken syndrome, SLS), while 15 causal genes have been identified in Bardet-Biedl syndrome (BBS) [Gascue et al., 2011]. BBS is a model ciliopathy exhibiting characteristic phenotypic pleiotropy, illustrated by 5 major features: retinitis pigmentosa (RP), obesity, polydactyly, urogenital abnormalities, and cognitive impairment [Beales et al., 1999]. BBS genes have been identified through various strategies ranging from traditional positional cloning, in some instances associated with comparative genomics [Katsanis et al., 2000; Slavotinek et al., 2000; Mykytyn et al., 2001, 2002; Nishimura et al., 2001, 2005; Chiang et al., 2004, 2006; Li et al., 2004; Stoetzel et al., 2006, 2007], to homology-based approaches [Ansley et al., 2003; Badano et al., 2003; Fan et al., 2004], evaluation of functional candidates [Kim et al., 2010] and overlap with other clinically distinct ciliopathies [Leitch et al., 2008]. The latest reported retinorenal ciliopathy gene, *SDCCAG8*, was identified by implementing an exon capture approach and homozygosity mapping with subsequent dedicated sequencing in phenotypically diverse cohorts that included patients with NPHP, including the associated SLS phenotype, as well as patients with phenotypes consistent with BBS [Otto et al., 2010]. We describe herein the genetic and phenotypic analysis of BBS patients from 2 large independent BBS cohorts (fig. 1). Overall, we identified 5 families fulfilling the diagnostic criteria for BBS and carrying causal mutations in *SDCCAG8*. We provide evidence that mutations in this locus are sufficient to cause this disorder, rendering *SDCCAG8*/*NPHP10* the sixteenth BBS gene *(BBS16)* and expanding the notion that mutations in some ciliary genes can drive the development of diverse clinical phenotypes [Zaghloul and Katsanis, 2009]. Moreover, we show that in each cohort, causal *SDCCAG8* mutations lead to absent polydactyly and fully penetrant renal disease, providing the first significant genotype-phenotype correlation and potentially offering predictive value for clinical genetic testing for any of the known BBS loci.

Patients and Methods

Patients and Related Phenotype

The patients analyzed in this study came from 2 independent BBS cohorts, 1 investigated in France (H.D.) and 1 in the United States (N.K.). Inclusion criteria have been defined elsewhere [Beales et al., 1999]. Informed consent was obtained from all participants in the study, in accord with the tenets of the Declaration of Helsinski.

The French BBS cohort includes more than 400 BBS families collected since 2002. Twenty of them with no apparent mutation in *BBS1–BBS14* were selected for a first screen by SNP array analysis because of consanguinity and/or because of the number of affected individuals. After the identification of mutations in

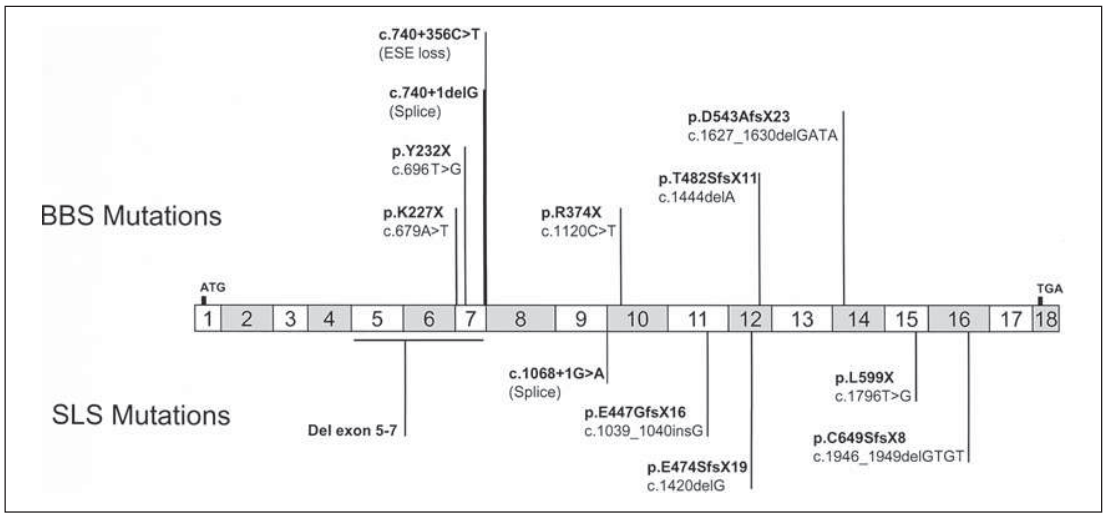


Fig. 2. Mutations reported in *SDCCAG8* to date. Mutations described in BBS patients are represented above the gene and mutations identified in SLS patients below. ESE = Exonic splicing enhancer.

SDCCAG8 in 3 families (I.2, II.22 and II.30), a second series of 40 patients, comprising mostly sporadic cases, was tested.

For the US BBS cohort, a series of 94 unrelated patients with BBS was tested irrespective of the presence of mutations at other loci.

Homozygosity Mapping (French Cohort)

Homozygosity mapping results for families I.2 and II.22 have been described elsewhere [Otto et al., 2010]. Family II.30, reported here for the first time, was genotyped with the Affymetrix GeneChip Human Mapping 250K Nsp Array (Affymetrix, Santa Clara, Calif., USA). Sample processing and labeling were conducted according to the manufacturer's instructions. Genotypes were called with Genotyping Console (GTC v4.0, Affymetrix).

Sequencing

Genomic DNA was extracted from blood samples according to the manufacturer's protocol (Flexigene DNA kit, Qiagen). PCR amplification was performed with 50 ng of genomic DNA template. The primers were designed with Primer 3 (http://frodo.wi.mit.edu/primer3); detailed protocols and primers are available on request. Bidirectional sequencing of purified PCR products was performed by the GATC Sequencing Facilities (http://www.gatcbiotech.com/en/about-us/gatc.html) or in-house on an ABI3730 Sequencer (Applied Biosystems, Santa Clara, Calif., USA).

Results

Homozygosity Mapping (French Cohort)

Results for families I.2 and II.22 have been reported previously [Otto et al., 2010]. In the consanguineous family II.30, we identified large homozygosity segments including 4 regions that contain known ciliopathy genes *(NPHP2/INVS*, *NPHP6/CEP290*, *MKS1,* and *BBS4)*. However, we found no mutations by direct sequencing of each locus. *SDCCAG8* was located in one of the smaller homozygosity regions (1.4 Mb) and, given its candidacy, we sequenced its coding regions. We identified a homozygous nonsense mutation in exon 10 (p.R374X, c.1120C>T) (fig. 2) that was absent from 192 ethnically-matched controls and segregated with the disorder under an autosomal recessive mode of inheritance.

Phenotype and Genotype

Family II.30 represents the fifth BBS family with mutations in *SDCCAG8*, in addition to the 6 reported SLS families [Otto et al., 2010] (fig. 1). The genotypic and phenotypic data for each family are summarized in table 1.

The first family, I.2, is a Gypsy consanguineous kinship with 5 affected members found in 2 related sibships. Affected sibs harbored a homozygous deep intronic mutation (c.740+356C>T) in intron 7 that was predicted to lead to the loss of an exonic splicing enhancer site, a prediction confirmed by cDNA sequencing [Zhang et al., 2005; Otto et al., 2010] (fig. 2). Although this family is classified as classic BBS, they were recruited initially because of the acute manifestation of chronic renal failure that was coupled to a variety of respiratory defects: patient I.2.21 was dialyzed between the age of 22 and 25 years, at which point the patient underwent a renal transplantation. Moreover, this patient developed subsequent-

Table 1. Main clinical manifestations and mutations in *SDCCAG8* for the patients reported in this study

	Family											
	I.2					II.22		II.30	AR37			1063
Parental consanguinity	yes					yes		yes	no			no
Ethnic origin	Gypsy					Europe		Turkish	North European			India
Individual	I.2.21	I.2.22	I.2.23	I.2.24	I.2.25	II.22.21	II.22.22		AR37-02	AR37-05	AR37-07	
Gender	M	F	F	M	M	M	F	F	M	M	F	ND
Age, years	37	22	14	8	31	19	15	6	27	22	11	ND
Obesity (BMI)	–	+ (38)	+ (34)	–	+ (54)	+ (41)	+ (31)	+ (22)	+ (31)	+ (30)	+ (28)	+
Cone-rod dystrophy (age)	+ (29)	+ (13)	+ (10)	+ (6)	+ (31)	+ (13)	+ (9)	+ (5)	+	+ (13)	+ (11)	+
Chronic renal failure (age)	+ (22)	+ (13)	+ (10)	+ (6)	+ (31)	+ (13)	+ (13)	+ (5)	+ (28)	+ (22)	+ (?)	ND
Polydactyly	–	–	–	–	–	–	–	–	–	–	–	–
Hypogonadism/hypogenitalism	ND	ND	ND	ND	+	+	+	ND	ND	+	ND	ND
Developmental delay	ND	+	+	+	+	+	+	+	–	–	+	+
Conductive hearing loss/ recurrent otitis	+	+	+	+	ND	+	+	+	ND	ND	ND	ND
Respiratory infection	+	+	+	+	+	+	+	+	ND	ND	ND	ND
Respiratory failure	+	+	–	–	ND	ND	ND	ND	ND	ND	ND	ND
Asthma	+	+	+	+	ND	+	+	ND	ND	ND	ND	ND
Mutation	c.740+356C>T/c.740+356C>T					c.679A>T/c.679A>T		c.1120C>T/ c.1120C>T	c.696T>G/c.740+1delG			c.1444delA/ c.1627_1630 delGATA
Location of mutation	intron 7					exon 7		exon 10	exon 7			exon 12/14
Protein modification	loss of ESE site (aberrant ins IVS7)					p.K227X/p.K227X		p.R374X/ p.R374X	p.Y232X/splice			p.T482SfsX11/ p.D543AfsX23
Familial segregation	yes					father htz/ mother ND		father htz/ mother htz	ND			ND

BMI = Body mass index; c = coding DNA; del = deletion; ESE = exonic splicing enhancer; fs = frameshift; htz = heterozygous; IVS = intervening sequence; ND = not determined; p = protein.

ly respiratory infections, bronchiectasis, and obstructive chronic respiratory failure, a combination of features seen uncommonly in BBS.

BBS was diagnosed 2 years later as RP was established both for him and his 13-year-old sister (I.2.22) (fig. 3A, fundus). In addition to RP, she exhibited obesity (BMI: 38 kg/m^2) (fig. 3A, whole body), mental retardation, and moderate deafness due to recurrent otitis. Renal failure was diagnosed and required dialysis and later transplantation. Like her brother, she presented early with regular respiratory infections and she developed bronchiectasis of the medium lobus and chronic obstructive respiratory failure. Although anosmia has been demonstrated in some 50% of BBS patients [Kulaga et al., 2004], she was normosmic, while a nasal biopsy showed ciliary normal beating.

The third affected sib (patient I.2.23) was regularly admitted to hospital since birth because of respiratory infections and delayed cognitive skills. The recurrent respiratory and ENT infections led to a chronic obstructive broncho-pneumopathy and to moderate conductive hearing loss. A CT scan revealed bronchiectasis. Immune deficiency, cystic fibrosis and allergy were excluded and a nasal biopsy showed normal motile cilia. An abdominal ultrasound showed hyperechogenic kidneys at 1 year of life; at age 9, kidneys were consistently hyperechogenic, with enlarged cortex and abnormal cortico-medullary differentiation. Renal scintigraphy showed mildly impaired renal function. RP was subsequently diagnosed, fulfilling the diagnostic criteria for BBS. One year later, the patient suffered from terminal renal failure and the ultrasound showed bilateral renal hypotrophy (55 × 23 mm) that required dialysis and renal transplantation 3 years later. Consistent with BBS, the patient became progressively obese (BMI: 34 kg/m^2). She was evaluated on the cognitive level by the Wechsler Intelligence Scale for Children® (WISC®-IV) and showed moderate cognitive impairment and mental retardation (data not shown).

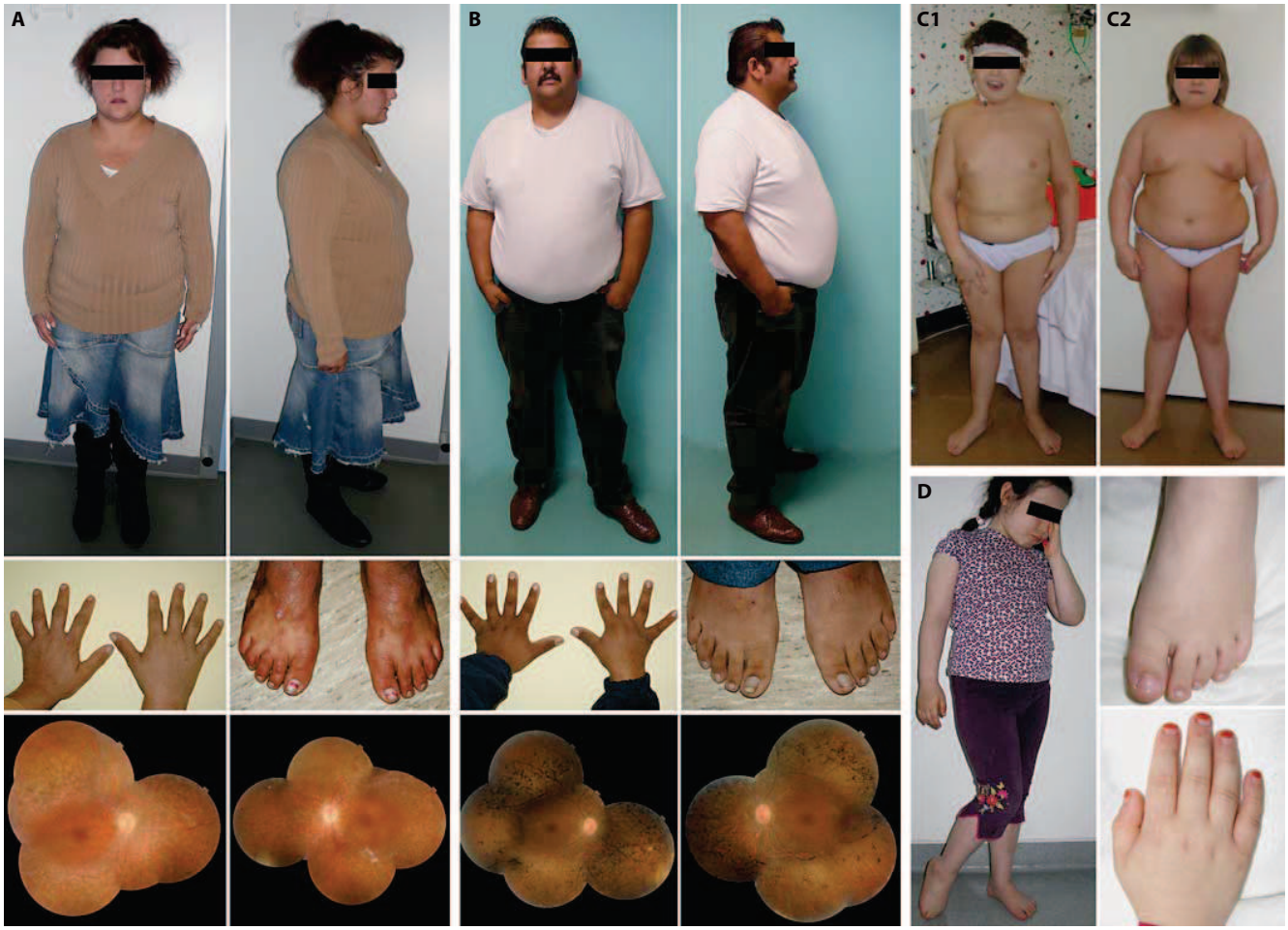


Fig. 3. A, **B** Photographs of patients I.2.22 (**A**) and I.2.25 (**B**) showing the obesity (photograph from head to toe), normal extremities and retinitis pigmentosa (lower fundus photographs), respectively. **C** Photographs of patients II.22.21 (**C1**) and II.22.22 (**C2**) showing obesity with genu valgum orthopedic complications. **D** Photographs of patient II.30 showing excess of weight and normal extremities.

Her brother (patient I.2.24) was diagnosed with terminal renal failure requiring dialysis at 6 years of age. An ultrasound revealed small kidneys (60 × 24 mm) with a complete undifferentiated parenchyma. Ophthalmic examination identified RP. Mental retardation was also diagnosed. Similar to the other affected members of this family, the patient was also treated regularly for bronchial and pulmonary infections and recurrent otitis with moderate conductive hearing loss.

Patient I.2.25 was the eldest at the time of BBS diagnosis. In fact, with the exception of several pulmonary infections and morbid obesity (BMI: 54 kg/m^2) (fig. 3B, whole body), the patient had no other complaint. At the age of 31 years, he presented with acute respiratory distress and edema of the lower limbs. Terminal renal failure was diagnosed and required dialysis. The ultrasound showed atrophic kidneys (100 mm of maximum height) with a thin cortex and a bad cortico-medullary differentiation. The diagnosis of RP was made at the same age, although the patient, with mild cognitive impairment, had not noticed that his visual fields were tubular (fig. 3B, fundus).

The combination of our clinical examination secured a diagnosis of BBS for this family. All affected sibs met the criteria for BBS although they all lacked polydactyly, one of the features of BBS that, in some populations, is present in >60% of patients [Beales et al., 1999] (fig. 3A, B, hands and feet; fig. 4).

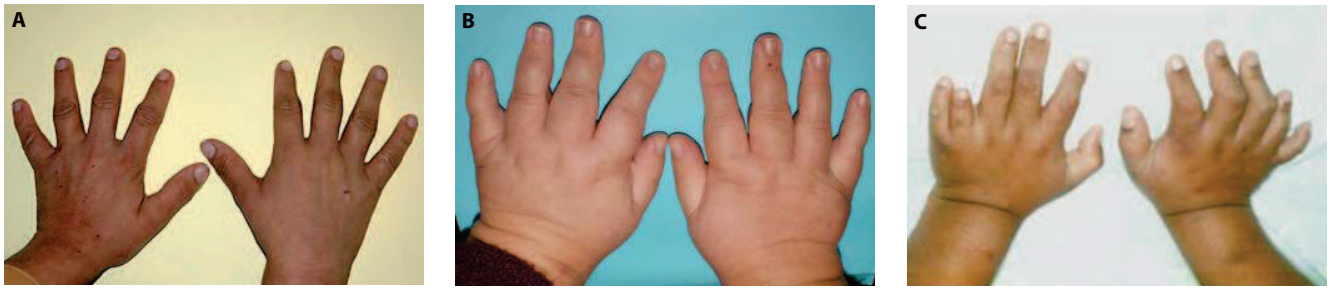


Fig. 4. Photographs showing different types of extremities found in Bardet-Biedl syndrome. **A** Normal extremities for an *SDCCAG8* mutated patient. **B** Typical brachydactyly found in BBS (here: *BBS10*-mutated patient) [Stoetzel et al., 2006]. **C** Typical polydactyly found in BBS (here: *BBS12*-mutated patient) [Stoetzel et al., 2007].

The second family II.22 is a French consanguineous family. The affected sibs carried a homozygous nonsense mutation (p.K227X, c.679A>T) in exon 7 of *SDCCAG8* [Otto et al., 2010] (fig. 2). Terminal renal failure was diagnosed for patient II.22.21 at 13 years of age. The kidney ultrasound was normal, except for several small cysts. Dialysis began immediately and he received a renal transplantation. The clinical association of renal disease with obesity (BMI: 28 kg/m^2; +5 SD) (fig. 3C1) and cognitive impairment led to a suspected diagnosis of BBS, despite the lack of polydactyly. This patient also had asthma and recurrent ENT infections with adenoidectomy and myringotomy at age 2. Follow-up showed a significant increase of obesity, with a BMI of 40 kg/m^2 at adult age. His 9-year-old sister II.22.22 had no polydactyly but obesity with a BMI of 31 kg/m^2 (+10 SD) (fig. 3C2), cognitive impairment and respiratory/ENT manifestations. For both of them, an extensive medical investigation also revealed RP, hypogenitalism (unilateral cryptorchism, small penis and testicles for the boy and labia minora hypoplasia for the girl), substantiating the BBS diagnosis. The ultrasound of patient II.22.22 showed kidneys of normal size but with bilateral cortico-medullary dedifferentiation. Follow-up detected progressive renal deterioration requiring dialysis, followed by subsequent transplantation.

The family II.30 proband is a sporadic case of consanguineous Turkish lineage, carrying a homozygous nonsense mutation (p.R374X, c.1120C>T) in *SDCCAG8* (identified in this study; fig. 2). Since the age of 1 month, she presented recurrent respiratory infections and bronchiolitis, leading to the diagnosis of infantile asthma. She also presented recurrent ENT infections and benefited from adenoidectomy and myringotomy. Moreover, she developed early-onset obesity (from month 6) and weighed 27.6 kg for a height of 111 cm at the age of 4 (BMI: 22.4 kg/m^2; +6 SD) (fig. 3D, whole body). At 5 years of age, she was hospitalized for general deterioration related to terminal renal failure. She was dialyzed and received a renal transplant a few months later. The histological renal evaluation suggested nephronophthisis. RP was diagnosed on the ERG, although the ocular fundus was unremarkable. Although an initial diagnosis of SLS was consistent with her findings (with obesity possibly being coincidental), she also had mild mental retardation and speech delay, which, in spite of the absence of polydactyly (fig. 3D, hand and feet), suggested BBS. She is, to date, the youngest patient affected in our cohort.

The sequencing of 96 unrelated patients of the US cohort led to the identification of 2 families with mutations in *SDCCAG8*. Family AR37 is a non-consanguineous family of North-European descent with 3 affected sibs and 3 unaffected sibs. Affected sibs harbored 2 compound mutations in exon 7: a nonsense mutation (p.Y232X, c.696T>G) and a splice site mutation (c.740+1delG) [Otto et al., 2010] (fig. 2).

The 3 sibs were obese, with a BMI of 28 kg/m^2 (+5 SD) for the youngest (AR37-07; 11 years old), 30 kg/m^2 for the second (AR37-05; 22 years old) and 31 kg/m^2 for the eldest (AR37-02; 27 years old). All of them had RP, diagnosed at 13 and 11 years old for AR37-05 and AR37-07, respectively (data on first diagnosis are not available for AR37-02). Similar to the patients in the French cohort, patients AR37-02 and -05 had terminal renal failure diagnosed at 28 and 22 years, respectively. The renal ultrasound of patient AR37-07 at 12 years showed early renal disease; the kidney biopsy revealed tubular and glomerular disease with persistent fetal glomeruli, prompting rigorous kidney follow-up. Notably, again reminiscent of the patients with *SDCCAG8* mutations in the French cohort, none of the 3 affected sibs had polydactyly. One patient

had hypogenitalism (AR37-05) and one had cognitive impairment (AR37-07).

Finally, 2 heterozygous deletions causing frameshifts and inducing a premature stop codon were identified in exon 12 (p.T482SfsX11, c.1444delA) and in exon 14 (p.D543AfsX23, c.1627_1630delGATA) in a sporadic case born from Indian non-consanguineous parents (F1063) [Otto et al., 2010] (fig. 2). This boy presented terminal renal failure in childhood, RP, and developmental delay, but no polydactyly. No more clinical data are available as the family is no longer accessible for follow-up studies.

First Evidence for Gene-Specific Phenotypes in BBS Patients

Despite the documented clinical variability of BBS, we were intrigued by the consistent presence of renal disease with a concomitant absence of polydactyly in all BBS patients in which *SDCCAG8* mutations were sufficient to explain the disorder. We therefore wondered whether this pattern was significant in our cohorts. Because the number of *SDCCAG8*-positive patients is modest, we compared: (a) the frequency of the single occurrence of either renal disease or polydactyly in *SDCCAG8* patients versus our cohorts; (b) the recurrence of these phenotypes in sib pairs with or without *SDCCAG8* mutations, and (c) the frequency of the combined occurrence of renal failure along with absence of polydactyly in *SDCCAG8* patients versus the cohorts.

In the French cohort, the 8 *SDCCAG8*-positive BBS patients all shared the same phenotype: renal failure (8/8), but no polydactyly (8/8), whereas in the 36 *SDCCAG8*-negative BBS patients only 18/36 had renal failure, and 15 of them (15/18) showed polydactyly. In the US cohort, we found a low incidence of renal disease: 12/92 *SDCCAG8*-negative patients had renal disease versus 4/4 *SDCCAG8*-positive, a significant difference (Fisher's exact test: $p < 0.0005$). These data reveal a significant difference of phenotype between *SDCCAG8*-positive patients and the rest of the BBS cohorts, at both levels of renal failure feature ($p < 0.0089$) or absence of polydactyly ($p < 0.0042$).

Consistent with the above data, in the US cohort the intrafamilial recurrence of renal disease was absolute in our 3 *SDCCAG8*-positive sib pairs, but true only in 2/37 negative sib pairs, an observation that was also significant ($p < 0.0011$) and which strengthened the hypothesis of renal disease recurrence being higher in *SDCCAG8*-positive sib pairs than in negative ones.

Finally, when we asked for the co-occurrence of renal disease and absence of polydactyly, we found an even more striking (and significant) difference. Once again, in the French PHRC 2002 cohort we identified 3 out of the 36 *SDCCAG8*-negative patients without polydactyly and with renal impairment, whereas 100% of the *SDCCAG8* patients with renal failure had no polydactyly ($p < 0.0001$). Identically, whereas all the US patients (and sib pairs) were fully concordant with these phenotypes, we only found 4/92 *SDCCAG8*-negative patients with renal disease and no polydactyly ($p < 0.00002$) and 0/37 sib pairs ($p < 0.0001$). The data for both cohorts only show a significant absence of polydactyly in renal affected *SDCCAG8* patients as compared to renal affected *SDCCAG8*-negative patients. In the French cohort, the co-occurrence of renal failure and absence of polydactyly in all *SDCCAG8* patients versus the rest of the BBS cohorts was also found to be significant ($p < 0.000004$).

Discussion

The ciliopathies represent useful models to study the architecture of human genetic disease, especially with regard to the mechanistic underpinnings of genetic heterogeneity and the molecular basis of phenotypic variability. Some exceptions notwithstanding, allelism at single ciliopathy loci is insufficient to explain phenotypic differences, an observation attributed in part to the concept of the variable activity of *trans*-acting modifiers within the biological system: the primary cilium [Khanna et al., 2009; Zaghloul and Katsanis, 2010; Davis et al., 2011]. For example, loss-of-function mutations in the pan-ciliopathy locus *CEP290* can cause isolated nephronophthisis (*NPHP6* [Sayer et al., 2006]), SLS (*SLSN6* [Sayer et al., 2006]), Leber congenital amaurosis (*LCA10* [den Hollander et al., 2006]), BBS (*BBS14* [Leitch et al., 2008]), Joubert syndrome (*JBTS5* [Sayer et al., 2006]) and Meckel-Gruber syndrome (*MKS4* [Baala et al., 2007]), with no evidence for single locus genotype-phenotype correlations.

The data emerging from the novel ciliopathy gene *SDCCAG8* suggest similar phenotypic overlap. While Otto et al. [2010], focusing on a cohort of patients with retinal-kidney phenotypes, identified mutations in *SDCCAG8* primarily in patients with an SLS diagnosis, we identified independently mutations in the same transcript in BBS patients. Given the pattern of variation detected in *CEP290*, as well as in other ciliopathy loci, we predict that *SDCCAG8* is an attractive functional/genetic candidate for other related ciliopathies, most notably

Meckel-Gruber syndrome and Joubert syndrome. However, in contrast to significantly divergent phenotypes seen upon loss of function of other ciliopathy transcripts, all null alleles in *SDCCAG8* are concomitant with early renal failure and RP. In most cases in our present study, renal failure was the first symptom of the disease, was discovered at the terminal stage and typically predated the diagnosis of BBS. Notably, although renal impairment represents one of the primary features of BBS, its penetrance ranges from 5 to 82% [Beales et al., 1999; Imhoff et al., 2011]. In the cohort described by Beales et al. [1999], 6 patients (5%) presented chronic renal failure and 4 of them were children. Moreover, although all individuals from a series of 20 BBS patients in Newfoundland (mean age: 30 years) presented structural and/or functional renal abnormalities, only 3 (15%) experienced terminal renal failure [Harnett et al., 1988]. As such, it is unusual that both the large US and French BBS cohorts used in the present study show a statistically significant and apparently fully penetrant end stage kidney disease in *SDCCAG8* patients (100%) as opposed to a low incidence of kidney failure in the *SDCCAG8*-negative BBS patients (13 and 8%, respectively).

In addition to the renal phenotype, 10 of 12 *SDCCAG8*-BBS patients (83%) presented early-onset obesity. Interestingly, a recent meta-analysis of genome-wide association studies for early-onset obesity identified 3 polymorphisms in introns 6, 9 and 10 of *SDCCAG8* in strong pairwise linkage disequilibrium [Scherag et al., 2010]. This association of *SDCCAG8* variants with pediatric obesity prior to the discovery of its involvement in BBS mirrors that of *BBS2*, *BBS4* and *BBS6* with the association of their polymorphisms with childhood and late-onset obesity reported after their BBS-causing mutations [Benzinou et al., 2006]. Taken together, these findings suggest that (a) BBS- and other ciliopathy-causing genes might be strong candidates for sporadic, genetically complex obesity, and (b) that the underlying obesity mechanism of BBS might be relevant to a broader subset of complex obesity.

All patients in the French cohort also presented recurrent pulmonary and ENT infections in early childhood, symptoms also found in Alström syndrome (ALMS). However, the *SDCCAG8* phenotype can be readily distinguished from ALMS as dilated or restrictive cardiomyopathy, diabetes and sensorineural component were not present in any of our patients, while, at the same time, ALMS patients are not known to have cognitive impairment [Alström et al., 1959].

The most striking differential clinical feature in *SDCCAG8*-BBS patients was the absence of polydactyly, one of the secondary features of BBS. Polydactyly is present in 58–74% of BBS patients [Bell, 1958; Green et al., 1989] and brachydactyly is often associated [Rudling et al., 1996]. In the French BBS cohort, polydactyly is significantly more common, with 98/110 patients with mutations in *BBS1–BBS12* having at least 1 extra digit, and at least 4 more families manifesting brachydactyly. This might represent a peculiarity of the genetic pool from which our patients are enrolled, or, more likely, might represent overly stringent criteria for securing a BBS diagnosis. Irrespective of the source of this difference, the high occurrence of polydactyly in the French cohort renders the absence of this phenotype from all *SDCCAG8*-BBS patients even more striking.

BBS, like most ciliopathies, is largely bereft of genotype-phenotype correlations. The only previously proposed correlation is the association of *BBS3* mutations with polydactyly of the 4 limbs [Sheffield et al., 1994] or of the lower limbs [Young et al., 1998]. However, this observation had not been supported by the description of additional patients mutated in *BBS3* presenting no polydactyly [Ghadami et al., 2000]. Therefore, our data potentially represent the first significant genotype-phenotype correlation in BBS. We suggest that in absence of polydactyly and brachydactyly in BBS patients *SDCCAG8* screening could be important after initial screening for recurrent BBS mutations [Muller et al., 2010], especially when such patients also present acute renal disease and recurrent respiratory/ENT infections. Moreover, the biological underpinning of these differences has the potential to apprehend the mechanism of the development and maintenance of several physiological systems.

Acknowledgements

We warmly thank the patients who accepted the publication of their clinical data. We thank in particular the French Bardet-Biedl support group (association Bardet-Biedl France). We also thank all medical contributors from genetic, pediatric, radiologic, biologic units as well as the CIC of Strasbourg University Hospital. We are grateful to the Centre National de Genotypage of Evry and the Affymetrix platform of IGBMC/Genopole de Strasbourg, especially Christelle Thibault and Violaine Alunni. The work of the French authors is funded by Retina France, UNADEV (Union Nationale des Aveugles et Déficients Visuels), FORMICOEUR, PHRC (Programme Hospitalier pour la Recherche Clinique) program 2007 and the Agence Nationale pour la Recherche call for rare diseases 2009. This study was supported in parts by the National Institutes of Health grants R01HD04260 from the National Institute of Child Health and Development and R01DK072301 from the National Institute of Diabetes, Digestive and Kidney disorders (N.K.). N.K. is a Distinguished Brumley Professor.

References

Alström CH, Hallgren B, Nilsson LB, Asander H: Retinal degeneration combined with obesity, diabetes mellitus and neurogenous deafness: a specific syndrome (not hitherto described) distinct from the Laurence-Moon-Bardet-Biedl syndrome: a clinical, endocrinological and genetic examination based on a large pedigree. Acta Psychiatr Neurol Scand Suppl 129:1–35 (1959).

Ansley SJ, Badano JL, Blacque OE, Hill J, Hoskins BE, et al: Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. Nature 425:628–633 (2003).

Baala L, Audollent S, Martinovic J, Ozilou C, Babron MC, et al: Pleiotropic effects of *CEP290 (NPHP6)* mutations extend to Meckel syndrome. Am J Hum Genet 81:170–179 (2007).

Badano JL, Ansley SJ, Leitch CC, Lewis RA, Lupski JR, Katsanis N: Identification of a novel Bardet-Biedl syndrome protein, BBS7, that shares structural features with BBS1 and BBS2. Am J Hum Genet 72:650–658 (2003).

Badano JL, Mitsuma N, Beales PL, Katsanis N: The ciliopathies: an emerging class of human genetic disorders. Annu Rev Genomics Hum Genet 7:125–148 (2006).

Beales PL, Elcioglu N, Woolf AS, Parker D, Flinter FA: New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J Med Genet 36:437–446 (1999).

Bell J: The Laurence-Moon syndrome; in Penrose LS (ed): The Treasury of Human Inheritance, pp 51–96 (Cambridge University Press, Cambridge 1958).

Benzinou M, Walley A, Lobbens S, Charles MA, Jouret B, et al: Bardet-Biedl syndrome gene variants are associated with both childhood and adult common obesity in French Caucasians. Diabetes 55:2876–2882 (2006).

Chiang AP, Nishimura D, Searby C, Elbedour K, Carmi R, et al: Comparative genomic analysis identifies an ADP-ribosylation factor-like gene as the cause of Bardet-Biedl syndrome (BBS3). Am J Hum Genet 75:475–484 (2004).

Chiang AP, Beck JS, Yen HJ, Tayeh MK, Scheetz TE, et al: Homozygosity mapping with SNP arrays identifies *TRIM32*, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene *(BBS11)*. Proc Natl Acad Sci USA 103:6287–6292 (2006).

Davis EE, Zhang Q, Liu Q, Diplas BH, Davey LM, et al: *TTC21B* contributes both causal and modifying alleles across the ciliopathy spectrum. Nat Genet 43:189–196 (2011).

den Hollander AI, Koenekoop RK, Yzer S, Lopez I, Arends ML, et al: Mutations in the *CEP290 (NPHP6)* gene are a frequent cause of Leber congenital amaurosis. Am J Hum Genet 79:556–561 (2006).

Fan Y, Esmail MA, Ansley SJ, Blacque OE, Boroevich K, et al: Mutations in a member of the Ras superfamily of small GTP-binding proteins causes Bardet-Biedl syndrome. Nat Genet 36:989–993 (2004).

Gascue C, Katsanis N, Badano JL: Cystic diseases of the kidney: ciliary dysfunction and cystogenic mechanisms. Pediatr Nephrol 26:1181–1195 (2011).

Gerdes JM, Davis EE, Katsanis N: The vertebrate primary cilium in development, homeostasis, and disease. Cell 137:32–45 (2009).

Ghadami M, Tomita HA, Najafi MT, Damavandi E, Farahvash MS, et al: Bardet-Biedl syndrome type 3 in an Iranian family: clinical study and confirmation of disease localization. Am J Med Genet 94:433–437 (2000).

Green JS, Parfrey PS, Harnett JD, Farid NR, Cramer BC, et al: The cardinal manifestations of Bardet-Biedl syndrome, a form of Laurence-Moon-Biedl syndrome. N Engl J Med 321:1002–1009 (1989).

Harnett JD, Green JS, Cramer BC, Johnson G, Chafe L, et al: The spectrum of renal disease in Laurence-Moon-Biedl syndrome. N Engl J Med 319:615–618 (1988).

Imhoff O, Marion V, Stoetzel C, Durand M, Holder M, et al: Bardet-Biedl syndrome: a study of the renal and cardiovascular phenotypes in a French cohort. Clin J Am Soc Nephrol 6:22–29 (2011).

Katsanis N, Beales PL, Woods MO, Lewis RA, Green JS, et al: Mutations in *MKKS* cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. Nat Genet 26:67–70 (2000).

Khanna H, Davis EE, Murga-Zamalloa CA, Estrada-Cuzcano A, Lopez I, et al: A common allele in *RPGRIP1L* is a modifier of retinal degeneration in ciliopathies. Nat Genet 41:739–745 (2009).

Kim SK, Shindo A, Park TJ, Oh EC, Ghosh S, et al: Planar cell polarity acts through septins to control collective cell movement and ciliogenesis. Science 329:1337–1340 (2010).

Kulaga HM, Leitch CC, Eichers ER, Badano JL, Lesemann A, et al: Loss of BBS proteins causes anosmia in humans and defects in olfactory cilia structure and function in the mouse. Nat Genet 36:994–998 (2004).

Leitch CC, Zaghloul NA, Davis EE, Stoetzel C, Diaz-Font A, et al: Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. Nat Genet 40:443–448 (2008).

Li JB, Gerdes JM, Haycraft CJ, Fan Y, Teslovich TM, et al: Comparative genomics identifies a flagellar and basal body proteome that includes the *BBS5* human disease gene. Cell 117:541–552 (2004).

Muller J, Stoetzel C, Vincent M, Leitch C, Laurier V, et al: Identification of 28 novel mutations in the Bardet-Biedl syndrome genes: the burden of private mutations in an extensively heterogeneous disease. Hum Genet 127:583–593 (2010).

Mykytyn K, Braun T, Carmi R, Haider NB, Searby CC, et al: Identification of the gene that, when mutated, causes the human obesity syndrome BBS4. Nat Genet 28:188–191 (2001).

Mykytyn K, Nishimura DY, Searby CC, Shastri M, Yen HJ, et al: Identification of the gene *(BBS1)* most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. Nat Genet 31:435–438 (2002).

Nishimura DY, Searby CC, Carmi R, Elbedour K, Van Maldergem L, et al: Positional cloning of a novel gene on chromosome 16q causing Bardet-Biedl syndrome (BBS2). Hum Mol Genet 10:865–874 (2001).

Nishimura DY, Swiderski RE, Searby CC, Berg EM, Ferguson AL, et al: Comparative genomics and gene expression analysis identifies *BBS9*, a new Bardet-Biedl syndrome gene. Am J Hum Genet 77:1021–1033 (2005).

Otto EA, Hurd TW, Airik R, Chaki M, Zhou W, et al: Candidate exome capture identifies mutation of *SDCCAG8* as the cause of a retinal-renal ciliopathy. Nat Genet 42:840–850 (2010).

Rudling O, Riise R, Tornqvist K, Jonsson K: Skeletal abnormalities of hands and feet in Laurence-Moon-Bardet-Biedl (LMBB) syndrome: a radiographic study. Skeletal Radiol 25:655–660 (1996).

Sayer JA, Otto EA, O'Toole JF, Nurnberg G, Kennedy MA, et al: The centrosomal protein nephrocystin-6 is mutated in Joubert syndrome and activates transcription factor ATF4. Nat Genet 38:674–681 (2006).

Scherag A, Dina C, Hinney A, Vatin V, Scherag S, et al: Two new loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. PLoS Genet 6:e1000916 (2010).

Sheffield VC, Carmi R, Kwitek-Black A, Rokhlina T, Nishimura D, et al: Identification of a Bardet-Biedl syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. Hum Mol Genet 3:1331–1335 (1994).

Slavotinek AM, Stone EM, Mykytyn K, Heckenlively JR, Green JS, et al: Mutations in *MKKS* cause Bardet-Biedl syndrome. Nat Genet 26:15–16 (2000).

Stoetzel C, Laurier V, Davis EE, Muller J, Rix S, et al: *BBS10* encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus. Nat Genet 38:521–524 (2006).

Stoetzel C, Muller J, Laurier V, Davis EE, Zaghloul NA, et al: Identification of a novel BBS gene *(BBS12)* highlights the major role of a vertebrate-specific branch of chaperonin-related proteins in Bardet-Biedl syndrome. Am J Hum Genet 80:1–11 (2007).

Young TL, Woods MO, Parfrey PS, Green JS, O'Leary E, et al: Canadian Bardet-Biedl syndrome family reduces the critical region of BBS3 (3p) and presents with a variable phenotype. Am J Med Genet 78:461–467 (1998).

Zaghloul NA, Katsanis N: Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy. J Clin Invest 119:428–437 (2009).

Zaghloul NA, Katsanis N: Functional modules, mutational load and human genetic disease. Trends Genet 26:168–176 (2010).

Zhang XH, Kangsamaksin T, Chao MS, Banerjee JK, Chasin LA: Exon inclusion is dependent on predictable exonic splicing enhancers. Mol Cell Biol 25:7323–7332 (2005).

Paper IV: Mesoaxial polydactyly is a major feature in Bardet-Biedl syndrome patients with *LZTFL1* (*BBS17*) mutations.

Schaefer E, Lauer J, Durand M, Pelletier V, Obringer C, Claussmann A, Braun JJ, **Redin C**, Mathis C, Muller J, Schmidt-Mutter C, Flori E, Marion V, Stoetzel C, Dollfus H. Clin Genet. 2013 May.

Aim of the study: Second report of BBS patients with mutations in *LZTFL1/BBS17*. Although only two families are reported so far with such mutations, an intriguing feature of atypical polydactyly (mesoaxial polydactyly) is observed in all patients, while rarely (if not never) in other BBS patients, proposing another hint for genotype-to-phenotype correlation. This findings need to be confirmed by additional reports of *BBS17* patients.

Contribution: Minor. Performed the statistical analyses for determining whether the mesoaxial polydactyly could be considered a significantly different feature for *LZTFL1*/*BBS17* patients compared to other BBS patients in spite of the very limited number of patients. Wrote the different conclusions accordingly.

CLINICAL GENETICS An International Journal of Genetics, Molecular and Personalized Medicine

Clin Genet 2014: 85: 476–481
Printed in Singapore. All rights reserved

CLINICAL GENETICS
doi: 10.1111/cge.12198

Short Report

Mesoaxial polydactyly is a major feature in Bardet–Biedl syndrome patients with *LZTFL1* (*BBS17*) mutations

Schaefer E., Lauer J., Durand M., Pelletier V., Obringer C., Claussmann A., Braun J.-J., Redin C., Mathis C., Muller J., Schmidt-Mutter C., Flori E., Marion V., Stoetzel C., Dollfus H. Mesoaxial polydactyly is a major feature in Bardet–Biedl syndrome patients with *LZTFL1* (*BBS17*) mutations.
Clin Genet 2014: 85: 476–481. © John Wiley & Sons A/S. Published by John Wiley & Sons Ltd, 2013

Ciliopathies are heterogeneous disorders sharing different clinical signs due to a defect at the level of the primary cilia/centrosome complex. Postaxial polydactyly is frequently reported in ciliopathies, especially in Bardet–Biedl syndrome (BBS). Clinical features and genetic results observed in a pair of dizygotic twins with BBS are reported. The following manifestations were present: retinitis pigmentosa, bilateral insertional polydactyly, cognitive impairment and renal dysfunction. X-rays of the hands confirmed the presence of a 4th mesoaxial extra-digit with Y-shaped metacarpal bones. The sequencing of *LZTFL1* identified a missense mutation (NM_020347.2: p.Leu87Pro; c.260T>C) and a nonsense mutation (p.Glu260*; c.778G>T), establishing a compound heterozygous status for the twins. A major decrease of LZTFL1 transcript and protein was observed in the patient's fibroblasts. This is the second report of *LZTFL1* mutations in BBS patients confirming *LZTFL1* as a BBS gene. Interestingly, the only two families reported in literature thus far with *LZTFL1* mutations have in common mesoaxial polydactyly, a very uncommon feature for BBS. This special subtype of polydactyly in BBS patients is easily identified on clinical examination and prompts for priority sequencing of *LZTFL1* (*BBS17*).

Conflict of interest

None.

E. Schaefer[a,†], J. Lauer[a,†], M. Durand[a], V. Pelletier[b], C. Obringer[a], A. Claussmann[a], J.-J. Braun[c], C. Redin[d], C. Mathis[e], J. Muller[d,f], C. Schmidt-Mutter[e], E. Flori[g], V. Marion[a], C. Stoetzel[a] and H. Dollfus[a,b]

[a]Laboratoire de Génétique Médicale, INSERM U1112, Faculté de Médecine de Strasbourg, Universitaires de Strasbourg, Strasbourg, France, [b]Centre de Référence pour les Affections Rares en Génétique Ophtalmologique (CARGO), [c]Service Oto-Rhino-Laryngologie et Chirurgie Cervico-Faciale, Hôpitaux Universitaires de Strasbourg, Strasbourg, France, [d]Département de Neurobiologie et Génétique, Laboratoire de Bioinformatique et Génomique Intégratives, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch-Graffenstaden, France, [e]Centre d'Investigation Clinique de Strasbourg, CIC-P 1002, [f]Laboratoire de Diagnostic Génétique, and [g]Laboratoire de Cytogénétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

[†]Both the authors contributed equally to this work

Key words: Bardet–Biedl syndrome – ciliopathy – *LZTFL1* gene – mesoaxial polydactyly

Corresponding author:
Hélène Dollfus, Laboratoire de Génétique Médicale, INSERM U1112, Faculté de Médecine de Strasbourg, Université de Strasbourg, Strasbourg, France.
Tel.: +33388128120;
fax : +33388128125;

e-mail: dollfus@unistra.fr

Received 6 February 2013, revised and accepted for publication 16 May 2013

Bardet–Biedl syndrome (BBS) is a well-recognized ciliopathy defined by the following major clinical features: retinitis pigmentosa, obesity, kidney anomalies, cognitive impairment and polydactyly (1). The digit anomalies are usually characterized as postaxial polydactyly and/or brachydactyly and more rarely syndactyly. We recently reported on a BBS family mutated in a novel BBS gene, *LZTFL1* (*BBS17*), for which mesoaxial polydactyly was a distinct manifestation (2). Herein, we report the second family mutated in *LZTFL1* with similar clinical features of mesoaxial polydactyly. This observation suggests a possible phenotype–genotype correlation for this very rarely mutated gene.

Materials and methods

Subjects

Two affected female twins aged 36 were referred for clinical investigations to the BBS national PHRC (Programme Hospitalier pour la Recherche Clinique) program held at the reference center for rare genetic eye disorders at the University Hospital of Strasbourg. Informed consent and ethical approval [Comité de Protection des Personnes Est IV (reference: 08/28)] were obtained according to the French legislation. Their parents were non-consanguineous and healthy, as was their eldest brother (Fig. 1a). A diagnosis of BBS was made out at childhood as three major clinical features in accordance with Beales's criteria were present: polydactyly, early-onset retinitis pigmentosa and kidney dysfunction (1).

Sequencing

Genomic DNA was extracted from blood samples according to the manufacturer's protocol (Flexigene DNA kit, Qiagen, Courtaboeuf, France). Oligonucleotide primer sequence details and polymerase chain reaction (PCR) protocols are available on request. Bidirectional sequencing of purified PCR products has been performed by the GATC Sequencing Facilities (http://www.gatcbiotech.com/en/about-us/gatc.html).

Zygosity evaluation

To determine the type of twining (monozygotic or dizygotic), 15 forensic markers were amplified from 2 ng of genomic DNA according to the manufacturer's protocol for the two sisters (PowerPlex® 16 HS System, Promega, Charbonnières, France). Detection of amplified fragments was performed on ABI3500 genetic analyzer (Applied Biosystems, Santa Clara, CA). The data were analyzed with GENEMAPPER ID software.

Real-time PCR determination and Western Blot

Total RNA and proteins were obtained from dermal fibroblasts following skin biopsy of the affected individuals. Real-time PCR determination and Western Blot were performed using the previously described protocols (2).

Results

Clinical evaluation

Patient 1

The first twin, II.2 (Fig. 1a), presented at birth with one postaxial extra-digit on each foot and one postaxial extra-digit plus one mesoaxial extra-digit on each hand. The extra-digits were removed at 3 months of life (initial X-rays are unavailable). Polyuria–polydypsia episodes occurred since birth and she developed chronic kidney failure at 4 years of age. Kidney transplantation was performed at the age of 23 years. As she gained weight easily, a strict diet was advocated from early childhood. At 8 years of age, advanced retinitis pigmentosa was diagnosed with non-recordable full-field electroretinogram (ERG) response. Learning difficulties, associated with slowness, were reported since childhood. She was directed to a specialized school at 14 years of age. Now aged 36 years, her educational level is equivalent to the elementary school level and she works in a sheltered environment.

On clinical examination, her height was 152 cm (−1.5 SD) and weight was 56.2 kg (+0.5 SD) with a body mass index (BMI) of 24.3 kg/m^2. Her vision was very poor as she could only perceive the difference between dark and light. On fundus examination, the retinal dystrophy was very advanced with macular atrophy as well as extensive retinal degeneration with widespread atrophic zones and bone-spicules. Olfactometry testing revealed complete anosmia (score of 9/40 to the UPSIT® test). Light microscopy analysis of a nasal biopsy showed abnormal enlarged cilia. The fourth digit of the hands was shorter than the fifth and both showed ulnar divergence (Fig. 1d). The X-ray revealed a synostosis between the third and the fourth metacarpal bones of both hands (Fig. 1f).

Patient 2

The second twin, II.3 (Fig. 1a), presented, like her sister, with six toes on each foot and seven fingers on each hand. She had polydypsia and polyuria and was diagnosed to have massively enlarged cystic kidneys that induced rapid chronic renal failure. She underwent dialysis for more than 10 years before receiving kidney transplantation. She followed a very strict hypocaloric diet during childhood. Retinitis pigmentosa was

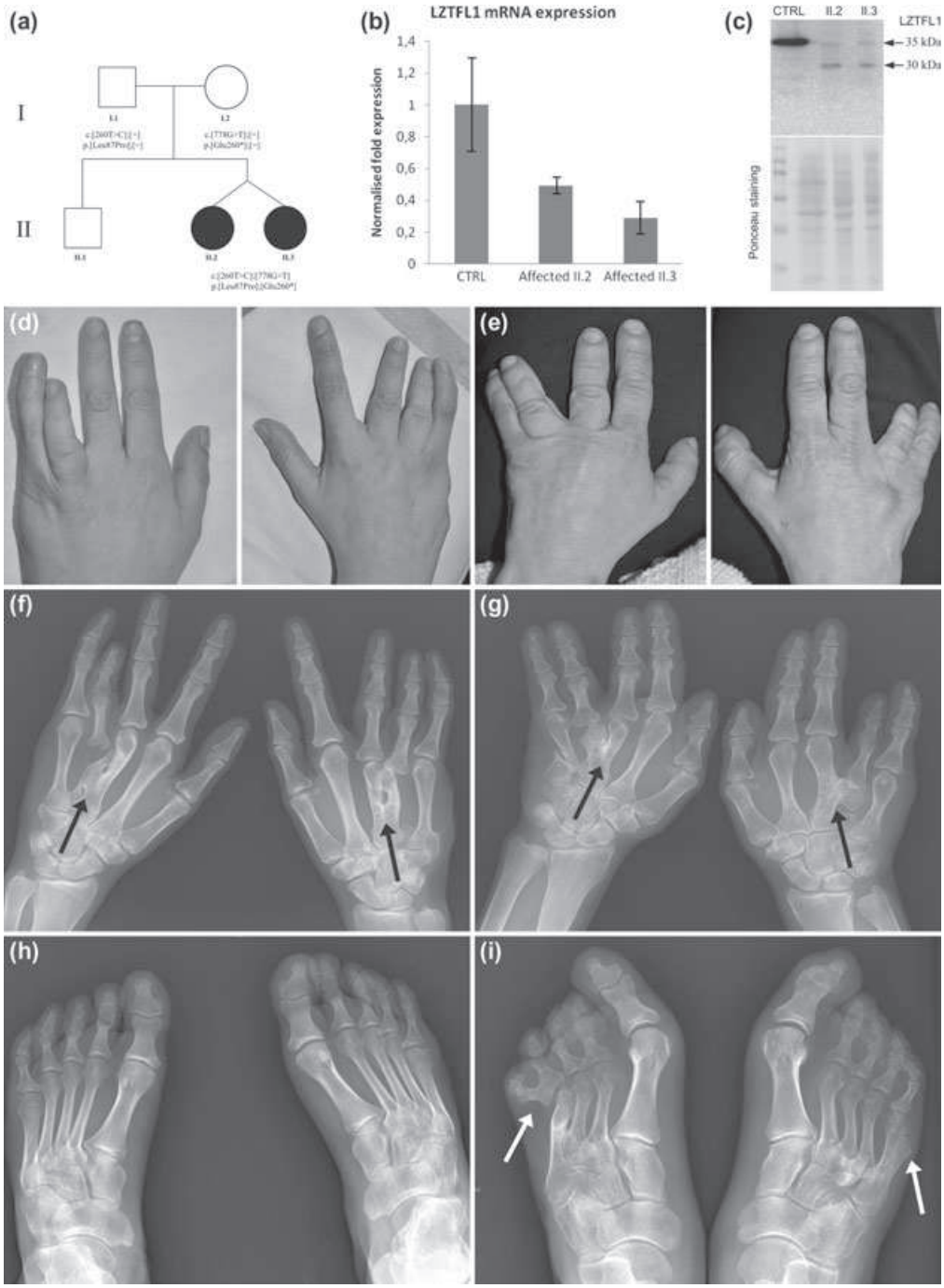


Fig. 1. Segregation of *LZTFL1* mutations, LZTFL1 expression, clinical and radiological description of the affected twins. (**a**) Pedigree of the family: affected twins (II.2 and II.3) harbored compound heterozygous mutations in *LZTFL1* : c.260T>C; p.Leu87Pro/c.778G>T; p.Glu260*. The father (I.1) is heterozygote for the missense allele and the mother (I.2) for the nonsense allele. (**b**) Real-time polymerase chain reaction (PCR) measurements of LTZFL1 expression in patient's dermal and control fibroblasts (technical triplicates). (**c**) Immunodetection of LTZFL1 in protein extracts from control and patient's fibroblasts (quantity loaded: 20 µg) with the membrane stained with Ponceau staining serving as total protein loading control identifying two bands: one at 35 kDa (arrow) and one at 30 kDa (arrow). (**d**) Photographs of the hands of Case 1/II.2. (**e**) Photographs of the hands of Case 2/II.3. (**f**) X-ray of the hands of Case 1. Left hand: synostosis between the third and the fourth4 metacarpal bones (arrow); distortion of the fourth metacarpal bone with a head merged with the third metacarpal neck; absence of metacarpal-phalangeal joint; hypotrophy of the proximal phalanx. Right hand: hypoplasia and fusion between the third and the fourth metacarpal bones (arrow). (**g**) X-ray of the hands of Case 2. Left hand: hypoplasia and fusion of the third and fourth metacarpal bones (arrow); abnormal hypoplasic fifth metacarpal bone with defect of fusion between the body and the basis and development of a new joint between the proximal part of the body and the inter-metacarpal spacing of fourth and fifth rays. Right hand: fourth metacarpal bone limited to a metacarpal basis developed on the medium and external part of the third metacarpal bone and articulated with the proximal phalanx of the fourth ray (arrow) resulting in very short fourth finger. (**h**) X-ray of the feet of Case 1: absence of malformations. (**i**) X-ray of the feet of Case 2. Left foot: fusion of third and fourth metatarsal bones at the basis and the proximal diaphyses (arrow); hypoplasia of fourth metatarsal bone with compensatory hypertrophy of fifth metatarsal bone; fusion of fourth and fifth proximal phalanges at the bases. Right foot: fusion of third and fourth metatarsal bones at the bases (arrow); hypertrophy of fifth metatarsal bone.

diagnosed at 8 years of age. She was also reported to have scholastic difficulties and slowness with an educational level actually equivalent to primary school.

On clinical evaluation, her height was 137.5 cm (−3.75 SD) and her weight was 43.6 kg (−2 SD) with a BMI of 23.1 kg/m^2. The ophthalmic examination showed a very poor visual function with only light perception and advanced retinal dystrophy on fundus examination. Severe hyposmia (score of 21/40 at the UPSIT®) was diagnosed but, unlike her sister, mobile cilia of the nose were abundant with normal morphology. On clinical examination, the fourth digit was shorter than the fifth (Fig. 1e). The X-ray showed synostosis of third and the fourth metacarpal bones (Fig. 1g). On left foot, extreme hypoplasia of the fourth metatarsal bone with compensative hypertrophy of the fifth metatarsal bone as well as synostosis of the fourth and the fifth proximal phalanges were recorded (Fig. 1i).

Molecular and functional analysis

Zygosity analysis

All the 15 STR loci analyzed to determine the zygosity type were informative. The copy number of the repeat sequence was identical for 5 loci and different for the other 10, showing that the twins were dizygotic.

Mutation screening

LZTFL1 gene (NM_020347.2) was screened *a priori* because of the unusual polydactyly (mesoaxial) noted in the twins akin to our previous BBS patient who had mutations in this gene (2). Two mutations were identified: a missense mutation (p.Leu87Pro; c.260T>C) and a nonsense mutation (p.Glu260*; c.778G>T). Parents were heterozygous for the changes (paternal mt/wt, maternal mt/wt).The p.Leu87Pro occurred in a well conserved site of *LZTFL1* and is predicted to be damaging with a score of 1.000 by POLY-PHEN-2. Both mutations are absent from the Exome Variant Server, 1000 genomes and dbSNP databases. Moreover, they were not found in 176 healthy controls analyzed by exome sequencing. Real-time PCR amplifying the cDNA sequence between the two mutations was performed *vs* a healthy control. LZTFL1 expression in the patients was halved (Fig. 1b). The LZTFL1 protein was detected at 30 kDa by Western blot in dermal fibroblasts and not at 35 kDa as expected (Fig.1c). We can make the assumption that the allele carrying the stop mutation leads to the production of a smaller size mRNA translated into a smaller size protein detected at 30 kDa. The second allele carrying the missense mutation probably leads to an abnormal splicing event with an aberrant mRNA with no protein translation as predicted by Human Splicing Finder website (www.umd.be).

Genotype-to-phenotype correlation

BBS is a highly pleiotropic and genetically heterogeneous disorder with poor genotype-to-phenotype correlations. The presence of mesoaxial polydactyly in subjects with *LZTFL1* mutations (the family previously reported (2) and the twins described herein) caught our attention. During the last 10 years, our laboratory has had access to well-documented clinical data of 110 patients known to carry mutation in the *BBS1* to *BBS16* genes; 89% (98/110) of them had at least one extra-digit. We therefore reviewed hands X-rays for the presence/absence of mesoaxial polydactyly in 41 BBS patients from 32 unrelated families of the 110 cases. None of those had mesoaxial polydactyly. Despite the limited number of *BBS17*-patients, the mesoaxial polydactyly may be associated with *LZTFL1*-related ciliopathy as it was observed in only *BBS17*-families (2/2) and none of other families with mutations in other BBS genes ($n = 32$; Fisher's Exact Test for small size samples: $p < 0.01$). Although this might just have occurred by chance, we do think it is a valuable hint to consider. The identification of additional *BBS17*-patients will definitively shed light on that hypothesis.

Discussion

Polydactyly is a frequent and well-recognized skeletal feature in ciliopathies (Table 1). In BBS, the polydactyly is a hallmark diagnostic feature (60% of patients (1)), mostly described as postaxial with normal metacarpal structure. To our knowledge, no BBS patients with mesoaxial polydactyly have been reported before our first family mutated in *LZTFL1* (2). Interestingly, two patients presenting with McKusick–Kaufman syndrome (a neonatal ciliopathy with mutation in BBS genes) with mesoaxial polydactyly have been reported in the literature but no molecular analysis was possible at the time (8). In our cohort, 89% of BBS patients presented at least one extra-digit, reported to be postaxial. Mesoxaxial polydactyly was only found in the two BBS families mutated in *LZTFL1-BBS17*: the family previously reported (2) and the twins described herein.

Mesoaxial polydactyly is a diagnostic feature for another ciliopathy namely Pallister–Hall syndrome with mutations in *GLI3*. Mutations in *GLI3* can be found both in isolated polydactylies, postaxial (types A1/B) or preaxial (type IV), and syndromic polydactylies (Pallister–Hall syndrome but also Greig syndrome) (6). *GLI3* encodes a protein which belongs to the C2H2-type zinc finger proteins subclass of the Gli family, characterized as DNA-binding transcription factors. GLI-3 has a dual function as an activator and a repressor of the sonic hedgehog (SHH) pathway and the activator/repressor ratio gradient specifies limb digit number and identity (16).

Interactions with many other proteins are involved in Gli3 trafficking at the level of the primary cilia. For example, KIF7 localizes to the base of the cilium and forms a complex with Gli proteins. Mutations were recently identified in *KIF7* in two ciliopathies: hydrolethalus syndrome and acrocallosal syndrome both characterized by postaxial polydactyly of the hands and preaxial polydactyly of the feet (7). In the same way, OFD1 is associated with a progressive loss of SHH

Table 1. Types and frequencies of polydactylies in different ciliopathies

Syndrome	Genes	Estimated frequency of polydactyly	Type of polydactyly	References
Short rib-polydactyly syndromes			Postaxial	
Type I: Saldino-Noonan	Unknown	++		
Type II: Majewski	*DYNC2H1*, *NEK1*	+++		
Type III: Verma-Naumoff	*IFT80*, *DYNC2H1*	++		
Type IV: Beemer–Langer	Unknown	+		
Type V	*WDR35*	(a)		(3)
EVC	*EVC*, *EVC2*	+++	Postaxial	(3)
Weyers	*EVC2*	++	Postaxial	(3)
Jeune syndrome	*IFT80*, *DYNC2H1*, *TTC21B*, *WDR19*	+	Postaxial	(3)
Joubert syndrome	*INPP5E*, *TMEM216*, *AHI1*, *NPHP1*, *CEP290*, *TMEM67*, *RPGRIP1L*, *ARL13B*, *CC2D2A*, *CXORF5*, *TTC21B*, *KIF7*, *TCTN1*, *TMEM237*, *CEP41*, *TMEM138*, *C5ORF42*, *TCTN3*, *ZNF423*, *TMEM231*	+ (8%)	Postaxial most often, preaxial rarely	(4)
Meckel syndrome	*MKS1*, *TMEM216*, *TMEM67*, *CEP290*, *RPGRIP1L*, *CC2D2A*, *NPHP3*, *TCTN2*, *B9D1*, *B9D2*	++ (83%)	Postaxial most often, preaxial rarely	(5)
Pallister–Hall syndrome	*GLI3*	++	Postaxial	(6)
		+++	Mesoaxial	
Greig syndrome	*GLI3*	+++	Preaxial or mixed postaxial hands and preaxial feet	(6)
Acrocallosal syndrome	*KIF7*	+++	Postaxial hands and preaxial feet	(7)
Hydrolethalus syndrome	*HLS1-S*, *KIF7*	+++	Postaxial hands and preaxial feet	(7)
Bardet–Biedl Syndrome	*BBS1*, *BBS2*, *ARL6*, *BBS4*, *BBS5*, *MKKS*, *BBS7*, *TTC8*, *PTHB1*, *BBS10*, *TRIM32*, *BBS12*, *MKS1*, *CEP290*, *C2Orf86*, *SDCCAG8*, *LZTFL1*	++ (60%)	Postaxial	(1)
McKusick–Kaufman Syndrome	*MKKS*	+++ (98%)	Postaxial	(8)
Oro-facial-digital syndromes				
Type I	*OFD1*	+ (1–2%) ++ (50%)	Pre/postaxial hands and preaxial feet	(9)
Type II: Mohr syndrome	Unknown	+++	Pre/postaxial polysyndactyly	
Type III	Unknown	(b)	Postaxial	
Type IV: Mohr-Majewski syndrome	*TCTN3*	++	Pre/postaxial	
Type V	Unknown	(c)	Postaxial	
Type VI	Unknown (one mutation reported in *OFD1*)	++	Post/mesoaxial hands and preaxial feet	(10)
Type VIII	Unknown	(a)	Pre/postaxial polysyndactyly	
Type IX	Unknown	++	Preaxial feet	
Type X	Unknown	(a)	Preaxial and oligodactyly	
Type XI	Unknown	(a)	Postaxial	
Cranioectodermal dysplasia type 3: Sensenbrenner syndrome	*IFT43*	(a)	Postaxial	(11)
Nephronophtisis type 15	*CEP164*	(a)	Postaxial	(12)
Retinitis pigmentosa				
Type 41	*PROM1*	(a)	Postaxial	(13)
Type64: CORD16	*C8Orf37*	(a)	Postaxial	(14)
Holoprosencephaly	*GLI2*	+	Postaxial	(15)

+, rare; ++, common; +++, consistently; (a), one family reported; (b), two families reported; (c), only Indian families reported.

signaling and an impaired processing of Gli3 leading to loss of digit identity (17). Mutations in *OFD1* are mainly described in oral-facial-digital (OFD) syndrome type I for which preaxial polydactyly is reported (9). One mutation was also found in a patient presenting with OFD VI and preaxial polydactyly (10). Finally, tectonic member family 3 (TCTN3) forms a complex at the ciliary transition zone with TCTN1 and TCTN2, both of which are also implicated in the transduction of SHH signaling. Mutations were identified in *TCTN3* in OFD IV syndrome but also in Meckel syndrome and Joubert syndrome for which polydactyly is mostly postaxial, rarely preaxial (18).

Interestingly, *LZTFL1* was recently identified as an important negative regulator of the BBSome ciliary trafficking and SHH pathway signaling (19). Our previous work confirmed these data showing a massive activation of the SHH signaling in the absence of LZTFL1 (2). The human phenotype we report suggests that LZTFL1 depletion impacts the development of the limbs and patterning of the digits through regulation of SHH pathway.

A mouse model would be of interest to pinpoint the action of LZTFL1 *in vivo* on limb patterning. The relationship between LZTFL1 and the developmental sequence implicated as well as the identification of LZTFL1 partners will be of major interest for the molecular understanding of the syndromes that include mesoaxial polydactyly.

In conclusion, herein we report a second BBS family with mutations in *LZTFL1* (*BBS17*) associated with mesoaxial polydactyly. This finding has to be confirmed on other cohorts worldwide, but seems to indicate a possible genotype–phenotype correlation. Thus, the presence of mesoaxial polydactyly in BBS patients, diagnosed on a clinical or X-ray basis, seems us to be an interesting feature to guide the molecular screening and to sequence directly *LZTFL1*.

Acknowledgements

We would like to thank the patients and their families for their very supportive involvement and all the patients' groups for their financial support. We would also like to thank Integragen, Evry, France, for sharing useful information concerning controls. This work was funded by RETINA France, UNADEV (Union Nationale des Aveugles et Déficients Visuels), FORMICOEUR, PHRC (Programme Hospitalier pour la Recherche Clinique) program 2007 and the Agence Nationale pour la Recherche GENOPATH call 2009.

References

1. Beales PL, Elcioglu N, Woolf AS, Parker D, Flinter FA. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J Med Genet 1999: 36 (6): 437–446.
2. Marion V, Stutzmann F, Gérard M et al. Exome sequencing identifies mutations in LZTFL1, a BBSome and smoothened trafficking regulator, in a family with Bardet–Biedl syndrome with situs inversus and insertional polydactyly. J Med Genet 2012: 49 (5): 317–321.
3. Huber C, Cormier-Daire V. Ciliary disorder of the skeleton. Am J Med Genet C Semin Med Genet 2012: 160C (3): 165–174.
4. Pellegrino JE, Lensch MW, Muenke M, Chance PF. Clinical and molecular analysis in Joubert syndrome. Am J Med Genet 1997: 72 (1): 59–62.
5. Alexiev BA, Lin X, Sun CC, Brenner DS. Meckel-Gruber syndrome: pathologic manifestations, minimal diagnostic criteria, and differential diagnosis. Arch Pathol Lab Med 2006: 130 (8): 1236–1238.
6. Biesecker LG. What you can learn from one gene: GLI3. J Med Genet 2006: 43 (6): 465–469.
7. Putoux A, Thomas S, Coene KL et al. KIF7 mutations cause fetal hydrolethalus and acrocallosal syndromes. Nat Genet 2011: 43 (6): 601–606.
8. Slavotinek AM, Biesecker LG. Phenotypic overlap of McKusick-Kaufman syndrome with Bardet-Biedl syndrome: a literature review. Am J Med Genet 2000: 95 (3): 208–215.
9. Gurrieri F, Franco B, Toriello H, Neri G. Oral-facial-digital syndromes: review and diagnostic guidelines. Am J Med Genet A 2007: 143A (24): 3314–3323.
10. Darmency-Stamboul V, Burglen L, Lopez E et al. Detailed clinical, genetic and neuroimaging characterization of OFD VI syndrome. Eur J Med Genet 2013: 56 (6): 301–308.
11. Arts HH, Bongers EM, Mans DA et al. C14ORF179 encoding IFT43 is mutated in Sensenbrenner syndrome. J Med Genet 2011: 48 (6): 390–395.
12. Chaki M, Airik R, Ghosh AK et al. Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling. Cell 2012: 150 (3): 533–548.
13. Maw MA, Corbeil D, Koch J et al. A frameshift mutation in prominin (mouse)-like 1 causes human retinal degeneration. Hum Mol Genet 2000: 9 (1): 27–34.
14. Estrada-Cuzcano A, Neveling K, Kohl S et al. Mutations in C8orf37, encoding a ciliary protein, are associated with autosomal-recessive retinal dystrophies with early macular involvement. Am J Hum Genet 2012: 90 (1): 102–109.
15. Roessler E, Du YZ, Mullor JL et al. Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features. Proc Natl Acad Sci USA 2003: 100 (23): 13424–13429.
16. Johnson RL, Tabin CJ. Molecular models for vertebrate limb development. Cell 1997: 90 (6): 979–990.
17. Bimonte S, De Angelis A, Quagliata L et al. Ofd1 is required in limb bud patterning and endochondral bone development. Dev Biol 2011: 349 (2): 179–191.
18. Thomas S, Legendre M, Saunier S et al. TCTN3 mutations cause Mohr-Majewski syndrome. Am J Hum Genet 2012: 91 (2): 372–378.
19. Seo S, Zhang Q, Bugge K et al. A novel protein LZTFL1 regulates ciliary trafficking of the BBSome and Smoothened. PLoS Genet 2011: 7 (11): e1002358.

IV- Transfer in routine to Strasbourg diagnostic laboratory: a one-year retrospective experience

Subsequently to the initial results, 67 other patients have been analyzed in research settings. The method was then transferred to the Strasbourg diagnostic laboratory in fall 2012, where it was integrated as a new routine test. Since the transfer, 47 new patients have been screened. I here provide the global results integrating all 152 patients (the previous 38 patients from Redin et al, plus 114 additional patients) that have been analyzed through research or diagnostic settings in order to give general insights about gene contribution and diagnostic efficiency.

1. Overall diagnostic yield in BBS patients

Among all patients screened through the targeted panel, some of them were addressed specifically for Alström (n=13), Joubert (n=4), Meckel (n=2), nephronophthisis (n=1) or 'ciliopathy' (n=1) patients and hence are not included in the overall diagnostic yield comprising patients strictly addressed for Bardet-Biedl syndrome. Out of the 152 patients, 131 were addressed for BBS and mutations could be identified for 60% of them (*i.e.* 79/131). This overall diagnostic yield is slightly lower than the one previously observed (68%) but does not reach significance ($p>0.2$). When we compare the proportion of patients meeting Beales' criteria and the others that do not (when clinical data was available), we observe that the cohort is roughly even for both categories: 50 patients that meet Beales' criteria, 53 that do not, which matches the proportions of our previous work (21 and 17 patients respectively). This does not explain the overall decrease in the general diagnostic yield.

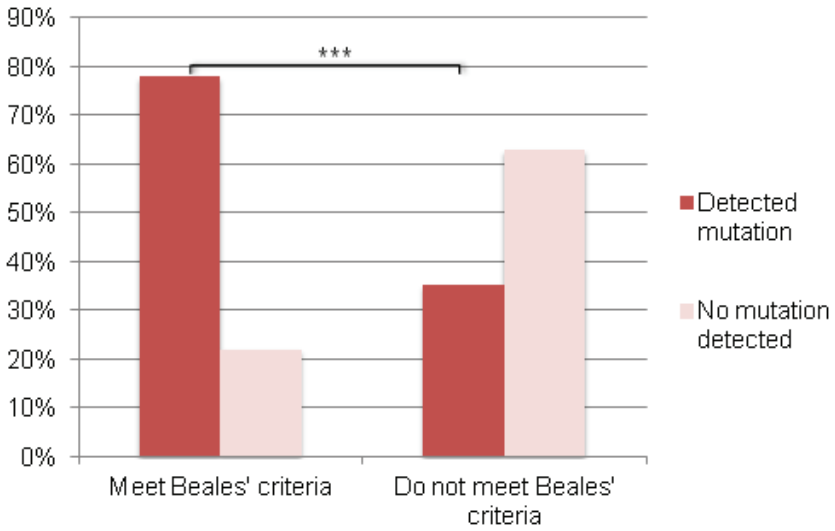


Figure 21: Diagnostic yield in patients is highly correlated to their compliance with strict BBS diagnostic criteria.

P-values were computed using the classical statistical test for comparison of two proportions.***: $p<0.002$.

When we look at the diagnostic yield per category (Beales' criteria vs not), we observe that the diagnostic yield among typical BBS patients is still significantly higher than the one in patients not meeting the diagnostic criteria (78% vs 35% respectively, p<0.002, see **Figure 21**) and very similar to the one previously reported (81%). The diagnostic yield among non-typical patients decreased (from 53% previously to 35% in the entire cohort), but the inclusion of even more highly atypical patients was encouraged within the past year encouraged by the still awkward 53% of success in non-typical BBS patients with our initial study. Importantly, we also had a few fetuses addressed for BBS testing, which do not meet Beales' criteria since many of them are features expressing in young childhood. The diagnostic yield from those patients was of 6/9, and no notable difference could be observed between the ones with mutations and the one without. Most fetuses were addressed since they presented with polydactyly (9/9), kidney abnormalities such as cystic or hyperechoic kidneys (7/9), hydrometrocolpos (3/9). Some other individual fetuses had vaginal atresia, hydramnios or encephalocele.

Finally, no obvious genotype to phenotype correlations could be drawn. However, it is interesting to note that most *BBS2* patients (8/9) and *BBS12* (4/4) comply with Beales' diagnostic criteria, which for both appear to be significant (p<0.05; test for comparison of two proportions; **Table 9**).

BBS gene	Meet Beales' criteria	Do not meet Beales' criteria	Fetuses
BBS1	7	4	2
BBS2	8	1	0
BBS3	2	2	0
BBS4	2	0	0
BBS5	3	4	0
BBS6	1	2	0
BBS7	2	3	1
BBS8	2	0	2
BBS9	2	1	1
BBS10	5	1	0
BBS12	4	0	0
ALMS1	1	1	0
Unknown	11	34	3
Total	**50**	**53**	**9**

Table 9: Mutation load of the different BBS genes and *ALMS1* in our cohort of patients according to their compliance or not to Beales' diagnostic criteria.

NB: Patients for which no clinical data was available are not reported here (but are in the overall **Table 11**).

2. Excess of *BBS1* heterozygous patients: towards being solved?

One early report linked to the recurrent *BBS1* (c.1169T>G, p.Met390Arg) mutation was that it was often detected at heterozygous state in BBS patients without a second heterozygous mutation in the gene on the second allele. One hypothesis was that the second allele carried a large deletion in *trans* in *BBS1*, which was not detectable by Sanger sequencing. Another hypothesis was that the mutations lied in another BBS gene, and that those patients were simply carrier for this relatively frequent variant (MAF=0.2% in Exome Variant Server). In one of the targeted sequencing runs, we included a majority of those heterozygous (c.1169T>G, p.Met390Arg) BBS patients, in order to test both hypotheses. A few of them indeed carried intragenic deletions on the second *BBS1* allele, but surprisingly most of the others carried the insertion of a repeat element (presumably of Alu type) on the second *BBS1* allele (c.1214_1215insAlu, p.?; **Figure 22**) that had been missed in Sanger sequencing presumably due to allele dropout. It appears to be somewhat recurrent in BBS patients, often in *trans* of the recurrent (c.1169T>G, p.Met390Arg) mutation in *BBS1*. Lastly, one the patient was only a carrier of the recurrent *BBS1* mutation, and carried biallelic splicing mutations in *BBS4*.

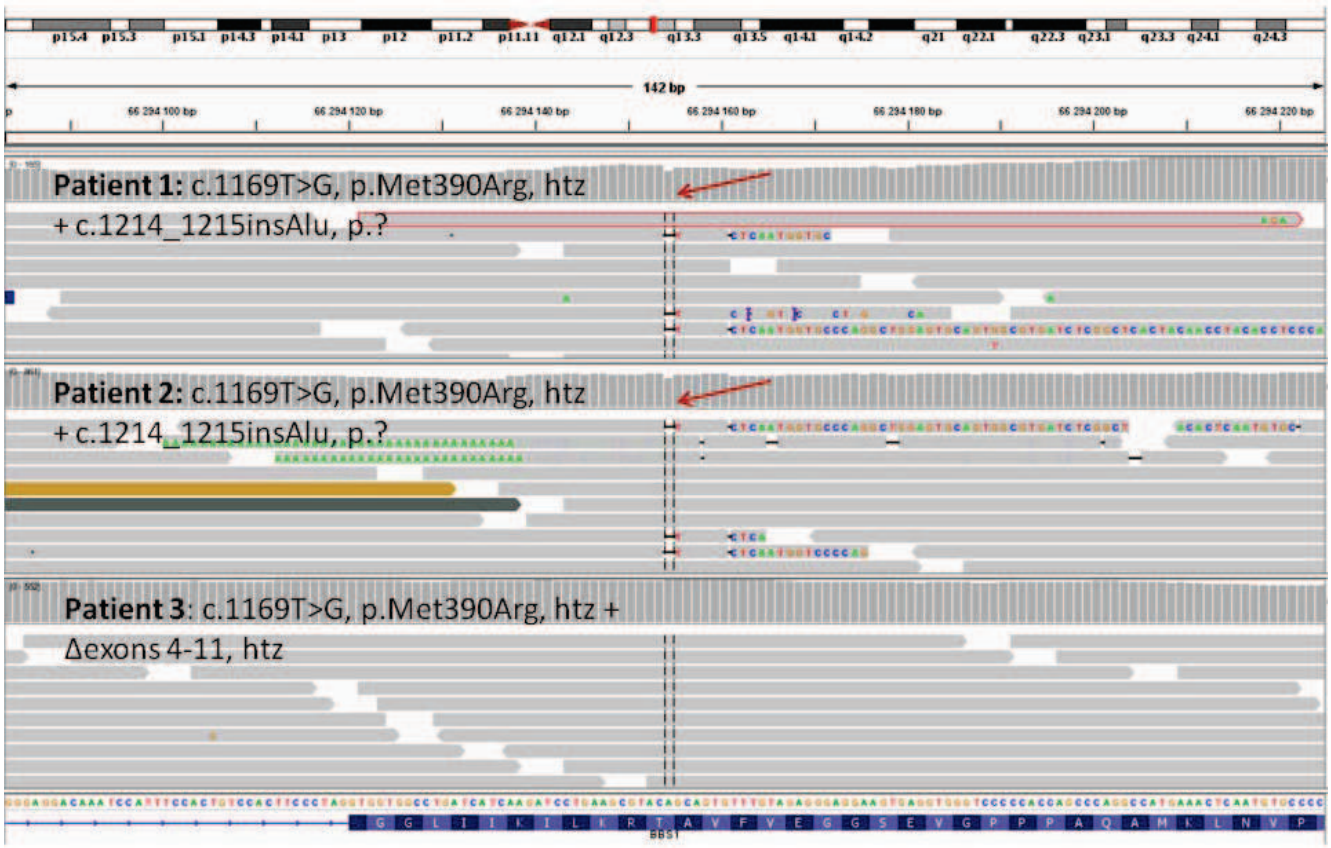


Figure 22: Visualization in IGV of three patients heterozygous for the recurrent (c.1169T>G, p.Met390Arg) mutation in *BBS1*. Two of them also (upper panels) also carry the novel *BBS1* recurrent mutation, detectable using targeted high-throughput sequencing.

3. Overall diagnostic yield in ALMS and other patients

Among the patients that were addressed for other syndromes, except for Alström patients no mutations could be detected. However, a large proportion of genes had not been included in our design, mainly focused on BBS. Indeed, only 7/20 of all Joubert syndrome genes and 5/11 of all Meckel syndrome genes were targeted, which represent less than the half. A somewhat larger proportion of Senior-Loken syndrome (4/5) and nephronophthisis (10/16) genes were included, but again the list was not exhaustive. Alström syndrome however is not genetically heterogeneous: a single gene *ALMS1* has been reported so far. Nonetheless, the coding sequence of this gene is so large than no diagnostic laboratory was proposing to screen it by Sanger sequencing in France (23 exons, 12 kb of coding sequences). A substantial number of ALMS-suspected patients were sent to Strasbourg diagnostic laboratory, and mutations could be found in 12/13 of them, suggesting that it might indeed be the only gene responsible for ALMS.

4. BBS and ALMS, true overlap or misdiagnosis?

As seen previously, almost all ALMS patients are diagnosed with an *ALMS1* mutation. Is the reverse true? Are all patients carrying *ALMS1* mutations ALMS patients? In our initial study, we reported two patients carrying *ALMS1* mutations that had been addressed for a BBS diagnostic test. Within the additional patients screened, a third case was reported amounting to a total of 3/131 BBS-suspected patients with *ALMS1* mutations. Among the patients with strict diagnostic criteria, that would lead to a frequency of 5% (3/90) of strict BBS patients with *ALMS1* mutations.

When looking back at phenotype of all three patients, it does not differ from the phenotype of the other 12 *ALMS1* patients suggesting they had simply been misdiganosed. When comparing the patients with *ALMS1* mutations with all patients with BBS mutations, two main clinical features appear significantly different from both cohorts: polydactyly is reported in 73% of patients with BBS mutations, as opposed to none with *ALMS1* mutations ($p<0.002$), and deafness is present in only 2% of patients with BBS mutations, as opposed to 50% with *ALMS1* mutations ($p<0.002$). All other major criteria (RP, obesity, hypogonadism, ID or learning disabilities and brachydactyly) are present in similar proportions (at least with no significant difference) within both cohorts (see **Table 10**). Other additional traits such as scoliosis and diabetes are reported in a fifth of ALMS patients while in few or none of the patients with BBS mutations may also be a hint to direct patients towards ALMS. A few (3/14) ALMS patients still may meet Beales' diagnostic criteria for BBS, specifically when they present with hypogonadism and renal abnormalities, in addition to the RP and obesity. Hence, both the absence of polydactyly and presence of deafness shall be taken under important consideration when discriminating BBS from ALMS patients (already reviewed in (Marshall *et al.*, 2013)). Also, the ophthalmic condition appears not to be exactly the same in both disorders.

Clinical feature	RP	Obesity	Deafness (or auditory defects)	Brachydactyly	Learning disabilities	Hypogonadism	Diabetes	Scoliosis	Renal abnormalities
Proportion in patients with *ALMS1* mutations	11/11	9/10	5/10	5/10	5/10	4/10	3/10	2/10	2/10

Table 10: Frequency of some clinical traits among patients with biallelic mutations detected in *ALMS1*.

5. Deciphering the patients with negative results

In order to determine the phenotypic differences between patients with a mutation detected in BBS genes from the others, we studied the frequency of the main clinical features among different populations (see **Figure 23).**

The overall differences from patients with or without mutations are the frequency of polydactyly, significantly higher in patients with a detected mutation (73% vs 23%, $p<0.002$), and the frequency of ID, significantly higher in patients without detected mutation (31% vs 53%, $p<0.05$). Specifically, ID is not one of the main features of BBS: patients are often reported with mild learning disabilities but rarely with moderate or severe cognitive defects. This explains the lower diagnostic yield in patients presenting with ID. Also, the overall differences between patients complying with Beales' criteria and patients that do not comply with Beales' criteria are significantly lower frequencies in polydactyly, hypogonadism, renal abnormalities and learning disabilities (**Figure 23**).

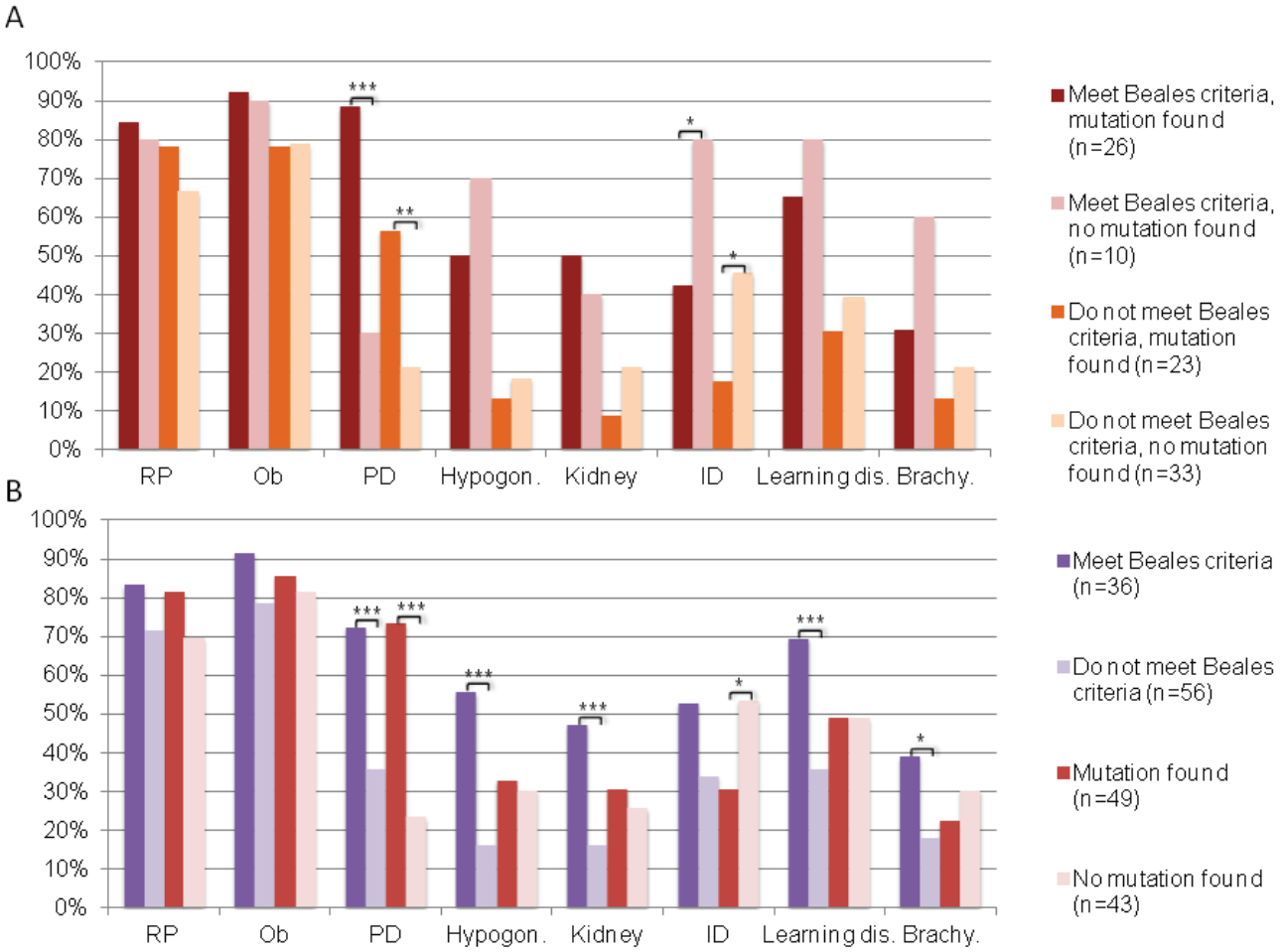


Figure 23: Frequency of BBS clinical features among different classes of patients, depending on whether a mutation could be identified, and/or whether they comply with Beales' diagnostic criteria.

A: Patients are separated according to their compliance to Beales' criteria and whether we were able to find biallelic mutations or not; the sum of all four categories represents the cohort of patients with available clinical data; **B**: Patients are separated only according to their compliance to Beales' criteria (purple colors),or to whether we were able to detect mutations or not (red colors): the sum of both purple categories (or both red categories) represents the cohort of patients with available clinical data; RP: retinitis pigmentosa; ob: obesity; PD: polydactyly; Hypogon.: hypogonadism; Kidney: renal abnormalities; ID: intellectual disability; learning dis: learning disabilities; brachy: brachydactyly. P-values were computed using the classical statistical test for comparison of two proportions. *: $p<0.05$; **: $p<0.01$; ***: $p<0.002$.

6. BBS genes load

Among all 79 BBS-suspected patients with detected mutations in BBS or ALMS genes, the frequency of mutated genes in decreasing order in the cohort is: *BBS1* (n=16), *BBS2* (n=10), *BBS10* and *BBS5* (n=8), *BBS9* (n=7), *BBS12* and *BBS7* (n=6), *BBS3* and *BBS6* (n=4), *BBS4*, *BBS8* and *ALMS1* (n=3), and *BBS14* (n=1). Our mutations were integrated into the vast BBS mutations reported in the literature, in order to get a better estimate of the contribution of mutations in each BBS gene in patients (**Table 11**). A total of 405 patients were included in this study (from (Muller

et al., 2010; Deveault *et al.*, 2011; Imhoff *et al.*, 2011; Abu-Safieh *et al.*, 2012; Aldahmesh *et al.*, 2014), all initial studies of each gene, and our cohort).

Gene	Literature	Strasbourg diagnostic laboratory – NGS experience
BBS1	23,7% (n=96)	18% (n=16)
BBS2	7,4% (n=30)	11% (n=10)
BBS3	2,0% (n=8)	4% (n=4)
BBS4	4,4% (n=18)	3% (n=3)
BBS5	4,7% (n=19)	9% (n=8)
BBS6	4,4% (n=18)	4% (n=4)
BBS7	3,0% (n=12)	7% (n=7)
BBS8	2,0% (n=8)	3% (n=3)
BBS9	3,7% (n=15)	8% (n=7)
BBS10	19,5% (n=79)	9% (n=8)
BBS11	0,2% (n=1)	0%
BBS12	9,1% (n=37)	7% (n=6)
BBS13	0,5% (n=2)	0%
BBS14	0,5% (n=2)	1% (n=1)
BBS15	0,2% (n=1)	0%
BBS16	1,0% (n=4)	0%
BBS17	0,5% (n=2)	NA
BBS18	0,2% (n=1)	NA
BBS19	0,2% (n=1)	NA
ALMS1	0,7% (n=3)	3% (n=3)
Unknown	14,6% (n=59)	12% (n=11)
Total	**100% (n=405)**	**90**

Table 11: Comprehensive report of mutations in BBS genes reported in patients from the literature.

For our cohort, the number of unknown cases includes only the patients meeting Beales' diagnostic criteria in which we did not find any mutations. Indeed, we did not want to include all of our unknown patients, since most of them did not present a typical BBS phenotype. We believe this could have introduced a bias by overestimating the proportion of patients that cannot be explained by a mutation in a BBS gene.

When comparing the entire cohort to the Strasbourg NGS cohort, the mutation load appears to be lower from what was previously reported in the literature for *BBS1* and *BBS10,* which is explained by the prior screening for the recurrent mutations (c.1169T>G, p.Met390Arg) in *BBS1*, and (c.271dup, p.Cys91fsLeu*95) in *BBS10* in most patients. Also the frequency appears higher in our cohort for some genes (*BBS2, BBS5, BBS7*, and *BBS9*) but without reaching significance. The

most frequent BBS genes still remain *BBS1*, *BBS10* and *BBS12*, all three accounting for over half of the mutations found in BBS patients.

7. Diagnosis of BBS in Strasbourg in the post-genomics era

Subsequently to the encouraging results obtained using targeted NGS, this approach has now been incorporated in routine in Strasbourg for the molecular diagnosis of BBS. The initial screening of BBS1 and BBS10 recurrent mutations is maintained (the novel *BBS1* recurrent mutation might also be added once its overall frequency will be assessed), but the screening of *BBS12* is not (not proficient enough in terms of positive patients subsequently to this screen).

The resulting new decisional tree is presented below (**Figure 24**).

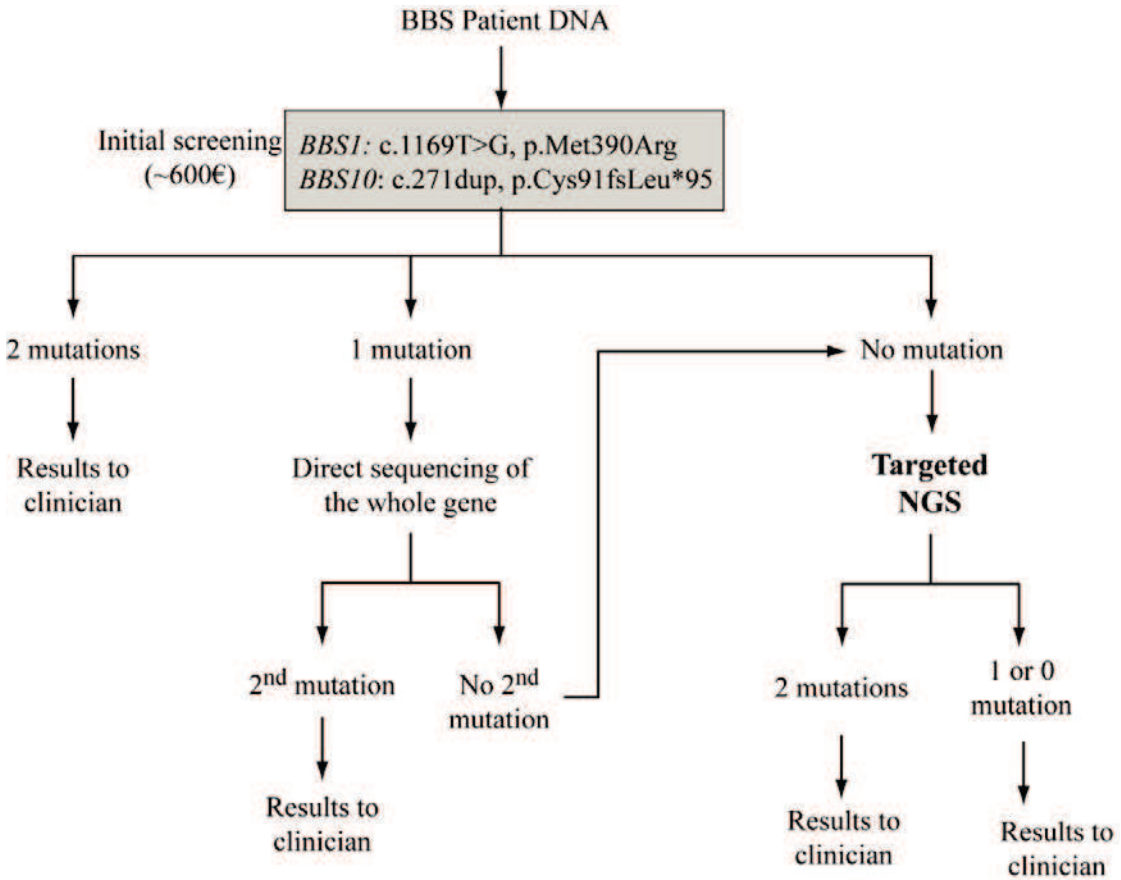


Figure 24: New decisional tree for the molecular diagnosis of BBS in Strasbourg incorporating the targeted NGS approach

V- Discussion

Bardet-Biedl syndrome is a relatively genetically heterogeneous disorder, with mutations that have been distributed across up to 19 different genes so far, precluding the implementation of an efficient diagnostic strategy by the classical strategies such as Sanger sequencing. Moreover, it shares overlapping clinical features with other ciliopathy disorders such as Joubert, Senior-Loken, Meckel, Alström syndrome or nephronophthisis. This phenotypic overlap reflects a genetic overlap, with many genes shared between those different syndromes. Alström syndrome might be the only exception, with a single gene reported so far and not shared with another ciliopathy. Nonetheless, the extensive size of its coding sequence has not facilitated the diagnostic offer, since Sanger sequencing of its 24 coding exons is not proposed in any French diagnostic laboratory for instance.

We developed a successful alternative method for the diagnosis of Bardet-Biedl and Alström syndromes based on targeted high throughput sequencing. This method can be incorporated relatively easily in diagnostic settings. Also, even if the acquisition costs of next-generation sequencing machines are significant, the relatively small size of the target design (200 kb) does not require an ultra high throughput sequencer in order to be implemented. Smaller throughput sequencers such as the Life Technologies PGM or the MiSeq from Illumina are perfectly suited for such applications of small-scale targeted sequencing.

The overall diagnostic yield across all BBS-suspected patients is of 60%, but much higher (81%) in the strict BBS patients (*i.e.* fulfilling Beales' diagnostic criteria) alone. The group of patients remaining without positive diagnostic is significantly enriched in patients presenting without polydactyly and with intellectual disability. However, polydactyly is one of the major BBS clinical features while ID is not. BBS patients generally face learning disabilities or in the worst-case mild cognitive impairment. The findings of a more severe degree of ID (moderate, severe or profound) in patients shall be an exclusion criterion for BBS. Hypogonadism, although reported among the major clinical feature, is reported in only 33% of all patients in whom a mutation is detected. The typical clinical features harbored by patients with a positive diagnostic are obesity (86%), retinitis pigmentosa (82%), polydactyly (73%) and learning disabilities (49%). Other features such as hypogonadism, renal defects, and mild cognitive impairment are reported in about a third of them.

Where do the mutations of the remaining 20% BBS patients lie? First, we recently identified a novel apparently recurrent mutation in *BBS1* (c.1214_1215insAlu, p.?) that had been previously missed by Sanger sequencing due to allelic dropout. Patients that remained negative prior to this discovery shall be screened for this recurrent mutation, often reported in trans with the other recurrent *BBS1* mutation. Also, at the time of the design, the most recent BBS genes (*BBS17-19*) had not been identified yet. Some chances exist that among the negative patients, a few carry mutations in those three genes. Since the cohort of negative patients is of significant size, testing

this hypothesis by screening the three genes with Sanger sequencing would be too expensive and time-consuming. We propose this could be more easily and rapidly verified by testing all three genes using the molecular inversion probes system for instance coupled with high throughput sequencing. Many of the most recent BBS genes (*BBS11*, *BBS13*, *BBS15, BBS18*, *BBS19*) have not been replicated so far in BBS patients. *BBS18* and *BBS19* having been reported very recently, this seems fair. Nonetheless, for others and especially *BBS11* for which the initial report was published in 2006, the waiting for replication studies in BBS patients persist. Regarding *BBS11*/*TRIM32*, mutations had initially been reported in patients presenting with limb girdle muscular dystrophy and other have been reported since (Frosk *et al.*, 2002; Saccone *et al.*, 2008). Interestingly, disruptions of *TRIM32/BBS11* were recently associated to a risk factor in autism spectrum disorders and other neurodevelopmental disorders (Lionel *et al.*, 2014). Evidences are thus lacking regarding its implication in BBS, and it would be interesting to perform exome sequencing on the initial family to see whether other candidate mutations are detected elsewhere. Concerning *BBS13/MKS1* and *BBS15/C2Orf86*, since they were initially reported in other ciliopathies we can legitimately ask whether the initially reported patients considered as BBS were not simply the tail of the corresponding ciliopathy spectrum rather than true BBS patients.

An alternative hypothesis is that missing mutations are located in not-yet-discovered BBS genes. Also, we may have missed mutations located in non-targeted regions (UTRs, deep intronic or promoter regulatory regions, etc). A recent report from Bachmann *et al.* (25th annual meeting of the German Society of Human Genetics, abstract) favors the second hypothesis by reporting a diagnostic yield of over 95% in classical BBS patients (including a few mutations in *ALMS1*), supporting the idea that not many BBS genes (if any) are left to discover. It would be interesting to see whether they also identified the novel *BBS1* mutation (c.1214_1215insAlu, p.?), and if not if it is present in some of the negative patients. Lastly, the oligogenic hypothesis could explain that we indeed found some pathogenic heterozygous variants without second mutated alleles, which along with another heterozygous variant in another gene would lead to the BBS phenotype. We did not find any evidence in favor of the triallelic model, and neither Bachmann and colleagues. The pure oligogenic model however – if true – will remain a herculean task to prove.

Concerning Alström syndrome, among patients that had been addressed to the diagnostic laboratory for an ALMS diagnosis the overall diagnostic yield is very high (93%, 95% confidence interval: 66-98%). A few additional patients that were initially addressed for BBS carried mutations in *ALMS1*. When comparing the phenotype of both classes of patients (ALMS versus BBS-suspected) they do not appear significantly different. We posit that the BBS-suspected patients carrying *ALMS1* mutations had been misdiagnosed. Indeed, BBS and ALMS although phenotypically overlapping are two distinguishable clinical entities. The ophthalmic defects are not exactly similar in both syndromes. Also, polydactyly is a major specific feature for BBS always absent in ALMS patients, while deafness and auditory defects are frequent in ALMS patients but

rare among BBS patients (rather reported as mild abnormalities of perception). Nonetheless, when ALMS patients present with retinal dystrophy, obesity, kidney defects and hypogonadism they might be falsely considered as BBS patients, depending on the area of expertise of the clinicians (*i.e.* the differential diagnosis will be more obvious for an ophthalmologist than for a pediatrician for instance). *ALMS1* should thus be systematically screened along with the other BBS genes. Finally, additional ALMS patients need to be screened to assess whether there is truly no other gene than *ALMS1* responsible for Alström syndrome, which would remain the exception of a non-genetically heterogeneous ciliopathy.

RESULTS - PART II

TARGETED SEQUENCING FOR THE DIAGNOSIS OF INTELLECTUAL DISABILITY

I- Introduction

Intellectual disability refers to a very heterogeneous group of conditions and belongs to the large family of neurodevelopmental disorders. This section aims at giving a broader background about intellectual developmental disorders, regarding the associated definitions, classifications, their epidemiology and their etiology.

1. A bit of History…

Intellectual disability refers to what was previously called Mental Retardation, or a plethora of other terminologies long before that (such as cretinism, feeblemindedness, idiocy, etc) that were eventually classified as disrespectful or politically incorrect. The first references to individuals with cognitive deficits date as far back as long before Christ, from Egyptian, Greeks and Romans civilizations as evidenced in papyrus or other writing traces. Early on, such individuals were considered as 'anger of Gods', and were abandoned and excluded from the cities (Harris, 2006). The concept of eugenics (even though the term was introduced in the 19th century by Francis Galton) started in antiquity, through a meticulous selection of strong and healthy newborns or a controlled mating between the elite of man and women in ancient Greece (Prof Jacques Battin, conference, 7ème assises de Génétique Humaine et Médicale).

Later on, individuals with mild ID were not discriminated from the others and considered as socially competent, while more affected ones were placed under protection of families or religious instances such as monasteries. The first allusion of ID as a disease dates back in the 17th century, when Dr Willis proposed it originated from structural brain abnormalities that could be congenital or acquired.

In the 19th century, instruction programs to rehabilitate and reintegrate ID people into societies started to develop in Europe (under the initiative of the French Dr Seguin) and in North America (under the lead of the Bostonian Drs Howe and Wilbur), with the spring of private and State schools dedicated to 'Idiotic Children'. The first dedicated residential institution was created in Switzerland in 1841 (Beirne-Smith *et al.*, 2006b).

Eugenics

In the second half of the century, ID individuals started to be massively segregated in self-sufficient residential institutions, due to the lack of proven efficiency of those instruction programs. That is about when Galton started to promote the eugenics movement, aiming at controlling reproduction of individuals with ID. ID individuals appeared as the scapegoats of the societies, blamed for poverty, illness and criminality. All of which were thought to be heritable traits that

were encouraged to develop through the improvements of medicine, interfering with natural selection by engendering 'regression towards mediocrity'. In response to this, countries such as Sweden, Norway and Denmark in Northern Europe, or the USA developed forced sterilization laws for 'feebleminded' at the beginning of the 20th century, which were largely supported by medical literature(Radford, 1991; Beirne-Smith *et al.*, 2006a). Specifically, in Germany under the nazi regim individuals with ID were initially sterilized (inspired from the American programs of forced sterilization) and even exterminated. In other countries such as the UK, segregation and massive custodial institutionalization was applied, less drastic but with nonetheless the same aim at limiting reproduction.

> *"Custodial institutions were most importantly the means by which the feebleminded were removed from a society in which they were perceived as a genetic threat and placed in isolated environments, completely segregated by gender"* (Radford, 1991).

It is important to mention that even though largely supported by the scientific community, institutionalization and segregation were not accepted at the unanimity, as illustrated by the early propos held by Dr Howe in the 1870s.

> *'The true, sound principles are: separation of idiots from each other; and then diffusion among the normal population. For these and other reasons it is unwise to organize establishments for teaching and training idiotic children, upon such principles as will tend to make them become asylums for life . . . Even idiots have rights which should be carefully considered!'*(Howe, 1874)

A hundred years later, Dr Wolf Wolfensberger raised again such disagreement in his work 'The origin and Nature of Our Institutional Models' in 1969, which marked the impulsion for the desegregation movement by governments. Such practices were gradually abolished by the end of the 19th century, and replaced by codified rights (including education) for people with disabilities. Among others, the main motivation for this was the implication of several other non-genetic factors contributing to ID as shown by extensive twin family studies, partially refuting the arguments for eugenics (Rainer, 1971). Tendency in the 21st century is now towards deinstitutionalization and community-based services (reflecting the ideal of Dr Howe and comparses), and reintegration into societies (reviewed in (Harbour and Maulik, 2010)).

2. Definition(s)

Several definitions have been proposed for ID, or formerly Mental Retardation (MR). Intellectual disability, or intellectual developmental disorders as called in the DSM-V, is defined by '*intellectual and adaptative functioning deficits in conceptual, social and practical domains originating during child development*'. Intellectual functioning reflects the ability to learn, reason,

memorize, and solve problems, while adaptative behavior includes the ability to communicate, interact socially and cope independently.

The definition provided by the World Health Organization in the ICD-10 is similar, defined as '*an arrested or incomplete development of the mind, which is characterized by impairment of skills manifested during the developmental period, which contribute to the overall level of intelligence (i.e. cognitive, language, motor and social abilities)'.*

Although the evaluation of intellectual functioning itself is complicated as intelligence is not a unitary characteristic but rather integrates a large number of different skills, ID is usually characterized by using the intelligence quotient (IQ) measure. Classically, individuals with an IQ below 70 are considered as presenting with ID. Adaptative functioning is assessed independently.

IQ is the final score of several combined psychometric tests based on reasoning, speed processing, memory, spatial ability, etc, and only reflects intellectual functioning abilities. Different IQ tests exist, either consisting of complex tasks involving on cognitive aspect such as reasoning (*e.g.* Ravens Progressive Matrices, or batteries of tasks interrogating different cognitive performances (*e.g.* the Wechsler Intelligence Scale for Children or WISC). All intelligence tests have been normalized to follow a normal distribution, with a mean score μ of 100 and a standard deviation σ of 15 (*i.e.,* 95% of the population have scores between 70 and 130; see **Figure 25**).

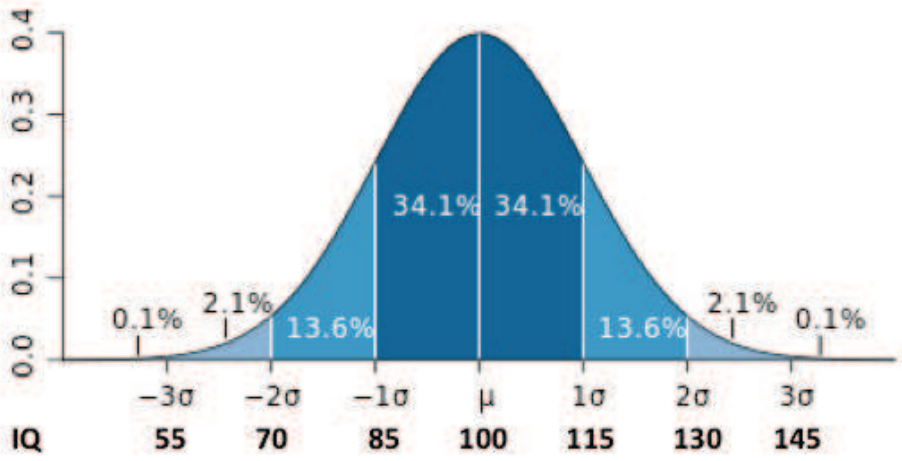


Figure 25: Theoretical distribution of IQ scores, which follows the normal distribution, within the general population.

Although the general tendency is for all intellectual functioning skills to develop to a similar level in each individual, there can be large discrepancies and especially in people with ID. Some individuals may present a severe impairment in one area (*e.g.* speech) while having another area of much higher skill (*e.g.* memory) complicating the evaluation. Nonetheless, whatever the intelligence test, it was shown that all tests appear strongly correlated and give consistent IQ results except for ASDs patients that may present frequently with such skill discrepancies (Deary *et al.*, 2010).

3. Classifications

ID can be reported in patients with a different degree of severity. Several classification methods exist to characterize the severity of the cognitive impairment in patients, based either on IQ values, the adaptative functioning deficit, or the intensity of medical and personal care required. All methods lead to a similar classification in four categories.

In the ICD-10, the four classes have been delineating using IQ thresholds criteria (see **Table 12**), with varying impairments evolving with age. The proportion of patients within the four classes of mild, moderate, severe and profound ID is assessed to be 85%, 10%, 3-4%, and 1-2% respectively (King, 2009). However, some epidemiological studies only divide patients into mild (IQ 50-70) and severe groups (IQ<50, regrouping the official moderate, severe and profound classes).

Degree of ID	**0-5 years** Maturation and development	**6-20 years** Training and education	**21 years and beyond** Social and vocational adequacy
Mild (IQ 50-70)	Generally develop communicative and social skills. May not be distinguishable until beginning school.	Can learn up to 4th/5th primary school grade skills when reaching the ages of 18 or 19 years. Can integrate into society.	Partially autonomous. Capable of acquiring social and work skills for integration into the work force at minimum wage
Moderate (IQ 35-50)	Can speak or learn to communicate. Some difficulties with motor skills.	Difficulty meeting 2nd primary school grade academic objectives.	May be able to partially maintain oneself economically in manual work under protected conditions
Severe (IQ 20-35)	Marked limitations in motor skills. Minimal language ability.	Can speak or learn to communicate. Can learn elemental self-care and health habits.	Can partially contribute to maintaining oneself economically under total supervision
Profound (IQ<20)	Significant delay, minimal functional ability in sensorimotor areas. Needs basic care.	Fully dependant, and need permanent care.	Some motor and language development. Can learn very limited personal care skills.

Table 12 (adapted from (Katz and Lazcano-Ponce, 2008)): Classification of different degrees of ID, with associated hallmarks depending on the age.

4. Epidemiology

ID is one of the most common neurodevelopmental disorder (with Autism Spectrum Disorders) existing in each socioeconomical class or country, with an overall prevalence of 1-3% showing an approximate 30% of male excess (Leonard and Wen, 2002; Maulik *et al.*, 2011). Its frequency (especially for milder forms) appears to be somewhat higher in lower social classes and developing countries, revealing the contribution of environmental factors (Leonard and Wen, 2002;

Emerson and Hatton, 2007). Although epidemiologic studies show discrepant results for the prevalence of mild ID worldwide (which appears to be highly influenced by environmental factors), most show consistent estimates for the prevalence of severe ID that is assessed to be of 0.3-0.5%, supposing a higher contribution of genetic factors in those forms (Durkin, 2002; Ropers, 2010).

5. Etiology of ID

The origins of ID are highly complex, numerous, and heterogeneous. They include genetic disorders, and acquired disorders that originate from developmental or environmental factors (**Table 13**).

I- Genetic Factors
Chromosomal anomalies, submicroscopic CNVs *e.g.: Down Syndrome, 16p11.2 microdeletion/duplication, 22q11 microdeletion, intragenic MECP2 duplications, etc…*
Single gene disorders
XLID genes (OMIM : 102 genes) *e.g. Fragile-X syndrome, Rett syndrome, ARX mutations, etc…*
Autosomal dominant genes (OMIM: 141 genes) *e.g.: Pitt-Hopkins syndrome, Smith-Magenis syndrome, mutations in epileptic genes etc…*
Autosomal recessive (OMIM: 330 genes) *e.g.: microcephaly genes, lissencephaly genes, etc…*
Oligogenism *e.g. CNVs' and rare variants' burden*
II- Acquired factors
Developmental or Environmental
1- Toxic *e.g.: poisoning, fetal alcohol syndrome, drugs, maternal phenylketonuria*
2- Infectious *e.g.: Rubella, cytomegalic inclusion body disease, toxoplasmosis, listeria …*
3- Traumas *e.g.: birth traumas, cerebral hemorrhage, ischemia*
4- Sociocultural

Table 13 (adapted from (Katz and Lazcano-Ponce, 2008)): Summarized proposed etiology of intellectual developmental disorders.

Whether ID was a heritable trait remained debatable for long. As mentioned in the previous part, it was first proposed to be heritable by Francis Galton in the 19th century. The heritability of ID was then largely refuted in the 20th century in favor of environmental factors, to be eventually proven as a major contributor as suggested by monozygotic twin studies.

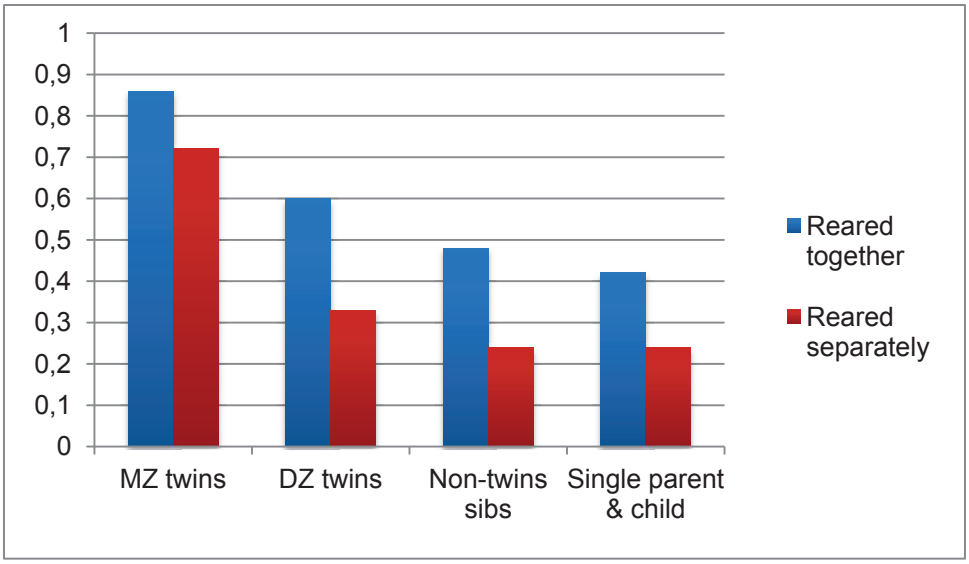


Figure 26 (adapted from (Baroff and Olley, 1999)): Familial correlations in general IQ depending on whether children are raised within the same environment or not.

(MZ: monozygotic twins, DZ: dizygotic twins)

For instance, as illustrated by the above figure, the correlation of IQ between family members shows the influence of both genetic and environmental factors on intellectual functioning. This graph supports a major role for genetic factors in IQ, as illustrated by the IQ correlation between monozygotic twins within a similar environment that is close to 1. Nonetheless, it also shows the substantial influence of environmental factors by the fact that this correlation between monozygotic twins is lesser than one when they are raised under the same roof, and is even lower when they are raised separately. The same observation is drawn for dizygotic twins and other siblings. The fact that monozygotic twins reared apart share a more similar IQ than dizygotic twins reared together confirms that the genetic background plays a preponderant role in intelligence bypassing the one of environment.

Genetic causes, which will be detailed in the following section, account for about 40% of all ID cases and include chromosomal abnormalities (either numerical or structural), and monogenic disorders caused by point or small DNA alterations. Monogenic disorders responsible for ID can be associated to an X-linked, an autosomal dominant or an autosomal recessive mode of inheritance. Lastly, some of those chromosomal abnormalities or DNA alterations are not associated to a full penetrance of ID and are rather considered as risk factors, or susceptibility loci.

Acquired causes of ID are thought to represent a smaller proportion of all ID cases (from 1 to 20%, depending on the sources (Curry *et al.*, 1997; van Karnebeek *et al.*, 2005; Rauch *et al.*, 2006). ID mostly originates from an exogenous insult that occurs during a critical step of embryonic brain development (either at the pre, peri or postnatal stage) resulting in irreversible cerebral damages. Acquired causes regroup toxic, infectious and traumatic conditions.

Altogether, the real distribution of IQ in the general population reflects the presence of those individuals with impaired intellectual functioning, for which IQ follows another normal distribution with a mean 35 and which primarily includes people with ID of both genetic and acquired causes (**Figure 27**).

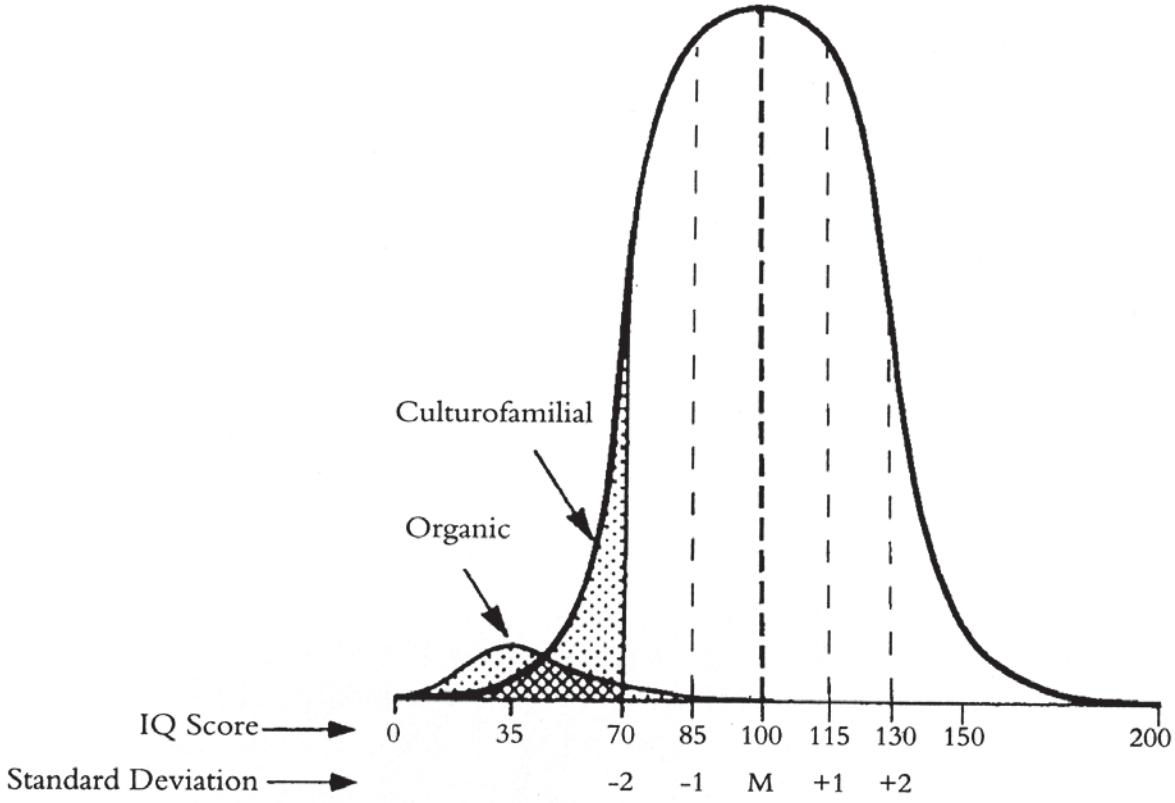


Figure 27 (from (Tarjan *et al.*, 1960)): Observed distribution of the IQ in the population as opposed to the theoretical normal distribution (Figure 25), showing an additional 'bump' curve of mean 35 primarily including individuals with ID of acquired or genetic etiology.

II- Genetics of intellectual disability

1. Chromosomal aberrations and copy number abnormalities

Chromosomal aberrations play a preponderant role in the etiology of ID, accounting for a majority of all ID cases. They include numerical abnormalities (supernumerary or missing chromosomes), large unbalanced or balanced structural variations (duplications, deletions, translocations, inversions) and submicroscopic CNVs (copy number variants, <5-Mb). About 10-15% of all ID cases are assessed to originate from large chromosomal anomalies (numerical or structural) and up to 20% when including submicroscopic CNVs, and are mostly reported in syndromic patients (Cooper *et al.*, 2011).

Numerical abnormalities

The first chromosomal cause that was identified as responsible for ID was the presence of a third copy of chromosome 21 in cells of patients with Down syndrome. It was discovered in 1959 by a French team (Marthe Gautier and Jérôme Lejeune; (Lejeune *et al.*, 1959)). Down syndrome, or trisomy 21, is the most frequent genetic cause of ID and accounts for about two thirds of all cases of ID caused by chromosomal aberrations with a prevalence of 1:750 to 1:800 (Ropers, 2010). Numerical chromosomal abnormalities of sex chromosomes (*e.g.* Turner or Klinefelter syndromes) may include ID as a secondary feature, but not as a major and consistent clinical trait so they are usually not considered as classical genetic causes of ID. Mosaic aneuploidies (*e.g.* of chromosome 8 and 9) have also been reported in patients with ID.

Large structural abnormalities and submicroscopic CNVs

A significant proportion of large structural variants (SVs) and submicroscopic CNVs reported in patients with ID are rare and private and can be dispersed throughout the genome, for which determining their contribution to the ID phenotype is highly complex. The larger they are, the higher the probability is for them to be pathogenic. A general burden of SVs and CNVs has been reported in ID cases as compared to controls, especially of large size (Cooper *et al.*, 2011). Most of those arise *de novo*, in favor of a pathogenic role. Several genes responsible for monogenic forms of ID have been uncovered subsequently to the identification of private reciprocal translocations/inversions (*e.g. MBD5, HUWE1*).

Nonetheless, many submicroscopic CNVs appear to be recurrent in patients with ID and are associated to specific clinical characteristic (*i.e.* some specific recognizable features are reported in patients such as facial dysmorphism or cardiovascular defects (Cooper *et al.*, 2011)). Some of them include a minimal 'critical' region, which is the smallest region that when deleted/duplicated results

in a characteristic phenotype in patients. This critical region may include one single gene. Point mutations in that gene may then be reported in patients presenting with a similar phenotype (*e.g.* Smith-Magenis syndrome, caused by 17p11.2 deletions or *RAI1* point mutations). Over 60 recurrent genomic intervals involved in submicroscopic deletions and/or duplications syndromes have been associated to ID and related neuropsychiatric disorders such as ASDs, schizophrenia and epilepsy (see **Table 14**; (Cooper *et al.*, 2011)). Telomeric CNVs are also a recurrent feature, with 12 chromosomes being particularly at risk (Cooper *et al.*, 2011).

CNV locus	Size (kb)	Del/Dup	Associated genomic disoder	ID degree	Assoc. to ASD	Assoc. to SCZ	Assoc. to EPI	Reported in controls/ unaffected parents
1q21.1	1,350	Del		Mild to Moderate	Yes	Yes	Yes	CT, UP
	1,350	Dup		Mild to Moderate	Yes	ND	ND	CT, UP
3q29	1,500	Del		Mild to Moderate	Yes	Yes	ND	-
	453 to 1,760	Dup		Mild to Moderate	No	ND	ND	UP
7q11.23	2,284	Del	Williams-Beuren Syndrome	Mild to severe	No	Yes	ND	-
	2,284	Dup		Mild to severe	Yes	No	Yes	-
15q11-q13	4,770 to 8,147	Del (mat.)	Prader-Willi syndrome	Mild to severe	No	ND	Yes	-
	4,770 to 8,147	Del (pat.)	Angelman syndrome	Mild to severe	Yes	ND	Yes	- (15q11.2 deletion: CT)
		Dup		Mild to severe	Yes	Yes	Yes	
15q13.3	1,930	Del		Mild to severe	Yes	Yes	Yes	UP, CT?
	1,930	Dup		Mild to Moderate	Yes	Yes	No	UP, CT
16p11.2	700	Del		Mild to Moderate	Yes	No	Yes	UP, CT
	700	Dup		Mild to Moderate	Yes	Yes	No	CT
17p11.2	3,775	Del	Smith-	Moderate	No	No	Yes	-

			Magenis syndrome	to severe				
	3,775	Dup	Potocki-Lupski syndrome	Mild to Moderate	Yes	No	No	-
22q11.2	1,500 to 3,000	Del	Velocardio-facial syndrome	Mild to severe	Yes	Yes	Yes	-
	1,500 to 3,000	Dup		None to mild	Yes	ND	ND	UP
22q13.3	1,000 to 9,000	Del	Phelan-McDermid syndrome	Mild to severe	Yes	Yes	Yes	-
	1,000 to 5,000	Dup		Mild to severe	Yes	ND	ND	-
Xq28		Del		Moderate to severe	Yes	No	Yes	-
	215-640	Dup	*MECP2* duplication	Moderate to severe	Yes	No	Yes	-

Table 14: List of most recurrent CNVs associated with ID showing the overlap with other neuropsychiatric and neurologic disorders.

Adapted from (Morrow, 2010; Cooper *et al.*, 2011; Grayton *et al.*, 2012)

Some of those recurrent CNVs and private SVs show nonetheless a reduced penetrance (*i.e.* might be inherited from apparently unaffected parents or reported in other apparently unaffected relatives, might be reported in other control populations). Their frequency in patients with ID appears significantly higher than in controls, and they are thus considered as susceptibility loci or risk factors.

2. Monogenic forms of ID: XLID

The 30% sex bias reported in ID was long thought to be originating from mutations that would involve genes located on the X-chromosome, making males more affected with this condition. Gene discovery research has long been focusing on X-linked ID (XLID), and X-linked genes through the study of large families with apparent X-linked mode of transmission of the disease (historically MRX families). Many international consortiums took part in those intense investigations based on linkage analysis, translocation studies, clinical examination, etc, reuniting teams mainly from Europe, Australia and the US. This led to the identification of over a hundred genes supposedly implicated in XLID (Lubs *et al.*, 2012). XLID is assessed to contribute to up to

10% of ID cases in males ((Lubs *et al.*, 2012)), which remains insufficient to explain entirely the observed sex bias.

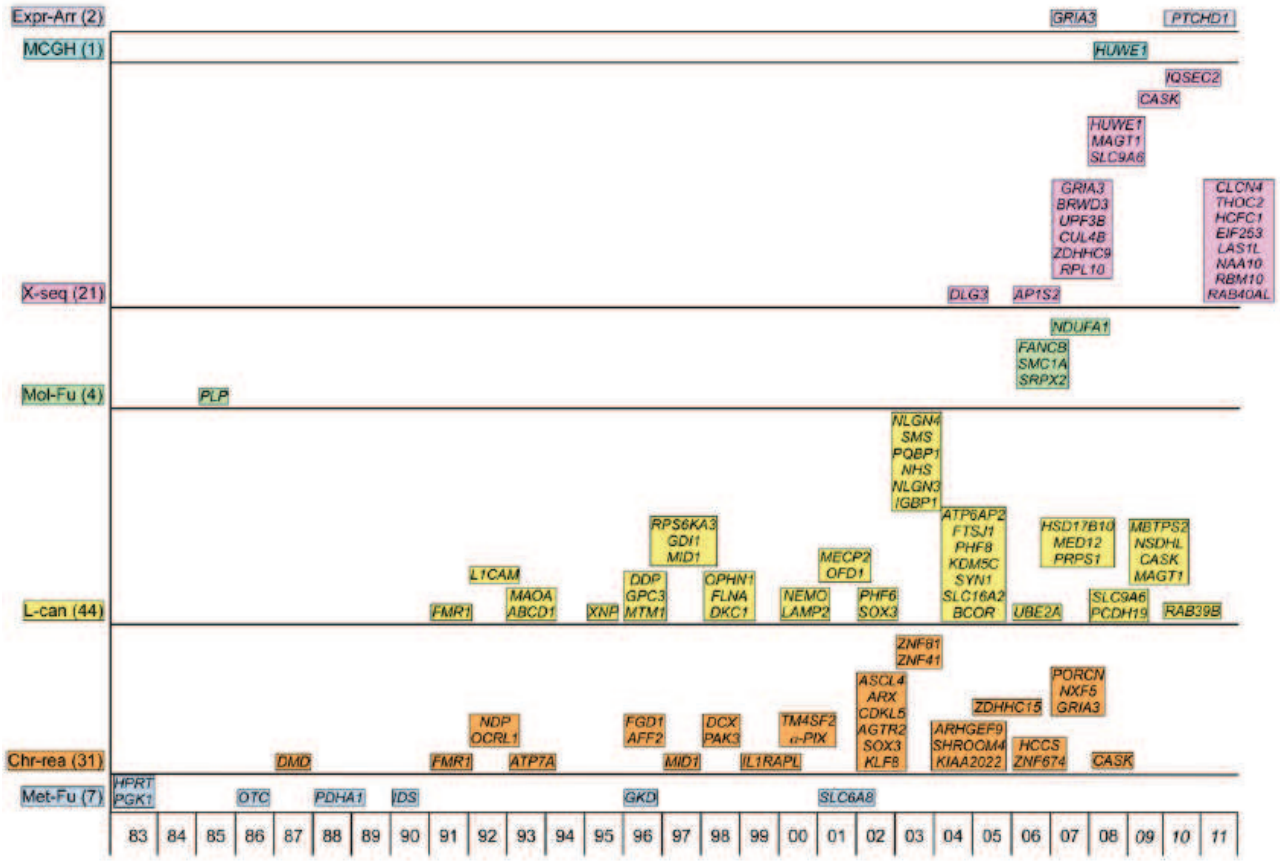


Figure 28 (adapted from (Lubs *et al.*, 2012)): Chronological history and evolution of methods for XLID novel genes discovery

NB: Genes that have been identified since 2011 and are therefore not mentioned in this figure are listed below (**Table 15**). For both **Figure 28** and **Table 15**, the following abbreviations are used: Exp-Arr = expression microarray; MCGH = genomic microarray; X-seq = gene sequencing; Mol-Fu = follow up of a known molecular pathway; L-can = candidate gene testing within a linkage interval; Chr-rea = positional cloning based on a chromosome rearrangement; Met-Fu = follow up of a known metabolic pathway.

The first evidence of a genetic abnormality of the X-chromosome leading to ID was the observation of a chromatid break at the extremity of the long arm the chromosome X, detectable cytogenetically in two brothers affected with ID in 1966. This was the premise of Fragile-X syndrome, characterized by moderate to profound ID, speech delay, facial dysmorphic traits (large ears, long face, and prominent jaw), macroorchidism and behavioral disorders. The corresponding gene – *FMR1,* encoding the FMRP protein – was identified later on by positional cloning in 1991. The most common mutation in patients is the expansion of a trinucleotide repeat in the 5'UTR region of the gene (>200 repeats), which leads to the creation of the fragile site detectable on karyotype and cytogenetic analysis. At the molecular level, the expansion mutation leads to the absence of FMRP preventing the regulation of downstream mRNA targets encoding proteins

involved in synaptic structure and function. It remains today the second most frequent cause of ID after Down syndrome, and the first cause of monogenic ID with an assessed incidence of about 0.02%, *i.e.* at most 1% of all ID cases ((Coffee *et al.*, 2009)).

Gene	Approach	Year	Ref
ALG13	X-seq, WES	2009, 2012	(Tarpey *et al.*, 2009; Timal *et al.*, 2012)
SYP	X-seq,	2009	(Tarpey *et al.*, 2009)
HDAC6	L-can	2010	(Simon *et al.*, 2010)
CLIC2	X-seq	2012	(Takano *et al.*, 2012)
HDAC8	X-seq, Mol-fu	2012	(Deardorff *et al.*, 2012; Harakalova *et al.*, 2012)
KDM6A	CGH	2012	(Lederer *et al.*, 2012)
WDR45	WES	2012	(Haack *et al.*, 2012)
BCAP31	X-seq	2013	(Cacciagli *et al.*, 2013)
SLC35A2	Mol-Fu	2013	(Kodera *et al.*, 2013; Ng *et al.*, 2013)
ZC4H2	X-seq, L-can	2013	(Hirata *et al.*, 2013)

Table 15: Novel XLID genes, identified within the last two years (OMIM source).

Subsequently to cytogenetic methods, a variety of approaches have been developed to map additional XLID genes. Familial study with linkage analysis followed by screening of candidate gene in the linkage region, or translocation studies followed by breakpoint mapping and screening of the disrupted genes in ID cohorts remained the gold standard for twenty years, as illustrated in **Figure 28**. Conversely, the use of know molecular functions or metabolic pathways led to the discovery of a more modest number of genes. With the later development of NGS, X-exome and exome sequencing are now the high-throughput preferred methods, which altogether led to the identification of about a fifth of all XLID genes (**Figure 28**; **Table 15** (Lubs *et al.*, 2012)). Following large-scale studies of ID cohorts, a few X-linked genes now appear to be somewhat recurrently mutated such as *MECP2*, *CUL4B*, *SLC16A2*, *IL1RAPL1* or *UPF3B* in either syndromic or non-syndromic forms (e.g. *IL1RAPL1*) of ID. However, despite the exponential growth in the number of XLID genes being identified, *FMR1* remains by far the most recurrently mutated XLID gene.

A few XLID syndromes

After *FMR1*, the most frequent gene mutated in patients with XLID is *ARX*, assessed to account for about 9% of all X-linked families (Poirier *et al.*, 2006). The *Aristaless*-related homeobox gene belongs to the large family of homeodomain transcription factors implicated in general development, including the brain. Mutations in this gene have been reported in a wide spectrum of conditions, including West and Partington syndrome, lissencephaly, myoclonic epilepsy with ID, but also fully non-syndromic ID (Bienvenu *et al.*, 2002; Kitamura *et al.*, 2002;

Stromme *et al.*, 2002). Lissencephaly/hydranencephaly with abnormal genitalia and non-syndromic ID seem to be the most frequent ARX-associated phenotypes.

MECP2 is the well-known gene that was identified in 1999 as implicated in Rett syndrome (initially described in 1966), a disorder that affects mostly females (Amir *et al.*, 1999). Rett syndrome is a severe neurodevelopmental disorder, characterized by an early regression of motor (*e.g.* crawling, walking, seating,…) and cognitive skills (*e.g.* speech) that occurs after a period of normal development. One of the particular features reported in Rett patients is a stereotypic movement of the hands, known as 'hand-flapping' or 'hand-washing'. Patients also present with ataxia, seizures and breathing abnormalities. The episode of regression is usually accompanied by undernourishment engendering poor growth, weight gain and cranial enlargement resulting in short stature and microcephaly. However, the clinical features can be highly variable, ranging from typical Rett syndrome, to Angelman syndrome, and even only mild cognitive impairment (Imessaoudene *et al.*, 2001; Watson *et al.*, 2001). Even though *MECP2* mutations were long thought to be lethal in males (Hagberg *et al.*, 1983), males were later reported with a distinct clinical picture of severe epileptic encephalopathy. Interestingly, the nature of mutations reported in *MECP2* in patients is highly diverse, ranging from point mutations leading to missense, nonsense or frameshifts, but also to submicroscopic events such as duplications, deletions or rearrangements. The presence of low copy repeats in *MECP2* seems to make it highly conducive to complex rearrangements (Carvalho *et al.*, 2009). Most *MECP2* cases are sporadic since very severe and hardly compatible with normal fitness. The few familial cases that have been reported so far are due to X-inactivation of the X-chromosome carrying the *MECP2* mutation in females, or to germline mosaicism.

CUL4B is another gene that seems to play a preponderant contribution in XLID. It is also linked to a rather syndromic form of ID, characterized by short stature, hypogonadism, ataxia or balance disorders and specific dysmorphic features (micro or macrocephaly, hypertelorism, macroglossia, micrognathia and macrostomia; reviewed in (Cabezas *et al.*, 2000)). Despite its early identification in 1993, it was only mapped in 2007 during the course of the large-scale X-chromosome resequencing study in ID, identifying mutations in 5/250 patients supposing a relatively elevated recurrence (Wei and Sulik, 1993; Tarpey *et al.*, 2007; Tarpey *et al.*, 2009).

SLC16A2 is associated to the severe Allan-Herndon-Dudley syndrome (MIM# 300523), characterized by major neonatal hypotonia, delayed motor skill acquisition, dyskinetic movements, spastic paraplegia and severe to profound ID. Most patients harbor white matter abnormalities on MRI. The gene was mapped in 2004, looking for candidates with known molecular function coherent with the thyroid defects reported in patients while fitting with linkage analyses (Dumitrescu *et al.*, 2004).

3. Monogenic forms: autosomal dominant forms of ID

Because the hypothetic transmission of a mutation from affected parents to offspring was incompatible with the limited fitness associated with ID (and especially with the more severe forms), the hypothesis of genes implicated in autosomal dominant forms of ID remained neglected for long apart from several well-known syndromes (*e.g.* Pitt-Hopkins, Smith-Magenis syndromes, etc).

It started to be even more seriously considered with the apparent major contribution of *de novo* CNVs in ID and related neurodevelopmental disorders. Even if supposedly lower than the *de novo* CNV rate, the rate of *de novo* variations remains quite consequent and leads to 50-100 novel acquired variations in each newborn genome. Only a few affect protein-coding sequences (*i.e.* 0 to 2) without necessarily leading to phenotypic consequences. The first successful studies that confirmed this hypothesis especially in non-syndromic ID performed direct testing of candidate genes with relevant function (Hamdan *et al.*, 2009a; Hamdan *et al.*, 2009b; Hamdan *et al.*). Nonetheless, a hypothesis-driven approach was required and precluded unbiased discovery of unexpected genes. D*e novo* mutations could not be interrogated genome-wide since the comparison of parental genomes with the one of the offspring was not realistically feasible by Sanger sequencing. The arrival of NGS technologies allowed much insight into the *de novo* hypothesis, allowing an unbiased comparison of the entire genomes of both parents and offsprings simultaneously in trio studies. The first trio exome sequencing study was performed in ten families with ID, proposing *DEAF1* and *DYNC1H1* as novel ID genes but most of all strengthening the evidence of the contribution of *de novo* mutations and autosomal dominant inheritance in neurodevelopmental disorders (Vissers *et al.*, 2010). Two other studies followed in larger cohorts of patients (total of 151 trios) further supporting this hypothesis, and proposed a few genes to be more recurrently affected by *de novo* mutations in sporadic cases such as *SCN2A*, *STXBP1*, *SYNGAP1* and *TCF4* (de Ligt *et al.*, 2012; Rauch *et al.*, 2012). The preponderant contribution in ID of *de novo* mutations in outbred Western populations shall however not restrict the studies to sporadic cases since extensive germline mosaicism still may occur. Almost 150 genes have been so far implicated in autosomal dominant forms of ID with full penetrance.

Many of those genes uncovered by NGS studies were accredited by the simple report of a *de novo* mutation in a single ID proband, while the *de novo* nature of a variation itself cannot be sufficient to prove the pathogenicity of a variant and consequently the implication of the associated gene in cognitive disorders. Not all *de novo* variations are pathogenic. One has to be aware that some false positive genes may lie in such proposed genes (see (Piton *et al.*, 2013) for a similar reflection on XLID genes). Also, the current WES and WGS trio strategies only permit discovering genes implicated in autosomal dominant forms of ID with mutations of full penetrance, since variants that are transmitted from an unaffected parent are usually filtered out. This might be a major limitation that needs to be tackled in the future, since many mutations in genes associated to

autosomal dominant disorders show incomplete penetrance (*e.g. SLC2A1* mutations responsible for glut1-deficiency syndrome, which may be inherited from asymptomatic parents).

4. Monogenic forms: autosomal recessive forms of ID

Inborn errors of metabolism are mostly associated to autosomal recessive mode of inheritance. With an assessed prevalence of about 1% in European countries, such disorders may contribute to an even more substantial part of ARID in regions of higher consanguinity. In particular, cases of phenylketonuria, medium-chain acyl-CoA dehydrogenase deficiency, carnitine palmitoyltransferase 1 deficiency and homocystinuria were reported in Iranian familial cases of ID (Ropers, 2010).

Also, most familial cases of ID that are not X-linked are thought to follow an autosomal recessive mode of inheritance. However, only a few studies focused on autosomal recessive ID (ARID), expected to be rare in populations with highly mixed genetic background and low levels of consanguinity. Conversely, the high level of consanguinity in countries of Northern Africa or Middle East increases the chances of autosomal recessive disorders caused by homozygous mutations. Moreover, the generally large size of families in those populations makes them a model of choice for studying ARID. The study of genetics of ARID was lagging for long, and especially in non-syndromic ID: only eight non-sydromic ARID genes had been identified laboriously in 2011 compared to over 100 XLID genes (Lubs *et al.*, 2012; Musante and Ropers, 2014). A few teams had been addressing the topic, with most studies consisting in homozygosity mapping combined with linkage analysis in large consanguineous families (Molinari *et al.*, 2002; Higgins *et al.*, 2004; Basel-Vanagaite *et al.*, 2006). Fostered by the emergence of NGS technologies, one of the largest studies included the high throughput targeted sequencing of coding exons from regions of homozygosity and linkage in 136 consanguineous families with probable ARID, confirming over 20 previously reported genes (among which some Joubert and BBS genes) and proposing an additional 50 candidate genes (Najmabadi *et al.*, 2011). Subsequently to the application of NGS approaches, 32 novel candidate genes have been identified, supposedly implicated in non-syndromic ARID (reviewed in (Musante and Ropers, 2014)). Nonetheless, similarly to the caution expressed regarding *de novo* mutations that are not all necessarily pathogenic, not all homozygous missense variants are neither.

5. Oligogenism, or the 'multiple hits' model

Despite the tremendous insights into the genetics of ID fostered by the systematic genome-wide studies made possible through NGS technologies, most cases remain without solved molecular etiology. One hypothesis is the 'polygenic' or 'oligogenic' model (that may or may not involve exogenous environmental factors), which has found valuable support in recent large-scale CNVs analyses. Indeed, some CNVs of demonstrated reduced penetrance such as the 16p12.1 locus show a much higher penetrance when accompanied by other large CNVs. Individuals with most severe neurodevelopmental phenotypes generally harbor a burden of large and small CNVs (Girirajan *et al.*, 2011).

This oligogenic hypothesis has been even more envisaged in other neurodevelopmental disorders such as ASDs, showing that probands harbor a burden of heterozygous variants (*de novo* and inherited) in autism susceptibility genes when compared to controls (Schaaf *et al.*, 2011). Another study reports ASDs probands carrying intragenic *de novo* deletions in *SHANK2* along with inherited CNVs at the 15q11-q13 locus affecting either *CHRNA7* or *CYFIP1*, supporting the 'multiple hit' hypothesis (Leblond *et al.*, 2012). Interestingly, in the recent large-scale exome sequencing study of ID families, no double or multiple hits was mentioned in any proband. This can presumably be explained because the scenario was simply not evaluated since the authors mostly focused on the contribution of *de novo* variants (de Ligt *et al.*, 2012; Rauch *et al.*, 2012).

6. Emerging interconnected gene networks

Among the several hundred genes that have been implicated in all forms of ID, many teams focused on understanding the relationship between those and establishing censed molecular networks, apart from the metabolic ones connecting most genes implicated in IEMs with associated ID. Also, a proportion of ID genes associated with cerebellar malformations have a preponderant in neurogenesis or neuronal migration.

In order to identify novel ID gene networks, several parameters are integrated such as common gene regulation pathways, physical protein interactions, expression patterns, transcription factor binding sites, etc. The main molecular networks that appear to be emerging from those gene networks are not unexpectedly 'nervous system development', and 'synaptic transmission' but also β-catenin/chromatin remodeling (O'Roak *et al.*, 2012b; Cristino *et al.*, 2014). This is coherent with the distinction between syndromic ID, where genetic mutations would perturb any early step of brain development (*e.g.* neural tube formation, neuronal migration and patterning) and would lead to gross brain abnormalities, with non-syndromic ID (*i.e.* without structural brain or other organ defects) in which the subjacent genetic mechanisms are more likely to be associated to the later steps of synaptogenesis or neuronal wiring (von Bohlen Und Halbach, 2010). Besides the emerging

gene network evidences, among the several hundred genes that have been implicated in monogenic forms of ID, a consequent proportion had already been shown to play a role in pre or post-synaptic formation or transmission (Guilmatre *et al.*, 2009; Perche *et al.*, 2010; Hamdan *et al.*, 2011; Melom and Littleton, 2011; Verpelli *et al.*, 2013).

Another group of genes lately implicated in ID have a role in major signaling cascades, such as the Ras-MAPK-ERK or the PI3K-AKT-mTOR pathways implicated in the global regulation of translation and transcription (*e.g. PTEN, HRAS, KRAS*, etc). Similarly, mutations in a novel class of genes with role in the regulation of transcription, chromatin remodeling, etc have been reported in patients with ID (*e.g. HDAC4, MED23, CHD7, EHMT1*, etc ; (Ronan *et al.*, 2013)). Interestingly, a number of such regulatory components have a role in either neurodevelopmental disorders or tumor developments and cancer (*e.g. PTEN, KRAS, CREBBP*).

III- Comorbidity: ID and associated clinical signs

Intellectual developmental disorders are not only highly heterogeneous at the genetic level, but also at the clinical level. One first needs to discriminate between syndromic ID, where patients can be generally classified according to specific physical, metabolic, behavioral features and involvement of other organs, with non-syndromic ID where no additional dysmorphic, metabolic or neuromuscular features are consistent between affected family members. Individuals with NS-ID are usually hardly clinically distinguishable (except for the severity of the cognitive impairment) and are considered as a homogeneous clinical entity.

In syndromic ID, a plethora of additional clinical features can be found in patients, with some comorbid traits being more frequent than others (**Table 16)**. Correct brain development depending on very tight orchestration of numerous crucial cellular and molecular events, any genetic or environmental insult occurring during neurulation, CNS patterning, embryonic/adult neurogenesis, myelination or synapse pruning can compromise the elaboration of normal cognitive function or brain structure. The consequences can be highly specific and localized (*e.g.* microcephaly, lissencephaly), or variable and diffuse (*e.g.* psychiatric, neurologic manifestations).

Co-morbid traits	Frequency
Neurologic features	
Epilepsy	22%
Balance disorders	20%
Neuropsychiatric features	
Autism	24-30%
Anxiety disorders	17%
ADHD	10%
Malformations	
Cerebellar	20-30%
Musculo-skeletal	4-8%
Cardiac	4-6%
Urogenital	2-3%
Gastrointestinal	2-4%

Table 16 (adapted from (Petterson *et al.*, 2007; Oeseburg *et al.*, 2011)): Frequency of several co-morbid features in patients with ID.

Comorbidity of structural defects involving other organs such as malformation syndromes will not be detailed here, since I will mostly focus on comorbid traits involving the central or peripheral nervous system.

1. Neuromuscular and neurologic manifestations

Gait and balance disorders

Gait and balance defects are frequently reported in patients with ID. Among them, ataxia refers to a loss of balance and movement coordination that are not attributable for muscle weakness or involuntary muscle activity. It usually originates from damage of the cerebellum or its projections leading to its dysfunction, and may manifest in multiple parts of the body (limb, trunk, etc). They are often accompanied by abnormal eye movements, nystagmus, dysarthria, and dysdiadochokinesia. Ataxias can be either hereditary (*i.e.*. of genetic origins) or acquired subsequently to infections (*e.g.* varicella) or brain injuries. Acquired ataxia may be only transient, while hereditary conditions are usually lifelong. The overall prevalence of childhood ataxias (combining both acquired and genetic forms) in Europe is assessed to be 0.026% (Musselman *et al.*, 2014). It is however more frequent in regions of higher degree of consanguinity since most genetic forms are associated to an autosomal recessive mode of inheritance (*e.g.* Friedreich's Ataxia, ARSAC, Joubert syndrome, etc). The most frequent forms of hereditary ataxia are metabolic ataxia and Freidreich's ataxia (Musselman *et al.*, 2014). Almost 60 different genes are implicated in forms of ataxia (Laing, 2012).

Some ID patients also present with specific muscular tone defects, such as hypotonia, dystonia, hypertonia, spasticity, or paraplegias. Such manifestations involve mainly the peripheral nervous system and muscles. Cognitive impairment and developmental delay are frequent features in disorders such as Duchenne or Becker muscular dystrophy, limb girdle muscular dystrophy or Allan-Herndon-Dudley syndrome.

Epilepsy

Epilepsy is a highly heterogeneous group of neurological disorders characterized by impaired electric brain activity. It is characterized by recurrent unpredictable seizures triggered by the rhythmic firing of groups of neurons. Current hypothesis is that the epileptic foci (regions of the cortex with abnormal activity) locally trigger the seizure that is then echoed to any synaptically connected region. Epileptic disorders are the most common chronic neurological disorder, with prevalence in the population of about 1-3% (Linehan *et al.*, 2010; Quet *et al.*, 2011). Epilepsy accompanies ID in 10-15% in the mild to moderate cases, 20-30% in the severe to profound cases, and in as much as 50% of XLID cases (Chapman *et al.*, 2011; Stevenson *et al.*, 2012). They can originate from a large variety of factors such as trauma, stroke, tumors, congenital vascular or cortical malformations, etc but in the remainder genetic factors are thought to play a major role. Most GWAS studies failed to identify major risk locus, and contrary to ID only a few CNVs have been shown to be responsible for the disease in patients (Kasperaviciute *et al.*, 2010; Consortium *et al.*, 2012; Heinzen *et al.*, 2012). Nonetheless, during the last few years, a few genes appeared to

contribute significantly to fully penetrant Mendelian forms of epilepsy in patients (such as *CDKL5, KCNQ2*, *KCNQ3, SCN1A, SCN2A, SCN8A, SLC2A1*, *STXBP1*…), revealed in several independent studies (Claes *et al.*, 2001; Kalscheuer *et al.*, 2003; Kamiya *et al.*, 2004; Saitsu *et al.*, 2008; Consortium *et al.*, 2012; Veeramah *et al.*, 2012). Some are associated to ion channel subunits, but not exclusively. Mutations were also recently reported in patients in genes usually associated to megalencephaly (Lee *et al.*, 2012), lissencephaly (*DCX*, (Gleeson *et al.*, 1998)) or periventricular heterotopia (*FLNA*, (Fox *et al.*, 1998)).

2. Cerebellar and structural brain abnormalities

Cephalic disorders mostly result from altered neuronal migration or proliferation.

Microcephaly

Microcephaly is a heterogeneous group of disorders characterized by a reduced occipitofrontal circumference (OFC) that is due either to general abnormal proliferation during embryonic brain development (congenital microcephaly) or to defective neuronal maturation during postnatal brain development (acquired microcephaly). Some are accompanied with abnormal convolution patterns (e.g. microcephaly with simplified gyral pattern or MSGP, agyria, microlissencephaly,) whilst others are not. Most genes associated to this condition are implicated in regulating cell proliferation (*ASPM, MCPH1, CDK5RAP2, CENPJ*, reviewed in (Francis *et al.*, 2006)). On the other extreme, macrocephaly (enlarged OFC) can be associated to megalencephaly that is mainly caused by activating mutations in genes of the PI3K/AKT pathway (Riviere *et al.*, 2012), which plays a preponderant role in brain growth and development.

Polymicrogyria

Polymicrogyria is characterized by a multitude of microscopic gyri at the brain surface, either diffuse or localized (focal), making it hard to detect by MRI. A thicker cortex, a smoother brain surface and enlarged ventricles are also often present near the affected region. All types are associated with disruption of the multilayered neocortex, most probably due to altered cell proliferation. Mutations in genes encoding for α or β-tubulin (*TUBA8* and *TUBB2B*), *GPR56* and rotatin (*RTT*) have been reported in patients with such condition resulting in microtubule assembly defects or aberrant ciliary function (Piao *et al.*, 2004; Robin *et al.*, 2006; Abdollahi *et al.*, 2009; Jaglin *et al.*, 2009; Kheradmand Kia *et al.*, 2012).

Schizencephaly

Schizencephaly consist in unilateral or bilateral clefts communicating with the ventricles, surrounded by grey matter and lined with polymicrogyric cortex. Some parts of the brain can be absent and filled with cerebrospinal fluid. Such condition supposes a very early defect of neuronal migration. Mutations in *SIX3*, *SHH* and *EMX2* have been reported to date in patients (Brunelli *et al.*, 1996; Roessler *et al.*, 1996; Schell-Apacik *et al.*, 2009; Hehr *et al.*, 2010). It shares genetic overlap with holoprosencephaly, a disorder where the prosencephalon fails to compartmentalize in two hemispheres resulting in abnormal brain structure and impaired facial development.

Lissencephaly

Lissencephaly is a heterogeneous condition defined as a smoothened brain surface, caused by defective neuronal migration leading to absence of cortical convolutions. It is also associated to thicker cortex, abnormal cortical lamination and diffuse neuronal heterotopia (accumulation of neurons that fail to migrate properly). Type I lissencephaly has to be discriminated from Type II or cobblestone lissencephaly, which most probably results from mutations in glycosyl transferase enzymes and not from migrational defects. Mutations in the genes *ARX*, *DCX*, *LAMB1*, *NDE1*, *PAFAH1B1*, *RELN*, or *TUBA1A* have been reported in patients with lissencephaly Type I (Lo Nigro *et al.*, 1997; Gleeson *et al.*, 1998; Hong *et al.*, 2000; Kitamura *et al.*, 2002; Keays *et al.*, 2007; Bakircioglu *et al.*, 2011; Radmanesh *et al.*, 2013).

Heterotopia

Gray matter heterotopias regroup heterogeneous conditions characterized by mislocalization of clumps of neurons, supposedly caused by an arrested migration of such cells that fail to cross the cortical layers and remain close to their site of generation. Such clumps can localize close to the ventricles (periventricular or subependymal), or within the white matter in a focal (focal subcortical) or diffuse (band form) pattern. Loss-of-function mutations in *FLNA* have been reported mainly in females with periventricular heterotopia, since there are generally embryonic-lethal in males (Fox *et al.*, 1998). Other proposed genes implicated in periventricular heterotopia include *ARFGEF2*, responsible for periventricular heterotopia with associated microcephaly (Sheen *et al.*, 2004), *DCHS1, FAT4, DCX, ERMARD* and *PAFAH1B1* (des Portes *et al.*, 1998; Pilz *et al.*, 1999; Cappello *et al.*; Conti *et al.*). Genetically, this condition appears to be a continuum of lissencephaly (on the milder side) since mutations in several genes were reported in patients with one or the other disorder (*PAFAH1B1, DCX*). For instance, males with *DCX* mutations present with lissencephaly while females mainly present with subcortical laminar heterotopia (Pinard *et al.*, 1994; Gleeson *et al.*, 1998).

Encephalocele

Encephalocele is proposed to originate from post-neurulation defects (reviewed in (Copp *et al.*)). It is characterized by a skull protuberance called meningeal sac, which may or may not contain brain tissue. It is compatible with life, and surgery is often proposed to remove the sac and close the lesion. Some ciliopathies like Meckel syndrome (see **Results - Part I**) may include encephalocele manifestation hence any Meckel syndrome gene can lead to such condition. A few other genes have been described as resulting in encephalocele such as *COL18A1* in Knobloch syndrome type 1, the homeobox gene *ALX3*, the lissencephaly gene *LAMB1,* or the ciliopathy gene *CSPP1* (Sertie *et al.*, 2000; Twigg *et al.*, 2009; Radmanesh *et al.*, 2013; Shaheen *et al.*, 2014; Tuz *et al.*, 2014).

3. Behavioral disorders

Autism Spectrum disorders

Autism spectrum disorders are defined in DSM-V by *'persistent deficits in social communication and social interaction and restricted, repetitive patterns of behavior, interests, or activities, which arise in the early developmental period'*. It is a generic term regrouping highly heterogeneous disorders, including autistic disorders, high-functioning autism, Kanner's autism, Asperger syndrome, pervasive developmental disorders not otherwise specified (PDD-NOS) and childhood disintegrative disorders. Their prevalence in the population is of about 1% (Fombonne, 2009), and is more frequent in males than females (4 to 1 ratio; (Newschaffer *et al.*, 2007)). ASDs' prevalence appears to be gradually increasing but it remains unclear whether it is because of our increasing awareness of the condition helping diagnosis, the novel DSM-V diagnostic criteria that regroup a plethora of conditions under the term of ASDs, or because of exogenous factors that are indeed responsible for an increasing number of cases.

Its origins are largely multifactorial, implicating environmental and genetic factors, but tremendous advances have been made in the past decade concerning the identification of monogenic genetic origins. Five major studies first reported an increased burden of *de novo* CNVs in ASDs children as compared to controls (Sebat *et al.*, 2007; Marshall *et al.*, 2008; Pinto *et al.*, 2010; Levy *et al.*, 2011; Sanders *et al.*, 2011). Later exome sequencing studies performed on a cohort of 20 trios (mostly sporadic patients) showed that pathogenic *de novo* mutations played a preponderant role in causing ASDs (O'Roak *et al.*). Four major studies published a year later confirmed the findings, subsequently to trio-exome sequencing in nearly a thousand of ASDs trios (965 in total; (Iossifov *et al.*, 2012; Neale *et al.*, 2012; O'Roak *et al.*, 2012b; Sanders *et al.*, 2012)). Affected genes were mostly brain-expressed genes and showed a high degree of connectivity. Most *de novo* mutations were also proven to have occurred in the father germ line and were correlated

with an increased paternal age (Iossifov *et al.*, 2012; Michaelson *et al.*, 2012; Neale *et al.*, 2012; O'Roak *et al.*, 2012b; Sanders *et al.*, 2012). As opposed to ID studies, among a thousand of ASDs patients screened by exome/genome sequencing (Michaelson *et al.*, 2012), only one or two genes were hit more than once (*SCN2A*, *KIRREL3* and *GPR98*) supposing a much higher degree of genetic heterogeneity. Exome sequencing combined with homozygosity mapping was also performed in consanguineous ASDs families to unravel autosomal recessive cause of ASDs, but with more modest success (Chahrour *et al.*, 2012). Altogether, CNVs and single gene disorders account for 15% of solved clinical genetic testing in ASDs cases (Shen *et al.*, 2010; Schaaf and Zoghbi, 2011). Among monogenic causes of ASDs, several hundred genes have now been described.

Attention deficit and hyperactivity disorders (ADHD)

ADHD is characterized by persistent inattention and/or hyperactivity that interfere with proper child functioning and development. Its prevalence is assessed to be of 5% in children with an increased frequency in males (2:1 ratio). Genetic causes of ADHD are substantial: no single gene variation could be clearly associated with full penetrance of the ADHD phenotype in children. Genetics factors are thought to be more risk factors, rather than sufficient causal factors. Many ID probands harbor ADHD.

4. ID associated with metabolic disorders

Inborn errors of metabolism (IEM) are a phenotypically and genetically heterogeneous group of disorders leading to malfunctioning metabolism and/or accumulation of toxic metabolites. IEMs are assessed to account for up to 14% of all ID cases, and over 80 IEMs have been reported for which ID is a major feature (Papavasiliou *et al.*, 2000; Engbers *et al.*, 2008; van Karnebeek and Stockler, 2012). A number of comorbid features among all the above-mentioned ones are found in IEMs, such as epilepsy and other neurologic features (spasticity, dyskinesia, or ataxias) and/or behavioral and psychiatric problems in nearly 100%, but also involvement of other organs in almost 70% (van Karnebeek and Stockler). Neurologic and neuropsychiatric features appear to originate either from defective late neurodevelopmental processes in progressive conditions (*e.g.* myelination, in metachromatic leukodystrophy or adrenoleukodystrophy) or via disruption of the synaptic and neurotransmitter systems in acute or episodic disorders (*e.g.* phenylketonuria, disorders of homocysteine metabolism). Interestingly, an increasing number of illness-modifying treatments have been developed for many of them, so it is of particular importance to detect them precociously. In a majority, the diagnostic of IEMs can be made on simple blood or urine metabolic screening tests (see **Figure 29**); some of them are currently implemented as newborn screening tests in specific countries.

The example of phenylketonuria

Phenylketonuria (PKU) is an autosomal recessive IEM resulting from deficiency of phenylalanine hydroxylase (PAH) leading to a neurotoxic accumulation of phenylalanine. If untreated, clinical manifestations include severe ID, light pigmentations, gait abnormalities, autistic features, and epilepsy (Brumm *et al.*, 2010). A phenylalanine-free diet was proposed in 1954 and proven to be highly effective, preventing the most severe manifestations when given in the presymptomatic phase (although psychiatric and behavioral problems may persist). With an assessed incidence of 1 in 15,000 newborns, PKU became a part of newborn screening tests in most European and Northern American (but only a few Asian countries) since the 1980s, and the phenylalanine-free diet has become a standard of care (Mak *et al.*, 2013). The corresponding *PAH* gene was only mapped in 1983 (Woo *et al.*, 1983).

A similar disorder primarily characterized by congenital heart disease can also be transmitted to the offspring by the mother during pregnancy, which is known as the 'maternal PKU syndrome' (Huntley and Stevenson, 1969).

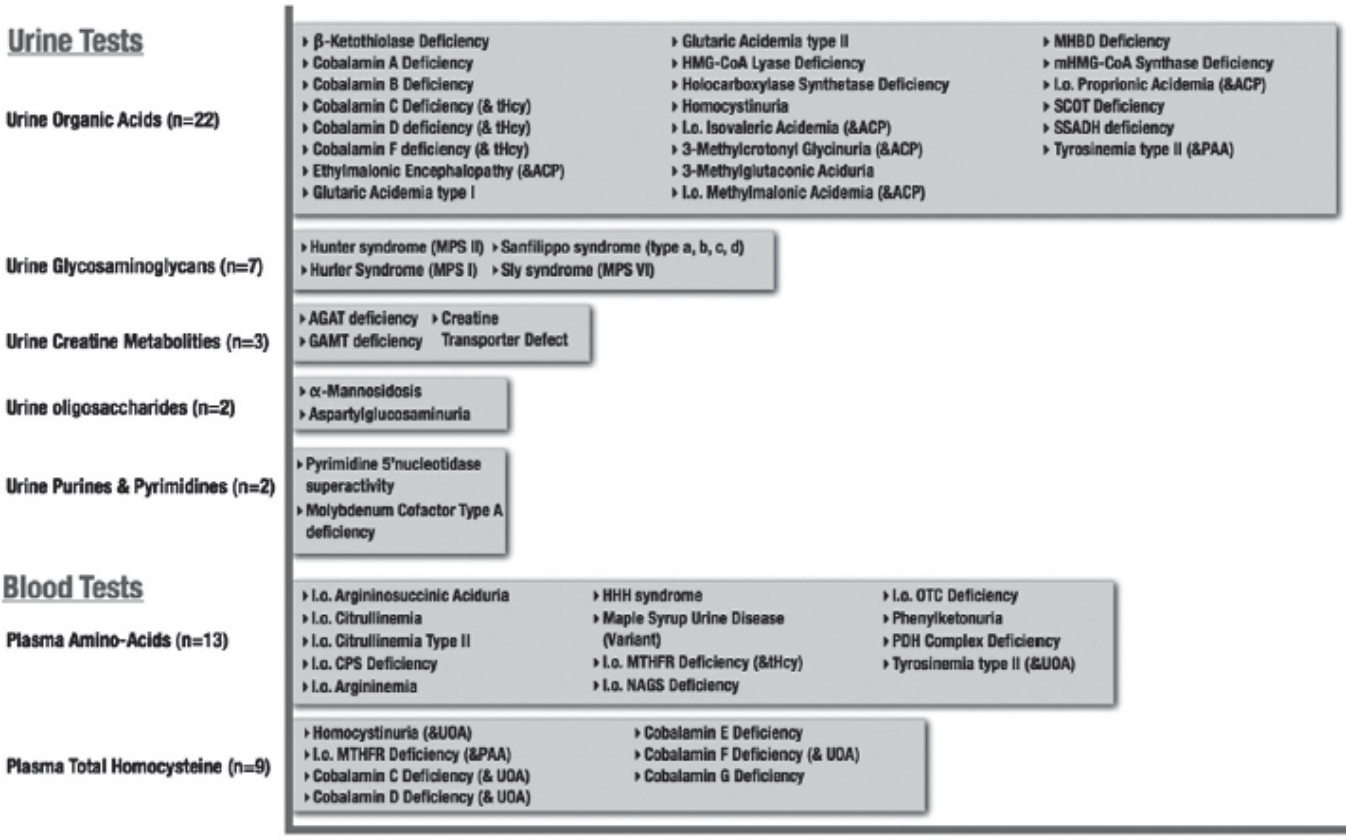


Figure 29 (from (van Karnebeek and Stockler, 2012)): Summary of the 50 IEMs that can be diagnosed using standard metabolic screening tests.

Length of bars is correlated with the number of IEMs a single test can identify. PAA: plasma aminoacids; tHcy: total homocystein; ACP: plasma cylcamitine profile; UOA: urine organic acids.

IV. Diagnosis of ID in France

Except for a few advices by expert clinicians (des Portes *et al.*, 2002), there are currently no strict guidelines regarding the diagnosis of patients with ID in France.

Classical, diagnosis of ID starts with extensive clinical investigations to better characterize the disorder of the patient, in search of additional comorbid features (*e.g.* hypotonia, epilepsy, microcephaly, etc), dysmorphic traits, etc. Several imaging, elecrophysiological, and biochemical/metabolic investigations (*e.g.* chromatography of organic and aminoacids, lactate, pyruvate and creatine levels in urine, …) are often performed subsequently, in order to detect specific cerebellar or metabolic defects. Nonetheless, metabolic investigations are not performed in a consistent and systematic manner, though we saw that for instance some simple metabolic tests can screen as many as 20 different IMEs. Some IMEs are systematically screened at birth as part as newborn screening tests such as PKU. Investigations regarding the history of the disorder in the patient's family are also primordial, to highlight putative genetic origins.

For an ID of suspected genetic etiology, most diagnostic laboratories generally follow a similar decisional tree. All patients are usually tested for structural and numerical chromosomal abnormalities and submicroscopic CNVs by karyotype and high-resolution array-CGH analysis. Most patients are also tested in parallel for the fragile-X *FMR1* expansion mutation. Altogether, array-CGH and fragile-X tests lead to positive diagnostic yields of 10-15% (with higher rates in highly syndromic patients and patients with moderate to profound ID) and 1% respectively.

Other specific genetic tests may be proposed to patients, such as FISH or telomeric rearrangement testing, and Sanger sequencing for one or several genes that would fit with a syndrome evoked by the clinical features of the patient (*e.g.: MECP2, UBE3A* and *CDKL5* for a patient with a phenotype Angelman syndrome-like; *DMPK* and *ZNF9* expansion repeats in a patient with considerable hypotonia).

Altogether, with this diagnostic pipeline over 60% of all patients with ID remain with no molecular diagnosis.

V- Results

Paper I: XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing.

Piton A*, **Redin C***, Mandel JL. Am J Hum Genet. 2013 Aug.

Aim of the study: The genetic heterogeneity in ID is such that it is difficult to find independent families carrying mutations in the same genes. Also, most mutations reported in patients are private mutations, so the challenge is to discriminate rare innocuous variants from true pathogenic mutations. Additional confirmation or replication studies are often needed - but generally missing – to further support the implication of a gene in ID. Some proposed mutations or even genes might be false positive calls, with no association to cognitive impairment. Our goal was to systematically reassess the evidences supporting the implication of each gene in ID, essentially based on the frequency of proposed ID mutations in a large general population (6,500 individuals). For this work we focused exclusively on the hundred published XLID genes, but the same approach should be performed on autosomal recessive and autosomal dominant ID genes, accordingly adapting the exclusion threshold for the frequency of a known mutation in the general population.

Contribution: Major. Performed the bioinformatic analyses (retrieving the mutations and variants from OMIM and EVS, and intersect the data), mined the literature for each XLID gene in delicate position, wrote the paper along with Dr Piton.

REPORT

XLID-Causing Mutations and Associated Genes Challenged in Light of Data From Large-Scale Human Exome Sequencing

Amélie Piton,[1,2,4,*] Claire Redin,[1,2,4] and Jean-Louis Mandel[1,2,3,*]

Because of the unbalanced sex ratio (1.3–1.4 to 1) observed in intellectual disability (ID) and the identification of large ID-affected families showing X-linked segregation, much attention has been focused on the genetics of X-linked ID (XLID). Mutations causing monogenic XLID have now been reported in over 100 genes, most of which are commonly included in XLID diagnostic gene panels. Nonetheless, the boundary between true mutations and rare non-disease-causing variants often remains elusive. The sequencing of a large number of control X chromosomes, required for avoiding false-positive results, was not systematically possible in the past. Such information is now available thanks to large-scale sequencing projects such as the National Heart, Lung, and Blood (NHLBI) Exome Sequencing Project, which provides variation information on 10,563 X chromosomes from the general population. We used this NHLBI cohort to systematically reassess the implication of 106 genes proposed to be involved in monogenic forms of XLID. We particularly question the implication in XLID of ten of them (*AGTR2*, *MAGT1, ZNF674, SRPX2, ATP6AP2, ARHGEF6, NXF5, ZCCHC12*, *ZNF41*, and *ZNF81*), in which truncating variants or previously published mutations are observed at a relatively high frequency within this cohort. We also highlight 15 other genes (*CCDC22*, *CLIC2*, *CNKSR2*, *FRMPD4*, *HCFC1*, *IGBP1*, *KIAA2022*, *KLF8*, *MAOA, NAA10*, *NLGN3*, *RPL10, SHROOM4*, *ZDHHC15*, and *ZNF261*) for which replication studies are warranted. We propose that similar reassessment of reported mutations (and genes) with the use of data from large-scale human exome sequencing would be relevant for a wide range of other genetic diseases.

Introduction

Intellectual disability (ID, formerly called mental retardation) is a developmental brain disorder commonly defined by an IQ below 70 and limitations in both intellectual functioning and adaptive behavior. ID can originate from environmental causes and/or genetic anomalies, and its incidence in children is estimated to be of 1%–2%.[1,2] As a result of an excess of males affected by ID (the male-to-female ratio is 1.3–1.4 to 1) and the identification of many families presenting with a clear X-linked segregation, much attention has been focused for the last 20 years on genes located on the X chromosome and thus responsible for X-linked ID (XLID, previously known as XLMR) when mutated.[3,4] One of the first genes identified as involved in XLID is *FMR1* (MIM 309550), a target of the unstable expansion mutation responsible for fragile X syndrome (MIM 300624); accounting for about 1%–2% of all ID cases, this mutation still remains the most common cause of XLID.[5,6] Since then, the number of genes involved in XLID when mutated has grown exponentially,[3,7,8] from only 11 in 1992 to 43 in 2002 and over 100 genes now identified thank to the efforts of various teams.[4,9,10] Half of the known genes carrying mutations responsible for XLID appear to be associated with nonsyndromic or paucisyndromic forms; the other half are associated with more syndromic forms (i.e., ID associated with defined clinical or metabolic manifestations), which facilitates the identification of causative mutations in the same gene because unrelated probands with comparable phenotypes can be more easily matched. However, the presence of "milder" mutations (in *RPS6KA3* [*RSK2*, MIM 300075] or *ARX* [MIM 300382], for instance) and/or incomplete penetrance of specific clinical signs in some individuals carrying mutations in genes associated with syndromic ID can blur the distinction between syndromic and nonsyndromic ID.[11]

Various approaches have been developed for the identification of genes and associated causative mutations responsible for XLID (see Lubs et al.[4] for a review): (1) positional cloning based on chromosomal rearrangements or copy-number variants (CNVs) affecting the X chromosome, (2) screening of genes located in candidate intervals identified via linkage analysis in large XLID-affected families, (3) direct sequencing of candidate genes with a function or expression pattern that suggests a role in cognition or that fits with metabolic or clinical observations in affected subjects, and (4) high-throughput sequencing allowing screening of mutations in all protein-coding regions of the genome or only in the X chromosome (exome versus X exome).[10,12–14]

The validation of potentially damaging mutations in a gene newly associated with XLID requires functional and/or genetic analyses, especially when the identification is based on reporting mutations in very few families or simplex cases. Functional studies are uneven in pertinence and strength. They can include direct assessment of the mutational impact at any of the protein, cellular, or

[1]Department of Translational Medicine and Neurogenetics, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique Unité Mixte de Recherche 7104, Institut National de la Santé et de la Recherche Médicale Unité 964, University of Strasbourg, 67404 Illkirch Cedex, France; [2]Chaire de Génétique Humaine, Collège de France, 75231 Paris Cedex 05, France; [3]Laboratoire de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, 67091 Strasbourg Cedex, France
[4]These authors contributed equally to this work
*Correspondence: piton@igbmc.fr (A.P.), jlmandel@igbmc.fr (J.-L.M.)
http://dx.doi.org/10.1016/j.ajhg.2013.06.013.

organism levels or functional connection to the disease (e.g., involvement in synaptic organization or plasticity). However, functional tests showing an effect of a candidate mutation at the protein or cellular level do not necessarily imply that this effect is inherently responsible for the disease. Indeed, some false positives can lie in mutations that have been "functionally" validated (see, for instance, *SRPX2* [MIM 300642] below).

Genetic validation of a mutation usually includes cosegregation analysis in the proband's family and testing for the absence of the mutation in a population of control chromosomes. Cosegregation analyses that validate mutations in simplex cases often lack statistical power and can lead to misleading conclusions, even in families with several affected individuals. Some noncausative variants will segregate by chance with the disease only because they lie in the linked candidate region, which often contains many genes. Recently, the identification of genes implicated in ID or autism spectrum disorders (ASDs) has extensively relied on the detection of de novo events, but it is clear now that such events can also be detected in unaffected siblings and hence that the de novo criterion is not sufficient to imply pathogenicity. Moreover, for X-linked genes, because de novo mutations occur more often in the male germline, probands' mothers are often silent carriers of a transmitted pathogenic mutation that might have occurred de novo in preceding generations. Investigating the maternal grandparents is thus very useful and informative but is often not possible. Lastly, because of the cost-efficiency limitations of Sanger sequencing, the number of control individuals sequenced is often limited to a few hundred individuals, which is therefore insufficient to allow the detection of rare innocuous variants. Some false-positive conclusions about the involvement of some genes in XLID might have arisen notably by the overinterpretation of the pathogenicity of missense variants, and this might concern genes proposed for mutation screening in XLID diagnostic panels.

A new powerful resource, the National Heart, Lung, and Blood Institute (NHLBI) data set of over 6,500 sequenced exomes (available on the Exome Sequencing Project Exome Variant Server [EVS]) can now be used for ascertaining the frequency of potential mutations in large cohorts of adults initially selected for cardiac, lung, or metabolic phenotypes but a priori not enriched with neurological or cognitive defects. This very large project originally aimed to identify genes underlying complex heart, lung, and metabolic disorders and provides detailed exome-variation information on unrelated African American and European American individuals (2,443 males and 4,060 females), amounting to a total of 10,563 X chromosomes.[15] Because participants had to provide written informed consent,[16] one can expect that these cohorts do not contain individuals with moderate or severe ID and that even the mild form is presumably underrepresented as compared to its incidence in the general population. This cohort can hence be considered as a "general population" for the analysis of rare variations in genes with a proposed implication in ID when mutated.

List of 106 Genes with Reported Mutations Involved in XLID

We established a list of 106 genes proposed to be associated with XLID by compiling data from reviews, lists of genes proposed for XLID diagnosis, and recent primary publications (Figure 1). Two of these 106 genes (*MTM1* [MIM 300415] and *GJB1* [MIM 304040]) appear to be misassociated with ID, given that very few cases were reported with both ID and centronuclear myopathy (MIM 310400) or Charcot-Marie Tooth neuropathy (MIM 302800) and that they carry deletions encompassing other genes in addition to *MTM1* or *GJB1*.[17,18] Both genes should thus be removed from XLID lists. For the remaining 104 genes, we systematically screened the NHLBI EVS data. We assumed that out of the ~2,500 males of the NHLBI population, there would be fewer than 50 males (2%) potentially affected by ID. Consequently, a previously published XLID mutation reported in two NHLBI males would imply that this mutation is responsible for 4% (2/50) of all ID cases and is therefore more common than the fragile X expansion mutation, which is very doubtful. Such a variant should therefore be considered a nonpathogenic (or low penetrance) rare single-nucleotide variant (SNV) with respect to monogenic forms of ID. The same logic can be applied if truncating variants (nonsense and splice variants) are identified in more than one male from the NHLBI population in a gene associated with ID; this would confer it a major role in ID, which is suspicious. This led us to further assess the association between some genes and ID when (1) truncating variants are observed in males within the NHLBI cohort (Table S1, available online), (2) previously described ID mutations are detected in this population at a frequency inconsistent with a causative effect on cognitive impairment (Tables S2 and S3), and (3) the implication of this gene in XLID relies on a single piece of evidence. For all such genes, we discuss the current evidence of their association with monogenic forms of ID on the basis of both the literature and our EVS data analysis (Supplemental Data and Table 1). All other genes were considered to be convincingly implicated in ID.

Presence in EVS of Truncating Variants Affecting Genes Involved in XLID

We first analyzed the NHLBI cohort data set to evaluate the presence, notably in males, of truncating SNVs in genes with a proposed implication in ID (Table S1). We did not include in our analysis the small indels that have been recently added to EVS (v.0.0.15, October 31, 2012) because the frequency of some of them indicates that they are presumably false-positive observations, either due to artifacts inherent from the sequencing technology or to difficulties in reliably annotating small indels (Table S4).[70] For instance, the frameshift c.1777_1780del (p.Leu593Phefs*7) has been reported in *NLGN4X* (MIM 300427) in 20

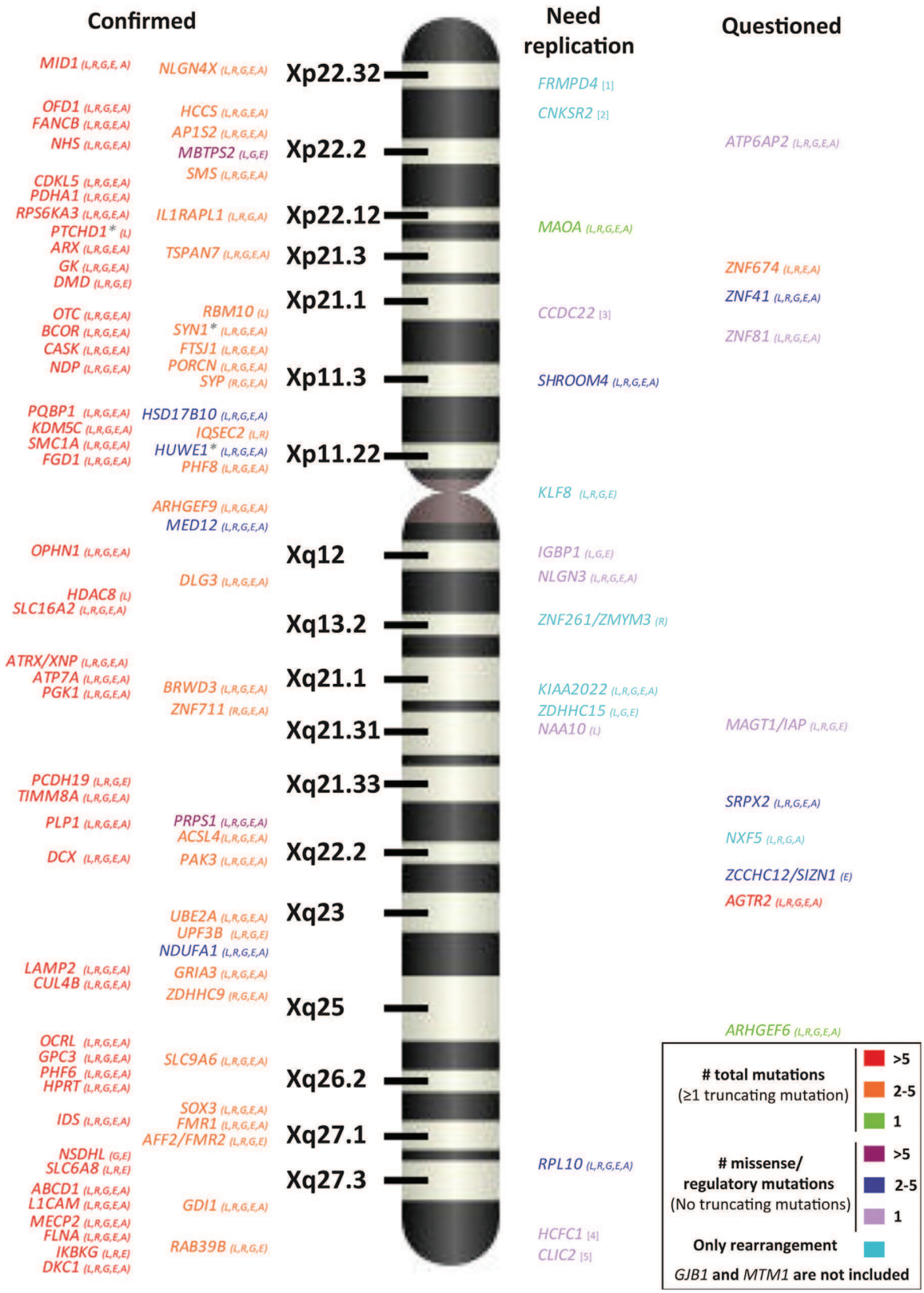


Figure 1. Representation along the X Chromosome of the 106 Genes in which Mutations Have Been Reported in XLID and Classification According to Both the Type and Number of Mutations Reported in OMIM

Genes for which involvement in XLID is convincing are listed on the left side of the chromosome, and genes for which implication in ID either is questionable or needs to be replicated are listed on the right side of the chromosome. *GJB1* and *MTM1* are linked to diseases with no real ID association and are thus not included in the figure. The color code was designed according to the total number of mutations reported in OMIM. If at least one truncating mutation (splice mutations, nonsense mutations, frameshifts, and large deletions that

(legend continued on next page)

homozygous females but only in a single heterozygous one, which departs from Hardy-Weinberg equilibrium. A total of 21 truncating SNVs are present in EVS and are distributed among 15 of the 104 genes in the XLID list (Table S1). Truncating variants found at the heterozygous state in females are not inconsistent with an implication of these genes in XLID. Similarly, hemizygous truncating variants that affect only some specific splicing isoforms do not question the implication of such genes either. When a unique truncating variant affecting the main isoform of a gene is identified in a single male (such as one in *MAGT1* [MIM 300715], *ZNF41* [MIM 314995], *ZNF81* [MIM 314998], or *ATP6AP2* [MIM 300556]), the interpretation is more ambiguous, and this alone cannot rule out the gene's role in ID. However, two different truncating variants have been reported in *NXF5* (MIM 300319) in eight males and in *ZNF674* (MIM 300573) in 19 males, and this seriously challenges the implication of these genes in monogenic XLID (see below and Supplemental Data).

Presence in EVS of Known XLID Mutations

We then investigated the presence among the NHLBI cohort of all XLID mutations identified in the 104 genes, which are listed in OMIM or were retrieved by a PubMed search (Table S2), and we found 22 such mutations (Table S3). All but two (c.1186C>T [p.Pro396Ser] in *CASK* and c.217C>T [p.Leu73Phe] in *PTCHD1*, identified in single heterozygous females) were detected in both males and females and are probably noncausative variants instead of real disease-causing mutations. The majority of these missense mutations (16/20) that now appear to be innocuous variants are predicted to have benign or unknown consequences on protein function. The pathogenicity of some of these missense mutations (c.515C>A [p.Pro172His] in *TSPAN7*, c.3872C>T [p.Pro1291Leu] in *FLNA*, and three missense mutations in *DMD*) was already questioned by the authors.[71–73] Concerning the *FGD1* (MIM 300546; involved in Aarskog-Scott syndrome [AAS (MIM 305400)]) c.935C>T (p.Pro312Leu) mutation, which was detected twice in NHLBI males, the phenotype of the reported family did not fit with classical AAS, already suggesting that it might not be a truly pathogenic mutation.[74] Even more surprising was the identification of 18 males carrying *NDUFA1* mutation c.94G>C (p.Gly32Arg), which was demonstrated to have functional consequences on protein activity.[75,76] Such observations question these specific mutations, but they do not raise doubt about the involvement of the cognate genes in ID given that several other convincing mutations have been reported in such genes (*TSPAN7* [MIM 300096], *FLNA* [MIM 300017], *FGD1*, *NDUFA1* [MIM 300078], and *PDHA1* [MIM 300502]). However, for other genes (*SRPX2*, *ZNF41*, *ARHGEF6* [MIM 300267], etc.), the presence of mutations at a relatively high frequency in the general population challenges their involvement in ID (see the following sections and Table 1).

Questioning Monogenic Involvement of 28 Specific Genes in XLID

Our investigation focused on 28 genes proposed to play a role in XLID when mutated and for which there is a paucity of confirmatory studies or some contradictory elements arising from the analysis of the EVS data (Table 1). This allowed us to classify them into five categories. We highlight five genes that appear very unlikely to play a role in ID with high penetrance in males (*AGTR2* [MIM 300034], *MAGT1*, *NXF5*, *ZNF674*, and *ZNF41*; we refer to these as "highly questionable") and five others for which implication in ID appears suspicious (*ZNF81*, *ARHGEF6*, *ATP6AP2*, *SRPX2*, and *ZCCHC12* [*SIZN1*, MIM 300701]; we refer to these as "questionable"). For 15 genes, additional confirmation studies are required, but no contradictory findings were found in EVS. These include some genes that have never been replicated since their first publication more than 8 years ago (*IGBP1* [MIM 300139], *KIAA2022* [MIM 300524], *KLF8* [MIM 300286], *NLGN3* [MIM 300336], *ZDHHC15* [MIM 300576], *ZMYM3* [MIM 300061], and *MAOA* [MIM 309850]; we call these "never replicated") or others that have been recently described but have not yet been replicated (*CCDC22* [MIM 300859], *CLIC2* [MIM 300138], *CNKSR2* [MIM 300724], *FRMPD4* [MIM 300838], *HCFC1* [MIM 300019], *NAA10* [MIM 300013], *RPL10* [MIM 312173], and *SHROOM4* [MIM 300579]; we refer to these as "awaiting replication"). Finally, for three genes—*HUWE1* (MIM 300697), *PTCHD1*, and *SYN1* (MIM 313440)—our analysis and/or recent confirmation reports support a real implication in ID (we refer to these genes as "likely"). We detail here current arguments and evidence for the five highly questionable genes and also present a few others as examples of the four other categories (*SRPX2* and *ARHGEF6* as questionable, *MAOA* as never replicated, *HCFC1* as awaiting replication, and *HUWE1* as likely). The detailed information for the remaining 18 genes is given in the Supplemental Data.

AGTR2: Highly Questionable

After a de novo balanced (X;7)(q24;q22) translocation disrupting *AGTR2* (angiotensin II receptor 2) was identified in one female with ID, it was screened in a cohort of 590 male ID probands, 38 of whom were in families affected by

might encompass several exons) was described, genes are in red (more than five mutations in total), orange (two to five mutations in total), or green (only one mutation). If only missense mutations or mutations involved in expression decrease were identified, genes are in violet (more than five mutations in total), blue (two to five mutations in total), or purple (only one mutation). Asterisks denote genes discussed in this paper but whose implication in ID remains very likely. In parentheses next to each gene are lists in which that gene is included (abbreviations are as follows: L, Lubs[4]; R, Ropers[8]; G, Greenwood Genetic Center; E, Emory Genetics Laboratory; and A, Ambry Genetics). References for the genes not reported in any list are as follows: [1] Honda et al., [2] Houge et al., [3] Voineagu et al., [4] Huang et al., and [5] Witham et al.

Table 1. List of All 28 Genes Discussed and Associated Evidence Supporting or Questioning Their Involvement in ID

Gene (RefSeq Accession Number)	Lists[a]	Reason for Highlight[b]	Identification Method[c]	ID Structural Variations	ID Point Mutations	Occurrence of Mutations in EVS[d]	Truncated Variants in EVS	dN/dS Ratio (#N/#S)	Protein Size	Implication in ID
AGTR2 (NM_000686.4)	L, R, G, E, A	T	1	(X; 7) translocation in a female[19]	c.402del (p.Phe134Leufs*5)[19]	(see Table S4)	yes (1 F)	0.731 (19/8)	363	highly questionable (already controversial)[20–22]
					c.62G>T (p.Gly21Val)[19,23]	10 M, 23 F (htz)				
					c.971G>A (p.Arg324Gln)[19]	4 M, 24 F (htz)				
					c.1009A>G (p.Ile337Val)[19]	NF				
					c.157A>T (p.Ile53Phe)[24]	NF				
					c.572G>A (p.Gly191Glu)[23]	NF				
MAGT1 (*IAP*) (NM_032121.5)	L, R, G, E	T	3	-	c.1028T>G (p.Val311_343Gly)[25]	5 M, 8 F (htz)	yes (1 M)	0.453 (15/10)	367	highly questionable
NXF5 (NM_032946.2)	L, R, G, A	T	1	inv(X)(p21.1;q22) in two males;[26,27] deletion disrupting the NXF cluster;[28,29] duplication at Xq22.1 disrupting[30] and duplicating[31] *NXF5* in one female and one male, respectively	-	-	yes (8 M)	0.783 (30/11)	365	highly questionable
ZNF674 (NM_001039891.2)	L, R, E, A	T	1	deletion encompassing *ZNF673*, *ZNF674*, *RP2*, *SLC9Z7*, and *CHST7* in one male;[32] duplication encompassing *ZNF673*, *ZNF674*, and *CHST7*[33]	c.352G>T (p.Glu118*)[32]	NF	yes (19 M)	0.750 (28/10)	581	highly questionable (already controversial)[34]
					c.1235C>A (p.Pro412Leu)[32]	NF				
					c.1028C>T (p.Met343Thr)[32]	49 M, 16 F (hmz), 223 F (htz)				
ZNF41 (NM_007130.2)	L, R, G, E, A	M	1	(X; 7) translocation disrupting *ZNF41*[35]	c.332C>T (p.Pro111Leu)[35]	2 M, 8 F (htz)	yes (1 M, 1 F)	0.696 (45/17)	779	highly questionable
					c.73−42A>C (p.?)[35]	5 M, 9 F (htz)				
ZNF81 (NM_007137.3)	L, R, G, E, A	M	1	(X; 9) translocation in one female;[36] 1.3 Mb duplication encompassing >30 genes at Xp11.23–p11.3 in one male;[37] 335 kb microduplication bearing two other genes at Xp11.2–p11.3 in one male[38]	c.536G>A (p.Ser179Asn)[36]	NF	yes (1 M)	0.595 (25/11)	661	questionable
ARHGEF6 (NM_004840.2)	L, R, G, E, A	M	1	(X; 21) translocation in one male[39]	c.166−11T>C (p.?)[39]	5 M, 13 F (htz)	no	0.456 (28/18)	776	questionable
ATP6AP2 (NM_005765.2)	L, R, G, E, A	T	2	-	c.321C>T (p.Asp107Asp) (affects splicing)[40]	NF	yes (1 M)	0.444 (14/10)	350	questionable

(*Continued on next page*)

Table 1. *Continued*

SRPX2 (NM_014467.2)	L, R, G, E, A	M	2	-	c.980A>G (p.Asn327Ser)[41]	3 M, 8 F (htz)	no	0.728 (31/14)	465	questionable
					c.215A>C (p.Tyr72Ser)[41]	1 M				
ZCCHC12 (*SIZN1*) (NM_173798.2)	E	M	3	-	c.19C>T (p.Arg7Cys)[42]	5 M, 28 F (htz)	no	0.406 (19/14)	402	questionable
					c.1031C>T (p.Thr344Ile)[42]	NF				
IGBP1 (NM_001551.2)	L, G, E	I	2	-	5′ UTR 2 bp substitution affecting *IGBP1* expression[43]	not covered	no	0.487 (19/11)	339	never replicated
KIAA2022 (NM_001008537.2)	L, R, G, E, A	I	1	inv(X)(q13;p22) in two related males[44]	-	NA	no	0.441 (54/34)	1,516	never replicated
KLF8 (NM_007250.4)	L, R, G, E	I	1	(X; 21) translocation leading to a loss of *KLF8* expression in a woman[45]	-	NA	no	0.831 (16/6)	359	never replicated
NLGN3 (NM_181303.1)	L, R, G, E, A	I	3	-	c.1411C>T (p.Arg451_471Cys) in two ASD brothers[46]	NF	no	0.209 (17/27)	848	never replicated
ZDHHC15 (NM_144969.2)	L, G, E	I	1	(X; 15) translocation leading to a loss of *ZDHHC15* expression in a female[47]	-	NA	no	0.628 (17/8)	337	never replicated
ZNF261 (*ZMYM3*) (NM_201599.2)	R	I	1	(X; 13) translocation disrupting *ZMYM3* in one female[48]	-	NA	no	0.232 (31/42)	1,370	never replicated
MAOA (NM_000240.3)	L, R, G, E, A	I	1	*MAOA-MAOB-NDP* deletion in subjects with severe ID[49]	c.886C>T (p.Gln296*)[50]	NF	no	0,359 (14/12)	527	never replicated
CCDC22 (NM_014008.3)	-	I	3	-	c.49A>G (p.Thr17Ala)[51]	NF	no	0.809 (40/17)	627	awaiting replication
CLIC2 (NM_001289.4)	-	I	4	-	c.303C>G (p.His101Gln)[52]	NF	no	0.346 (6/5)	247	awaiting replication
CNKSR2 (NM_014927.3)	-	I	1	partial deletion of *CNKSR2*[53]	-	NA	no	0.340 (22/19)	1,034	awaiting replication
FRMPD4 (NM_014728.3)	-	I	1	partial duplication of *FRMPD4*[54]	-	NA	no	0.383 (57/46)	1,322	awaiting replication
HCFC1 (NM_005334.2)	-	I	2	-	chrX: 152890455A>G[55]	NA	no	0.227 (50/80)	2,035	awaiting replication
					c.674G>A (p.Ser225Asn)[55]	NF				
NAA10 (NM_003491.2)	L	I	4	-	c.109T>C (p.Ser37Pro)[56]	NF	no	0.074 (2/8)	235	likely in Ogden syndrome but needs further replication in nonsyndromic ID
					c.346C>T (p.Arg116Trp)[13]	NF				

(*Continued on next page*)

Table 1. *Continued*

Gene (RefSeq Accession Number)	Lists[a]	Reason for Highlight[b]	Identification Method[c]	ID Structural Variations	ID Point Mutations	Occurrence of Mutations in EVS[d]	Truncated Variants in EVS	dN/dS Ratio (#N/#S)	Protein Size	Implication in ID
RPL10 (NM_006013.3)	L, R, G, E, A	I	3	-	c.616C>A (p.Leu206Met) (ASD)[57]	NF	no	0.154 (2/4)	214	awaiting replication
					c.639C>G (p.His213Gln) (ASD)[57,58]	NF				
SHROOM4 (NM_020717.3)	L, R, G, E, A	I	1	X-autosome translocations in two females;[59] Xp11.22 deletion bearing *SHROOM4* in one family[54]	c.3266C>T (p.Ser1089Leu)[59]	NF	yes (1 F)	0.918 (86/28)	1,493	awaiting replication
					c.1422A>G (p.Glu474Glu)[59]	NF				
HUWE1 (NM_031407.5)	L, R, G, E, A	I	1	microduplications encompassing *HSD17B10* and *HUWE1*[60,61] or *HUWE1* alone[62] in a total of 16 unrelated families	c.12037C>T (p.Arg4013Trp)[61]	NF	no	0.201 (67/105)	4,374	likely
					c.8942G>A (p.Arg2981His)[61]	NF				
					c.12559C>T (p.Arg4187Cys)[61]	NF				
					c.2849T>A (p.Val950Asp) (ASD)[63]	NF				
PTCHD1 (NM_173495.2)	L	M	1	160 kb deletion leading to a null *PTCHD1* in dizygotic ASD twin brothers;[64] 50–390 kb inherited deletions disrupting *PTCHD1* in six ASD male probands;[65] 90 kb deletion disrupting *PTCHD1* in three related ID males;[60,66] 200 kb deletion in two ID brothers[67]	c.517A>G (p.Ile173Val)[66]	2 M, 2 F (htz)	no	0.366 (29/25)	888	likely
					c.583G>A (p.Val195Ile)[66]	NF				
					c.1008_1009delinsTA (p.MetLeu336_337IleIle)[66]	NF				
					c.217C>T (p.Leu73Phe)[66]	1 F (htz)				
					c.1436A>G (p.Glu479Gly)[66]	NF				
					c.1409C>A (p.Ala470Asp)[66]	NF				
					c.1076A>G (p.His359Arg)[66]	NF				
SYN 1 (NM_006950.3)	L, R, G, E, A	M	2	-	c.1067G>A (p.Trp356*)[68]	NF	no	0.347 (12/12)	705	likely
					c.1663C>T (p.Gln555*)[69]	not covered				
					c.152C>G (p.Ala51Gly)[69]	11 M, 30 F (htz)				
					c.1648G>A (p.Ala550Thr)[69]	not covered				
					c.1699A>G (p.Thr567Ala)[69]	not covered				

Abbreviations are as follows: L, Lubs et al.;[4] R, Ropers et al.;[8] G, Greenwood Genetic Center (see Web Resources); E, Emory Genetics Laboratory (see Web Resources); A, Ambry Genetics (see Web Resources); T, truncating variants reported in EVS; M, ID mutations reported in EVS; I, one piece of evidence or study implicating this gene in ID; htz, heterozygous; hmz, homozygous; M, number of males carrying the mutation; F, number of females carrying the mutation; NF, not found; NA, not applicable; N, number of nonsynonymous variants; and S, number of silent variants.

[a]Lists in which genes are included.

[b]Reason why we discuss this gene in the paper.

[c]Approach used for the initial identification, consistent with the information from the text: 1, positional candidate found through rearrangement or CNV; 2, positional candidate located in a linked region; 3, functional candidate; and 4, through exome or X-exome sequencing.

[d]Occurrence of ID mutations in EVS.

probable or possible XLID.[19] This led to the identification of (1) a 1 bp deletion causing a frameshift (c.402del [p.Phe134Leufs*5]) in a family affected by probable XLID and in another simplex case and (2) three missense mutations (c.62G>T [p.Gly21Val], c.971G>A [p.Arg324Gln], and c.1009A>G [p.Ile337Val]) absent from 510 control X chromosomes. A further study of 57 males with ID detected two individuals with the p.Gly21Val or p.Ile53Phe changes.[24] Other extensive studies (25 XLID-affected families showing coherent linkage to the locus, 116 affected sibling pairs, and 224 simplex cases,[9,20] as well as another set of families[10]) failed to detect any additional mutations (besides the p.Gly21Val variant and a missense variant that did not show full cosegregation in a family that was later shown to carry a *SOX3* [MIM 313430] mutation). More recently, screening of 203 Japanese males with ID detected an additional inherited missense change (c.572G>A [p.Gly191Glu]) in a boy showing severe ID, pervasive developmental disorder, and epilepsy.[23] Meanwhile, both the p.Phe134Leufs*5 frameshift and the p.Gly21Val variant were reported—in rarely cited publications—in males from control cohorts, suggesting that they are unlikely to be causative.[21,22] This is now supported by the NHLBI data, given that two of the three initially reported missense variants (p.Gly21Val and p.Arg324Gln) are present in ten and four males, respectively. Moreover, although we did not include the indel data from EVS, we noted that the initial frameshift mutation was observed in 7 of 1,770 European American males (minor allele frequency [MAF] = 0.4%), but not in African American individuals, thus excluding a possible sequencing artifact (Table S4). Consequently, EVS data support the serious doubts previously raised about the implication of *AGTR2* in ID, despite the suggestive phenotype of the *AGTR2*-knockout mouse showing altered memory capacities and abnormal dendritic-spine morphology.[77]

NXF5: Highly Questionable

Mutant *NXF5* (nuclear RNA export factor 5) was suspected to cause ID after pericentric inversion inv(X)(p21.1;q22), leading to the loss of NXF5 expression, was identified in an ID-affected male.[26,27] The observation of two nonsense variants (c.162G>T [p.Cys54*] and c.958G>A [p.Arg320*]) affecting a total of eight males in EVS raises serious doubts about the involvement of this gene in ID, despite some indication of its function in the brain. This gene was more recently found to be deleted or duplicated in other ID individuals, but never alone.[28,29,31]

MAGT1: Highly Questionable

After the identification of an ID-causing mutation in *TUSC3* (encoding a putative subunit of an oligosaccharyltransferase complex [MIM 601385]), the sequencing of its X-linked paralog gene, *MAGT1* (magnesium transporter 1, formerly called *IAP*), in XLID-affected families revealed one missense change segregating with the ID status (LOD score 1.8) and absent from more than 267 control chromosomes.[25] This variant (p.Val311Gly but annotated c.1028T>G [p.Val343Gly] in EVS) was detected in 13 out of 10,557 X chromosomes from the NHLBI cohort. Moreover, one male from the same population carries a nonsense variant in this gene. Although the implication of *TUSC3* in recessive ID appears well supported from the identification of additional consanguineous families affected by mutations in this gene,[78] no ID-related mutation has been reported in *MAGT1* since 2008, and its disruption has been recently associated with several forms of X-linked immunodeficiency without neurological manifestations.[79] In light of these observations, there is no current support for the implication of *MAGT1* in ID.

ZNF674: Highly Questionable

After the identification of one ID-associated CNV encompassing *ZNF674* (zinc-finger protein 674) and four other genes (*RP2* [MIM 300757], *SLC9A7* [MIM 300368], *CHST7* [MIM 300375] and *ZNF673* [MIM 300585]), its screening in additional families led to the identification of a nonsense mutation (c.352G>T [p.Glu118*]) segregating with the ID phenotype in a large family (LOD score 2.51).[32] Further screening of 28 families with nonsyndromic XLID exhibiting coherent linkage to the region and 309 simplex cases detected one inherited missense change (c.1235C>A [p.Pro412Leu]) in two related males and a second (c.1028C>T [p.Thr343Met]) in another small family.[32] Males carrying the p.Pro412Leu missense variant also displayed partial trisomy 21 and monosomy 18 as a result of an unbalanced translocation, which precluded conclusions about the causative effect of this variant. Also, the authors concluded that the p.Thr343Met missense variant was unlikely to be pathogenic (despite its being absent from 354 controls) because it did not affect a highly conserved residue and because the replacing methionine was present in the *ZNF674* chimpanzee ortholog. EVS data indicate that this variant is present in 49 males, 223 carrier females, and 16 homozygous females, fully excluding its implication in ID. The implication of *ZNF674* in ID was recently questioned when it was found to be deleted together with *RP2* (in retinitis pigmentosa [MIM 312600]) in two families with X-linked retinal dystrophy without cognitive impairment.[34] Lastly, given that two nonsense mutations (c.1324C>A [p.Asp442*] and c.601G>A [p.Arg201*]) have been reported in the NHLBI cohort and that one is present in 19 hemizygous males, *ZNF674* seems unlikely to be involved in monogenic ID.

ZNF41: Highly Questionable

ZNF41 (zinc-finger protein 41) was found to be disrupted by a de novo balanced (X;7) translocation in a girl with severe developmental delay.[35] Subsequent screening of 210 probands from families affected by probable or possible XLID revealed one missense change (c.332C>T [p.Pro111Leu]) and one intronic variant (c.479−42A>C) causing a splicing alteration; both are absent from over 400 control

X chromosomes, and both show cosegregation in the families. Although the authors concluded carefully that *ZNF41* could be an ID candidate, it is now included in most lists of ID-associated genes. However, two truncating variants have been reported in one male and one female in EVS, and the two putative point mutations initially described by Shoichet et al. have been detected in a total of seven males at a frequency too high to be fully penetrant ID mutations (see Table 1).[35] No other *ZNF41* mutation has been reported since the initial 2003 report.

SRPX2: Questionable

In 2006, Roll et al. reported a family affected by probable dominant XLID, rolandic epilepsy, and speech dyspraxia.[41] Sequencing the genes of the linkage region (LOD score 3.01) revealed only a missense variant (c.980A>G [p.Asn327Ser]) in *SRPX2* (sushi-repeat-containing protein, X-linked 2). Another missense variant (c.215A>C [p.Tyr72Ser]) was detected in one epileptic male and his female relatives, all of whom showed additional bilateral perisylvian polymicrogyria. This latter variant was reported to increase the affinity of the encoded SRPX2 for the urokinase-type plasminogen activator receptor.[80] The initial p.Asn327Ser alteration was shown to affect protein glycosylation and to cause partial retention of the altered protein within the endoplasmic reticulum in a transfected-cell model.[41] Despite their observed functional consequences, both variants are present in EVS in three males and eight females (all of European American origin; p.Asn327Ser) and in one male (p.Tyr72Ser). Accordingly, 0.28% of the EVS European American population carries one of these two variants, which considerably weakens the evidence supporting a role for *SRPX2* in epilepsy, speech, and cognition. It is most unlikely that mutations in *SRPX2* are a major cause of epilepsy given that no mutation has been reported in this gene since 2006, despite its sequencing in 100 individuals with a similar phenotype of epilepsy and speech disorder (D. Sanlaville, G. Lesca, and P. Szepetowski, personal communication). Because SRXP2 seems to play a role during brain development and was recently shown to influence neuronal migration in the developing cerebral cortex in rats (P. Szepetowski, personal communication), we cannot, however, exclude that other drastic alterations of SRXP2 function might cause cognitive impairment.

ARHGEF6: Questionable

ARHGEF6 (*α-PIX* or *Cool-2*; rho guanine nucleotide exchange factor 6) was identified as disrupted via an inherited balanced (X;21) translocation in a male with ID.[39] In the same study, subsequent screening of 119 unrelated males with nonsyndromic ID allowed the identification of a cosegregating intronic mutation (c.166−11T>C) affecting splicing in a large XLID-affected family; it was absent from 170 control chromosomes. However, this intronic variant has been reported in EVS in five males, all of European origin (MAF = 0.26%). An inherited duplication of *ARHGEF6* was also reported in two brothers with moderate ID, but later analysis showed that the duplication spanned 1.8 Mb and encompassed 24 genes.[81,82] In the Tarpey et al. study, two missense changes were detected in single families: c.362G>A (p.Arg121His), which is found in 14 NHLBI males, and c.992A>G (p.Tyr331Cys), which is not observed in EVS. However, the latter missense variant is not in the list of likely pathogenic mutations and was not further commented on by the authors. Engineered *αPix* (*Arhgef6*)-deficient mice present altered hippocampal neuronal connectivity, impaired synaptic function, and additional cognitive deficits.[83] However, because no other pathogenic mutation has been described since 2003 and given the unlikely pathogenicity of the initially reported splice variant, the implication of *ARHGEF6* in ID still remains unclear.

MAOA: Never Replicated

The combined loss of *MAOA* (monoamine oxidase A) and *MAOB* (MIM 309860), along with the deletion of the gene associated with Norrie disease, *NDP* (MIM 300658), has been described in some probands with a continuous syndrome and presenting with severe ID.[49] In 1993, Brunner et al. described a large Dutch family affected by X-linked mild ID and prominent behavioral abnormalities associated with disturbances in monoamine metabolism and segregating with a truncating mutation (c.886C>T [p.Gln296*]) in *MAOA*.[30,50] More recently, deletions of *MAOA* with *MAOB* alone were described in two males with severe developmental delay or ID.[60,84] Mouse models with total *Maoa* inactivation or with a spontaneous mutation present indeed enhanced aggressive behavior and additional "autistic-like" features.[85] To our knowledge, no other clearly pathogenic mutation has been reported in *MAOA* in other individuals in the past 20 years,[86] even though one missense variant (c.812A>T [p.Asn271Ile]) was reported in a proband with ASD.[87] However, we recently identified a pathogenic missense mutation (c.797_798GC>TT [p.Cys266Phe]) associated with deficient MAOA activity in one family affected by autism, ID, and abnormal behavior, confirming the findings from 1993 (A.P., unpublished data). A third family harboring a *MAOA* mutation was recently identified (M. Field, J. Gecz, and V.M. Kalscheuer, personal communication).

HCFC1: Awaiting Replication

The mutation responsible for ID in the first family reported as linked to the MRX3 locus remained unsolved for a long time after the initial linkage analysis of 1991. A targeted massive resequencing of the genomic linkage interval recently revealed a regulatory point mutation in a functional binding site for the YY1 transcription factor in the *HCFC1* (host-cell factor C1) promoter; this mutation leads to an upregulation of the gene's expression in lymphoblastoid cells. The screening of extra unsolved XLID-affected families identified an additional missense mutation (c.674G>A [p.Ser225Asn]) segregating with the disease.[55]

Both mutations are absent from EVS, and other unique missense variants were also described in a female with schizophrenia and a boy with ASD.[87] However, because *HCFC1* encodes a large protein of 2,035 amino acids, the probability of identifying a missense change is high (22 missense variants predicted to be possibly or probably damaging are present in EVS). Additional evidence would therefore be useful for definitely confirming the implication of *HCFC1* in ID.

HUWE1: Likely

HUWE1 (Hect, UBA, and WWE domains-containing protein 1) encodes a very large protein (4,375 amino acids) with an E3 ubiquitin ligase function. Array comparative genomic hybridization analysis initially detected duplications of *HUWE1* and *HSD17B10*, both of which showed overexpression in affected probands in six families afflicted with predominantly nonsyndromic XLID; two of these families were large XLID-affected families.[61] Four similar cases were reported in an independent replication study.[60] A further CNV analysis showed that de novo duplications covering *HUWE1* alone were also associated with simplex cases of ID, indicating that increased expression of this gene is deleterious.[62] In the initial study, three different missense mutations affecting highly conserved residues and cosegregating with the ID status were identified in three unrelated XLID-affected families (c.12037C>T [p.Arg4013Trp] in a large Australian family [LOD score = 3.31][88] and c.8942G>A [p.Arg2981His] and c.12559C>T [p.Arg4187Cys] in two smaller families; all three mutations were absent from 750 controls). More recently, a de novo missense variant (c.2849T>A [p.Val950Asp]) was observed in one ASD male but was absent in his less severely affected brother, suggesting either a different etiology or complex inheritance.[63] None of these variants is present in EVS. Functional studies, notably in mouse models, indicate that *HUWE1* plays a role in the control of neurogenesis in the developing brain via the N-Myc pathway.[89–91] In mice, targeted inactivation of *Huwe1* in the CNS or only in cerebellar granule neuron precursors and radial glia resulted in high neonatal lethality, which could explain the absence of truncating mutations reported in individuals.[89,90] All these observations further support the role of mutations in *HUWE1* in XLID. However, one should be very cautious in ascribing a pathogenic role to a HUWE1 missense change identified in an individual with ID because 21 different such variants were observed in males in the NHLBI cohort (ten of these variants are qualified as possibly or probably damaging); this number is consistent with the large target size of the protein-coding sequence.

dN/dS Ratio Distribution among the Genes Involved in XLID

The latest publication on EVS data reported a total of 285,960 nonsynonymous and 188,975 silent variants on 2,440 exomes.[16] Interestingly, compared to the full genome, genes implicated in XLID tend to accumulate fewer nonsynonymous variants (nonsynonymous = 2,623; silent = 2,195; $p = 0.04$). In order to estimate the evolutionary pressure exerted on those different genes, we used EVS to compute an intrahuman dN/dS ratio, which we calculated by using DnaSP.[92] For this ratio, the number of nonsynonymous variants observed in EVS over the number of total putative nonsynonymous positions in the studied transcript is divided by the analogous ratio for synonymous variants (Table S5). Accordingly, a strong selection pressure is expected to engender a low dN/dS ratio, and a truly associated gene can be expected to have such a low ratio in a general population. In contrast, genes that show a high dN/dS ratio are less likely to be disease related, especially if the disease affects fitness. We generated box plots of the dN/dS ratios for the 104 selected genes with supposed involvement in XLID according to the category we classed them into (Figure S1). Interestingly, the mean dN/dS ratios for the ten genes in which mutations associated with ID are questioned was significantly higher than that for the validated genes ($\mu = 0.604 \pm 0.150$ versus $\mu = 0.359 \pm 0.255$; Student's t test $p = 0.0004$). Conversely, no significant difference was observed between the validated and the "awaiting replication" pools ($\mu = 0.427 \pm 0.259$, Student's t test $p = 0.35$) that might contain genes with either real or no implication in ID. Among the 11 genes with a high dN/dS ratio above the 90th percentile (0.723; Table S5), four belong to the "questionable" class (*AGTR2*, *ZNF674*, *SRPX2*, and *NXF5*) and three are considered to be "awaiting replication" (*CCDC22*, *KLF8*, and *SHROOM4*).

Discussion

This work demonstrates that among the 100 genes described in the past as involved in XLID (on the basis of available data sets and sequencing technologies), at least 10 (10%) appear doubtful and 15 others should be considered with caution until validation by replication studies. The involvement in ID of two of these genes, *AGTR2* and *ZNF674*, was already questioned in specific publications that were seldom cited.[22,34] Curiously, some fully validated genes are missing from several XLID diagnostic panels (*NSDHL* [MIM 300275], *MBTPS2* [MIM 300294], *SLC6A8* [MIM 300036], etc., Figure 1), whereas others that have been described in only a few studies and whose implication in ID is not fully convincing are systematically included (*AGTR2*, *MAGT1*, *ZNF41*, and *SRPX2*). This work also highlights that even in a well-known ID-associated gene, the pathogenicity of a novel nonsynonymous variant should be inferred with caution given that some reported nonsynonymous "mutations" in *NDUFA1*, *TSPAN7*, or *PTCHD1* are likely to be nonpathogenic (or at least not fully penetrant), and this is of particular relevance for diagnostic applications (see also *HUWE1* section).[93]

The number of genes involved in ID when mutated is rapidly growing as a result of the exome-sequencing analysis of cohorts of affected individuals.[12–14] Although the

Table 2. Proposed Guidelines for the Assessment of a Variant or a Gene Potentially Involved in ID or in Another Rare or Genetically Heterogeneous Disease[a]

	X-Linked or Dominant Form	Recessive Form[b]
Involvement of a Variant in a Disease		
Supports	not found in any male[c] or individual of the GP	found at an allele frequency $\leq \sqrt{i}$
	segregates with the disease in the family[c]	
	appears de novo in a simplex case	
Inconclusive	found in one male[c] or individual of the GP	
Does not support	found in more than one male[c] or individual of the GP	found in GP at an allele frequency $> \sqrt{i}$
Association between a Gene and a Disease		
Supports	additional candidate variant(s) in unrelated individuals with similar specific phenotype	
	If candidate mutation(s) = protein truncation(s): - no truncating variant is reported in males[c] or individuals of the GP	-
	If candidate mutation(s) = missense(s): - a low level of missense mutations predicted to be damaging is reported in males[a] or individuals of the GP	-
Inconclusive	*If candidate mutation(s) = protein truncation(s):* - one truncating variant in one male[c] or individual of the GP - truncating variants in the GP affect only the last amino acids or one of several functional isoform(s)	-
Does not support	*If candidate mutation(s) = protein truncation(s):* - several truncating variants in males[c] or individuals of the GP	homozygous truncating or damaging variants are reported in individuals of the GP
	If candidate mutation(s) = missense(s): - high level of missenses in the GP (high dN/dS ratio, e.g., >0.7)	truncating or damaging variants are reported in the GP at allele frequency $> \sqrt{i}$

The ID example (incidence = 2%): we do not expect one mutation responsible for more than 0.1% of ID cases, so $i = 0.02 \times 0.001 = 1 / 50{,}000$ in the GP. Therefore, the variation should not be found at a frequency $> \sqrt{1/50,000} = 0.0045$, meaning no more than 38 times in 8,600 European American chromosomes. Most true mutations should be even rarer given mutation heterogeneity at a given locus; for instance, the most frequent mutation in phenylketonuria (c.1222C>T [p.Arg408Trp]) is found 15 times in 8,600 European American chromosomes. The following abbreviations are used: GP, general population (e.g., EVS); and *i*, estimated disease incidence that can be caused by mutations in a single gene after postulated nonallelic genetic heterogeneity is taken into account.
[a]For a phenotype not enriched in the cohort used as a GP.
[b]Caution: geographic origin of the GP should match that of individuals with the candidate mutation.
[c]For an X-linked disease.

discovery of mutations in additional genes represents an important step for further understanding the underlying genetics of ID, there is a clear need for replication studies to validate these findings (such as the recent work by O'Roak et al.).[94] Recently, *RAB40AL* (MIM 300405), not included in our list, was found to be associated with Martin-Probst syndrome (MPS [MIM 300519]), a disorder characterized by deafness, cognitive impairment, short stature, and distinct craniofacial dysmorphisms, with the identification (by parallel whole-genome, whole-exome, and X-exome sequencing) of a 2 nt substitution (c.176_177AC>GA) leading to a missense change (p.Asp59Gly) shown to disrupt protein localization.[95] However, this variation is present in two males in the EVS population. Also, a few months after this initial publication, Iqbal et al. reported it in four other ID-affected families; however, it was not related to the disease in at least three of the families, questioning the involvement of *RAB40AL* in the MPS phenotype.[96] Some other recent publications of exome sequencing in ID and ASD simplex cases identified de novo truncating variants suggested by the authors to be most likely involved in autosomal-dominant ID or ASD in genes not previously associated with these diseases. However, for *ADAM33*, which was initially considered a candidate gene for dominant forms of ASD after a de novo truncating mutation was identified in an ASD proband, three nonsense variants and two splice variants were reported in a total of five carrier individuals from the NHLBI cohort.[97] Similarly, six nonsense and three splice variants were detected in a total of 11 carriers in *ACACB*, which was also proposed as a likely candidate for dominant ASD.[98] This illustrates that an analysis similar to ours should be performed on all proposed autosomal genes for both dominant and recessive forms of ID.

We believe that this approach should be extended to other Mendelian diseases with high genetic heterogeneity, but the way to assimilate the NHLBI data has to be adapted carefully and specifically for each disease by the integration of both its frequency and proposed transmission mode (see Table 2). While this work was being completed, a thorough similar review of mutations previously associated with cardiomyopathy revealed the presence of about 15% of them in EVS, showing a cumulative frequency inconsistent with an implication in monogenic forms of the disease.[99] However, the NHLBI cohort cannot really be considered a control cohort with respect to

cardiomyopathies or other cardiac diseases given that many individuals were recruited through the NHLBI because of cardiac phenotypes.

Because data from projects such as the NHLBI Exome Sequencing Project are available, one should check for the absence of each candidate mutation, especially if it is a missense change. In the case of a novel candidate mutation in a gene never associated with a disease, one should also analyze other information provided by EVS, such as the presence of truncating variants and the dN/dS ratio. Indeed, this ratio is a good indicator of the selection pressure exerted on a particular gene, and if the candidate mutation results in a missense change, one should remember that there is a higher likelihood of identifying a neutral nonsynonymous change in a gene with a high dN/dS ratio.

However, one important limitation of the EVS is the impossibility of going back to the cognitive or neurologic phenotype of individuals in whom rare potentially pathogenic variants were detected. Similarly to existing CNV databases, such as Decipher,[100] a public database of rare exome variants linked to some phenotypic descriptions would be warranted, especially given the overwhelming amount of such sequencing information that is currently being generated in research and that will be soon engendered for diagnostic purposes of ID and autism.

Supplemental Data

Supplemental Data include an additional description of the evidence supporting the role of the remaining 19 ID-associated genes included in Table 1, one figure, and five tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

We thank Stephanie Le Gras and Jean Muller for bioinformatics assistance and more particularly for helping download data from the National Heart, Lung, and Blood Institute Exome Sequencing Project Exome Variant Server. Our work on intellectual disability genetics is supported by grants from the Fondation Jerome Lejeune and the Agence de Biomédecine, a postdoctoral fellowship from the Fondation pour la Recherche Médicale to A.P., and a student fellowship from the APLM and Groupement d'Intérêt Economique - Centre Européen de Recherche en Biologie et en Médecine to C.R.

Received: February 19, 2013
Revised: May 29, 2013
Accepted: June 8, 2013
Published: July 18, 2013

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://www.1000genomes.org/
Ambry Genetics, http://www.ambrygen.com
dbSNP, http://www.ncbi.nlm.nih.gov/SNP/
Emory Genetics Laboratory, http://genetics.emory.edu/egl/tests
Euro-MRX Consortium, http://www.euromrx.com
Greenwood Genetic Center, http://www.ggc.org
Human Genome Variation Society, http://www.hgvs.org/mutnomen/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
PubMed, http://www.ncbi.nlm.nih.gov/pubmed/
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq

References

1. Ellison, J.W., Rosenfeld, J.A., and Shaffer, L.G. (2013). Genetic basis of intellectual disability. Annu. Rev. Med. *64*, 441–450.
2. Leonard, H., and Wen, X. (2002). The epidemiology of mental retardation: challenges and opportunities in the new millennium. Ment. Retard. Dev. Disabil. Res. Rev. *8*, 117–134.
3. Gécz, J., Shoubridge, C., and Corbett, M. (2009). The genetic landscape of intellectual disability arising from chromosome X. Trends Genet. *25*, 308–316.
4. Lubs, H.A., Stevenson, R.E., and Schwartz, C.E. (2012). Fragile X and X-linked intellectual disability: four decades of discovery. Am. J. Hum. Genet. *90*, 579–590.
5. Biancalana, V., Beldjord, C., Taillandier, A., Szpiro-Tapia, S., Cusin, V., Gerson, F., Philippe, C., and Mandel, J.L. (2004). Five years of molecular diagnosis of Fragile X syndrome (1997-2001): a collaborative study reporting 95% of the activity in France. Am. J. Med. Genet. A. *129A*, 218–224.
6. Coffee, B., Keith, K., Albizua, I., Malone, T., Mowrey, J., Sherman, S.L., and Warren, S.T. (2009). Incidence of fragile X syndrome by newborn screening for methylated FMR1 DNA. Am. J. Hum. Genet. *85*, 503–514.
7. Kaufman, L., Ayub, M., and Vincent, J.B. (2010). The genetic basis of non-syndromic intellectual disability: a review. J Neurodev Disord *2*, 182–209.
8. Ropers, H.H. (2010). Genetics of early onset cognitive impairment. Annu. Rev. Genomics Hum. Genet. *11*, 161–187.
9. de Brouwer, A.P., Yntema, H.G., Kleefstra, T., Lugtenberg, D., Oudakker, A.R., de Vries, B.B., van Bokhoven, H., Van Esch, H., Frints, S.G., Froyen, G., et al. (2007). Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium. Hum. Mutat. *28*, 207–208.
10. Tarpey, P.S., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C., O'Meara, S., Latimer, C., Dicks, E., Menzies, A., et al. (2009). A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. Nat. Genet. *41*, 535–543.
11. Frints, S.G., Froyen, G., Marynen, P., and Fryns, J.P. (2002). X-linked mental retardation: vanishing boundaries between non-specific (MRX) and syndromic (MRXS) forms. Clin. Genet. *62*, 423–432.
12. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. *367*, 1921–1929.
13. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated

with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet *380*, 1674–1682.
14. Vissers, L.E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., et al. (2010). A de novo paradigm for mental retardation. Nat. Genet. *42*, 1109–1112.
15. Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al.; NHLBI Exome Sequencing Project. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216–220.
16. Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science *337*, 64–69.
17. Dahl, N., Hu, L.J., Chery, M., Fardeau, M., Gilgenkrantz, S., Nivelon-Chevallier, A., Sidaner-Noisette, I., Mugneret, F., Gouyon, J.B., Gal, A., et al. (1995). Myotubular myopathy in a girl with a deletion at Xq27-q28 and unbalanced X inactivation assigns the MTM1 gene to a 600-kb region. Am. J. Hum. Genet. *56*, 1108–1115.
18. Hodgson, S.V., Robertson, M.E., Fear, C.N., Goodship, J., Malcolm, S., Jay, B., Bobrow, M., and Pembrey, M.E. (1987). Prenatal diagnosis of X-linked choroideremia with mental retardation, associated with a cytologically detectable X-chromosome deletion. Hum. Genet. *75*, 286–290.
19. Vervoort, V.S., Beachem, M.A., Edwards, P.S., Ladd, S., Miller, K.E., de Mollerat, X., Clarkson, K., DuPont, B., Schwartz, C.E., Stevenson, R.E., et al. (2002). AGTR2 mutations in X-linked mental retardation. Science *296*, 2401–2403.
20. Bienvenu, T., Poirier, K., Van Esch, H., Hamel, B., Moraine, C., Fryns, J.P., Ropers, H.H., Beldjord, C., Yntema, H.G., and Chelly, J. (2003). Rare polymorphic variants of the AGTR2 gene in boys with non-specific mental retardation. J. Med. Genet. *40*, 357–359.
21. Erdmann, J., Dähmlow, S., Guse, M., Hetzer, R., and Regitz-Zagrosek, V. (2004). The assertion that a G21V mutation in AGTR2 causes mental retardation is not supported by other studies. Hum. Genet. *114*, 396, author reply 397.
22. Huang, D., Sun, W., and Strom, C.M. (2005). Sequence variations in AGTR2 are unlikely to be associated with X-linked mental retardation. Am. J. Med. Genet. A. *139*, 243–244.
23. Takeshita, E., Nakagawa, E., Nakatani, K., Sasaki, M., and Goto, Y. (2012). Novel AGTR2 missense mutation in a Japanese boy with severe mental retardation, pervasive developmental disorder, and epilepsy. Brain Dev. *34*, 776–779.
24. Ylisaukko-oja, T., Rehnström, K., Vanhala, R., Tengström, C., Lähdetie, J., and Järvelä, I. (2004). Identification of two AGTR2 mutations in male patients with non-syndromic mental retardation. Hum. Genet. *114*, 211–213.
25. Molinari, F., Foulquier, F., Tarpey, P.S., Morelle, W., Boissel, S., Teague, J., Edkins, S., Futreal, P.A., Stratton, M.R., Turner, G., et al. (2008). Oligosaccharyltransferase-subunit mutations in nonsyndromic mental retardation. Am. J. Hum. Genet. *82*, 1150–1157.
26. Frints, S.G., Jun, L., Fryns, J.P., Devriendt, K., Teulingkx, R., Van den Berghe, L., De Vos, B., Borghgraef, M., Chelly, J., Des Portes, V., et al. (2003). Inv(X)(p21.1;q22.1) in a man with mental retardation, short stature, general muscle wasting, and facial dysmorphism: clinical study and mutation analysis of the NXF5 gene. Am. J. Med. Genet. A. *119A*, 367–374.
27. Jun, L., Frints, S., Duhamel, H., Herold, A., Abad-Rodrigues, J., Dotti, C., Izaurralde, E., Marynen, P., and Froyen, G. (2001). NXF5, a novel member of the nuclear RNA export factor family, is lost in a male patient with a syndromic form of mental retardation. Curr. Biol. *11*, 1381–1391.
28. Chen, C.P., Su, Y.N., Lin, H.H., Chern, S.R., Tsai, F.J., Wu, P.C., Lee, C.C., Chen, Y.T., and Wang, W. (2011). De novo duplication of Xq22.1→q24 with a disruption of the NXF gene cluster in a mentally retarded woman with short stature and premature ovarian failure. Taiwan J. Obstet. Gynecol. *50*, 339–344.
29. Grillo, L., Reitano, S., Belfiore, G., Spalletta, A., Amata, S., Bottitta, M., Barone, C., Falco, M., Fichera, M., and Romano, C. (2010). Familial 1.1 Mb deletion in chromosome Xq22.1 associated with mental retardation and behavioural disorders in female patients. Eur. J. Med. Genet. *53*, 113–116.
30. Brunner, H.G., Nelen, M.R., van Zandvoort, P., Abeling, N.G., van Gennip, A.H., Wolters, E.C., Kuiper, M.A., Ropers, H.H., and van Oost, B.A. (1993). X-linked borderline mental retardation with prominent behavioral disturbance: phenotype, genetic localization, and evidence for disturbed monoamine metabolism. Am. J. Hum. Genet. *52*, 1032–1039.
31. Froyen, G., Van Esch, H., Bauters, M., Hollanders, K., Frints, S.G., Vermeesch, J.R., Devriendt, K., Fryns, J.P., and Marynen, P. (2007). Detection of genomic copy number changes in patients with idiopathic mental retardation by high-resolution X-array-CGH: important role for increased gene dosage of XLMR genes. Hum. Mutat. *28*, 1034–1042.
32. Lugtenberg, D., Yntema, H.G., Banning, M.J., Oudakker, A.R., Firth, H.V., Willatt, L., Raynaud, M., Kleefstra, T., Fryns, J.P., Ropers, H.H., et al. (2006). ZNF674: a new kruppel-associated box-containing zinc-finger gene involved in nonsyndromic X-linked mental retardation. Am. J. Hum. Genet. *78*, 265–278.
33. Ramaswamy, V., Castillo, M., and Bolduc, F.V. (2010). Developmental disability: duplication of zinc finger transcription factors 673 and 674. Pediatr. Neurol. *43*, 209–212.
34. Delphin, N., Hanein, S., Taie, L.F., Zanlonghi, X., Bonneau, D., Moisan, J.P., Boyle, C., Nitschke, P., Pruvost, S., Bonnefont, J.P., et al. (2012). Intellectual disability associated with retinal dystrophy in the Xp11.3 deletion syndrome: ZNF674 on trial. Guilty or innocent? Eur. J. Hum. Genet. *20*, 352–356.
35. Shoichet, S.A., Hoffmann, K., Menzel, C., Trautmann, U., Moser, B., Hoeltzenbein, M., Echenne, B., Partington, M., Van Bokhoven, H., Moraine, C., et al. (2003). Mutations in the ZNF41 gene are associated with cognitive deficits: identification of a new candidate for X-linked mental retardation. Am. J. Hum. Genet. *73*, 1341–1354.
36. Kleefstra, T., Yntema, H.G., Oudakker, A.R., Banning, M.J., Kalscheuer, V.M., Chelly, J., Moraine, C., Ropers, H.H., Fryns, J.P., Janssen, I.M., et al. (2004). Zinc finger 81 (ZNF81) mutations associated with X-linked mental retardation. J. Med. Genet. *41*, 394–399.
37. El-Hattab, A.W., Bournat, J., Eng, P.A., Wu, J.B., Walker, B.A., Stankiewicz, P., Cheung, S.W., and Brown, C.W. (2011). Microduplication of Xp11.23p11.3 with effects on cognition, behavior, and craniofacial development. Clin. Genet. *79*, 531–538.
38. Alesi, V., Bertoli, M., Barrano, G., Torres, B., Pusceddu, S., Pastorino, M., Perria, C., Nardone, A.M., Novelli, A., and Serra,

G. (2012). 335.4 kb microduplication in chromosome band Xp11.2p11.3 associated with developmental delay, growth retardation, autistic disorder and dysmorphic features. Gene *505*, 384–387.
39. Kutsche, K., Yntema, H., Brandt, A., Jantke, I., Nothwang, H.G., Orth, U., Boavida, M.G., David, D., Chelly, J., Fryns, J.P., et al. (2000). Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. Nat. Genet. *26*, 247–250.
40. Ramser, J., Abidi, F.E., Burckle, C.A., Lenski, C., Toriello, H., Wen, G., Lubs, H.A., Engert, S., Stevenson, R.E., Meindl, A., et al. (2005). A unique exonic splice enhancer mutation in a family with X-linked mental retardation and epilepsy points to a novel role of the renin receptor. Hum. Mol. Genet. *14*, 1019–1027.
41. Roll, P., Rudolf, G., Pereira, S., Royer, B., Scheffer, I.E., Massacrier, A., Valenti, M.P., Roeckel-Trevisiol, N., Jamali, S., Beclin, C., et al. (2006). SRPX2 mutations in disorders of language cortex and cognition. Hum. Mol. Genet. *15*, 1195–1207.
42. Cho, G., Bhat, S.S., Gao, J., Collins, J.S., Rogers, R.C., Simensen, R.J., Schwartz, C.E., Golden, J.A., and Srivastava, A.K. (2008). Evidence that SIZN1 is a candidate X-linked mental retardation gene. Am. J. Med. Genet. A. *146A*, 2644–2650.
43. Graham, J.M., Jr., Wheeler, P., Tackels-Horne, D., Lin, A.E., Hall, B.D., May, M., Short, K.M., Schwartz, C.E., and Cox, T.C. (2003). A new X-linked syndrome with agenesis of the corpus callosum, mental retardation, coloboma, micrognathia, and a mutation in the Alpha 4 gene at Xq13. Am. J. Med. Genet. A. *123A*, 37–44.
44. Cantagrel, V., Lossi, A.M., Boulanger, S., Depetris, D., Mattei, M.G., Gecz, J., Schwartz, C.E., Van Maldergem, L., and Villard, L. (2004). Disruption of a new X linked gene highly expressed in brain in a family with two mentally retarded males. J. Med. Genet. *41*, 736–742.
45. Lossi, A.M., Laugier-Anfossi, F., Depetris, D., Gecz, J., Gedeon, A., Kooy, F., Schwartz, C., Mattei, M.G., Croquette, M.F., and Villard, L. (2002). Abnormal expression of the KLF8 (ZNF741) gene in a female patient with an X;autosome translocation t(X;21)(p11.2;q22.3) and non-syndromic mental retardation. J. Med. Genet. *39*, 113–117.
46. Jamain, S., Quach, H., Betancur, C., Råstam, M., Colineaux, C., Gillberg, I.C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C., and Bourgeron, T.; Paris Autism Research International Sibpair Study. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat. Genet. *34*, 27–29.
47. Mansouri, M.R., Marklund, L., Gustavsson, P., Davey, E., Carlsson, B., Larsson, C., White, I., Gustavson, K.H., and Dahl, N. (2005). Loss of ZDHHC15 expression in a woman with a balanced translocation t(X;15)(q13.3;cen) and severe mental retardation. Eur. J. Hum. Genet. *13*, 970–977.
48. van der Maarel, S.M., Scholten, I.H., Huber, I., Philippe, C., Suijkerbuijk, R.F., Gilgenkrantz, S., Kere, J., Cremers, F.P., and Ropers, H.H. (1996). Cloning and characterization of DXS6673E, a candidate gene for X-linked mental retardation in Xq13.1. Hum. Mol. Genet. *5*, 887–897.
49. Collins, F.A., Murphy, D.L., Reiss, A.L., Sims, K.B., Lewis, J.G., Freund, L., Karoum, F., Zhu, D., Maumenee, I.H., and Antonarakis, S.E. (1992). Clinical, biochemical, and neuropsychiatric evaluation of a patient with a contiguous gene syndrome due to a microdeletion Xp11.3 including the Norrie disease locus and monoamine oxidase (MAOA and MAOB) genes. Am. J. Med. Genet. *42*, 127–134.
50. Brunner, H.G., Nelen, M., Breakefield, X.O., Ropers, H.H., and van Oost, B.A. (1993). Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science *262*, 578–580.
51. Voineagu, I., Huang, L., Winden, K., Lazaro, M., Haan, E., Nelson, J., McGaughran, J., Nguyen, L.S., Friend, K., Hackett, A., et al. (2012). CCDC22: a novel candidate gene for syndromic X-linked intellectual disability. Mol. Psychiatry *17*, 4–7.
52. Witham, S., Takano, K., Schwartz, C., and Alexov, E. (2011). A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. Proteins *79*, 2444–2454.
53. Houge, G., Rasmussen, I.H., and Hovland, R. (2012). Loss-of-Function CNKSR2 Mutation Is a Likely Cause of Non-Syndromic X-Linked Intellectual Disability. Mol Syndromol *2*, 60–63.
54. Honda, S., Hayashi, S., Imoto, I., Toyama, J., Okazawa, H., Nakagawa, E., Goto, Y., and Inazawa, J. (2010). Copy-number variations on the X chromosome in Japanese patients with mental retardation detected by array-based comparative genomic hybridization analysis. J. Hum. Genet. *55*, 590–599.
55. Huang, L., Jolly, L.A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., Shoubridge, C., Wieczorek, D., Tzschach, A., Cohen, M., et al. (2012). A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. Am. J. Hum. Genet. *91*, 694–702.
56. Rope, A.F., Wang, K., Evjenth, R., Xing, J., Johnston, J.J., Swensen, J.J., Johnson, W.E., Moore, B., Huff, C.D., Bird, L.M., et al. (2011). Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. Am. J. Hum. Genet. *89*, 28–43.
57. Klauck, S.M., Felder, B., Kolb-Kokocinski, A., Schuster, C., Chiocchetti, A., Schupp, I., Wellenreuther, R., Schmötzer, G., Poustka, F., Breitenbach-Koller, L., and Poustka, A. (2006). Mutations in the ribosomal protein gene RPL10 suggest a novel modulating disease mechanism for autism. Mol. Psychiatry *11*, 1073–1084.
58. Chiocchetti, A., Pakalapati, G., Duketis, E., Wiemann, S., Poustka, A., Poustka, F., and Klauck, S.M. (2011). Mutation and expression analyses of the ribosomal protein gene RPL10 in an extended German sample of patients with autism spectrum disorder. Am. J. Med. Genet. A. *155A*, 1472–1475.
59. Hagens, O., Dubos, A., Abidi, F., Barbi, G., Van Zutven, L., Hoeltzenbein, M., Tommerup, N., Moraine, C., Fryns, J.P., Chelly, J., et al. (2006). Disruptions of the novel KIAA1202 gene are associated with X-linked mental retardation. Hum. Genet. *118*, 578–590.
60. Whibley, A.C., Plagnol, V., Tarpey, P.S., Abidi, F., Fullston, T., Choma, M.K., Boucher, C.A., Shepherd, L., Willatt, L., Parkin, G., et al. (2010). Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability. Am. J. Hum. Genet. *87*, 173–188.
61. Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C., Bauters, M., Govaerts, K., Vandeleur, L., Van Esch, H., et al. (2008). Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. Am. J. Hum. Genet. *82*, 432–443.

62. Froyen, G., Belet, S., Martinez, F., Santos-Rebouças, C.B., Declercq, M., Verbeeck, J., Donckers, L., Berland, S., Mayo, S., Rosello, M., et al. (2012). Copy-number gains of HUWE1 due to replication- and recombination-based rearrangements. Am. J. Hum. Genet. *91*, 252–264.
63. Nava, C., Lamari, F., Héron, D., Mignot, C., Rastetter, A., Keren, B., Cohen, D., Faudet, A., Bouteiller, D., Gilleron, M., et al. (2012). Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. Transcult. Psychiatry *2*, e179.
64. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. *82*, 477–488.
65. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature *466*, 368–372.
66. Noor, A., Whibley, A., Marshall, C.R., Gianakopoulos, P.J., Piton, A., Carson, A.R., Orlic-Milacic, M., Lionel, A.C., Sato, D., Pinto, D., et al.; Autism Genome Project Consortium. (2010). Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. Sci. Transl. Med. *2*, 49ra68.
67. Filges, I., Röthlisberger, B., Blattner, A., Boesch, N., Demougin, P., Wenzel, F., Huber, A.R., Heinimann, K., Weber, P., and Miny, P. (2011). Deletion in Xp22.11: PTCHD1 is a candidate gene for X-linked intellectual disability with or without autism. Clin. Genet. *79*, 79–85.
68. Garcia, C.C., Blair, H.J., Seager, M., Coulthard, A., Tennant, S., Buddles, M., Curtis, A., and Goodship, J.A. (2004). Identification of a mutation in synapsin I, a synaptic vesicle protein, in a family with epilepsy. J. Med. Genet. *41*, 183–186.
69. Fassio, A., Patry, L., Congia, S., Onofri, F., Piton, A., Gauthier, J., Pozzi, D., Messa, M., Defranchi, E., Fadda, M., et al. (2011). SYN1 loss-of-function mutations in autism and partial epilepsy cause impaired synaptic function. Hum. Mol. Genet. *20*, 2297–2307.
70. Lescai, F., Bonfiglio, S., Bacchelli, C., Chanudet, E., Waters, A., Sisodiya, S.M., Kasperavičiūtė, D., Williams, J., Harold, D., Hardy, J., et al. (2012). Characterisation and validation of insertions and deletions in 173 patient exomes. PLoS ONE *7*, e51292.
71. Lenk, U., Hanke, R., and Speer, A. (1994). Carrier detection in DMD families with point mutations, using PCR-SSCP and direct sequencing. Neuromuscul. Disord. *4*, 411–418.
72. Unger, S., Mainberger, A., Spitz, C., Bähr, A., Zeschnigk, C., Zabel, B., Superti-Furga, A., and Morris-Rosendahl, D.J. (2007). Filamin A mutation is one cause of FG syndrome. Am. J. Med. Genet. A. *143A*, 1876–1879.
73. Zemni, R., Bienvenu, T., Vinet, M.C., Sefiani, A., Carrié, A., Billuart, P., McDonell, N., Couvert, P., Francis, F., Chafey, P., et al. (2000). A new gene involved in X-linked mental retardation identified by analysis of an X;2 balanced translocation. Nat. Genet. *24*, 167–170.
74. Lebel, R.R., May, M., Pouls, S., Lubs, H.A., Stevenson, R.E., and Schwartz, C.E. (2002). Non-syndromic X-linked mental retardation associated with a missense mutation (P312L) in the FGD1 gene. Clin. Genet. *61*, 139–145.
75. Liang, G.S., de Miguel, M., Gómez-Hernández, J.M., Glass, J.D., Scherer, S.S., Mintz, M., Barrio, L.C., and Fischbeck, K.H. (2005). Severe neuropathy with leaky connexin32 hemichannels. Ann. Neurol. *57*, 749–754.
76. Potluri, P., Davila, A., Ruiz-Pesini, E., Mishmar, D., O'Hearn, S., Hancock, S., Simon, M., Scheffler, I.E., Wallace, D.C., and Procaccio, V. (2009). A novel NDUFA1 mutation leads to a progressive mitochondrial complex I-specific neurodegenerative disease. Mol. Genet. Metab. *96*, 189–195.
77. Maul, B., von Bohlen und Halbach, O., Becker, A., Sterner-Kock, A., Voigt, J.P., Siems, W.E., Grecksch, G., and Walther, T. (2008). Impaired spatial memory and altered dendritic spine morphology in angiotensin II type 2 receptor-deficient mice. J. Mol. Med. *86*, 563–571.
78. Garshasbi, M., Kahrizi, K., Hosseini, M., Nouri Vahid, L., Falah, M., Hemmati, S., Hu, H., Tzschach, A., Ropers, H.H., Najmabadi, H., and Kuss, A.W. (2011). A novel nonsense mutation in TUSC3 is responsible for non-syndromic autosomal recessive mental retardation in a consanguineous Iranian family. Am. J. Med. Genet. A. *155A*, 1976–1980.
79. Li, F.Y., Lenardo, M.J., and Chaigne-Delalande, B. (2011). Loss of MAGT1 abrogates the Mg2+ flux required for T cell signaling and leads to a novel human primary immunodeficiency. Magnes. Res. *24*, S109–S114.
80. Royer-Zemmour, B., Ponsole-Lenfant, M., Gara, H., Roll, P., Lévêque, C., Massacrier, A., Ferracci, G., Cillario, J., Robaglia-Schlupp, A., Vincentelli, R., et al. (2008). Epileptic and developmental disorders of the speech cortex: ligand/receptor interaction of wild-type and mutant SRPX2 with the plasminogen activator receptor uPAR. Hum. Mol. Genet. *17*, 3617–3630.
81. Madrigal, I., Fernández-Burriel, M., Rodriguez-Revenga, L., Cabrera, J.C., Martí, M., Mur, A., and Milà, M. (2010). Xq26.2-q26.3 microduplication in two brothers with intellectual disabilities: clinical and molecular characterization. J. Hum. Genet. *55*, 822–826.
82. Madrigal, I., Rodríguez-Revenga, L., Badenas, C., Sánchez, A., Martinez, F., Fernandez, I., Fernández-Burriel, M., and Milà, M. (2007). MLPA as first screening method for the detection of microduplications and microdeletions in patients with X-linked mental retardation. Genet. Med. *9*, 117–122.
83. Ramakers, G.J., Wolfer, D., Rosenberger, G., Kuchenbecker, K., Kreienkamp, H.J., Prange-Kiel, J., Rune, G., Richter, K., Langnaese, K., Masneuf, S., et al. (2012). Dysregulation of Rho GTPases in the αPix/Arhgef6 mouse model of X-linked intellectual disability is paralleled by impaired structural and synaptic plasticity and cognitive deficits. Hum. Mol. Genet. *21*, 268–286.
84. O'Leary, R.E., Shih, J.C., Hyland, K., Kramer, N., Asher, Y.J., and Graham, J.M., Jr. (2012). De novo microdeletion of Xp11.3 exclusively encompassing the monoamine oxidase A and B genes in a male infant with episodic hypotonia: a genomics approach to personalized medicine. Eur. J. Med. Genet. *55*, 349–353.
85. Bortolato, M., Godar, S.C., Alzghoul, L., Zhang, J., Darling, R.D., Simpson, K.L., Bini, V., Chen, K., Wellman, C.L., Lin, R.C., and Shih, J.C. (2013). Monoamine oxidase A and A/B knockout mice display autistic-like features. Int. J. Neuropsychopharmacol. *16*, 869–888.
86. Haavik, J., Blau, N., and Thöny, B. (2008). Mutations in human monoamine-related neurotransmitter pathway genes. Hum. Mutat. *29*, 891–902.
87. Piton, A., Gauthier, J., Hamdan, F.F., Lafrenière, R.G., Yang, Y., Henrion, E., Laurent, S., Noreau, A., Thibodeau, P.,

Karemera, L., et al. (2011). Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. Mol. Psychiatry *16*, 867–880.
88. Turner, G., Gedeon, A., and Mulley, J. (1994). X-linked mental retardation with heterozygous expression and macrocephaly: pericentromeric gene localization. Am. J. Med. Genet. *51*, 575–580.
89. D'Arca, D., Zhao, X., Xu, W., Ramirez-Martinez, N.C., Iavarone, A., and Lasorella, A. (2010). Huwe1 ubiquitin ligase is essential to synchronize neuronal and glial differentiation in the developing cerebellum. Proc. Natl. Acad. Sci. USA *107*, 5875–5880.
90. Zhao, X., D' Arca, D., Lim, W.K., Brahmachary, M., Carro, M.S., Ludwig, T., Cardo, C.C., Guillemot, F., Aldape, K., Califano, A., et al. (2009). The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. Dev. Cell *17*, 210–221.
91. Zhao, X., Heng, J.I., Guardavaccaro, D., Jiang, R., Pagano, M., Guillemot, F., Iavarone, A., and Lasorella, A. (2008). The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. Nat. Cell Biol. *10*, 643–653.
92. Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics *25*, 1451–1452.
93. Bell, C.J., Dinwiddie, D.L., Miller, N.A., Hateley, S.L., Ganusova, E.E., Mudge, J., Langley, R.J., Zhang, L., Lee, C.C., Schilkey, F.D., et al. (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci. Transl. Med. *3*, ra4.
94. O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., et al. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science *338*, 1619–1622.
95. Bedoyan, J.K., Schaibley, V.M., Peng, W., Bai, Y., Mondal, K., Shetty, A.C., Durham, M., Micucci, J.A., Dhiraaj, A., Skidmore, J.M., et al. (2012). Disruption of RAB40AL function leads to Martin—Probst syndrome, a rare X-linked multisystem neurodevelopmental human disorder. J. Med. Genet. *49*, 332–340.
96. Kalscheuer, V.M., Iqbal, Z., Hu, H., Haas, S.A., Shaw, M., Lebrun, N., Seemanova, E., Voesenek, K., Hobson, L., Ropers, H.H., et al. (2013). RAB40AL loss-of-function mutation does not cause X-linked intellectual disability. J. Med. Genet. Published online March 25, 2013. http://jmg.bmj.com/content/49/5/332/reply#jmedgenet_el_1730.
97. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.
98. Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. Neuron *74*, 285–299.
99. Andreasen, C., Nielsen, J.B., Refsgaard, L., Holst, A.G., Christensen, A.H., Andreasen, L., Sajadieh, A., Haunsø, S., Svendsen, J.H., and Olesen, M.S. (2013). New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. Eur. J. Hum. Genet.
100. Swaminathan, G.J., Bragin, E., Chatzimichali, E.A., Corpas, M., Bevan, A.P., Wright, C.F., Carter, N.P., Hurles, M.E., and Firth, H.V. (2012). DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. Hum. Mol. Genet. *21*(R1), R37–R44.

Paper II: Efficient Strategy for the Molecular Diagnosis of Intellectual Disability using Targeted High Throughput Sequencing.

Redin C*, Gérard B, Lauer J, Herenger Y, Muller J, et al. JAMA, *submitted*. 2014 Feb.

Aim of the study: Implementation of a targeted high throughput sequencing approach of 220 genes (100 XLID genes, 120 autosomal ID genes) in a cohort of over a hundred patients with ID, in order to assess its relevance in terms of diagnostic yield. The approach would be considered as successful if leading do diagnostic yield at least similar to the array-CGH and fragile-X testing altogether (above 15%).

Contribution: Major. Performed part of the sequencing libraries, analyzed and interpreted the processed data, performed co-segregation validations, gave back reports to clinicians, wrote the paper along with Dr Piton.

Efficient Strategy for the Molecular Diagnosis of Intellectual Disability using Targeted High Throughput Sequencing

Claire Redin[1,2], Bénédicte Gérard[3], Julia Lauer[3], Yvan Herenger[3], Jean Muller[1,3], Angélique Quartier[1,2], Alice Masurel-Paulet[4], Marjolaine Willems[5], Gaétan Lesca[6], Salima El-Chehadeh[4], Stéphanie Le Gras[7], Serge Vicaire[7], Muriel Philipps[7], Michaël Dumas[7], Véronique Geoffroy[8], Claire Feger[7], Nicolas Haumesser[1,2], Yves Alembik[9], Magalie Barth[10], Dominique Bonneau[10], Estelle Colin[10], Hélène Dollfus[11], Bérénice Doray[9], Marie-Ange Delrue[12], Valérie Drouin-Garraud[13], Elisabeth Flori[9], Mélanie Fradin[14], Christine Francannet[15], Alice Goldenberg[13], Serge Lumbroso[16], Michèle Mathieu-Dramard[17], Dominique Martin-Coignard[18], Didier Lacombe[12], Gilles Morin[17], Anne Polge[16], Sylvie Sukno[19], Christel Thauvin-Robinet[4], Julien Thevenon[4], Martine Doco-Fenzy[20], David Genevieve[5], Pierre Sarda[5], Patrick Edery[6], Bertrand Isidor[21], Bernard Jost[7], Laurence Olivier-Faivre[4], Jean-Louis Mandel[1,2,3@], Amélie Piton.[1,2@]

[1]Département de Médicine translationnelle et Neurogénétique, IGBMC, CNRS UMR 7104/INSERM U964/Université de Strasbourg, Illkirch, France
[2]Chaire de Génétique Humaine, Collège de France, Illkirch, France
[3]Laboratoire de diagnostic génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
[4]Centre de Génétique et Centre de Référence Anomalies du développement et Syndromes malformatifs, Hôpital d'Enfants, CHU de Dijon, Dijon, France
[5]Département de Génétique Médicale, Centre de Référence Maladies Rares Anomalies du Développement et Syndromes Malformatifs Sud-Languedoc Roussillon, Hôpital Arnaud de Villeneuve, Montpellier, France
[6]Département de Génétique Médicale, Hospices Civils de Lyon, Bron, France
[7]Plateforme de Biopuces et Séquençage, IGBMC, CNRS UMR 7104/INSERM U964/Université de Strasbourg, Illkirch, France
[8]Plateforme de Bioinformatique de Strasbourg (BIPS), IGBMC, CNRS UMR 7104/INSERM U964/Université de Strasbourg, Illkirch, France
[9]Département de Génétique, CHU de Hautepierre, Strasbourg, France
[10]Department de Biochimie et de Génétique, CHU d'Angers, Angers, France
[11]Laboratoire de Génétique Médicale, INSERM U1112, Faculté de Médecine de Strasbourg, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
[12]CHU de Bordeaux, Génétique Médicale, Université de Bordeaux, Laboratoire MRGM, Bordeaux, France
[13]Department de Génétique Médicale, CHU de Rouen, Rouen, France
[14]Service de Génétique Médicale, Centre De Référence Anomalies du Développement, CHU de Rennes, Rennes, France
[15]Service de Génétique Médicale, Hôtel Dieu, Clermont-Ferrand, France
[16]Laboratoire de Biochimie, CHU de Nîmes, Nîmes, France
[17]Unité de Génétique Clinique, CHU d'Amiens, Amiens, France
[18]Service de Génétique, Centre Hospitalier, Le Mans, France
[19]Service de Neuropédiatrie, Hôpital Saint Vincent de Paul, Groupe Hospitalier de l'Institut Catholique Lillois, Faculté Libre de Médecine, Lille, France
[20]Service de Génétique, CHU de Reims, EA3801, Reims, France
[21]Service de Génétique Médicale, CHU de Nantes, Nantes, France
@Correspondence: amelie.piton@igbmc.fr and jlmandel@igbmc.fr

ABSTRACT

BACKGROUND: Intellectual disability (ID) is characterized by an extreme genetic heterogeneity. Several hundred genes have been associated to monogenic forms of ID, considerably complicating molecular diagnosis offers. Trio-exome sequencing was recently proposed as a diagnostic approach, yet remains costly for a general implementation.

METHODS: We report the alternative strategy of targeted high-throughput sequencing of 217 genes in which mutations had been reported in patients with ID or autism as the major clinical concern. We analyzed 106 patients (mostly males and sporadic cases) with ID of unknown etiology following array-CGH analysis and other genetic investigations.

RESULTS: We identified 26 causative mutations: 16 in X-linked genes (*ATRX*, *CUL4B*, *DMD*, *FMR1*, *HCFC1, IL1RAPL1, IQSEC2, KDM5C, MAOA, MECP2, SLC9A6, SLC16A2, PHF8*) and 10 de novo in autosomal dominant genes (*DYRK1A, GRIN1, MED13L, TCF4, RAI1, SHANK3, SLC2A1, SYNGAP1*). We also detected four possibly-causative mutations (e.g. in *NLGN3*) requiring further investigations. We present detailed reasoning for assigning causality for each mutation, and associated patients' clinical information. Some genes were hit more than once in our cohort suggesting they correspond to more frequent ID-associated conditions (*KDM5C, MECP2, DYRK1A, TCF4*). We highlight some unexpected genotype to phenotype correlations, with causative mutations being identified in genes associated to defined syndromes in patients deviating from the classic phenotype (*DMD, TCF4, MECP2*). We also bring additional supportive (*HCFC1, MED13L*) or unsupportive (*SHROOM4, SRPX2*) evidences for the implication of previous candidate genes or mutations in cognitive disorders.

CONCLUSIONS: With a diagnostic yield of 25% targeted sequencing appears relevant as a first intention test for the diagnosis of ID, but importantly will also contribute to a better understanding regarding the specific contribution of the many genes implicated in ID and autism.

Key words: intellectual disability, mutation, targeted high-throughput sequencing, autism, causative

Word count : 4688

Introduction

Intellectual disability (ID, formerly known as Mental Retardation [1]) is a common neurodevelopmental disorder reported in 1-2% of the newborns [2]. ID is defined by significant limitations in both intellectual functioning and adaptative behavior with onset before the age of 18. Different classes of ID are conventionally defined upon intelligence quotient (IQ) values (severe or profound, < 35; moderate, 35-49 and mild, 50-70). However, in routine genetic practice, clinical assessment mainly based on records of developmental history, speech acquisition and patient's autonomy is used for classification in such subcategories.

Causes of ID can be environmental, genetic, or multifactorial. Single genetic events are thought to account for a majority of cases, varying vary from large chromosomal anomalies or copy number variants (CNVs) affecting several genes to point mutations in single genes. These latter monogenic forms are characterized by an extreme genetic heterogeneity, with a hundred genes described as implicated in X-linked ID (XLID), and more associated to autosomal recessive or dominant forms. Altogether there are more than 400 genes proposed to cause ID with high penetrance when mutated [3-7], underlying a phenotypic heterogeneity of the same extent in both severity and associated symptoms. This genetic heterogeneity has long limited the diagnosis offer for patients and families, which was often restricted to fragile-X (MIM 300624) testing, array-CGH analysis, and generic metabolic tests (**Figure S1**). It may be complemented by sequencing a few genes associated to a specific syndrome evoked by patient's phenotype, yet the diagnostic yield remains low (1-2% for the recurrent fragile-X mutation; 10-15% for array-CGH and chromosomal analyses, higher in highly syndromic patients)[8 9]. A majority of patients remain therefore without molecular diagnosis, while it is of crucial importance for establishing recurrence risks and providing genetic counseling in the family. Moreover, such diagnosis often has direct consequences for the medical prognosis of patients or their optimized health care, and even (yet in still a minority of cases) can indicate specific therapeutic options.

To obviate this low diagnostic yield, we developed the simultaneous targeted-sequencing of protein-coding exons of 217 genes associated with ID or autism spectrum disorders (ASDs) as primary clinically significant

feature: 99 located on the X-chromosome, 118 on the autosomes. We report here the results of such strategy on a cohort of 106 ID-patients with or without associated autistic-like features, negative for array-CGH, fragile-X and other specific genetic analyses. A causal mutation was detected in 25% of these patients, regardless the severity of their cognitive impairment. We illustrate cases in which the molecular diagnosis was immediately established, as well as other more complex situations. This highlights the challenge of interpreting variants generated by NGS-technologies, already from targeted approaches restricted to a few hundred genes. This work demonstrates that a targeted sequencing approach is highly efficient for the diagnosis of ID, but also allows refining the clinical spectrum associated with mutations in certain genes, and confirming or questioning the involvement of other genes in cognitive disorders.

Methods

Cohort of patients

DNA samples from 106 patients were addressed for testing through clinical geneticists from 16 public hospitals in France. ID patients with polymalformative syndromes or suspected mitochondrial/peroxisomal disorders were excluded. All patients had been previously excluded for the recurrent fragile-X mutation, for pathogenic CNVs via array-CGH analysis, and for several other specific genetic investigations (on average: two genes tested per patient, **Table 1**). Clinical data was recorded following a standardized clinical record highlighting prenatal history, developmental history, neurological and behavioral disorders. ID severity was assessed as explained previously by geneticists upon clinical evaluation. This study was approved by the local Ethics Committee of the Strasbourg University Hospital (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale, CCPPRB). For all patients a written informed consent for genetic testing was obtained from their legal representative.

Targeted genes and capture design

The 217 selected genes include 99 genes associated to XLID and 118 genes located on autosomes, implicated in dominant (45), recessive (66) or complex (7) forms of ID (**Table S8**, further justification on gene selection is given in **Supplementary Methods)**. We targeted all protein-coding exons of these genes, including 20-bp of intronic flanking sequences. Corresponding 120-bp RNA baits were designed using SureDesign (https://earray.chem.agilent.com/suredesign/).

Library preparation and sequencing

DNA samples were extracted from peripheral blood or saliva. Sequencing libraries were prepared as described previously [10], performing individual in-solution SureSelect capture reaction for each DNA sample (Agilent, Santa Clara, CA, USA). Paired-end sequencing (2x101-bp) was performed on an Illumina HiSeq 2000/2500, multiplexing up to 16 samples per sequencing lane.

Bioinformatic Pipeline and Variant Ranking

Read mapping, variant calling and annotation was performed as described previously [10]. Detected variants (short indels and single nucleotide variants, SNVs) are ranked by VaRank (an in-house developed script), which incorporates the annotations retrieved by alamut-HT (putative effect on the protein, conservation scores, splice site predictions, allelic frequency in the 106 patients and in control cohorts such as Exome Variant Server: EVS, or 1000genomes). Candidate variants were selected when: 1/ present in EVS in <50 controls (for autosomal recessive genes), <2 controls (for autosomal dominant genes), or <2 males (for X-linked genes) and 2/ present in the cohort in ≤2 patients [11]. Remaining variants predicted as potentially pathogenic and fitting with the mode of inheritance associated to the affected gene were tested for validation (**Table S9**).

CNVs detection pipeline

Putative heterozygous/homozygous/hemizygous structural or CNVs were highlighted using the previously described method based on a depth-of-coverage comparison between the index sample and eight other random samples from the same sequencing lane [10]. For the X-chromosome, coverage was normalized according to the number of X-chromosomes of the patient.

Mutation validation

All candidate mutations were validated by Sanger sequencing and co-segregation analyses were performed as extensively as possible. Putative splicing mutations were confirmed either using a minigene *in vitro* assay with the SPL3B plasmid as described previously [12], or using patients' fibroblasts or blood RNA when available. For apparent *de novo* variants, pedigree concordance was checked using polymorphic microsatellite markers (PowerPlex® 16HS System, Promega). Mutations were considered as certainly-causative when no doubt remained regarding their pathogenicity and a diagnosis could be established without any error risk. Such mutations co-segregated with the disease status in the family and were either: truncating mutations or missense mutations that had been previously convincingly published or that we confirmed with functional analyses. Mutations were considered as potentially-causative when they co-segregate with ID and were predicted to be damaging, but further functional studies were needed to prove unambiguously their pathogenicity.

Results

High quality sequencing data ensures low rates of false-positive/negative calls for SNVs, indels and CNVs

Our strategy allowed generating a high quality sequencing dataset, with a mean depth-of-coverage of 350X and an average per patient of 97.7% of targeted regions being well-covered (> 40X; **Table S1**). Such coverage ensures a sensitivity of 99-99.9% of detecting SNVs and indels at any allelic state (Illumina Technical Note). We further assessed the sensitivity of SNV detection by comparing allelic states of single nucleotide polymorphism (SNPs) detected by SNP-array (Affymetrix® SNP Array 6.0) with the corresponding sites located in the targeted sequencing data of two patients, and found no false-negatives (452 SNPs analyzed). Interestingly, among all 6 SNPs that show discrepant allelic states between both methods, Sanger sequencing results were always in favor of targeted sequencing data suggesting a much higher sensitivity (data not shown). No false positive was detected out of the 80 candidate variants located in well-covered regions that were tested for confirmation by Sanger sequencing.

This high sequencing depth also ensured reliable CNVs calling. All CNVs detected by our pipeline were validated by Sanger sequencing, qPCR and/or confirmed retrospectively when looking at array-CGH data (**Table S2**). Some were not initially mentioned in the array-CGH report because covered by only a few SNP probes (≤2 deleted probes, under the detection threshold detection).

Very few regions (a total of 3.9-kb, only 1.8-kb being protein-coding) appear consistently poorly covered (coverage <40X in >90% of the samples; **Table S3**). Those are mainly first exons or highly GC-rich regions that are a well-known burden in such capture strategy.

Cohort description and diagnostic yield

Patients harbored various degrees of cognitive impairment, albeit with a higher proportion of moderate or severe forms (46% and 42% respectively; **Table 1)**. The cohort was highly enriched in males. Among male probands most cases were sporadic, the remaining had familial history of cognitive impairment mainly evocative of an X-linked mode of transmission (**Table 1**). We detected certainly-causative mutations in 26/106 patients (**Table 2**; **Figures S2-19**), leading to an overall diagnostic yield of 25% for the entire cohort (from 23% for sporadic cases to 31% for familial cases). Unexpectedly, the diagnostic yield appears unrelated to the severity of ID in patients (**Table 1**).

Sixteen mutations are located in genes of the X-chromosome: 14 point mutations (in *ATRX*, *CUL4B*, *DMD*, *HCFC1*, *IL1RAPL1, IQSEC2*, *KDM5C*x2*, MECP2*x2, *MAOA*, *PHF8*, *SLC9A6*, *SLC16A2*), as well as two larger pathogenic events (one hemizygous complex rearrangement in *MECP2,* one hemizygous exon deletion in *FMR1*; **Figures S17** and **S13)**. We identified 10 *de novo* point mutations in genes involved in autosomal dominant/haploinsufficient forms of ID (in *DYRK1A*x2, *GRIN1, MED13L*, *RAI1*, *SHANK3*, *SYNGAP1*, *SLC2A1*, *TCF4*x2). In four other patients we identified potentially-causative mutations (in *NLGN3*, *PQBP1*, *SLC2A1* and *TCF4;* **Figures S20, S21, S7, S9** respectively), whose implication in cognitive impairment has to be further confirmed. Finally, missense variants that appeared at first likely to be pathogenic, notably based on the very high evolutionary conservation of the affected residues or on previous publications, appeared excluded as causal after further segregation analysis (in *FLNA*, *FMR1*, *HUWE1* or *MECP2*, see **Figure S22**)

Atypical type of mutations

Among the 26 certainly-causative mutations identified, some were surprising by the nature of the mutation itself. In a boy with severe encephalopathy, epilepsy, hypotonia and microcephaly we detected a highly complex rearrangement in exon #4 of *MECP2*, involving a 139-bp deletion flanked by the insertion of two sequences in inverted orientation derived from intron #2 still keeping the reading frame downstream of the rearrangement. Such event is inherited from the proband's mother who presents with speech delay and dyslexia (**Figure S17**). This exon is known to be the one accumulating most mutations in patients, especially the 3' half, which is a recombination hotspot and the target of several deletions/duplications and less frequently inversions [13-15].

We report here for the first time an intragenic deletion affecting *FMR1* outside of the promoter/exon #1 region that is the target of the fragile-X syndrome CGG-expansion. Very few point mutations have been reported in coding regions. We identified full deletion of the last exon of *FMR1* in one patient and his two affected brothers presenting with uneven clinical features of fragile-X syndrome (**Figure S13**).

We also identified a patient carrying a maternally-inherited 10-bp deletion causing a frameshift in exon #7 of *IL1RAPL1*, while unexpectedly his affected brother bears a de novo deletion of the full exon highlighting that affected relatives may carry distinct mutations. We propose that small 10-bp deletion may have created a sequence conformation favoring further instability and leading to the larger deletion observed in the second brother (**Figure S14**). Indeed, a large proportion of *IL1RAPL1* causative mutations are intragenic exon deletions or pericentric inversions [16], supporting that *IL1RAPL1* is highly susceptible to recombination events.

Lastly, we identified a de novo missense mutation in *GRIN1* in a male proband with severe ID, hypotonia, feeding disorders and very poor speech but no epilepsy, a phenotype similar to that associated to *GRIN1* missense mutations in some patients [17 18]. A maternal uncle presented with similar features except for the poor speech and hypotonia (**Figure S3**), strongly suggesting an associated X-linked mode of inheritance. Unlike expected, a de novo missense mutation was identified further highlighting the prevalence of phenocopies in cognitive disorders.

240 *Genotype-phenotype correlations: from expected to unexpected*

241 The majority of certainly-causative mutations were identified in patients whose clinical phenotype
242 was retrospectively consistent with previous reports **(Table 2).** For instance, the proband carrying a
243 truncating mutation in *IQSEC2* presents with severe ID, no speech, motor developmental delay, severe
244 epilepsy, strabismus and autistic features (**Figure S15**), which matches the recently proposed clinical
245 spectrum associated to mutations in this gene even though initially reported in non-syndromic patients [19
246 20]. Likewise, *DYRK1A* was originally found disrupted by translocations or deleted in several patients with
247 ID and microcephaly [21-23]. More recently, truncating mutations in this gene were shown to cause
248 syndromic ID, characterized by primary microcephaly, growth retardation, developmental delay, facial
249 dysmorphic traits, seizures and major feeding difficulties, with or without associated autism [24-28]. We
250 report here two novel *de novo* truncating mutations in patients with similar clinical features but no epilepsy
251 (**Figure S2**). Nonetheless in a few other cases (e.g. mutations in *RAI1* or *MECP2*) the probands lacked some
252 clinical features, thus corresponding diagnosis of Smith-Magenis (MIM 18229) or Rett (MIM 312750)
253 syndrome was not evoked by experienced clinical geneticists (**Figures S5** and **S17**). For instance, among the
254 three patients with *MECP2* mutations, two present with classical phenotype (severe encephalopathy and Rett
255 syndrome respectively) while the third female proband presents with non-classical Rett phenotype (no
256 regression episode, no hand-flapping features).

257 A few detected mutations were unexpected because detected in patients whose phenotype did not
258 match previous descriptions. For instance, mutations in *TCF4* mainly cause Pitt-Hopkins syndrome (PHS,
259 MIM 610954) characterized by severe motor retardation, absence of speech, characteristic dysmorphic traits,
260 autistic features, intestinal problems and hyperventilation [29]. We here describe truncating *TCF4* mutations
261 in two patients, one with clinical manifestations highly suggestive of PHS, the other with less syndromic
262 manifestations and no dysmorphic traits (**Figure S9**). *TCF4* mutations were already reported in patients with
263 non-syndromic ID suggesting that such mutations were likely to be underdiagnosed [30].
264 Another patient and his affected brother both carry a distal frameshift mutation in *DMD* affecting the major
265 muscle transcript encoding the dystrophin protein associated to Duchenne or Becker muscular dystrophy
266 (DMD, MIM 310200; BMD, MIM 300376), and the brain-specific isoform Dp71. Index case presents with
267 moderate ID, psychomotor retardation, no speech, behavioral disorders, dysmorphic traits but strikingly no

muscular phenotype (**Figure S12**). His brother harbors a milder phenotype with additional hypotonia and cerebellar dysplasia. Both harbor borderline-high CPK-levels. The association of cognitive impairment with DMD/BMD has been extensively reported and correlated to truncating mutations affecting Dp71, yet never in the absence of a muscular phenotype [31-36]. Our findings extend the recent report of a large family with affected males carrying an in-frame single aminoacid deletion associated to mild ID and no muscular phenotype [37].

Confirmation of candidate genes for cognitive disorders

Some selected genes were only candidate ID or ASD-genes at the time of the design, with single pieces of evidence in the literature. The identification of additional mutations in patients with similar phenotype definitively confirms their implication in cognitive disorders.

We reported a damaging missense affecting the function of the Monoamine Oxidase A enzyme (MAOA) [38 39], which replicated for the first time in 20 years the implication of *MAOA* in autism/ID associated to significant behavioral disorders [38 39]. We also identified a probably-pathogenic missense variant in *NLGN3* in a male and his cousin, both presenting with ID and autism (**Figure S20).** *In silico* predictions, high conservation of the mutated residue across all neuroligin paralogs, and familial analysis are altogether in favor of a pathogenicity of this missense change. A definitive functional effect of this missense still needs to be demonstrated to clearly establish the diagnosis. The implication of this gene was never replicated since the initial publication [40], albeit screened in several cohorts with comparable phenotypes [41-47].

Similarly, we identified a novel truncating point mutation in *MED13L* confirming the implication of this gene in ID (**Figure S4**). Disruption of *MED13L* was initially associated with transposition of the great arteries (TGA), associated to ID in a single case with a chromosomal translocation [48]. A homozygous missense mutation was then identified in two siblings from a consanguineous family presenting with non-syndromic ID, suggesting an implication of the gene in autosomal recessive forms of ID justifying its selection in our panel [5]. A total of five patients were more recently described with de novo intragenic CNVs or point mutations affecting *MED13L,* delineating a recognizable *MED13L*-haploinsufficiency syndrome characterized by hypotonia, moderate ID, uneven cardiac defects, facial hypotonia and dysmorphic

traits [49 50]. Our patient with the *MED13L* mutation presents with concordant phenotype such as an open mouth appearance and muscular hypotonia but no cardiac defects. Due to the initially proposed autosomal recessive mode of inheritance associated to ID, at first we did not consider this heterozygous truncating mutation as causative, as it may also have been the case for a heterozygous splicing mutation identified earlier in a large-scale exome sequencing study in one male with ASD [25]. Altogether, these findings suggest that ID associated to *MED13L*-haploinsufficiency syndrome is a rather frequent condition [5 48].

Ambiguous mode of inheritance in ID-associated genes: the example of DEAF1

As for *MED13L*, the mode of inheritance associated to some genes is ambiguously described in literature. We identified an heterozygous variant affecting splicing in *DEAF1* inherited from the asymptomatic mother in a patient presenting with severe ID, developmental delay, poor speech, pain resistance, dysmorphic features and aggressive behavior (**Figure 1**), while the gene had been proposed as associated to autosomal dominant forms of ID [7 51]. The recent report of two additional individuals carrying de novo missense mutations narrowed the associated phenotype to moderate/severe ID, speech impairment, behavioral problems, high pain threshold, dysmorphic features and abnormal walking pattern, hence highly similar to the one of our proband [52]. Although in vitro validation studies suggest that the reported missense variants lead to an impaired function of DEAF1, the authors concluded that they presumably act as dominant-negatives incapacitating both normal and mutant proteins since truncating variants had been observed in asymptomatic individuals [52]. In parallel, a homozygous missense mutation clustering in the same SAND-domain with all three de novo missense mutations was reported in members of a consanguineous family presenting with ID, microcephaly and white matter abnormalities therefore suggesting a possible autosomal recessive mode of inheritance [53]. The pathogenic mechanism associated to *DEAF1* mutations is therefore unclear. Due to the highly similar clinical features of the herein reported proband and of probands carrying de novo missense mutations, the splice variant detected here may contribute to the phenotype of our patient, possibly through a recessive mode of inheritance (*i.e.* acting in *trans* with another heterozygous variant) since haploinsufficiency appears tolerated in healthy individuals. Altogether those findings either suppose a similar phenotype for autosomal dominant and autosomal recessive mode of inheritance associated to *DEAF1* mutations, or a universal autosomal recessive mode of

inheritance with a second variant that has not been yet identified, alike what was finally proven for TAR syndrome for instance [54].

Unsupportive evidences for proposed ID-associated mutations and genes

As already warned, the identification of previously-reported mutations described as pathogenic should be considered with caution as there are a consequent proportion of "false-positive" mutations [10]. We here question the causative effect of a few previously proposed mutations (**Table S4**). For instance, one missense variant identified in *FLNA* (c.3872C>T , p.Pro1291Leu) was previously reported in a patient with FG syndrome (MIM 300321) [55]. This variant was identified here in a patient with a different phenotype (**Figure S22**) and is also reported in one male in EVS (presumably not presenting with cognitive disorders), raising doubt about its pathogenicity.

The identification of non-segregating truncating variants (*i.e.* detected both in patients and healthy relatives) can also challenge the implication of genes in X-linked and autosomal dominant forms of ID. In one family we identified a frameshift variant in *SRPX2* in a male proband. It is most likely inherited from the deceased asymptomatic maternal grandfather since also detected in the mother and in three maternal aunts yet absent from the maternal grandmother. Despite recent functional evidences regarding the role of SRPX2 in brain development [56 57], its definitive implication in cognitive disorders has already been questioned following the presence of the initially proposed mutations in EVS, and subsequently to the identification of a missense mutation in *GRIN2A* co-segregating with the epileptic status in the initial *SRPX2* family [17]. In another family, we identified a nonsense variant in *SHROOM4* in a male proband but also in his unaffected brothers, which also further defies its implication in X-linked cognitive disorders (**Table S5, Figure 2**) [11].

Patients carrying probably-pathogenic variants in two ID-associated genes

In three unrelated patients, we identified candidate variants in two separate genes, requiring the evaluation of different scenarios: a) one single contributor while the second variant is innocuous, b) one major contributor while the second variant acts as a modifier c) both variants are implicated in the phenotype and have a synergistic effect.

In one male proband, we identified a maternally-inherited splice variant leading to a frameshift in the X-linked gene *PHF8*, together with a de novo truncating variant in *DOCK8* (**Figure S18)**. Considering both the weak evidences implicating *DOCK8* in autosomal dominant ID (two probands reported with translocations disrupting the gene), the clinical features consistent with a *PHF8* mutation (even in absence of cleft lip/palate) and the X-inactivation bias identified in the mother, the *PHF8* variant alone is most likely to be responsible for the phenotype leaving the *DOCK8* variant as probably innocuous [58 59].

In another male proband, two possibly-causative missense variants were identified in the XLID genes *ATRX* and *HCFC1* (**Figure 3**). Patient's phenotype was not evocative of an *ATRX* mutation (no dysmorphic traits, no urogenital abnormalities, absence of Heinz bodies), but perfectly matched the recent description of cobalamin-X metabolic disorder (*cblX*, MIM 309541; [60]). We concluded to a causative effect of the *HCFC1* mutation (being one of the *cblX* recurrent mutations), but cannot exclude a contributor effect of the *ATRX* variant based on current co-segregation data.

In the third family, the proband carries one distal truncating variant in *SLC2A1* and one missense variant in *ANKRD11*, each inherited from one asymptomatic parent. Mutations in *SLC2A1* have already been associated to GLUT-1 deficiency (MIM 606777/612126) with incomplete penetrance [61]. The proband presents with evocative symptoms of both *SLC2A1* and *ANKRD11* mutations (major hypotonia along with skeletal abnormalities) that might suggest a synergistic contribution of both variants to the phenotype (**Figure S7)**. No conclusion could be unambiguously drawn in this case, but such a di- or oligogenic mode of inheritance has already been proposed in neurodevelopmental disorders [62-64].

Discussion

Targeted sequencing of 217 genes in a cohort of 106 patients with unknown genetic etiology of ID led to a positive diagnostic yield of 25%. A majority of causative mutations identified are located in XLID genes, which can be partially explained by our male-enriched cohort. When excluding familial cases, X-linked mutations are found in 7/73 sporadic male cases, matching the proposed figure that 10% of males with sporadic ID carry mutations in X-linked genes [65]. Combining the results of recent trio-exome analyses leads to similar proportions (X-linked causative mutations in 7/66 males with sporadic ID; [3 51]). The

detection rate of X-linked mutations in our cohort is - as expected - higher in males with familial history of ID, but is also significant in females (2/10, both with *MECP2* mutations). Also unsurprisingly, mutations in autosomal dominant/haploinsufficient ID genes are mostly found within sporadic cases.

Although we did not expect patients to harbor mutations in the same genes because of the small size of the cohort and the extensive genetic heterogeneity of ID, we had a few genes hit more than once: *MECP2* (despite being screened in 11% of patients prior to inclusion), *KDM5C, DYRK1A* and *TCF4*. Our results, intersected with the ones from other studies, highlight mutations within common genes (*MECP2, CUL4B, IL1RAPL1*, *IQSEC2, KDM5C*, *SLC9A6, SLC16A2, DYRK1A*, *SLC2A1*, *SYNGAP1* or *TCF4*) suggesting that they are more frequently mutated in ID patients [3 7 17 25-27 51 66 67]. If mutations in such ID genes are confirmed to account for a substantial number of patients, introducing for diagnosis massive multiplexed re-sequencing of such genes in large cohorts shall be considered [26].

The extensive genetic and phenotypic heterogeneity of ID is the major hindrance for obtaining a precise molecular diagnosis. Direct sequencing strategies consisting in sequentially screening candidate genes are being replaced in diagnostic laboratories by more high-throughput NGS-based strategies: multiplex targeted sequencing of a few genes in large cohorts, selective targeted sequencing of up to several hundred genes, exome sequencing, and full-genome sequencing. Exome or full-genome strategies are more exhaustive alternatives and very attractive issues in the field of genetic diagnosis. Their universal approaches, whatever the clinical features, enhance the technical management of the workflow. However, the actual coverage proposed with the exome or full-genome strategies is still frequently insufficient which may lead to missed mutations [26]. Also, the finding of a putative mutation in a novel gene never associated to a given pathology marks the start of a research endeavor to validate the finding but is not *per se* a diagnostic result.

Targeted sequencing appears more appealing for well-defined pathologies in which most implicated genes have been uncovered or in clinically-homogeneous entities (*i.e.* Bardet-Biedl and related ciliopathies [10], retinal dystrophies [68], hearing loss [69], …). We show here that it is also a powerful alternative for diagnostic purposes in ID (**Table S6**). The positive diagnostic yield in our cohort of 106 patients is of 25% overall and 21% for sporadic cases, which is similar to what was reported with the trio-exome strategy (32 highly-likely causative mutations in a total of 151 patients, 21%; **Table S7**; [3 51]). Most of these mutations

406 identified by exome sequencing affect genes included in our panel, and would thus have been detected with
407 our strategy. If restricting the findings of the exome studies to genes of our panel, the resulting diagnostic
408 yield on the same cohorts would be of 16% (24/151) which is lesser than the one reported in this study
409 (**Table S7)**. Our high diagnostic yield is not due to an overestimation regarding the pathogenicity of the
410 identified variants as we were highly stringent regarding the classification of certainly-causative mutations
411 (*e.g.* we did not include all probably-damaging missense variants located in XLID genes that were detected
412 in males). The high depth-of-coverage and the smaller portion of poorly-covered regions achieved with our
413 strategy ensure a high sensitivity and specificity of detecting pathogenic events in the regions of interest
414 (SNVs, indels, but also larger exon deletions). Also, the relative ease of variant analysis and smaller number
415 of follow-up studies for candidate variants may contribute to this significant yield. As fewer candidate
416 variants are identified per patient with this approach, they can also be analyzed more thoroughly (putative
417 effect on splicing for variants not affecting canonical splice sites, predictive impact on the protein through
418 structural modeling, etc).
419 The limited number of sequenced genes with the targeted approach - restricted to those involved in cognitive
420 disorders - also avoids the controversial issues raised by incidental findings. Nonetheless, the targeted
421 sequencing approach will miss newly identified genes and theoretically prevents data re-analysis (*i.e.*
422 incorporating novel findings regarding genetics of ID). The major value of targeted sequencing for ID is that
423 it should allow the application of such test to a much higher proportion of patients awaiting molecular
424 diagnosis given the significantly lower cost of sequencing, but especially of data analysis, storage and
425 interpretation. It should thus generate much more promptly large amounts of data regarding the spectrum of
426 mutations and phenotypes associated to the many genes implicated in ID, and thus considerably increase our
427 knowledge on the specific condition associated with each gene.
428 The substantial proportion of patients that remain without molecular diagnosis with either strategy
429 raises several issues: whether a large fraction of ID-associated genes remain to be discovered, whether many
430 mutations are missed because located in non-coding regions, or lastly whether more complex genetic
431 scenarios are implicated such as variants with reduced penetrance, oligogenic and/or multifactorial modes of
432 inheritance [62-64]. Particularly, many disorders associated with autosomal dominant inheritance are
433 associated with incomplete penetrance and high intra-familial phenotypic variability, although such scenario

is mostly excluded from trio-studies that focus on de novo mutations. Full-genome analyses of large cohorts of patients with still unexplained ID may shed some light regarding the contribution of each of those hypotheses.

Acknowledgements

We thank the students who were involved in this project: Sébastien Kirsch, Audrey Creppy, Inès Bekkour and Grace Gan. We thank Nadège Calmels, Valérie Biancalana, Elsa Nourisson and all other members of the Genetic Diagnostic Laboratory of the Nouvel Hopital Civil (Strasbourg) for their help with patients' DNA sample selection and preparation. We thank Cecile Pizot for the development of VaRank. We thank Damien Sanlaville, Christine Coubes, Delphine Héron, Sophie Naudion, James Lespinasse, and Marie-Line Bichon for their contribution to the recruitment of patients, and all medical interns or genetic counselors who participated to this project. Finally, we warmly thank all patients and their families for their implication in this study.

REFERENCES

1. Salvador-Carulla L, Bertelli M. 'Mental retardation' or 'intellectual disability': time for a conceptual change. Psychopathology 2008;**41**(1):10-6 doi: 000109950 [pii]

10.1159/000109950[published Online First: Epub Date]|.

2. Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. Ment Retard Dev Disabil Res Rev 2002;**8**(3):117-34 doi: 10.1002/mrdd.10031[published Online First: Epub Date]|.

3. de Ligt J, Willemsen MH, van Bon BW, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 2012;**367**(20):1921-9 doi: 10.1056/NEJMoa1206524[published Online First: Epub Date]|.

4. Lubs HA, Stevenson RE, Schwartz CE. Fragile X and X-linked intellectual disability: four decades of discovery. Am J Hum Genet 2012;**90**(4):579-90 doi: S0002-9297(12)00103-6 [pii]

10.1016/j.ajhg.2012.02.018[published Online First: Epub Date]|.

5. Najmabadi H, Hu H, Garshasbi M, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 2011;**478**(7367):57-63 doi: nature10423 [pii]

10.1038/nature10423[published Online First: Epub Date]|.

6. Ropers HH. Genetics of early onset cognitive impairment. Annu Rev Genomics Hum Genet 2010;**11**:161-87 doi: 10.1146/annurev-genom-082509-141640[published Online First: Epub Date]|.

7. Vissers LE, de Ligt J, Gilissen C, et al. A de novo paradigm for mental retardation. Nat Genet 2010;**42**(12):1109-12 doi: ng.712 [pii]

10.1038/ng.712[published Online First: Epub Date]|.

8. Biancalana V, Beldjord C, Taillandier A, et al. Five years of molecular diagnosis of Fragile X syndrome (1997-2001): a collaborative study reporting 95% of the activity in France. Am J Med Genet A 2004;**129A**(3):218-24 doi: 10.1002/ajmg.a.30237[published Online First: Epub Date]|.

9. van Karnebeek CD, Jansweijer MC, Leenders AG, et al. Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness. Eur J Hum Genet 2005;**13**(1):6-25 doi: 5201279 [pii]

10.1038/sj.ejhg.5201279[published Online First: Epub Date]|.

10. Redin C, Le Gras S, Mhamdi O, et al. Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alstrom syndromes. Journal of medical genetics 2012;**49**(8):502-12 doi: 10.1136/jmedgenet-2012-100875[published Online First: Epub Date]|.

11. Piton A, Redin C, Mandel JL. XLID-Causing Mutations and Associated Genes Challenged in Light of Data From Large-Scale Human Exome Sequencing. Am J Hum Genet 2013;**93**(2):368-83 doi: S0002-9297(13)00282-6 [pii]

10.1016/j.ajhg.2013.06.013[published Online First: Epub Date]|.

12. Piton A, Jouan L, Rochefort D, et al. Analysis of the effects of rare variants on splicing identifies alterations in GABAA receptor genes in autism spectrum disorder individuals. European journal of human genetics : EJHG 2013;**21**(7):749-56 doi: 10.1038/ejhg.2012.243[published Online First: Epub Date]|.
13. Archer HL, Whatley SD, Evans JC, et al. Gross rearrangements of the MECP2 gene are found in both classical and atypical Rett syndrome patients. J Med Genet 2006;**43**(5):451-6 doi: jmg.2005.033464 [pii]

10.1136/jmg.2005.033464[published Online First: Epub Date]|.
14. Lebo RV, Ikuta T, Milunsky JM, et al. Rett syndrome from quintuple and triple deletions within the MECP2 deletion hotspot region. Clin Genet 2001;**59**(6):406-17 doi: cge590605 [pii][published Online First: Epub Date]|.
15. Ravn K, Nielsen JB, Skjeldal OH, et al. Large genomic rearrangements in MECP2. Hum Mutat 2005;**25**(3):324 doi: 10.1002/humu.9320[published Online First: Epub Date]|.
16. Behnecke A, Hinderhofer K, Bartsch O, et al. Intragenic deletions of IL1RAPL1: Report of two cases and review of the literature. Am J Med Genet A 2011;**155A**(2):372-9 doi: 10.1002/ajmg.a.33656[published Online First: Epub Date]|.
17. Epi KC, Epilepsy Phenome/Genome P, Allen AS, et al. De novo mutations in epileptic encephalopathies. Nature 2013;**501**(7466):217-21 doi: 10.1038/nature12439[published Online First: Epub Date]|.
18. Hamdan FF, Gauthier J, Araki Y, et al. Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. Am J Hum Genet 2011;**88**(3):306-16 doi: 10.1016/j.ajhg.2011.02.001[published Online First: Epub Date]|.
19. Shoubridge C, Tarpey PS, Abidi F, et al. Mutations in the guanine nucleotide exchange factor gene IQSEC2 cause nonsyndromic intellectual disability. Nat Genet 2010;**42**(6):486-8 doi: 10.1038/ng.588[published Online First: Epub Date]|.
20. Tran Mau-Them F, Willems M, Albrecht B, et al. Expanding the phenotype of IQSEC2 mutations: truncating mutations in severe intellectual disability. Eur J Hum Genet 2013 doi: ejhg2013113 [pii]

10.1038/ejhg.2013.113[published Online First: Epub Date]|.
21. Moller RS, Kubart S, Hoeltzenbein M, et al. Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. Am J Hum Genet 2008;**82**(5):1165-70 doi: 10.1016/j.ajhg.2008.03.001[published Online First: Epub Date]|.
22. van Bon BW, Hoischen A, Hehir-Kwa J, et al. Intragenic deletion in DYRK1A leads to mental retardation and primary microcephaly. Clin Genet 2011;**79**(3):296-9 doi: 10.1111/j.1399-0004.2010.01544.x[published Online First: Epub Date]|.
23. Yamamoto T, Shimojima K, Nishizawa T, et al. Clinical manifestations of the deletion of Down syndrome critical region including DYRK1A and KCNJ6. Am J Med Genet A 2011;**155A**(1):113-9 doi: 10.1002/ajmg.a.33735[published Online First: Epub Date]|.
24. Courcet JB, Faivre L, Malzac P, et al. The DYRK1A gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. J Med Genet 2012;**49**(12):731-6 doi: jmedgenet-2012-101251 [pii]

10.1136/jmedgenet-2012-101251[published Online First: Epub Date]|.
25. Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. Neuron 2012;**74**(2):285-99 doi: S0896-6273(12)00340-6 [pii]

10.1016/j.neuron.2012.04.009[published Online First: Epub Date]|.
26. O'Roak BJ, Vives L, Fu W, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 2012;**338**(6114):1619-22 doi: 10.1126/science.1227764[published Online First: Epub Date]|.
27. O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012;**485**(7397):246-50 doi: 10.1038/nature10989[published Online First: Epub Date]|.

559 28. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian
560 disorders. N Engl J Med 2013;**369**(16):1502-11 doi: 10.1056/NEJMoa1306555[published Online
561 First: Epub Date]|.
562 29. Amiel J, Rio M, de Pontual L, et al. Mutations in TCF4, encoding a class I basic helix-loop-helix
563 transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy
564 associated with autonomic dysfunction. Am J Hum Genet 2007;**80**(5):988-93 doi: S0002-
565 9297(07)60955-0 [pii]

566 10.1086/515582[published Online First: Epub Date]|.
567 30. Hamdan FF, Daoud H, Patry L, et al. Parent-child exome sequencing identifies a de novo truncating
568 mutation in TCF4 in non-syndromic intellectual disability. Clin Genet 2013;**83**(2):198-200 doi:
569 10.1111/j.1399-0004.2012.01890.x[published Online First: Epub Date]|.
570 31. Bushby KM, Appleton R, Anderson LV, et al. Deletion status and intellectual impairment in Duchenne
571 muscular dystrophy. Dev Med Child Neurol 1995;**37**(3):260-9
572 32. Daoud F, Angeard N, Demerre B, et al. Analysis of Dp71 contribution in the severity of mental
573 retardation through comparison of Duchenne and Becker patients differing by mutation
574 consequences on Dp71 expression. Human molecular genetics 2009;**18**(20):3779-94 doi: ddp320
575 [pii]

576 10.1093/hmg/ddp320[published Online First: Epub Date]|.
577 33. Emery AE, Skinner R, Holloway S. A study of possible heterogeneity in Duchenne muscular dystrophy.
578 Clin Genet 1979;**15**(5):444-9
579 34. Kerr TP, Sewry CA, Robb SA, et al. Long mutant dystrophins and variable phenotypes: evasion of
580 nonsense-mediated decay? Hum Genet 2001;**109**(4):402-7 doi: 10.1007/s004390100598[published
581 Online First: Epub Date]|.
582 35. Taylor PJ, Betts GA, Maroulis S, et al. Dystrophin gene mutation location and the risk of cognitive
583 impairment in Duchenne muscular dystrophy. PLoS One;**5**(1):e8803 doi:
584 10.1371/journal.pone.0008803[published Online First: Epub Date]|.
585 36. Zellweger H, Niedermeyer E. Central nervous system manifestations in childhood muscular dystrophy
586 (CMD). I. Psychometric and electroencephalographic findings. Ann Paediatr 1965;**205**(1):25-42
587 37. de Brouwer AP, Nabuurs SB, Verhaart IE, et al. A 3-base pair deletion, c.9711_9713del, in DMD results in
588 intellectual disability without muscular dystrophy. European journal of human genetics : EJHG
589 2014;**22**(4):480-5 doi: 10.1038/ejhg.2013.169[published Online First: Epub Date]|.
590 38. Brunner HG, Nelen M, Breakefield XO, et al. Abnormal behavior associated with a point mutation in the
591 structural gene for monoamine oxidase A. Science 1993;**262**(5133):578-80
592 39. Piton A, Poquet H, Redin C, et al. 20 ans apres: a second mutation in MAOA identified by targeted high-
593 throughput sequencing in a family with altered behavior and cognition. Eur J Hum Genet 2013 doi:
594 ejhg2013243 [pii]

595 10.1038/ejhg.2013.243[published Online First: Epub Date]|.
596 40. Jamain S, Quach H, Betancur C, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and
597 NLGN4 are associated with autism. Nat Genet 2003;**34**(1):27-9 doi: 10.1038/ng1136

598 ng1136 [pii][published Online First: Epub Date]|.
599 41. Blasi F, Bacchelli E, Pesaresi G, et al. Absence of coding mutations in the X-linked genes neuroligin 3 and
600 neuroligin 4 in individuals with autism from the IMGSAC collection. Am J Med Genet B
601 Neuropsychiatr Genet 2006;**141B**(3):220-1 doi: 10.1002/ajmg.b.30287[published Online First: Epub
602 Date]|.
603 42. Gauthier J, Bonnel A, St-Onge J, et al. NLGN3/NLGN4 gene mutations are not responsible for autism in
604 the Quebec population. Am J Med Genet B Neuropsychiatr Genet 2005;**132B**(1):74-5 doi:
605 10.1002/ajmg.b.30066[published Online First: Epub Date]|.

606 43. Liu Y, Du Y, Liu W, et al. Lack of association between NLGN3, NLGN4, SHANK2 and SHANK3 gene
607 variants and autism spectrum disorder in a Chinese population. PLoS One 2013;**8**(2):e56639 doi:
608 10.1371/journal.pone.0056639

609 PONE-D-12-33239 [pii][published Online First: Epub Date]|.
610 44. Vincent JB, Kolozsvari D, Roberts WS, et al. Mutation screening of X-chromosomal neuroligin genes: no
611 mutations in 196 autism probands. Am J Med Genet B Neuropsychiatr Genet 2004;**129B**(1):82-4
612 doi: 10.1002/ajmg.b.30069[published Online First: Epub Date]|.
613 45. Wermter AK, Kamp-Becker I, Strauch K, et al. No evidence for involvement of genetic variants in the X-
614 linked neuroligin genes NLGN3 and NLGN4X in probands with autism spectrum disorder on high
615 functioning level. Am J Med Genet B Neuropsychiatr Genet 2008;**147B**(4):535-7 doi:
616 10.1002/ajmg.b.30618[published Online First: Epub Date]|.
617 46. Yan J, Oliveira G, Coutinho A, et al. Analysis of the neuroligin 3 and 4 genes in autism and other
618 neuropsychiatric patients. Mol Psychiatry 2005;**10**(4):329-32 doi: 4001629 [pii]

619 10.1038/sj.mp.4001629[published Online First: Epub Date]|.
620 47. Ylisaukko-oja T, Rehnstrom K, Auranen M, et al. Analysis of four neuroligin genes as candidates for
621 autism. European journal of human genetics : EJHG 2005;**13**(12):1285-92 doi: 5201474 [pii]

622 10.1038/sj.ejhg.5201474[published Online First: Epub Date]|.
623 48. Muncke N, Jung C, Rudiger H, et al. Missense mutations and gene interruption in PROSIT240, a novel
624 TRAP240-like gene, in patients with congenital heart defect (transposition of the great arteries).
625 Circulation 2003;**108**(23):2843-50 doi: 10.1161/01.CIR.0000103684.77636.CD[published Online
626 First: Epub Date]|.
627 49. Asadollahi R, Oneda B, Sheth F, et al. Dosage changes of MED13L further delineate its role in congenital
628 heart defects and intellectual disability. European journal of human genetics : EJHG
629 2013;**21**(10):1100-4 doi: 10.1038/ejhg.2013.17[published Online First: Epub Date]|.
630 50. M MvH, G RM, Duran K, et al. Further confirmation of the MED13L haploinsufficiency syndrome.
631 European journal of human genetics : EJHG 2014 doi: 10.1038/ejhg.2014.69[published Online First:
632 Epub Date]|.
633 51. Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic
634 sporadic intellectual disability: an exome sequencing study. Lancet 2012;**380**(9854):1674-82 doi:
635 S0140-6736(12)61480-9 [pii]

636 10.1016/S0140-6736(12)61480-9[published Online First: Epub Date]|.
637 52. Vulto-van Silfhout AT, Rajamanickam S, Jensik PJ, et al. Mutations Affecting the SAND Domain of DEAF1
638 Cause Intellectual Disability with Severe Speech Impairment and Behavioral Problems. Am J Hum
639 Genet 2014 doi: 10.1016/j.ajhg.2014.03.013[published Online First: Epub Date]|.
640 53. Faqeih EA, Al-Owain M, Colak D, et al. Novel homozygous DEAF1 variant suspected in causing white
641 matter disease, intellectual disability, and microcephaly. Am J Med Genet A 2014 doi:
642 10.1002/ajmg.a.36482[published Online First: Epub Date]|.
643 54. Albers CA, Paul DS, Schulze H, et al. Compound inheritance of a low-frequency regulatory SNP and a
644 rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat Genet
645 2012;**44**(4):435-9, S1-2 doi: 10.1038/ng.1083[published Online First: Epub Date]|.
646 55. Unger S, Mainberger A, Spitz C, et al. Filamin A mutation is one cause of FG syndrome. Am J Med Genet
647 A 2007;**143A**(16):1876-9 doi: 10.1002/ajmg.a.31751[published Online First: Epub Date]|.
648 56. Salmi M, Bruneau N, Cillario J, et al. Tubacin prevents neuronal migration defects and epileptic activity
649 caused by rat Srpx2 silencing in utero. Brain : a journal of neurology 2013;**136**(Pt 8):2457-73 doi:
650 10.1093/brain/awt161[published Online First: Epub Date]|.
651 57. Sia GM, Clem RL, Huganir RL. The human language-associated gene SRPX2 regulates synapse formation
652 and vocalization in mice. Science 2013;**342**(6161):987-91 doi: 10.1126/science.1245079[published
653 Online First: Epub Date]|.

58. Griggs BL, Ladd S, Saul RA, et al. Dedicator of cytokinesis 8 is disrupted in two patients with mental retardation and developmental disabilities. Genomics 2008;**91**(2):195-202 doi: 10.1016/j.ygeno.2007.10.011[published Online First: Epub Date]|.
59. Laumonnier F, Holbert S, Ronce N, et al. Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. J Med Genet 2005;**42**(10):780-6 doi: 42/10/780 [pii]

10.1136/jmg.2004.029439[published Online First: Epub Date]|.
60. Yu HC, Sloan JL, Scharer G, et al. An X-linked cobalamin disorder caused by mutations in transcriptional coregulator HCFC1. Am J Hum Genet 2013;**93**(3):506-14 doi: S0002-9297(13)00342-X [pii]

10.1016/j.ajhg.2013.07.022[published Online First: Epub Date]|.
61. Leen WG, Klepper J, Verbeek MM, et al. Glucose transporter-1 deficiency syndrome: the expanding clinical and genetic spectrum of a treatable disorder. Brain : a journal of neurology 2010;**133**(Pt 3):655-70 doi: 10.1093/brain/awp336[published Online First: Epub Date]|.
62. Girirajan S, Rosenfeld JA, Cooper GM, et al. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet 2010;**42**(3):203-9 doi: 10.1038/ng.534[published Online First: Epub Date]|.
63. Leblond CS, Heinrich J, Delorme R, et al. Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. PLoS genetics 2012;**8**(2):e1002521 doi: 10.1371/journal.pgen.1002521[published Online First: Epub Date]|.
64. Schaaf CP, Sabo A, Sakai Y, et al. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. Human molecular genetics 2011;**20**(17):3366-75 doi: 10.1093/hmg/ddr243[published Online First: Epub Date]|.
65. Mandel JL, Chelly J. Monogenic X-linked mental retardation: is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations. Eur J Hum Genet 2004;**12**(9):689-93 doi: 10.1038/sj.ejhg.5201247

5201247 [pii][published Online First: Epub Date]|.
66. Nava C, Lamari F, Heron D, et al. Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. Transl Psychiatry 2012;**2**:e179 doi: tp2012102 [pii]

10.1038/tp.2012.102[published Online First: Epub Date]|.
67. Tarpey PS, Smith R, Pleasance E, et al. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. Nat Genet 2009;**41**(5):535-43 doi: ng.367 [pii]

10.1038/ng.367[published Online First: Epub Date]|.
68. Simpson DA, Clark GR, Alexander S, et al. Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa. Journal of medical genetics 2011;**48**(3):145-51 doi: 10.1136/jmg.2010.083568[published Online First: Epub Date]|.
69. Brownstein Z, Friedman LM, Shahin H, et al. Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in Middle Eastern families. Genome biology 2011;**12**(9):R89 doi: 10.1186/gb-2011-12-9-r89[published Online First: Epub Date]|.
70. Huang L, Jolly LA, Willis-Owen S, et al. A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. Am J Hum Genet 2012;**91**(4):694-702 doi: S0002-9297(12)00423-5 [pii]

10.1016/j.ajhg.2012.08.011[published Online First: Epub Date]|.
71. Gibbons RJ, Wada T, Fisher CA, et al. Mutations in the chromatin-associated protein ATRX. Hum Mutat 2008;**29**(6):796-802 doi: 10.1002/humu.20734[published Online First: Epub Date]|.

Figure Legends

Figure 1: Inherited disrupting splice variant in *DEAF1*: what contribution to the phenotype?

(A) Pedigree showing the maternally-inherited splice variant in *DEAF1* (c.290-3C>G); **(B)** Prediction scores for the effect of the herein described variant on splicing (prediction scores for acceptor splice sites (ASS) as computed by SpliceSite Finder, MaxEntScan, NNsplice, GeneSplicer and Human Splicing Finder for the consensus ASS with either the wild-type or the mutated allele); **(C)** Localization of the variant, and its resulting effect on splicing in vitro (minigene construct): 90% of abnormal transcripts: 80% with entire exon #2 skipped, and 10% using the alternative ASS c.290-16 both leading to a frameshift and a premature stop codon.

Figure 2: Truncating variants not or ambiguously co-segregating with ID

(A) Pedigree of patient APN-13 carrying a frameshift variant in *SRPX2* (c.602del, p.Ala201Valfs*10) demonstrating the likely inheritance from the asymptomatic (deceased) maternal grandfather, yet a putative germinal mosaicism of a de novo mutation cannot be excluded.; (**B)** Predicted functional domains of SRXP2 from Pfam indicating locations of the herein identified mutation and those previously described. DUF4174: Domain of unknown function; (**C)** Pedigree showing the non-segregating nonsense variant in *SHROOM4* (c.3772C>T; p.Gln1258*) in the family of patient APN-86; (**D**) Location of the premature stop codon, which would disrupt the ASD2 domain of the protein.

Figure 3: Patient carrying two probably damaging missense variants in *HCFC1* and *ATRX*: one causative mutation and one modifier variant?

(A) Family tree of patient APN-113: proband carries a missense mutation in *HCFC1* (c.218C>T; p.Ala73Val) already reported in two patients with *cblX* as well as another missense variant in *ATRX* (c.1013C>G, p.Ser338Cys). Both variants are maternally-inherited, absent from the unaffected brother, but carried by the younger brother who died of sudden death at 2 months-old; **(B)** Associated predictions for the recurrent pathogenic missense mutation in *HCFC1* [60] and the possible modifier in *ATRX*, showing putative pathogenicity and moderate nucleotide conservation; **(C)** Representation of HCFC1 and its domains: kelch domains (K1–K5), Fn3 (fibronectin type 3), basic domain, HCF-proteolysis repeats (HCF-pro), acidic domain, and NLS (nuclear localization signal) domain. The initial mutations involved in milder non-syndromic ID are indicated above: a regulatory variant in the 5'UTR of *HCFC1* was identified by targeted massive parallel re-sequencing in a family with probable X-linked ID (MRX3), which had for long remained unsolved. This variant was disrupting the functional binding site of the transcription factor YY1 within the *HCFC1*promoter region, leading to an up-regulation of its expression in lymphoblastoid cells [70].

Subsequent screening of additional unsolved XLID families identified one single co-segregating missense variant (c.674G>A; p.Ser225Asn) in *HCFC1*. The phenotype of both patients was rather mild: non-syndromic mild to moderate ID. In the bottom are indicated the mutations recently described in *cblX* patients [60]; **(D)** The *ATRX* missense variant is located close but outside of the hotspot for disease-causing missense mutations (in the zinc-finger binding domain, in green) reported in patients with *ATRX* mutations, in red: mutations reported independently in at least 2 patients. Mutations are indicated when affecting residues 1-1500 (reported in OMIM, ClinVar or in [71]), the rest of the protein is not represented; **(E)** Clinical information regarding APN-113. *CblX* patients present with the same very severe phenotype: severe ID, early infantile epilepsy, choreoathetosis, microcephaly and more variable muscular hypotonia. Three of them are reported with early death in infancy [60]; **(F)** Biochemical abnormalities observed in the two affected brothers are similar to those previously observed [60].

Tables

Table 1. Description of the cohort of 106 patients with ID and global diagnostic results

	Cohort (n=106)	**With positive diagnostic (n=26)**			**Yield (per category)**
Gender		**XLID**	**ADID**	**Total**	
Male	96 (91%)	14	8	**22**	22/96 (23%)
Female	10 (9%)	2	2	**4**	4/10 (40%)
Total	*106*	*16*	*10*	***26***	*26/106 (**25%**)*
Age					
]0-10]	57 (54%)	9	5	**14**	14/57 (25%)
]10-20]	31 (29%)	5	4	**9**	9/31 (29%)
>20	18 (17%)	2	1	**3**	3/18 (17%)
Sporadic cases [a]					
Female	8 (7%)	2	2	**4**	4/8 (50%)
Male	72 (68%)	7	7	**14**	14/72 (19%)
Total	*80 (75%)*	*9*	*9*	***18***	*18/80 (**23%**)*
Familial History					
Male sib-pairs	12	5	0	**5**	5/12 (42%)
Possible XLID [b]	8	2	1	**3**	3/8 (38%)
Other (non X-linked)	6	0	0	**0**	0/6 (0%)
Total	*26 (25%)*	*7*	*1*	***8***	*8/26 (**31%**)*
Consanguinity	3 (3%)	0	0	**0**	0/3 (0%)
ID severity					
Mild/Borderline	12 (11%)	3	0	**3**	3/12 (25%)
Moderate	49 (46%)	4	6	**10**	10/49 (20%)
Severe	45 (42%)	9	4	**13**	13/45 (29%)
Co-morbidity					
Microcephaly (<-2SD)	14 (13%)	3	2	**5**	5/14 (36%)
Epilepsy	28 (26%)	5	2	**6**	6/28 (21%)
Autistic traits	34 (32%)	7	2	**9**	9/34 (26%)
Hypotonia	36 (34%)	4	4	**8**	8/36 (22%)
Previous exploration					
CGH	106 (100%)				
Fragile-X	105 (99%)				
Karyotype	95/99 (96%)				
# Other genetic tests (mean per patient)	2				
IRM	53 (50%)				
Metabolism [c]	82 (77%)				

[a] No familial first degree ID

[b] Affected male relatives

[c] At least one biochemical test performed

Table 2. List of all causative/probably causative mutations identified in our cohort

Patient ID	Sex	Gene	Mutation	Inheritance	Mode of inheritance	Degree of ID	Consistency with classic phenotype	See Figure #
Certainly-causative mutations								
APN-58	M	*DYRK1A*	chr21:g.38858865C>T; c.613C>T; p.Arg205*; htz	*de novo*	AD	++	Yes	S2
APN-87	M	*DYRKA1*	chr21:g.38858873del; c.621_624delinsGAA; p.Glu208Asnfs*3; htz	*de novo*	AD	++	Yes	S2
APN-63	M	*GRIN1*	chr9:g.140056661C>G; c.1733C>G; p.Pro578Arg; htz	*de novo*	AD	+++	Yes	S3
APN-14	M	*MED13L*	chr12:g.116406845_116406852del; c.6118_6125del; p.Gly2040Asnfs*32; htz	*de novo*	AD	++	Partially	S4
APN-46	M	*RAI1*	chr17:g.17698594_17698598del; c.2332_2336del; p.Gly778Glnfs*7; htz	*de novo*	AD	++	Partially	S5
APN-122	F	*SHANK3*	chr:g.51159168_51159183dup; c.2955_2970dup; p.Pro992Argfs*325; htz	*de novo*	AD	+++	Yes	S6
APN-38	M	*SLC2A1*	chr1:g.43395407G>A; c.724C>T; p.Gln242*; htz	*de novo*	AD	+++	Yes	S7
APN-139	M	*SYNGAP1*	chr6:g.33414346G>A; c.3583-6G>A; p.?; htz; splice decreased	*de novo*	AD	?	Yes	S8
APN-41	M	*TCF4*	chr18:g.53017622_53017625del; c.514_517del; p.Lys172Phefs*61; htz	*de novo*	AD	+++	Yes	S9
APN-117	F	*TCF4*	chr18:g.53017619G>A; c.520C>T; p.Arg174*; htz	*de novo*	AD	++	No	S9
APN-138	M	*ATRX*	**chrX: g.76972632G>A; c.109C>T; p.Arg37* (rs122445108); hemz**	Inherited (Ma)	XL	+++	Yes	S10
APN-137	M	*CUL4B*	chrX: g.119681009_119681010del; c.811_812del; p.Gln271Aspfs*11; hemz	Inherited (Ma)	XL	+++	Partially	S11
APN-42	M	*DMD*	chrX:g.31164440del; c.10889del; p.Arg3630Glnfs*27; hemz	Inherited (Ma)	XL	++	No	S12
APN-26	M	*FMR1*	last exon deletion; hemz	Inherited (Ma)[b]	XL	+++	Partially	S13
APN-113	M	*HCFC1 (ATRX)*	**chrX:g.153230153G>A; c.218C>T; p.Ala73Val; hemz** (chrX:g.76939735G>C, c.1013C>G, p.Ser338Cys)	Inherited (Ma) Inherited (Ma)	XL (XL)	+++	Yes (Partially)	3
APN-82	M	*IL1RAPL1*	chrX:g.29935696_29935705del; c.894_903del; p.Trp299Thrfs*18; hemz	Inherited (Ma)	XL	++	Yes	S14
APN-68	M	*IQSEC2*	chrX:g.53268395G>A; c.3097C>T; p.Gln1033*; hemz	*de novo*	XL	+++	Yes	S15
APN-34	M	*KDM5C*	chrX:g.53228250C>G; c.2152G>C; p.Ala718Pro; hemz	*de novo*	XL	++	Partially	S16
APN-135	M	*KDM5C*	chrX:g.53240784dup; c.1296dup; p.Glu433*; hemz	Inherited (Ma)	XL	++	Partially	S16
APN-16	M	*MAOA*	chrX:g.43590942_43590943delinsTT; c.797_798delinsTT; p.Cys266Phe; hemz	Inherited (Ma)	XL	+/-	Yes	[39]
APN-130	F	*MECP2*	**chrX: g.153296363G>A; c.952C>T; p.Arg318Cys (rs28935468); htz**	*de novo*	XL	+++	Partially	S17

Table 2. List of all causative/probably causative mutations identified in our cohort (continued)

APN-142	F	*MECP2*	**chrX:g.153296777G>A; c.538C>T; p.Arg180* (rs61748421); htz**	*de novo*	XL	+++	Partially	S17
APN-3	M	*MECP2*	Complex rearrangement of exon 4; hemz	Inherited (Ma)	XL	+++	Yes	S17
APN-105	M	*PHF8* (*DOCK8*)	chrX:g.54028583C>G; c.1249+5G>C; p.Tyr406Phefs*24; hemz; (chr9:g.407035G>T; c.3496G>T; p.Glu1166*; htz)	Inherited (Ma); (*de novo*)	XL; (AD)	+	Partially (No)	S18
APN-43	M	*SLC9A6*	chrX:g.135080258_135080262del; c.526-9_526-5del; p.?; splice disrupted; hemz	Inherited (Ma)	XL	+	Yes	Masurel-Paulet et al., *in preparation*
APN-110	M	*SLC16A2*	**chrX:g.73749067T>C; c.1412T>C ; p.Leu471Pro (rs122455132); hemz**	Inherited (Ma)	XL	+++	Yes	S19
Possibly-causative mutations								
APN-131	M	*SLC2A1* *ANKRD11*	chr1:g.43392779del; c.1412delG; p.Gly471Glufs*37; htz chr16:g.89348867G>T; c.4083C>A; p.His1361Gln; htz	Inherited (Pa); Inherited (Ma)	AD; AD	+++	Partially	S7
APN-101	M	*TCF4*	chr18:g.52899907C>T; c.1487-5G>A; p.Arg495_Gly496insAla?,; htz	*de novo ?* [a]	AD	++	No	S9
APN-99	M	*NLGN3*	chrX:g.70389249C>T ; c.1849C>T; p.Arg617Trp; hemz	Inherited (Ma)	XL	+++	Yes	S20
APN-70	M	*PQBP1*	chrX:g.48760294C>T ; c.731C>T ; p.Pro244Leu; hemz	Inherited (Ma)	XL	++	No	S21

[a]: absent from the mother, deceased father (untested).

[b]: present in the 3 brothers. Mother untested, but most probably maternally inherited.

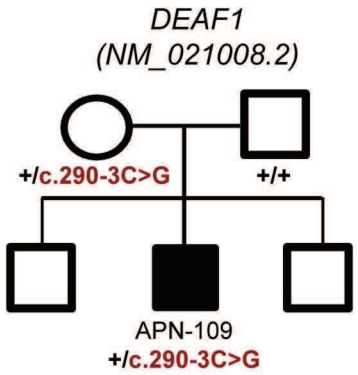
In bold: mutations previously reported in other patients

M: male, F: female, htz: heterozygous, hemz: hemizygous, Pa: paternally-inherited, Ma: maternally-inherited

AD: Autosomal dominant; XL: X-linked.

-: no ID, +: mild ID, ++: moderate ID, +++: severe ID

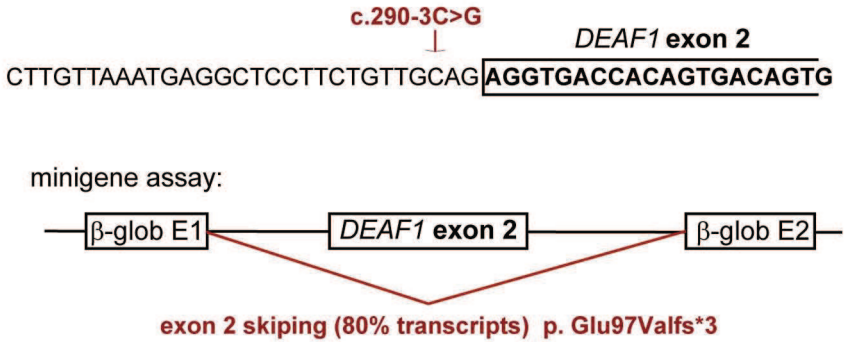
A.

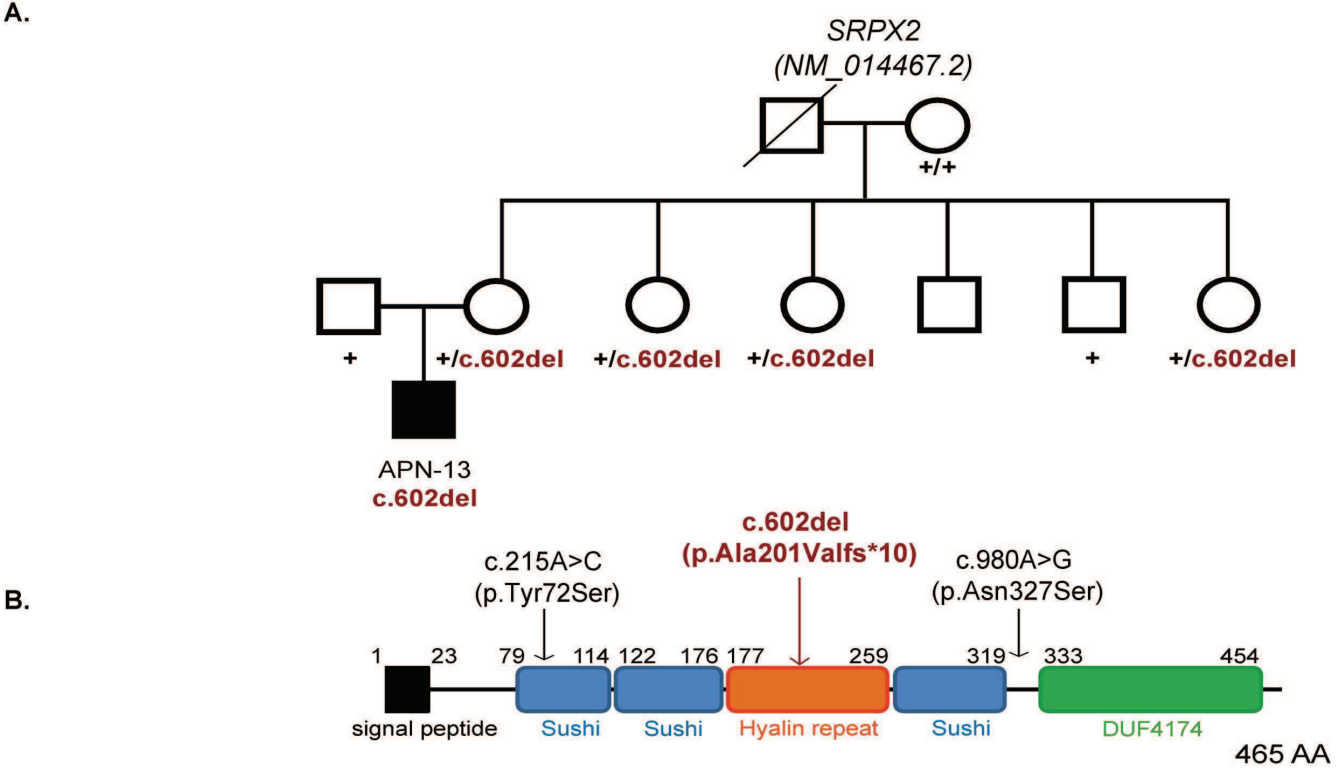


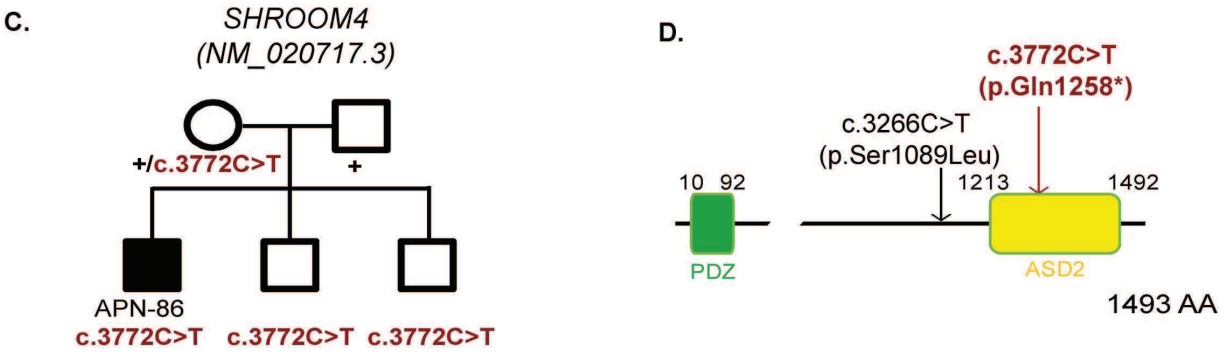
B.

	c.290-3C Wild-type ASS	c.290-3C>G Mutant ASS
SpliceSiteFinder	89.5	78.6 (-12%)
MaxEntScan	8	0 (-100%)
NNSplice	0.7	0 (-100%)
GeneSplicer	8.9	0 (-100%)
HumanSplicingFinder	89.3	79.0 (-12%)

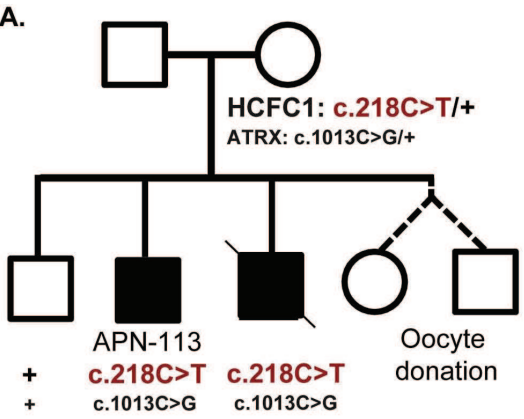
C.




A.
SRPX2
(NM_014467.2)
+/+
+
+/c.602del
+/c.602del
+/c.602del
+
+/c.602del
APN-13
c.602del
B.
c.602del
(p.Ala201Valfs*10)
c.215A>C
(p.Tyr72Ser)
c.980A>G
(p.Asn327Ser)
1
23
79
114
122
176
177
259
319
333
454
signal peptide
Sushi
Sushi
Hyalin repeat
Sushi
DUF4174
465 AA


C.
SHROOM4
(NM_020717.3)
+/c.3772C>T
+
APN-86
c.3772C>T
c.3772C>T
c.3772C>T
D.
c.3772C>T
(p.Gln1258*)
c.3266C>T
(p.Ser1089Leu)
10
92
1213
1492
PDZ
ASD2
1493 AA

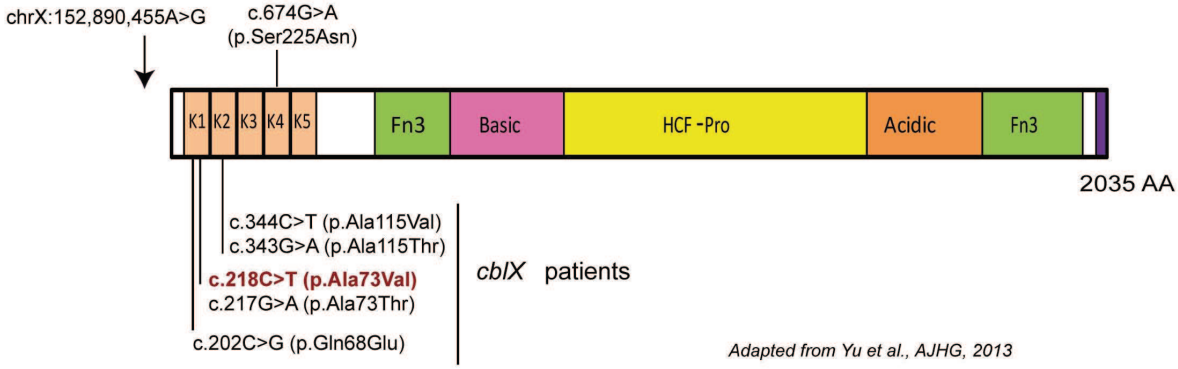
A.



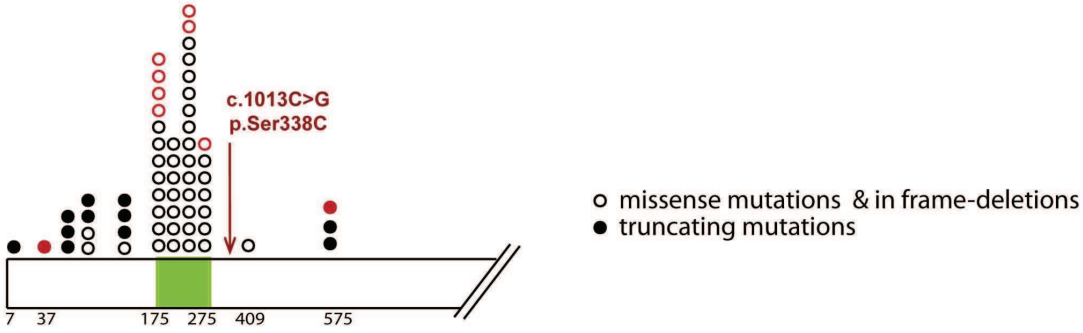
B.

	HCFC1 NM_005334.2 c.218C>T p.Ala73Val (Pathogenic, [68] & this study)	ATRX NM_000489.3 c.1013C>G p.Ser338Cys (Possible modifier, this study)
SIFT	Deleterious (0.02)	Deleterious (0.00)
PolyPhen2	Deleterious (HumDiv: 0.991)	Deleterious (HumDiv: 0.999)
Mutation T@ster	Disease causing (1)	Polymorphism (1)
Grantham [0;215]	64	112
PhyloP [-14.1;6.4]	5.451	1.981
PhastCons [0;1]	1	0.425

C.



D.



E.

Patient	Sex	Birth	Preliminary genetic tests	Clinical information
APN-113	M	2003	- Fragile-X test: negative - Karyotype: negative - Array-CGH: negative - *FOXG1, CDKL5, MECP2, ARX* direct sequencing: negative - X-inactivation bias: 92:8	Severe ID, no acquired walking/seating/speech, microcephaly (-4SD), myoclonic epilepsy since 3 months-old, 4-5 crisis/week (abnormal EEG), hypotonia, abnormal gait, choreoathetosis (dyskinetic movements). Autistic disorders, attention/concentration disorders, behavioral disorder, sleep disturbances, altered sleep structure MRI: normal Spectroscopy: abnormal peaks of choline and creatine in frontal regions

F.

	APN-113				
	1 year-old	**3 years-old**	**5 years-old**	**affected brother**	*normal value*
homocystein level (in plasma)	11 µmol/L	29.4 µmol/L	27 µmol/L	N.A.	*<9 µmol/l*
methylmalonic acid level (in urine)	51 mmol/mol creatinin	90 mmol/mol creatinin	68.2 mmol/mol creatinin	82.5 mmol/mol creatinin	*<5 mmol/mol creatinin*

SUPPORTING INFORMATION

SUPPORTING METHODS

Gene selection justification and design strategy

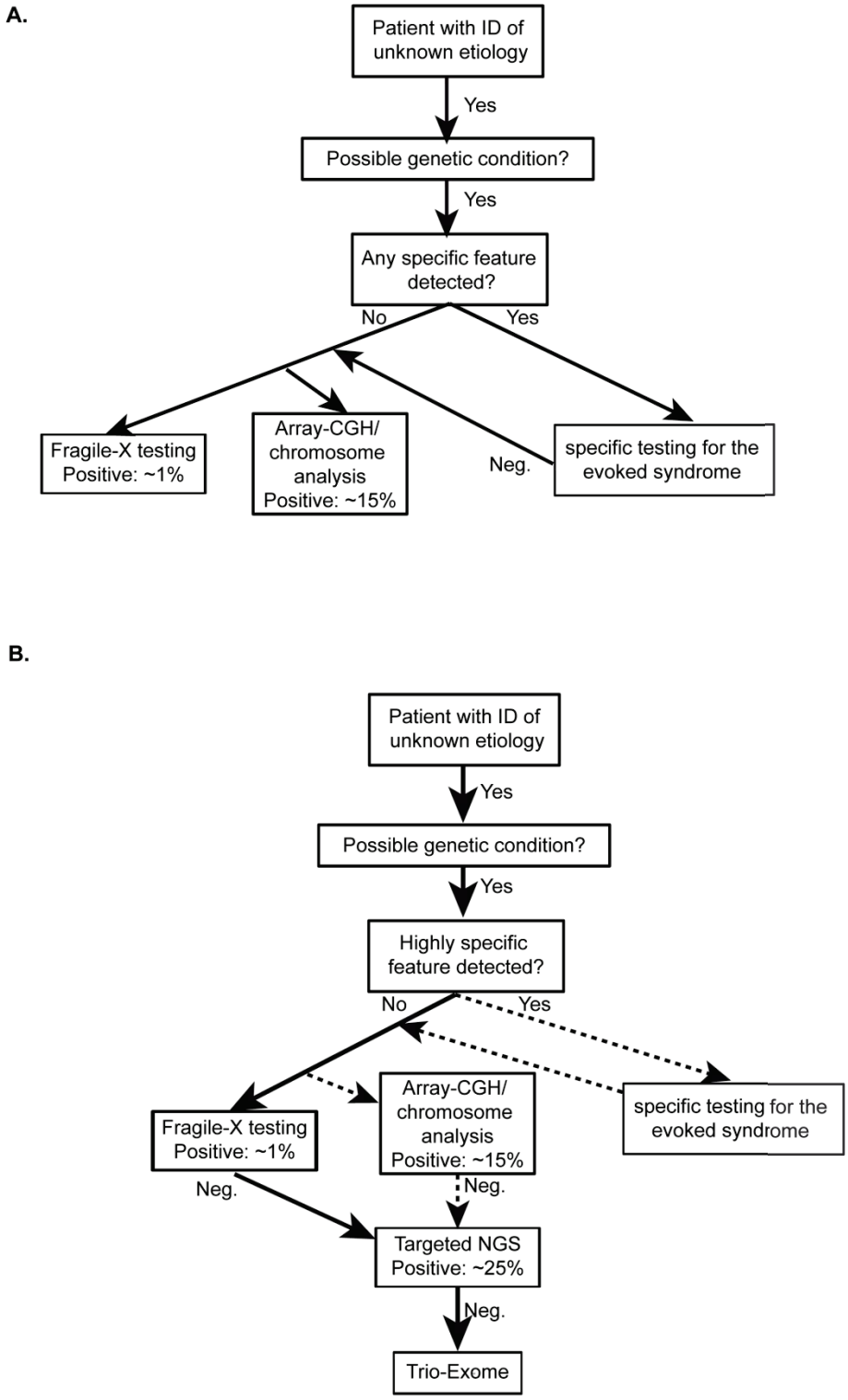
The design of the capture baits in terms of target genes was performed in November 2011 and therefore omits genes for which implication in ID was identified subsequently. Our strategy for including genes was based on the following scientific and economic considerations. First, we were limited to a total of 1-Mb of sequences as this size corresponded to a major price threshold for the manufacture of the array. Next, as we wanted to estimate the contribution of X-linked mutations in affected males and since we assumed that the vast majority of XLID genes had already been identified [1-3], we included almost all X-linked ID or ASD genes known at that time (99 in total, seven being excluded subsequently to our critical analysis of XLID genes, namely *ABCD1*, *AGTR2*, *FANCB*, *IKBKG*, *MTM1*, *PGK1* and *RAB40AL*; [4]). For autosomal genes, we favored those involved in non-syndromic ID/ASD but also included some implicated in syndromic forms. We included most of the genes described as involved in autosomal dominant/haploinsufficient forms of ID or ASD. Considering autosomal recessive genes, we only included those for which convincing mutations had been reported in at least two families or in single families but with additional evidences supporting their implication in ID (functional validations, CNV encompassing the gene in patients with ID, etc; [5-7]). Selected genes had to be associated to conditions in which intellectual disability (or ASD) is the major clinical concern, hence excluding diseases with major involvement of other organs (ex: genes encoding subunits of mitochondrial electron transfer complex since their involvement can be suspected on the basis of metabolic investigations, most genes responsible for severe brain anomalies detectable by MRI such as Joubert syndrome, leukodystrophies…). Conversely, we did include the distal part of *DMD* encoding the brain expressed DP71 transcript, as a few reported patients carrying mutations affecting DP71 presented with mild muscle involvement while had been initially ascertained on the basis of ID/autism [8]. In total, we selected 45 genes associated to an autosomal dominant/haploinsufficiency type of ID, 66 to an autosomal recessive one, and 7 to other complex or unclear mode of transmission (**Table S1**).

Capture baits tiling

RNA capture baits were designed via eArray to achieve a tiling frequency of 5x (i.e. number of independent probes covering each nucleotide of a targeted region; a probe being considered as an orphan if it is >100-bp away from any neighboring probe). Some segments were left without bait due to multi-mapping thus requiring manual design. The number of probe replicates was then customized according to their GC-content or associated region complexity in order to maximize the pull-down efficiency (60<%GC<65: x4, 65<%GC<70: x5, 70<%GC: x7, orphan baits: x10, custom baits for uncovered regions: x5, total: 53,000 baits).

SUPPORTING FIGURES

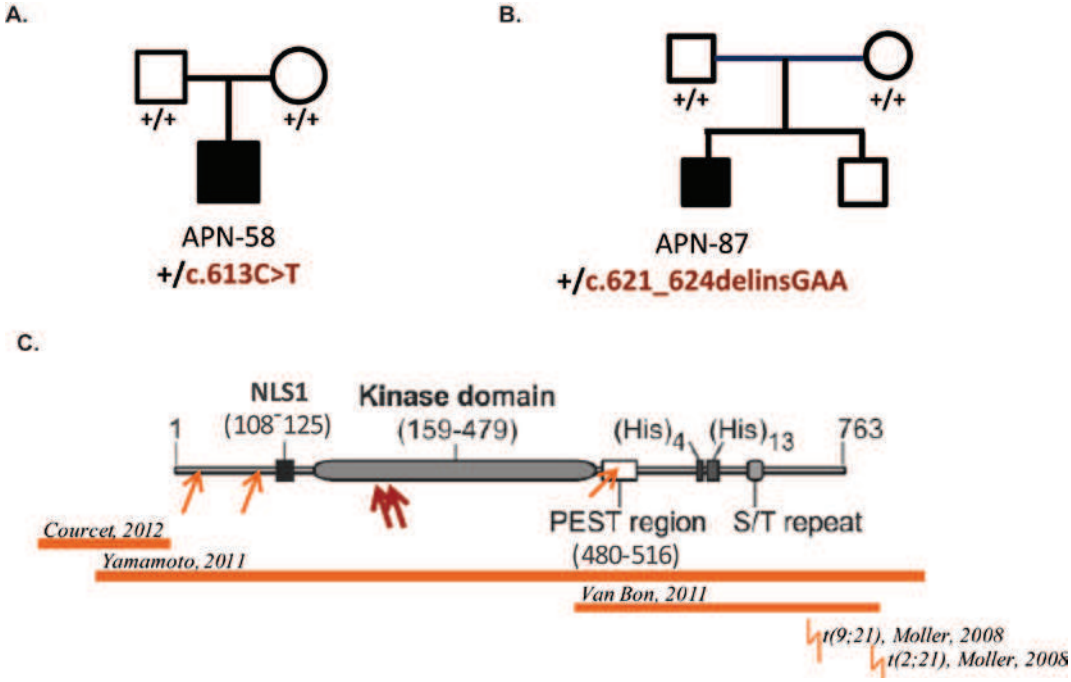
Figure S1. Decisional trees for the molecular diagnosis of intellectual disability.



A: Current tree used in routine in most diagnostic centers; **B:** Proposed new decisional tree integrating NGS-approaches. Dashed-lines: steps that might become obsolete upon analysis of cost effectiveness and reliability (targeted sequencing may replace array-CGH for reliable detection of CNVs).

Figure S2. *DYRK1A*, two de novo truncating causative mutations

APN-58: c.613C>T (NM_001396.3), p.Arg205*, heterozygous *de novo*
APN-87: c.621_624delinsGAA (NM_001396.3), p.Glu208Asnfs*3, heterozygous *de novo*



A. Pedigree of patient APN-58; **B.** Pedigree of patient APN-87; **C.** Representation of the protein DYRK1A with its putative domains, localization of causative mutations identified in this study (red arrows) and of previously reported mutations (orange arrows), deletions (orange bars), translocations (orange lightning) in patients [9-13].

Previous implication of *DYRK1A* in cognitive disorders:

Trisomy of *DYRK1A* was for long proposed having a major role in the cognitive impairment observed in Down syndrome patients. Recently, several mutations were identified in patients with ID or ASD, and very similar clinical phenotype: microcephaly, growth retardation, developmental delay, major speech impairments, feeding difficulties, stereotypic movements, large low-set ears, abnormal hair growth and specific facial features (micrognathia, hypotelorism, thin lips and large nose with long philtrum; [9,10,12-16]).

Patient APN-58 (male, born in 1998)

Preliminary genetic tests:

Fragile-X: negative; Array-CGH: negative; Karyotype: normal

Telomeres MLPA: normal; *MED12*, *ZDHHC9*, *FBN1, TGFBR1* and *TGFBR2* direct sequencing: negative

Clinical information:

Moderate ID, bordeline microcephaly (-2 SD), developmental delay (delayed sitting/walking/speech acquisition), hyperactivity, feeding disorders (sucking difficulties). Pectus excavatum, arachnodactyly, dysmorphic traits.

Patient APN-87 (male, born in 2009)

Preliminary genetic tests:

Fragile-X: negative; Array-CGH: negative

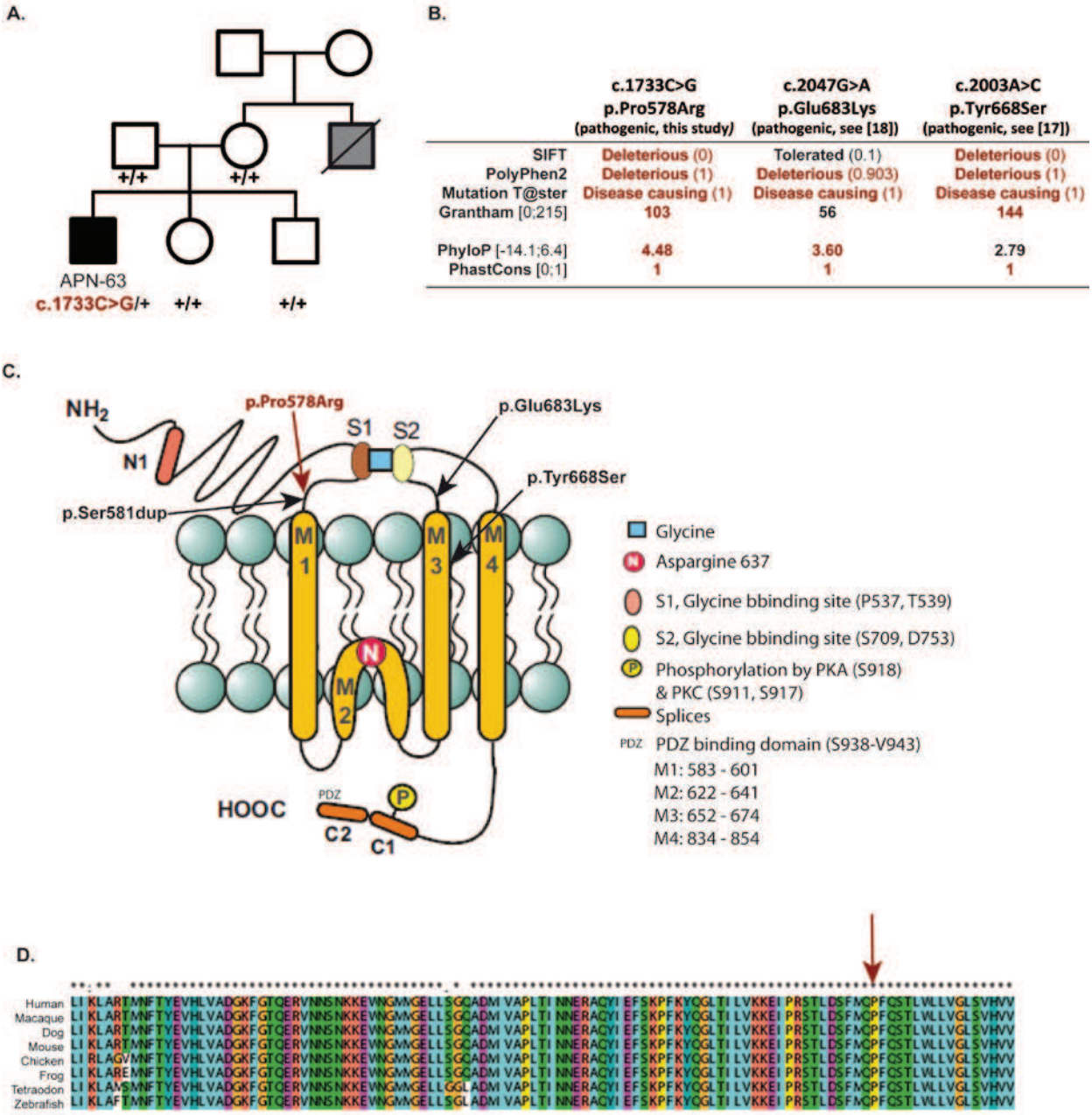
Clinical information:

Moderate ID, developmental delay (late sitting/walking/speech acquisition). Microcephaly (-4 SD) Dysmorphic traits. Hyperopia, astigmatism. Sleep disturbances, stereotypic movements. Hypertonia, opisthotonos episodes, feeding disorders, constipation.

MRI: discrete hydrocephaly, dilatation of lateral ventricles.

Figure S3. *GRIN1*, one causative *de novo* missense mutation

APN-63: c.1733C>G (NM_001185090.1), p.Pro578Arg, heterozygous *de novo*



A. Pedigree of patient APN-63; **B.** Associated prediction scores comparing the pathogenic missense mutation identified in this study with two pathogenic previously described missense mutations [17,18], showing a predicted deleterious status and a high conservation at the nucleotide level for the missense p.Pro578Arg; **C.** Representation of GRIN1 protein with the different domains. Mutations that were reported in ID patients are also indicated (black arrows), with the herein detected mutation in red (adapted from Parsons et al., 2007; [19]); **D.** Alignment of GRIN1 protein orthologs using Clustalw showing the high conservation of the affected residue Pro578.

Previous implications of *GRIN1* in cognitive disorders

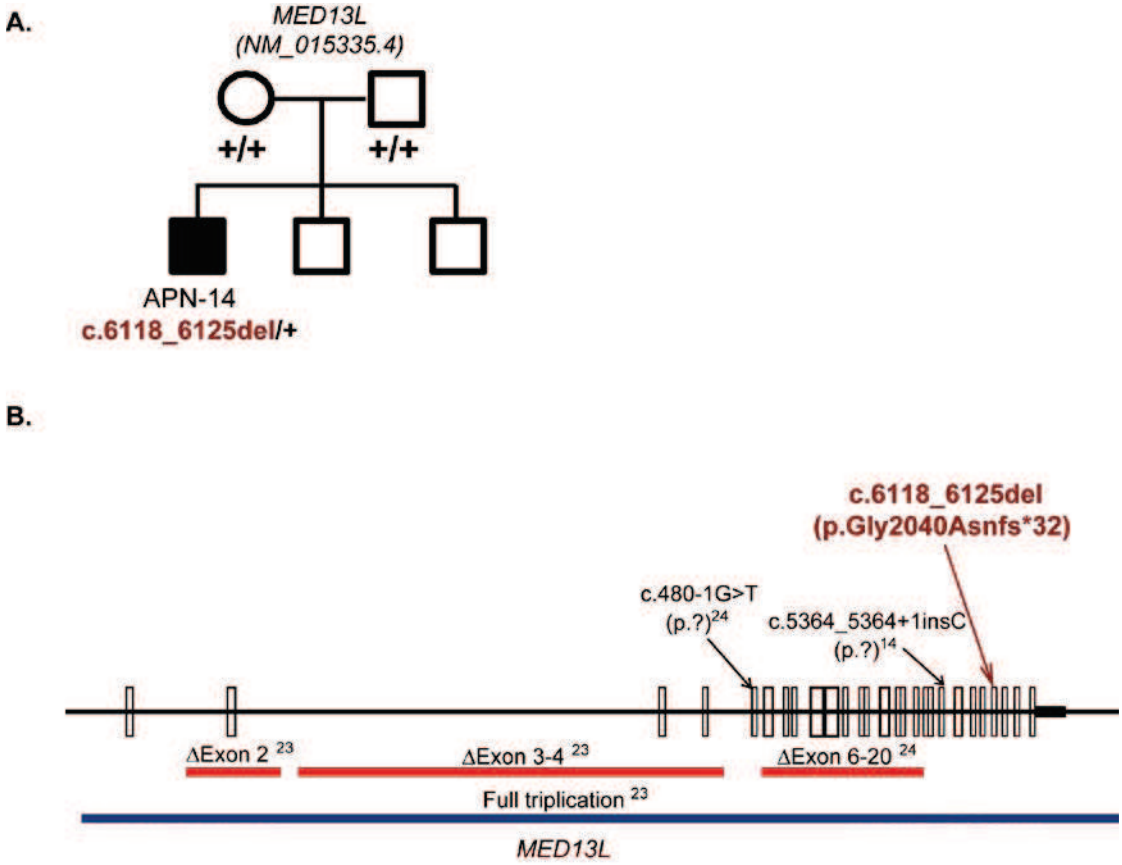
Following the evidence for the implication of NMDA (N-methyl-D-aspartate) type ionotropic glutamate receptors in schizophrenic-like behavior in mouse models, various studies have screened cohorts of schizophrenic patients but failed to identify any *GRIN1* mutations [20-22]. Later on, *de novo* mutations (one missense, p.Glu683Lys and one in-frame duplication, p.Ser581dup) were reported in patients with moderate to severe non-syndromic ID [18]. Both mutations resulted in decreased efficiency of the NMDAR channel. More recently, one patient was lately identified as carrying a *de novo* missense (p.Tyr668Ser) and presenting with early infantile epilepsy, no regression, severe ID, acquired relative microcephaly and no speech at 18 [17].

Patient APN-63 (male, born in 1972)
Preliminary genetic tests:
Fragile-X: negative; Array-CGH: negative; Karyotype: negative
MED12, *ZDHCC19*, *UPF3B*, *FBN1*, *TGFBR2* direct sequencing: negative
FraxA/FraxE: negative

Clinical information:
Proband: Severe ID, poor speech. Severe hypotonia and feeding disorders in early infancy, developmental delay (walking/speech), behavioral and mood disorders. Arachnodactyly, mild *pectus excavatum*.
MRI, scanner, echocardiography: normal
Skin biopsy: normal fibrilin
Plasma homocysteine levels: normal

Maternal Uncle: Developmental delay, behavioral and mood disorders (aggression), but normal speech and no hypotonia. Within the last 10 years of life: cognitive decline, anxiety, significant swallowing disorders. Deceased in 2010 from a heart attack.

Figure S4. *MED13L*, one causative *de novo* truncating mutation
APN-14: c.6118_6125del (NM_015335.4), p.Gly2040Asnfs*32, heterozygous *de novo*



A. Pedigree of patient APN-14**; B.** Visualization of the previously reported intragenic deletions and point mutations associated or probably-associated to *MED13L*-haploinsufficiency syndrome [14,23,24], and the point mutation identified in our patient.

Previous implication of *MED13L* in ID

A chromosomal balanced translocation disrupting *MED13L* was described in a patient with transposition of the great arteries (TGA) and intellectual disability. Three heterozygous missense mutations were then identified in patients with isolated TGA and a homozygous missense mutation in two siblings with non-syndromic ID from a consanguineous family [6,25]. More recently, Asadollahi et al. described three patients with copy number changes affecting *MED13L* and delineated a recognizable *MED13L*-haploinsufficiency syndrome characterized by hypotonia, moderate ID, variable degrees of conotruncal heart defect, facial hypotonia and dysmorphic traits (upslanting palpebral fissures, flat nasal root with bulbous tip, deep philtrum, micrognathia, large low-set ears, and broad forehead; [23]). This haploinsufficiency syndrome delineated by ID, uneven cardiac defects, developmental and speech delays, and hypotonic open mouth appearance was further confirmed by the report of two additional patients, carrying a de novo splice site mutation and a de novo exon deletion respectively[24].

Patient APN-14 (male, born in 1996)

Preliminary genetic tests:

Fragile-X test: negative; Array-CGH: duplication in 15q11.2 of 2.1-Mb (encompassing *CYFIP1*), considered as non pathogenic

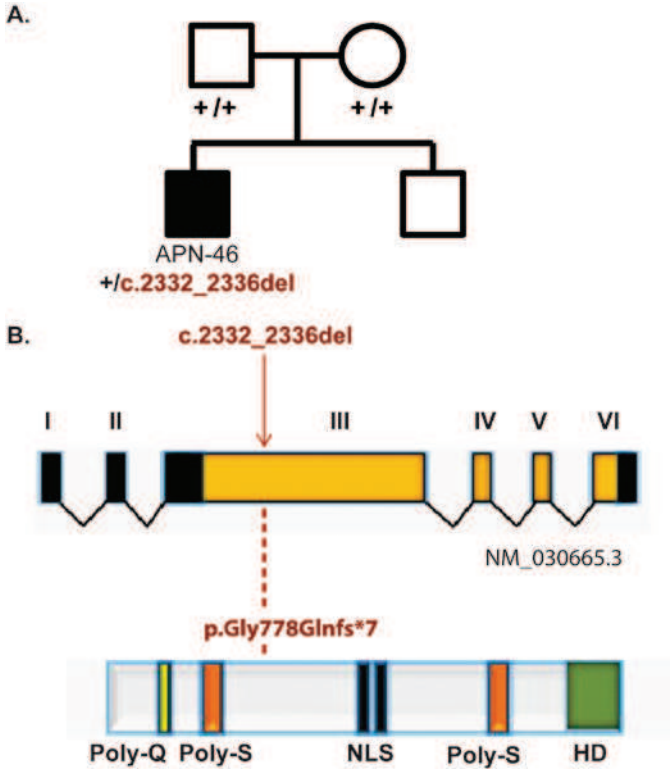
Myotonic dystrophy: negative for DM1/*DMPK* repeat expansions

PAX6 direct sequencing: negative

Clinical information: Moderate ID, motor development delay, muscular hypotonia, dysarthria, facial dysmorphic traits (round face, hypertelorism, eversed lower lip, low set ears), buccal dysprasia**,** coloboma. Echocardiography (2010): normal.

Figure S5. *RAI1*, one causative *de novo* truncating mutation

APN-46: c.2332_2336del (NM_030665.3), p.Gly778Glnfs*7, heterozygous *de novo*



A. Pedigree of patient APN-46; **B.** Representation of RAI1 at both gene and protein levels: polyglutamine (Poly-Q) and polyserine (Poly-S) tracts, bipartite nuclear localization signals (NLS), and C-terminal plant homeodomain (PHD), (adapted from [26]).

Previous implication of *RAI1* in cognitive disorders

Smith-Magenis syndrome (SMS; OMIM #182290, *607642) is a complex disorder characterized by variable degree of intellectual disability, sleep disturbances, craniofacial and skeletal anomalies, self-injurious and attention-seeking behaviors, and speech and motor delay. SMS is caused by either a heterozygous ~4-Mb deletion in 17p11.2 or heterozygous mutations in *RAI1*[26-28].

Patient APN-46 (male, born in 2009)

Preliminary genetic tests:

Fragile-X test: negative; Array-CGH array: negative

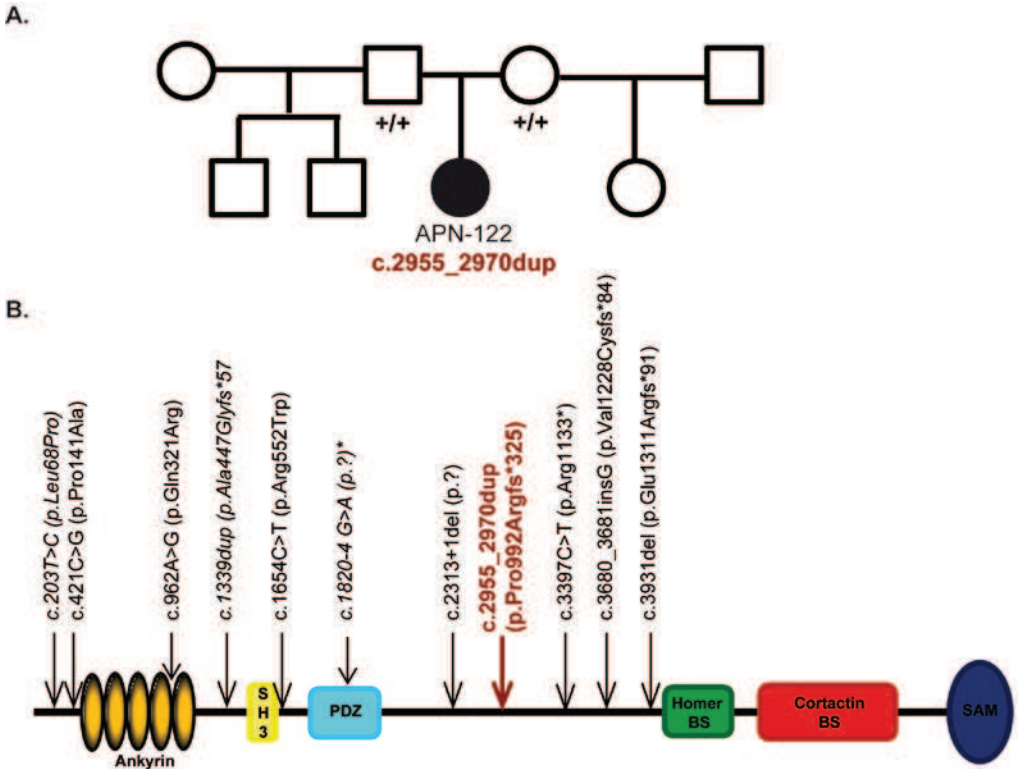
ARX, Prader-Willi, *GAMT* direct sequencing: negative

Clinical information: Moderate ID, **developmental delay**, **sleep disorders**, **severe behavioral disorders** (auto and hetero aggressivity), **no speech**. **Facial hypotonia**, **frontal bossing**, **upslanting palpebral fissures**, **midface hypoplasia**, a broad square-shaped face with depressed nasal bridge. Normal stature and OFC but overweight. Sleep disturbances after reevaluation.Hypermetropia, **strabismus** and astigmatism.

(in bold: classical features of Smith Magenis syndrome, from [26])

Figure S6. *SHANK3*, one causative *de novo* truncating mutation

APN-122: c.2955_2970dup (NM_001080420.1), p.Pro992Argfs*325, heterozygous *de novo*



A. Pedigree of patient APN-122; **B.** Representation of protein SHANK3 and its putative functional domains, and distribution of all reported point mutations in ASDs, ID and SCZ patients (deletions, translocations and duplications are not represented; adapted from [29,30]). Italic: mutations inherited from an unaffected parent. *: variant also detected in EVS (MAF=0.02%, 2/6237 individuals). SH3: Src homology-3 domain; PDZ: PDZ domain; SAM: sterile α-motif domain.

Previous implication of *SHANK3* in cognitive disorders

SHANK3 is the major haploinsufficient gene implicated in Phelan-McDermid syndrome (PMS, OMIM #606232) that is caused by a deletion in 22q13 encompassing several other genes and whose characteristic phenotype is: moderate to severe ID, absent or severe speech delay, neonatal hypotonia, and minor facial dysmorphic traits [31]. Heterozygous *de novo* point mutations in *SHANK3* have also been described in patients with autism spectrum disorders and schizophrenia associated to moderate to severe ID and poor language[29,30,32-35]. Shank3 knock-out mice harbor concordant phenotype, with self-injurious repetitive grooming, and deficits in social interactions and communication [36-39].

Patient APN122 (female, born in 2003)

Preliminary genetic tests:

Fragile-X: negative; Array-CGH: negative; Karyotype: negative

RAI1 and *MECP2* direct sequencing: one variant paternally inherited, considered as non-pathogenic

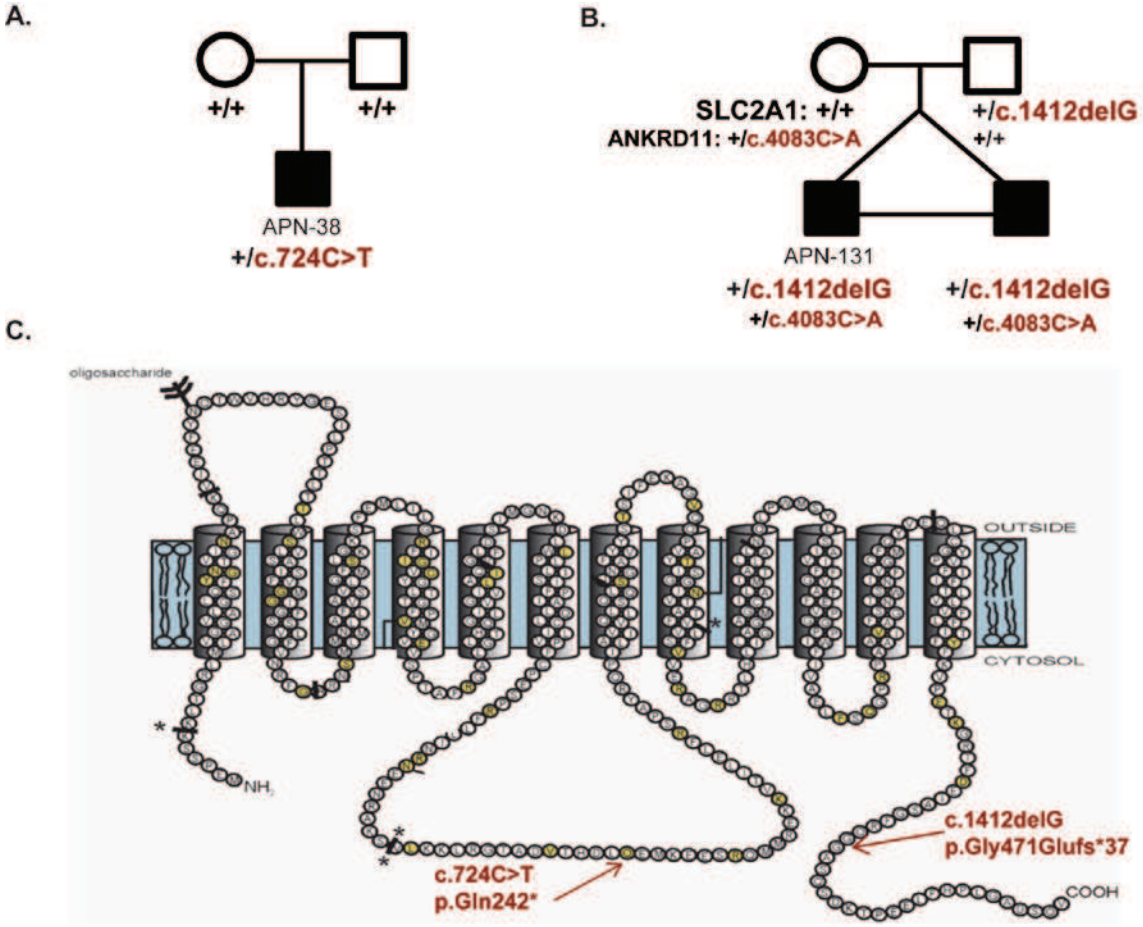
Clinical information:

Severe ID, developmental delay (acquired walking at 18 months-old, first words at 3 years-old followed by speech regression), no acquired reading/writing. Abnormal EEG without epilepsy, autistic traits, attention deficit/concentration disorder, behavioral disorders, sleep disturbances, distal spasticity, facial dysmorphic traits (upward philtrum, laterally-inserted eyebrows, bulbous nose).

Figure S7. *SLC2A1,* one causative *de novo* truncating mutation and one potentially-causative inherited truncating mutation

APN-38 : c.724C>T (NM_006516.2), p.Gln242*, heterozygous *de novo*

APN-131: c.1412delG (NM_006516.2), p.Gly471Glufs*37, heterozygous, inherited from unaffected father and *ANKRD11*: c.4083C>A (NM_013275.5), p.His1361Gln, inherited form unaffected mother



A. Pedigree of patient APN-38; **B.** Pedigree of patient APN-131; **C.** Structure of the SLC2A1 transporter and localization of the two herein described mutations reported in patients (in red; adapted from [40]).

Previous implication of *SLC2A1* in ID

Heterozygous mutations in GLUT1 transporter (SLC2A1) are responsible for GLUT1 deficiency syndrome-1 (GLUT1DS1, MIM #606777), a neurologic disorder characterized by infantile-onset epileptic encephalopathy associated with cognitive impairment (from learning disabilities to severe intellectual disability), delayed development, acquired microcephaly, hypotonia, motor incoordination (ataxia and dystonia), and spasticity [41,42]. Other forms of GLUT1DS can include ID, dysarthria, intermittent ataxia but no clinical seizures. Some clinical features such as sleep disturbances and headache can unevenly be described [43]. Low CSF glucose (less than 40 mg/dl) and low CSF lactate are detected in patients with GLUT1DS1. A ketogenic or modified Atkins diet often result in marked clinical improvement of the motor and seizure symptoms [41].

Previous implication of *ANKRD11* in ID

Mutations in *ANKRD11* are responsible for KBG syndrome (MIM #148050), characterized by macrodontia, distinctive craniofacial features (triangular face, ptosis, hypertelorism, prominent nasal bridge, anteverted nostrils, long philtrum, large and prominent ears, …), short stature, costovertebral anomalies, clinodactyly, cryptorchidism and neurologic involvement that includes global developmental delay, seizures, and mild to moderate intellectual disability [44]. A mouse model carrying a missense mutation in *ankrd11* presents with important craniofacial anomalies [45].

Patient APN-38 (male, born in 2005)
Preliminary genetic tests:
Fragile-X test: negative; Array-CGH: amplification 18q21.31 and 18q21.32 inherited from the mother and concluded as non-pathogenic
ARX direct sequencing: negative; Prader-Willi/Angelman test: negative

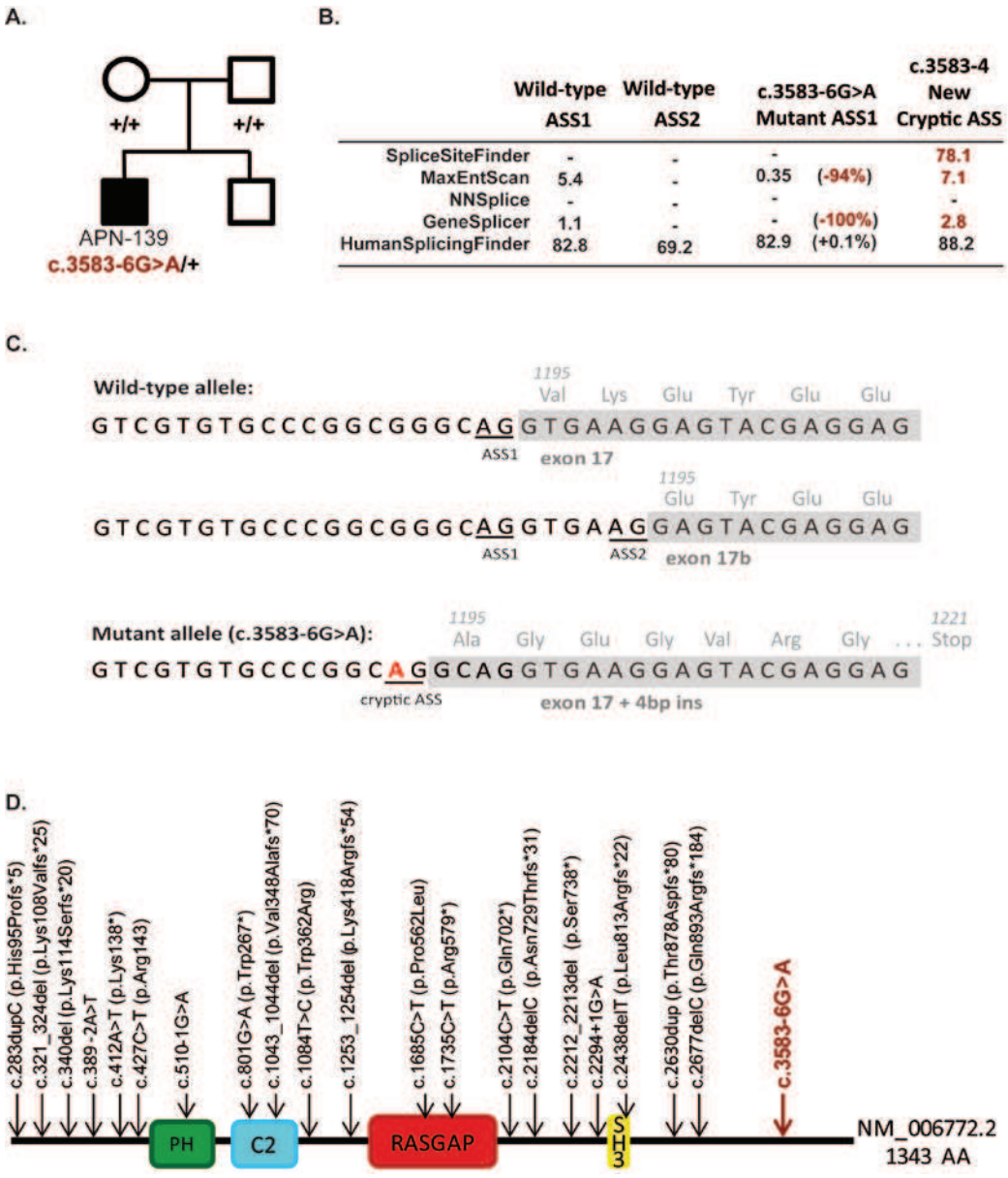
Clinical information: Severe ID, no speech. Hypotonia with ataxia, epilepsy starting from 14-months of age.

Patient APN-131 (male, born in 2003)
Preliminary genetic tests:
Fragile-X test: negative; Array-CGH: negative
ARX, DMPK, CREBBP, EP300, ZEB2, PYCR1 direct sequencing: negative
Prader-Willi/Angelman test: negative

Clinical information: Severe ID, hypotonia, hyperlaxity, developmental delay. Dorsal hemivertebra, crypotorchidism, neonatal hypoglycemia, dysmorphic features (hypertelorism, asymmetrical eyelids, thin upper lip). OFC: -2SD. No epilepsy.
Clinical phenotype of the monozygotic twin brother: highly similar.

Figure S8. *SYNGAP1*, one causative *de novo* mutation affecting splicing

APN-139: c.3583-6G>A (NM_006772.2), p.Val1195Alafs*27, heterozygous *de novo*



A. Family pedigree of patients APN-139; **B.** Prediction scores for acceptor splice sites (ASS) determined by MaxEnt, NNsplice, GeneSplicer or Human splicing Finder (HSF): prediction scores for the normal acceptor sites ASS1 (exon17) and ASS2 (exon17b) with the wild-type allele are given. The variations of scores for ASS1 with the c.3583G>A mutation are also indicated, as well as the prediction score for the new alternative cryptic ASS created by the mutation. **C.** Patient blood mRNA was analyzed and confirmed the use of this alternative cryptic ASS at c.3583-4 when the mutation is present instead of the usual ASS1 or ASS2 sites. Four nucleotides are inserted, causing a frameshift that leads to a premature truncated protein (1210 aminoacids instead of 1343); **D.** Schematic representation of protein SYNGAP1 with its different domains (PH: Pleckstrin homology-like, C2, RASGAP: Ras-GTPase activating and SH3 domains) and distributions of all *SYNGAP1* mutations identified in patients since 2009.

Previous implication of *SYNGAP1* in cognitive disorders

To date, about 20 or more heterozygous truncating or missense mutations have been identified in *SYNGAP1* in patients with moderate to severe ID, with or without associated severe epilepsy or autism [46-50].

Patient APN-139 (male, born in 2008)

Preliminary genetic tests:

Fragile-X test: negative; Array-CGH: negative

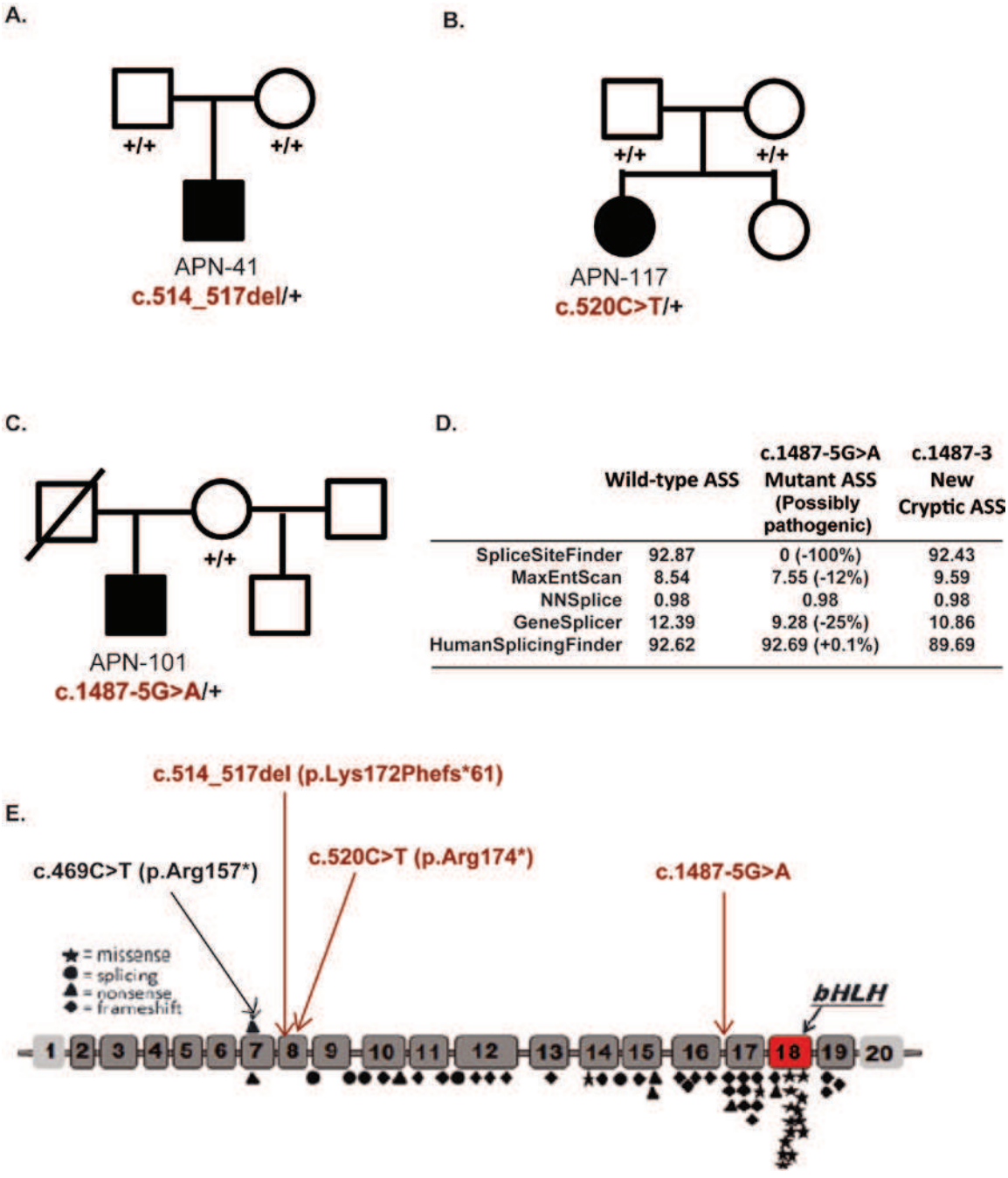
Clinical information: Moderate ID, no speech, psychomotor delay, stereotypic movements, behavioral disorders with hetero and auto-aggressivity, hypotonia, cerebellar syndrome.

Figure S9. *TCF4*, two *de novo* truncating causative mutations and one heterozygous potentially-causative splice site mutation

APN-41: c.514_517del (NM_001083962.1), p.Lys172Phefs*61, heterozygous *de novo*

APN-117: c.520C>T (NM_001083962.1), p.Arg174*, heterozygous *de novo*

APN-101: c.1487-5G>A (NM_001083962.1), p.Arg495_Gly496insAla?, heterozygous putative splice site mutation, which may create a new cryptic splice site that would result in the in-frame addition of a glutamine codon



	Wild-type ASS	c.1487-5G>A Mutant ASS (Possibly pathogenic)	c.1487-3 New Cryptic ASS
SpliceSiteFinder	92.87	0 (-100%)	92.43
MaxEntScan	8.54	7.55 (-12%)	9.59
NNSplice	0.98	0.98	0.98
GeneSplicer	12.39	9.28 (-25%)	10.86
HumanSplicingFinder	92.62	92.69 (+0.1%)	89.69

A-C. Pedigrees of patients carrying mutations in *TCF4*; **D.** Prediction scores for acceptor splice sites (ASS) as computed by SpliceSite Finder, MaxEntScan, NNsplice, GeneSplicer and Human Splicing Finder for the consensus ASS with either the wild-type or the mutated allele, and for the newly created cryptic ASS (in -3) that adds one in frame alanine residue (tested by minigene assay); **E.** Representation of *TCF4* with location of the previously-described Pitt-Hopkins mutations (diamonds, triangles, stars and circles) and the one associated with non-syndromic ID (black arrow; adapted from [51]).

Previous implication of *TCF4* in cognitive disorders

Missense and truncating mutations in *TCF4* cause Pitt-Hopkins syndrome (MIM #610954) whose classical phenotype is: severe motor delay and intellectual disability, no speech, characteristic dysmorphic features (wide mouth), stereotypic hand movements, constipation and hyperventilation (later onset: after 6-7 years old; [52]). Recently, a truncating mutation was identified in *TCF4* in a patient with non-syndromic ID [51].

Patient APN-41 (male, born in 2008)

Preliminary genetic tests:

Fragile-X test: negative; Array-CGH: 8p23.2 deletion inherited from the mother

Myotonic dystrophy: negative for DM1/*DMPK* repeat expansions

Prader-Willi test: negative

Clinical information: Severe ID with no speech. Constipation, stereotypic movements, frequent ear infections, motor development delay, hypotonia, no epilepsy, no breathing abnormality. No familial history.

Patient APN-117 (female, born in 2002)

Preliminary genetic tests:

Fragile-X test: negative; Array-CGH: negative; Telomeric rearrangement: negative

17p11.23 (SMS), 22q11.2: negative

UBE3A, *MECP2* direct sequencing: negative

Myotonic dystrophy: negative for DM1/*DMPK* and DM2/*ZN9* repeat expansions

Clinical information: Moderate ID, motor development delay, poor speech, partial autonomy. Attention and sleep disorder, hypotonia, cerebellar syndrome. No epilepsy. Left epicanthus, hemangioma. Echocardiography, electromyography and EEG: normal.

Patient APN-101 (male, born in 1986)

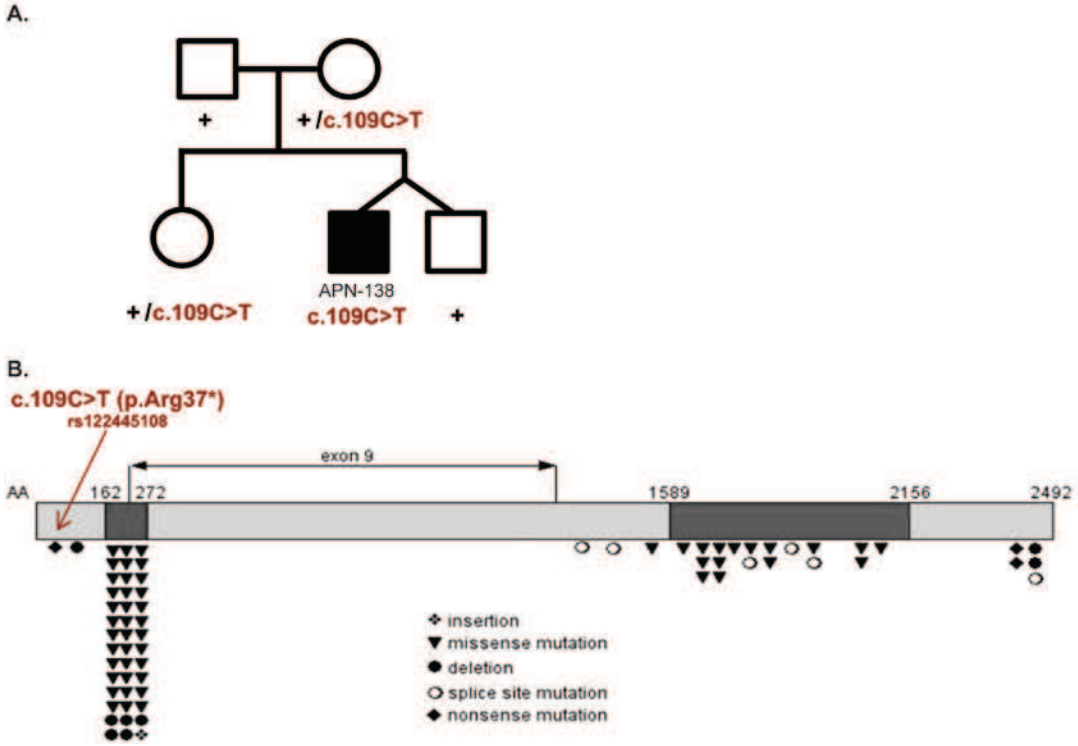
Preliminary genetic tests:

Fragile-X test: negative; Array-CGH: negative; Karyotype: negative

Clinical information: Moderate ID with normal speech, partial autonomy. Autistic features, attention deficit disorders. Facial dysmorphic traits: facial asymmetry, malar hypoplasia and earlobe hypoplasia.

Figure S10. *ATRX*, one maternally-inherited causative nonsense mutation

APN-138: c.109C>T (NM_000489.3), p.Arg37* (rs122445108), hemizygous, maternally inherited



A. Pedigree of patient APN-138; **B.** Representation of the ATRX protein with distribution of mutations identified in patients with ID (adapted from [53]).

Previous implication of *ATRX* in cognitive disorders

Mutations in this gene are reported in patients with Alpha-thalassemia/mental retardation syndrome (MIM #301040) or Mental retardation-hypotonic facies syndrome (MIM #309580) [53,54]. These mutations are mainly missense mutations but truncating mutations have also been identified (81 missense mutations, 18 mutations affecting splicing, 8 nonsense mutations and 12 small indels; [55]). Clinical features of ATR-X syndrome include: ID, moderate to severe motor delay, characteristic dysmorphic facies (microcephaly, hypertelorism, epicanthus, anteverted nostrils, carp-shaped mouth), hypotonia, mild-to-moderate anemia secondary to alpha-thalassemia, and genital anomalies. Most patients carrying the c.109C>T, p.Arg37* mutation present with classical ATR-X syndrome, including characteristic facies [56,57].

Patient APN-138 (male, born in 2001)

Preliminary genetic tests:

Fragile-X test: negative; Array-CGH: 15q15.1 duplication paternally inherited; Karyotype: negative

RSK2, *EHMT1* direct sequencing: negative

Clinical information: Moderate ID, mild motor delay (acquired walking: 20 months-old), microcephaly (-4SD) facial dysmorphic traits (epicanthus, short palpebral fissures, short nose with anteverted nostrils, carp-shaped mouth, triangular nose), protruding tongue, attention/concentration disorders, stereotypic movements, anxiety. Cryptorchidism, birthmark (on the leg) and minor hand anomalies.

Detection of Heinz bodies (2003): negative

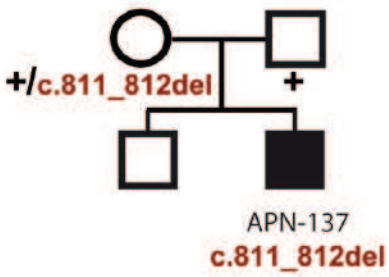
New test performed subsequently to NGS results (2013): 1.2% Heinz bodies detected.

X-inactivation bias detected in the mother (99:1) and the sister (98:2), subsequently to NGS results (2013).

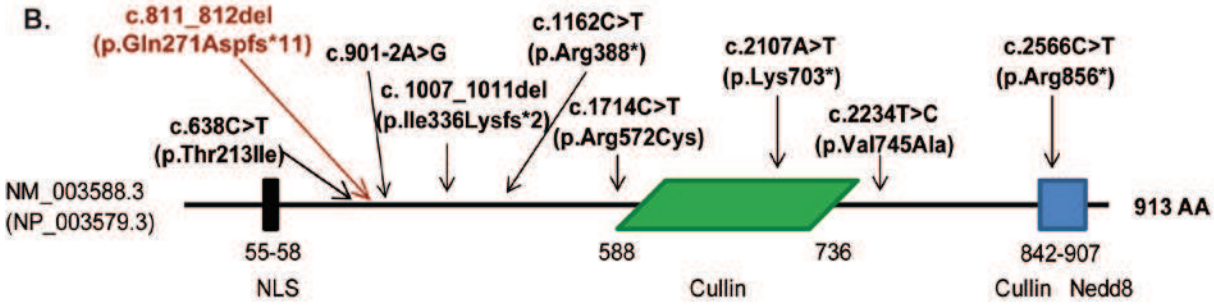
Figure S11. *CUL4B*, one causative frameshift mutation

APN-137: c.811_812del (NM_003588.3), p.Gln271Aspfs*11, hemizygous, maternally-inherited

A.



B.



C.

Clinical findings	Patients	
	Number	%
Intellectual disability	29/29	100%
Motor delay	12/12	100%
Speech delay	25/25	100%
Short stature	14/18	78%
Seizures	12/17	71%
Tremors	13/19	68%
Gait ataxia	13/19	68%
Aggressive outbursts	13/22	59%
Macrocephaly	8/18	44%
Obesity	15/25	60%
Pes cavus	8/14	57%
Smal testes	12/20	60%
Prominent lower lip	12/23	52%
Kyphosis	4/19	21%
Abnormal toes	11/14	79%

A. Family tree of patient APN-137; **B.** Representation of protein CUL4B with its different domains. Patient's mutation (in red) and mutations previously identified in ID (in black) are also indicated **C.** Table of recurrent clinical findings and their incidence in patients carrying *CUL4B* mutations (adapted from [58]).

Previous implication of *CUL4B* in ID

Missense and truncating mutations in *CUL4B* were reported in several families with ID, short stature, hypogonadism, abnormal gait, and more variable features such as speech delay, prominent lower lip and tremor (Cabezas type) [58-60].

Patient APN-137 (male, born in 2009)

Preliminary genetic tests:

Fragile-X test: negative; Array-CGH: negative; Karyotype: negative

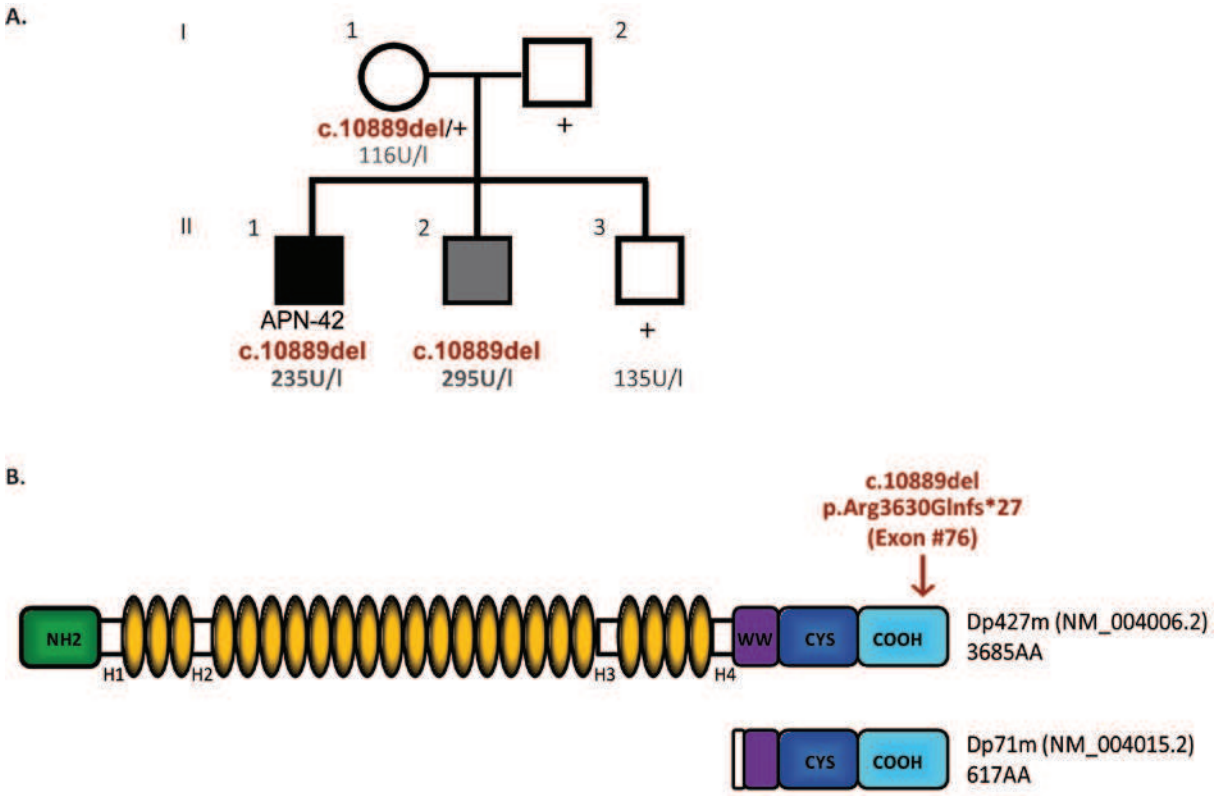
Angelman/Prader Willi test: negative

RSK2, ZEB2 (Mowat-Wilson syndrome) direct sequencing: negative

Clinical information: Severe ID, motor delay (no acquired walking at 4 years-old), absence of speech, autistic traits, sleep disturbances, behavioral disorders, constipation. Facial dysmorphic traits (wide mouth with prominent lower lip, depressed nasal bridge).

Figure S12. *DMD*, one causative frameshift mutation in two affected brothers without muscular phenotype

APN-42: c.10889del (NM_004006.2), p.Arg3630Glnfs*27, hemizygous, maternally-inherited



A. Pedigree of patient APN-42: his brother II-2 presents with important hypotonia, cerebellar dysplasia and developmental delay, while the youngest brother II-3 is unaffected. Serum Creatin PhosphoKinase (CPK) levels are indicated for all three brothers and the mother, showing slightly elevated levels for affected probands (normal values: <300 U/l); **B.** The protein encoded by the longest DMD transcript, DP427m, is represented with the different domains: N-terminal domain involved in actin binding (green), spectrin repeats (yellow), WW and EF/ZZ (CYS) domains (purple) involved in β-dystroglycan binding and the C-terminal domain (blue). The protein encoded by the shortest DMD transcript Dp71, which is the most abundant in brain is also represented. Localization of the mutation is indicated in red, and is one of the most distal truncating variant ever described in the gene (premature stop codon lies at the beginning of exon #77). The absence of muscle phenotype in our patient may be explained by an inefficient nonsense-mediated mRNA decay, allowing the translation of C-terminal truncated dystrophin and Dp71 that still contain the domain associated to the sarcoglycan/dystroglycan complex in the muscle, but lack the distal part of Dp71 that may be necessary for brain function.

Previous implication of *DMD* in ID

Mutations in *DMD* can cause Duchenne muscular dystrophy (DMD, MIM #310200) or the milder form Becker muscular dystrophy (BMD, MIM #300376). Muscular dystrophy is often accompanied by intellectual disability, especially when the mutation affects also the DP71 isoform and is located within the C-ter of the protein [8,61,62]. Depending on the specific mutation in the C-ter region, the muscular phenotype varies from mild BMD to DMD. The variability of phenotypes reported in patients with distal frameshift mutations (affecting exon #70 and higher) most probably implicates variable efficiency of nonsense mediated mRNA decay (NMD) [63]. More recently, a single amino-acid deletion in the EF domain was identified in DMD within the linkage region of one large family with non-specific mild to moderate XLID [64]. There was no muscular phenotype in the affected members of the family and CPK levels in the proband were just above the normal threshold (279U/l *vs*. normal range in Netherland: 30-200U/l).

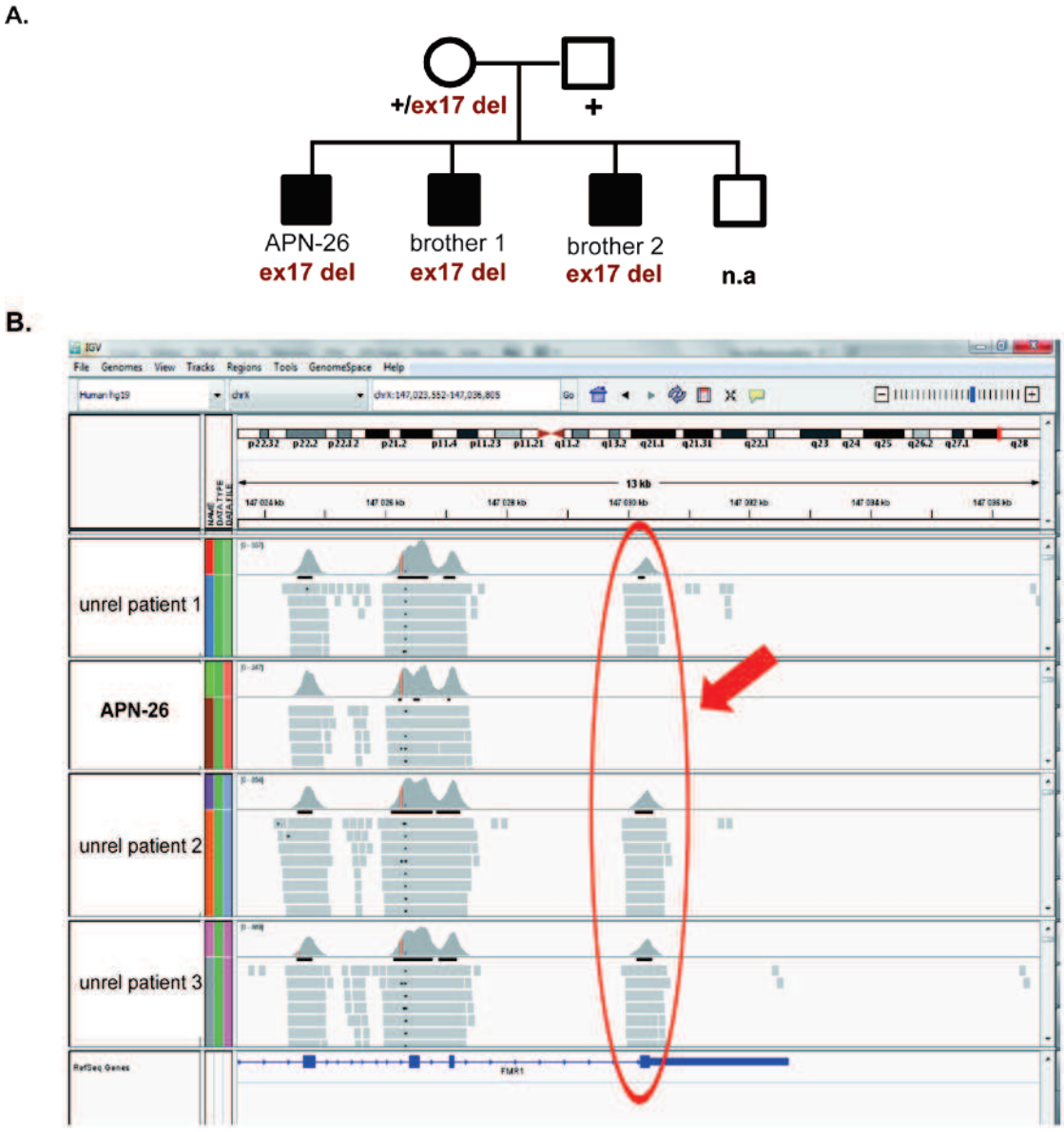
Patient APN-42 (male, born in 2003)
Preliminary genetic tests:
Fragile-X test: negative; Array-CGH: heterozygous 8p23.2 deletion (exons 2-5 of *MCPH1*) inherited from the mother; Karyotype: negative
ARX direct sequencing: negative

Clinical information: Moderate ID, absence of speech, developmental delay, important behavioral disorders, facial dysmorphic traits (narrow forehead, pointed nose, hanging columella, low-set ears with posterior rotation). Slightly elevated CPK-levels 235U/L but within normal range (20-300U/l)

Affected brother II-2 (male,born in 2009): Important hypotonia, maybe due to fetal distress (APGAR Score: 4 at 1 minute and 10 at 5 minutes). Cerebellar dysplasia and developmental delay. Slightly levated CPK-levels 295U/L but within normal range (20-300U/l).

Figure S13. *FMR1*: one causative exon deletion in 3 affected brothers

APN-26: exon 17 hemizygous deletion, maternally inherited



A. Pedigree of patient APN-26; **B.** Visualization of the sequencing data in Integrative Genome Viewer (IGV) showing deletion of the last coding exon in proband,

Previous implication of *FMR1* in ID

Mutations in *FMR1* are responsible for fragile-X syndrome (MIM #300624). The vast majority of reported patients carry the trinucleotide $(CGG)_n$ repeat expansion of greater than 200 repeats. Fragile-X syndrome is characterized by moderate to severe ID, macroorchidism and characteristic facial features (long face, large ears, and prominent jaw). Additional clinical features may include hyperlaxity, skeletal manifestations (scoliosis, pectus excavatum), strabismus and epilepsy.

Other reported mutations (besides the CGG expansion) in patients are: 1) a *de novo* missense mutation (p.Ile304Asn reported in a single patient with a severe form of fragile-X syndrome [65]; 2) a *de novo* frameshift (c.373delA; p.Thr125Leufs*35) in a patient with developmental delay and typical facies, 3) an inherited splice mutation in an adult male with classical fragile-X syndrome [66]; 4) an inherited stop mutation (p.Ser27*) in a man with classic features of fragile-X syndrome [67]; and 5) a potentially-pathogenic missense variant (c.413G>A; p.Arg138Gln) in a patient with intellectual disability [68].

Family of patient APN-26 (male, born in February, 1954)

Preliminary genetic tests:

Fragile-X: negative; Array-CGH: negative

Direct sequencing of *MED12, ZDHHC9, UPF3B, FBN1, TGFBR2, MECP2, ARX*: negative

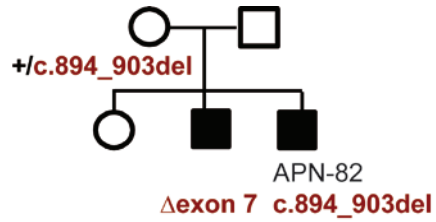
Clinical information:

- APN26: Moderate to severe ID with epilepsy, scoliosis, pectus excavatum, strabismus
- Brother 1: Moderate ID without epilepsy, facial features evocative of Fragile-X syndrome (long face and large ears), pectus excavatum, arachnodactyly.
- Brother 2: Mild ID without epilepsy.

No other familial history of ID reported in families of the three maternal aunts or of the two maternal uncles.

Figure S14. *IL1RAPL1*, one causative frameshift mutation and exon 7 deletion in two affected brothers

APN-82: c.894_903del (NM_014271.3), p.Trp299Thrfs*18, hemizygous mutation maternally-inherited
APN-82's brother: exon 7 deletion, hemizygous *de novo*



Pedigree of patient APN-82, showing the two different hemizygous mutations in *IL1RAPL1* carried by both affected brothers: one deletion of 10-bp in exon 7 identified in the proband and maternally inherited, and a *de novo* deletion of the entire exon 7 in the affected brother. PCR confirmed that exons 6 and 8 are not deleted in *IL1RAPL1* in APN-82's brother. Variants in genes located on each side of *IL1RAPL1* confirmed that the deletion of exon 7 appeared on the chromosome originally carrying the small 10-bp deletion, suggesting that this small deletion might have provoked a genomic instability leading to the larger deletion observed in the second brother.

Previous implication of *IL1RAPL1* in ID

Mainly pericentric inversions and numerous intragenic deletions have been reported in patients with ID or ASD, supporting that *IL1RAPL1* is highly susceptible to abnormal recombination events [69]. Two small deletions and two truncating point mutations have also been described.

Nature of reported mutations	References
Pericentric inversions	
Pericentric inversion (inv(X)(p21.3q27.1)	[70]
Pericentric inversion (inv(X)(p21.3q27.1)	[71]
Pericentric inversion (inv(X)(p22.1q13)	[72]
Intragenic deletions	
del exons 1-5	[73]
del exon 2	[74]
del exons 2-5	[75]
del exons 2-5	[73]
del exons 3-5	[74]
del exons 3-5	[76]
del exons 3-7	[77]
del exon 7	*present study*
del exons 3-11	[78]
Small deletions	
del 7nts in exon 9	[77] (ASD patient)
del 10 nts in exon 7	*present study*
Point mutations	
c.1377C>A p.Y459*	[76]
c.1460G>A p.W487*	[79]

Patient APN-82 (male, born in 2009)

Preliminary genetic tests:
Fragile-X: negative; Array-CGH: negative

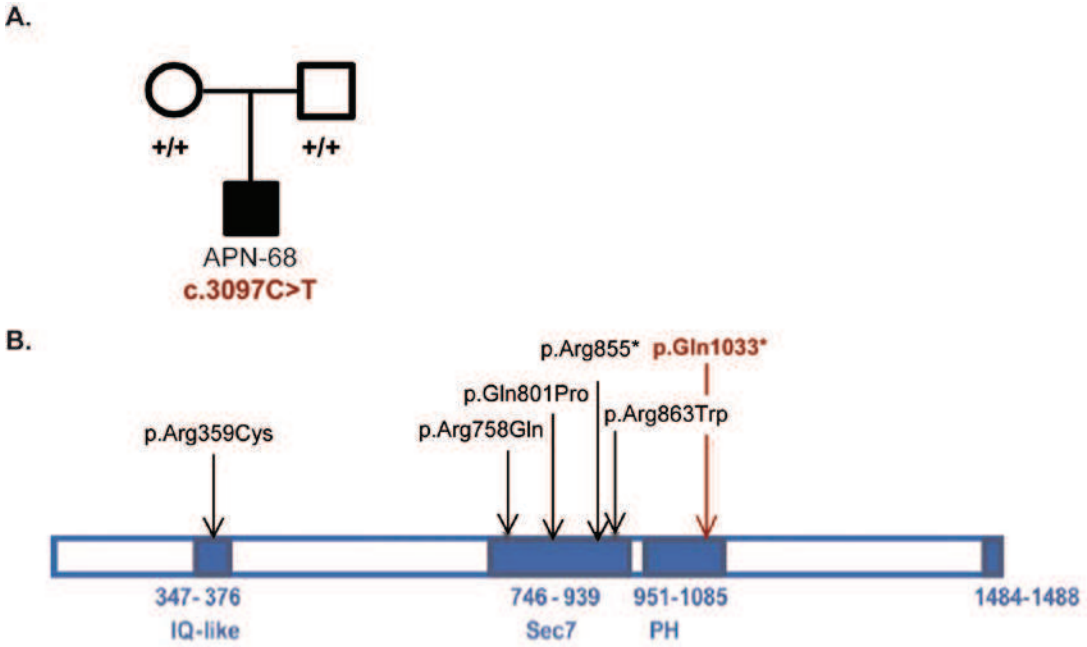
Clinical information: Moderate ID, speech delay, hypotonia. At 29 months: weight: 12kg, height: 88.cm (-1SD), OFC: 48.5cm (-1SD).
MRI: some anomalies detected.

Affected brother (male, born in 2006)

Clinical information: Moderate ID with pervasive developmental disorders, delayed motor development, poor speech, anxiety, hypotonia.

Figure S15. *IQSEC2*, one causative *de novo* nonsense mutation

APN-68: c.3097C>T (NM_001111125.1), p.Gln1033*, hemizygous *de novo*



A. Pedigree of patient APN-68; **B.** Representation of IQSEC2 and distribution of the mutations previously reported in ID patients (adapted from [80]).

Previous implication of *IQSEC2* in ID

Missense mutations in *IQSEC2 (*p.Asn801Pro; p.Arg758Asn; p.Arg359Cys, p.Arg863Trp) were identified in four large families with NS-ID (some individuals had ASD or epilepsy) [80]. All four mutations affect the GEF (guanine nucleotide exchange factor) activity of IQSEC2. Recently, a *de novo* nonsense mutation (p.Arg855*) was identified in a boy with ID and unspecific facial features with convergent strabismus [50]. Tran Mau-Them F et al., reported two additional male patients with duplications disrupting *IQSEC2* [81]. They compared the phenotype of all patients with *IQSEC2* truncating mutations, describing a severe syndromic XLID phenotype characterized by severe ID, progressive microcephaly with postnatal onset, severe delayed motor and speech development, midline hand stereotypic traits, additional autistic traits and seizures.

Patient APN-68 (male, born in 2005)

Preliminary genetic tests:

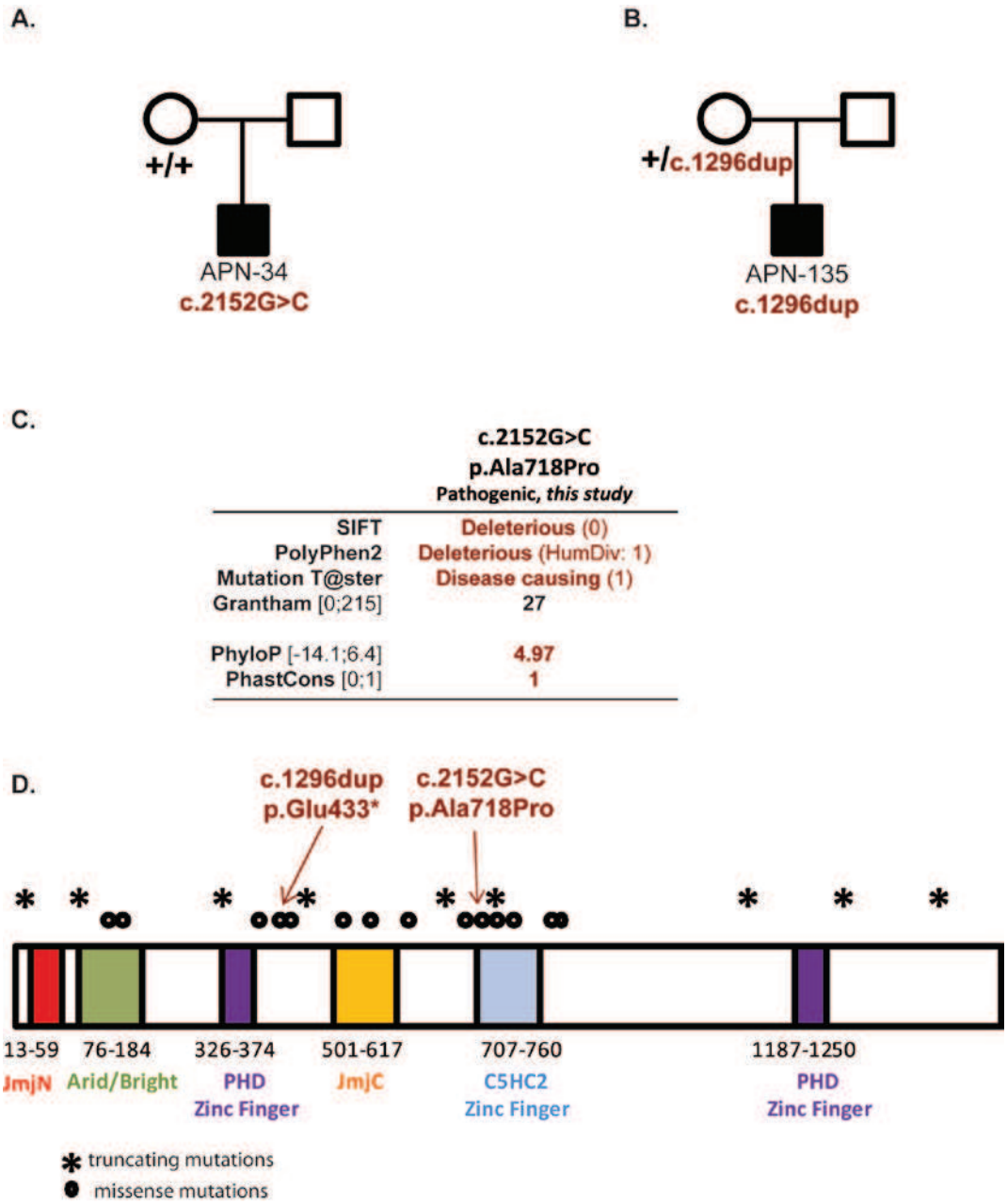
Fragile-X test: negative; Array-CGH: dupXp22.3 inherited from the mother, concluded as non-pathogenic

MECP2, *ARX* and *CDKL5* direct sequencing: negative.

Clinical information: Severe ID, no acquired language, reduced autonomy, motor developmental delay, severe epilepsy, strabismus, stereotypic features and behavioral disorder. The phenotype overlaps that was reported by Tran Mau-Them et al. for patients carrying truncating *IQSEC2* mutation but without microcephaly [81].

Figure S16. *KDM5C*, two causative mutations

APN-34: c.2152G>C (NM_004187.3), p.Ala718Pro, hemizygous *de novo*
APN-135: c.1296dup (NM_004187.3), p.Glu433*, hemizygous, maternally inherited



A-B. Family trees of both patients with *KDM5C/JARID1C* mutations detected; **C.** Associated prediction scores of the pathogenic missense mutation identified in this study showing a predicted deleterious status and a high conservation at the nucleotide level; **D.** Representation of KDM5C and its domains with localization of these detected mutations and those previously reported in ID patients.

Previous implications of *KDM5C* in ID

The resequencing of 47 genes in the Xq28 region in MRX families led to the identification of 7 mutations in *KDM5C* [82]. Many other mutations have been described afterwards,[83-86] in other patients with intellectual disability. Recurrent clinical symptoms include hyperreflexia, short stature, aggressive behavior and seizures. Microcephaly is reported unevenly [87]. Mutations in *JARID1C/KDM5C* appear to be a frequent cause of intellectual disability.

Patient APN-34 (male, born in 2004)
Preliminary genetic tests:
Fragile-X test: negative; Array-CGH: negative; Karyotype: negative
Subtelomeric rearrangement: negative
FOXP2 direct sequencing: negative

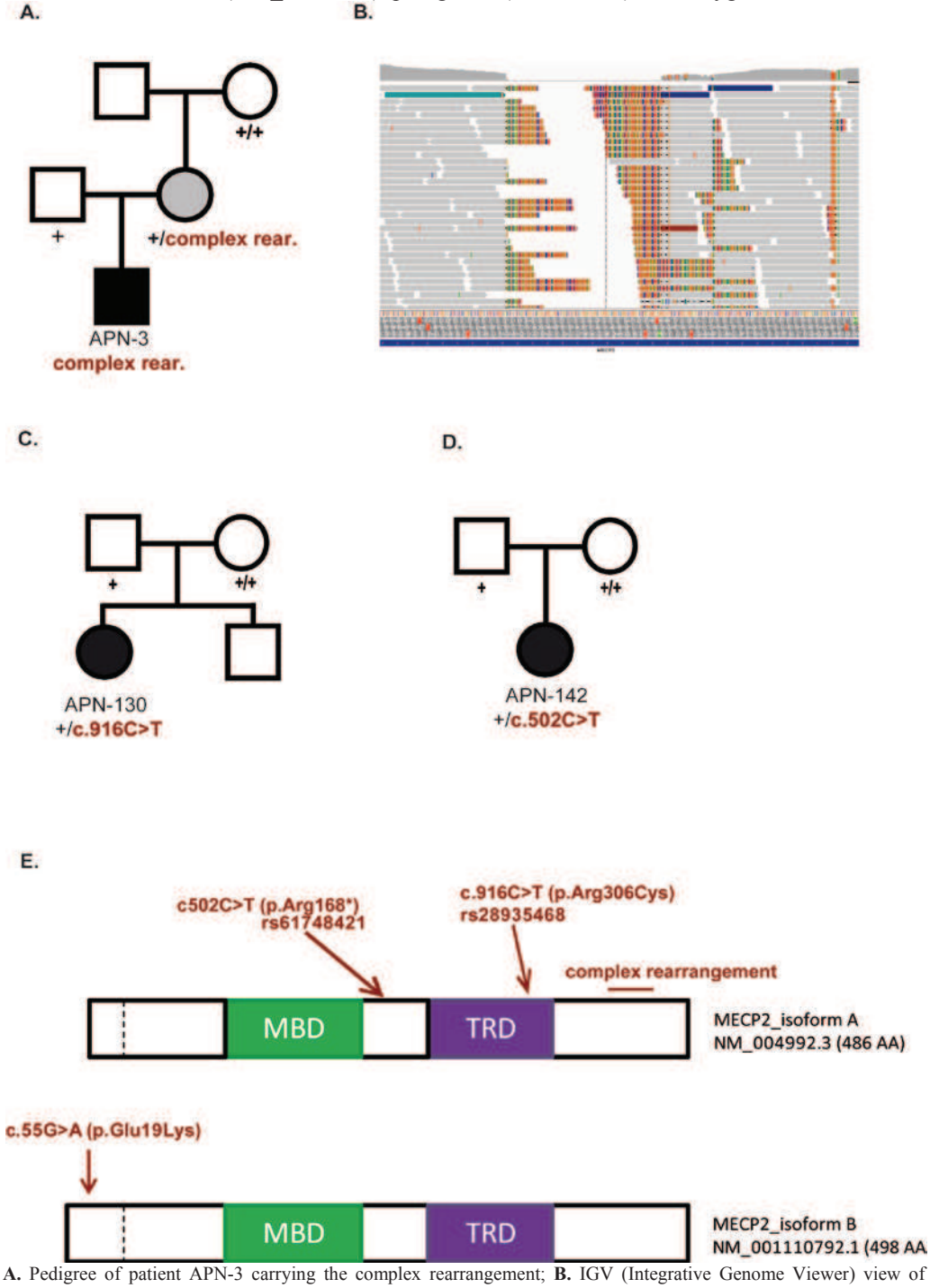
Clinical information: Moderate ID with hyperactivity and friendly behavior despite some aggressive bursts, poor language with speech apraxia, pyramidal syndrome, strabismus and hypermetropia. At 8 years-old: OFC=53 cm (mean), height=125 cm (mean), weight = 29 kg (90th centile).

Patient APN-135 (male, born in 2004)
Preliminary genetic tests:
Fragile-X test: negative; Array-CGH: negative; Karyotype: negative
ARX direct sequencing: negative

Clinical information: Moderate ID, poor language (only a few words), sleep disturbances, PDD diagnosed. Behavioral disturbances (aggression). Minor facial features, clinodactyly, astigmatism.
At birth, weight: 2600 g, height: 45 cm, OFC: 34 cm. At 9 years-old, weight: 25 kg, height: 119 cm (-2SD), OFC: 51 cm. Treatment with growth hormone since November, 2009.

Figure S17. *MECP2*, three causative mutations

- APN-3: c.[954A>T; 957_960delins4bp;1060delCins 6bp; 1097_1235delins50bp](NM_004992.3); (p.[Glu318Asp; Val320His; Arg354_Val412delins41]) hemizygous complex rearrangement
- APN-130: c.916C>T (NM_004992.3); p.Arg306Cys (rs28935468), heterozygous *de novo,* known mutation
- APN-142 : c502C>T (NM_004992.3); p.Arg168* (rs61748421), heterozygous *de novo,* known mutation



A. Pedigree of patient APN-3 carrying the complex rearrangement; **B.** IGV (Integrative Genome Viewer) view of the sequencing data showing complex rearrangement in proband; **C., D.** Pedigrees of APN-130 and APN-142 carrying *de novo* mutations, previously reported in Rett patients; **E.** Representation of both *MECP2* isoforms (A and B) and localization of the four variants/mutations identified in patients.

Previous implication of *MECP2* in cognitive disorders

MECP2 is mutated in Rett syndrome (MIM #312750), a severe neurodevelopmental disorder that affects mostly females, and characterized by an early-stage arrested development and regression of skills, speech loss, specific stereotypic features ("hand-flapping"), microcephaly, seizures and intellectual disability. Since the discovery of *MECP2* as the cause of Rett syndrome in 1999 [88], MECP2 mutations have also been reported in males [89-95]. These males phenotypically have classical Rett syndrome when the mutation arises as somatic mosaicism or when they have an extra X-chromosome [89,96]. Otherwise, males with MECP2 mutations show diverse phenotypes: from severe early-onset congenital encephalopathy associated with microcephaly, hypotonia, seizures, respiratory irregularities, constipation and intellectual disability with various neurological symptoms to mild cognitive impairment [95]. The last exon of *MECP2* is known to be the one accumulating most mutations in patients with the first 5' half mainly subject to point mutations, while the 3' half is a hotspot for rearrangements [97-99].

Patient APN-3 (male, born in 2001)

Preliminary genetic tests:

Fragile-X test: negative; Array-CGH: negative

FISH 22q13, 1qter, 1p36, 15q11-q13, Coffin-lowry, CDG, Angelman/Prader-Willi, Telomeric rearrangement screening: negative

X-inactivation bias: negative

Clinical information: Severe ID, epilepsy, myotonia, ataxia, hypotonia. Progressive myoclonic encephalopathy with cerebellar and pyramidal syndrome. Overweight. Microcephaly (-3SD), dysmorphic traits, photophobia.

His mother presents with speech delay, dyslexia, sleep and attention disorders, and overweight.

Proband has a young maternal half-brother born in September 2009 who presents with similar clinical features (severe encephalopathy, ataxia, hypotonia, sleep and behavioral disorder) and also carries the mutation.

Patient APN-130 (female, born in 2003)

Preliminary genetic tests:

Fragile-X test: negative; Karyotype: negative; Array-CGH: negative;

FISH 17p11.2: negative

Angelman/Prader-Willi, *GNAS1* direct sequencing: negative

Telomeric rearrangements screening by MLPA: negative

Clinical information: Normal development during the first year. Walking at 12 months. Stagnation of development starting at 14 months-old, with walking denial during 20 months subsequently to a fall. Severe ID, absence of speech. Autistic features, which appeared between 1-2 years-old, but no typical Rett phenotype. At 10 years, presence of some median hand movements. Behavioral disorders including hyperactivity and sleep disorders. Hyperactivity and aggressive behaviors progressively improved starting from 8 years-old.

Overweight (+5SD) in a context of familial obesity. Facial dysmorphic traits (narrow forehead, prominent cheeks, small and deep-set eyes, turned-up nose). Birth measurements: weight: 4100g, height: 53cm, OFC: 33 cm. At 9 year-old: weight: 66 kg (+5SD), height: 141cm (+1.5SD), OFC : 53.5cm (+1SD).

Patient APN-142 (female, born in 2006)

Preliminary genetic tests:

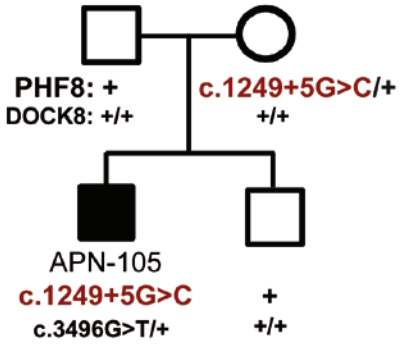
Fragile-X test: negative; Array-CGH: negative; Karyotype: negative

Clinical information: Severe ID, no speech. Developmental delay followed by a stagnation and regression (3 years-old). Autistic and stereotypic features ('hand-flapping'). Measurements at birth, weight: 2640g, height: 48cm, OFC: 33cm. At 6 year-old, weight: 24.3 kg (+2SD), height: 118 cm (+1SD), OFC: 51cm.

Figure S18. *PHF8*, one causative mutation affecting splicing

APN-105: c.1249+5G>C (NM_001184896.1), p.Tyr406Phefs*24 hemizygous, maternally inherited
(+ *DOCK8* : c.3496G>T (NM_203447.3), p.Glu1166*, heterozygous, *de novo)*

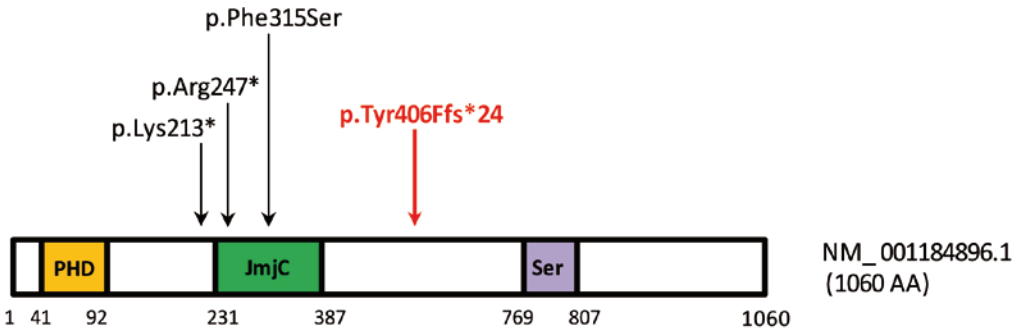
A.



B.

	Usual DSS wild-type DSS	Usual DSS mutant (c.1249+5G>C)	Cryptic DSS c.1214G
MaxEnt splice	7,7	4,5 (- 42%)	7,1
NnSplice	0,90	- (- 100%)	0,74
GeneSplicer	4,9	2,7 (- 45%)	2,3
HSF	82,8	70,7 (- 15%)	84,5

C.



A. Pedigree of patient APN-105 carrying both the *PHF8* mutation and a *de novo* variant in *DOCK8*; **B.** Prediction scores for donor splice sites (DSS) as computed by MaxEnt, Nnsplice, GeneSplicer or Human splicing Finder (HSF): prediction scores for the normal DSS with the wild-type allele and the 1249+5G>C mutation are given as well as the percentage of score decrease created by this mutation. Prediction score for the alternative cryptic DSS is also indicated. Patient blood mRNA was analyzed and confirmed the use of this alternative cryptic DSS instead of the usual DSS in over 4/5 of the transcripts; **C**. Representation of PHF8 domains: JmjC, PHD-type and Ser-rich domains with the present (red) and previously (black) reported mutations. **D**. Pictures of patient APN-105

Previous implication of *PHF8* in ID

Mutations in *PHF8* (a 12-bp deletion affecting exon 8 splicing and a nonsense mutation p.Arg247*) were described in two families with Siderius-Hamel syndromic ID (MIM #300263, moderate ID along with cleft/lip palate) [100]. Phenotypes vary in these families from ID with cleft lip/cleft palate to non-specific ID. Screening of additional patients with cleft lip/cleft palate identified an additional nonsense mutation (p.Lys213*) in PHF8 and a missense (p.Phe315Ser) mutation [101,102].

Patient APN-105 (male, born in 2010)

Preliminary genetic tests:

Fragile-X: negative; Array-CGH: negative

MEF2C direct sequencing: negative

12p tetrasomy: negative

X-inactivation bias in APN-105's mother: 91-9%

Clinical information: Mild ID, motor delay (sitting: 18 months, walk: 31 months), language delay. Hypotonia with hyperlaxity, stereotypic movements, behavioral disorders. Facial dysmorphic traits (hypertelorism, wide mouth, high hairline with golf temporal, clefts ears).

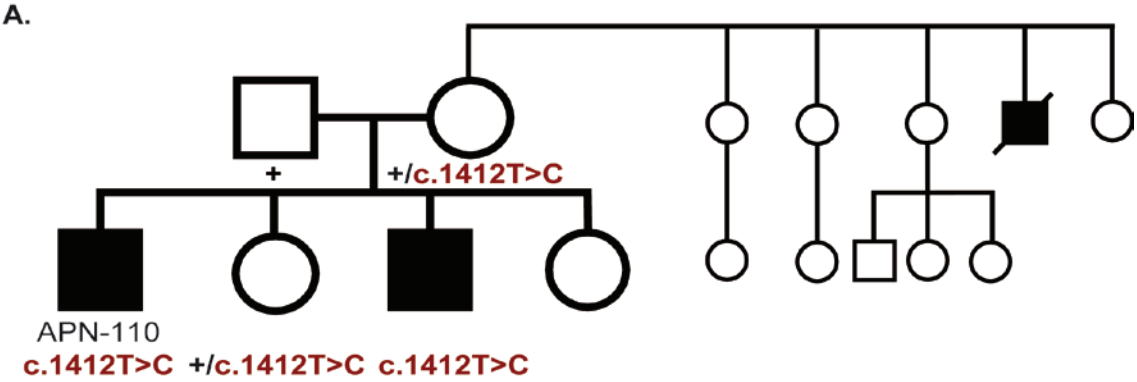
CAA, CAO blood/urine, lactate/pyruvate, glycosylation, creatine, ceruloplasmine: normal

MRI: ventricular enlargement

Figure S19. *SLC16A2/MCT8*, one causative missense mutation

APN-110: c.1412T>C (NM_006517.3), p.Leu471Pro (rs122455132), hemizygous, known mutation

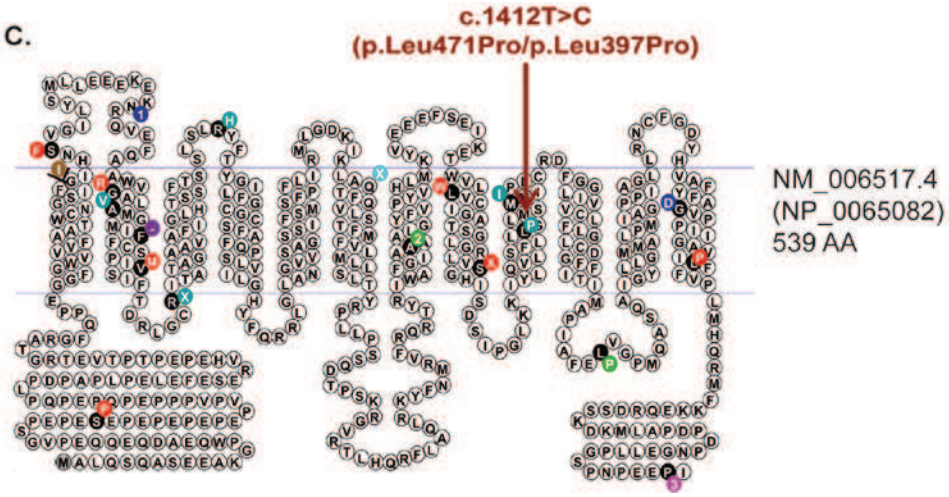
A.



B.

	c.1412T>C p.Leu471Pro Pathogenic ([103] & this study)
SIFT	Deleterious (0.01)
PolyPhen2	Deleterious (HumDiv: 0.997)
Mutation T@ster	Disease causing (1)
Grantham [0;215]	98
PhyloP [-14.1;6.4]	4.73
PhastCons [0;1]	1

C.



A. Pedigree of APN-110 showing segregation of the missense mutation with the disease; **B.** Associated prediction scores of the pathogenic missense mutation identified in APN-110 in this study (initially reported in [103]), showing a predicted deleterious status and a high conservation at the nucleotide level; **C.** Representation of the protein and its transmembrane domains, and localization of the affected aminoacid within the MCT8 protein (adapted from [104]).

Previous implication of *SLC16A2/MCT8* in ID

Patients with mutations in *SLC16A2/MCT8* show severe psychomotor retardation combined with elevated level of thyroid hormone T3 (triiodothyronine) in serum. The mutation p.Leu397Pro/ *p.Leu471Pro* was initially reported in a patient with severe psychomotor retardation, axial hypotonia, spastic quadriplegia, dystonic movements and absence of speech [103]. This mutation was shown to reduce SLC16A2/MCT8 protein levels and affect T3 uptake in JEG3 cells [105].

Patient APN-110 (male, born in April, 1986)

Preliminary genetic tests:

Fragile-X test: negative; Array-CGH: negative

ARX, *MECP2* direct sequencing: negative.

X-inactivation bias identified in patient's mother.

Clinical information: Severe ID, no sitting position (kyphosis), no walking, absence of speech. Poor autonomy. Major hypotonia without epilepsy. No facial dysmorphic traits. No dystonic movement.

Affected brother: (male, born in November, 1997): Kyphosis, no acquired walking, absence of speech. Mild growth retardation.

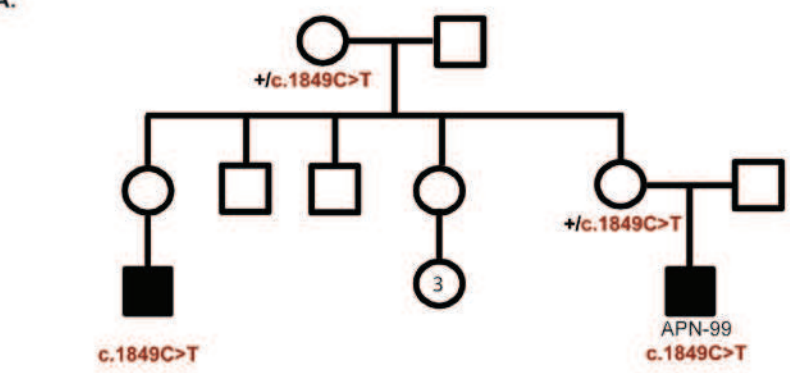
Serum thyroid hormone T3L levels confirmed the genetic diagnostic: 11.2 pmol/l (APN-110) and 14.5 pmol/l for his affected brother (normal range values: 3.1-6.8).

Maternal uncle: death at 23 year-old, with clinical features highly similar to his nephews.

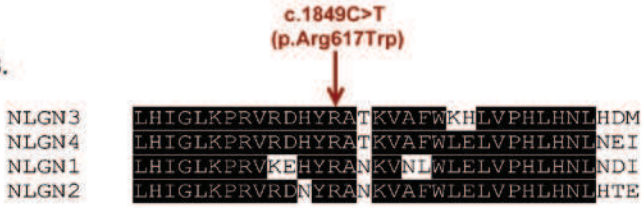
Figure S20. *NLGN3*, one potentially-causative missense mutation

APN-99: c.1849C>T (NM_181303.1); p.Arg617Trp, hemizygous

A.



B.



C.

	c.1849C>T p.Arg617Trp (probably pathogenic)	c.1411C>T p.Arg471Cys (pathogenic, see [106])
SIFT	Deleterious (0.00)	Deleterious (0.00)
PolyPhen2	Deleterious (HumDiv: 1)	Deleterious (HumDiv: 1)
Mutation T@ster	Disease causing (1)	Disease causing (1)
Grantham [0;215]	101	180
PhyloP [-14.1;6.4]	3.840	3.600
PhastCons [0;1]	1	1

A. Pedigree of patient APN-99 showing a probable X-linked mode of transmission of the disease; **B.** Alignment of the human neuroligin paralogs showing conservation of the affected Arg617 residue; **C.** Associated prediction scores comparing the probably pathogenic missense variant identified in this study with the initial pathogenic missense mutation described in [106], showing both a predicted deleterious status and a very high conservation at the nucleotide level**.**

Previous implication of *NLGN3* in ID

Screening of *NLGN3* in ASD individuals led to the identification of a missense c.1411C>T (p.Arg471Cys) mutation in two brothers (one with typical autism the other with Asperger syndrome)[106] localized in the carboxylesterase domain. The *NLGN3* knock-in mouse displayed an increase in inhibitory synaptic transmission and an "autistic-like" phenotype [107,108]. Talebizadeh et al. identified an alternative transcript of *NLGN3* in lymphoblastoid cells that lacked exon 7 and encoded a new truncated protein, present in all 30 control individuals and in all but one of the 10 ASD females tested [109]. The authors speculated that the lack of the specific truncated isoform in this female patient might be implicated in her autistic phenotype.

Patient APN-99 (male, born in 2007)

Preliminary genetic tests:

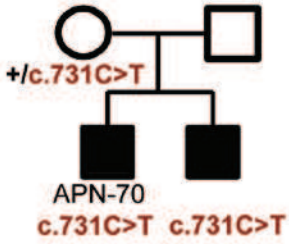
Fragile-X test: negative; Array-CGH: del 5q15 inherited from the father; Karyotype: negative

Clinical information: Severe ID with ASD and sleep disorders. Absence of speech. One maternal cousin with ID and autism, also carrying the variant.

Figure S21. *PQBP1*, one potentially-causative missense mutation in two affected brothers

APN-70: c.731C>T (NM_005710.2), p.Pro244Leu, hemizygous, maternally inherited

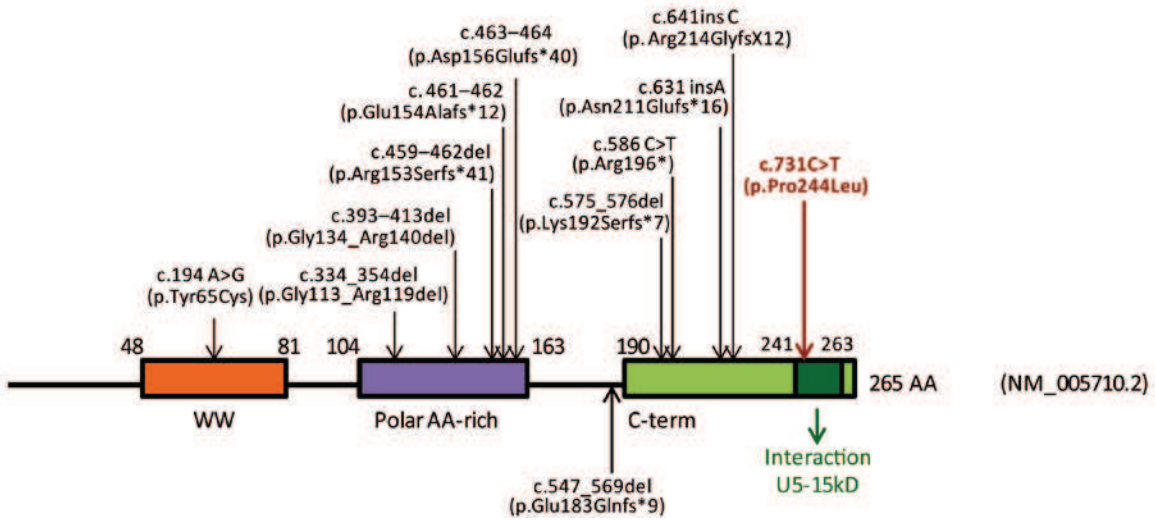
A.



B.

	c.731C>T p.Pro244Leu (probably pathogenic, this study)	c.194A>G p.Tyr65Cys (pathogenic, see [110])
SIFT	Deleterious (0.00)	Deleterious (0)
PolyPhen2	Deleterious (HumDiv: 1)	Deleterious (HumDiv: 1)
Mutation T@ster	Disease causing (1)	Disease causing (1)
Grantham [0;215]	98	194
PhyloP [-14.1;6.4]	5.13	3.76
PhastCons [0;1]	0.99	1

C.



A. Pedigree of patient APN-70; **B.** Associated prediction scores comparing the probably pathogenic missense variant identified in this study with the only pathogenic missense mutation described in [110], showing for both a similar predicted deleterious status and a high conservation at the nucleotide level; **C.** Representation of the different PQBP1 protein domains: the WW domain is responsible for the interaction with RNA polymerase, the polar amino acid-rich domain interacts with polygutamine-interacting proteins and the C-terminus domain contains a region involved in the interaction with the spliceosomal protein U5-15kD [111].

Previous implication of *PQBP1* in ID

Truncating mutations (and one variation leading to a missense p.Tyr65Cys) were reported in *PQBP1* in several X-chromosome-linked intellectual disability (XLID) disorders, such as Renpenning (MIM #309500), Sutherland-Haan, Hamel, Porteous, and Golabi-Ito-Hall (GIH) syndromes [110,112]. These syndromes share similar clinical features: in addition to severe intellectual disability, patients also have a lean body, short stature, microcephaly, and are frequently diagnosed with cardiac abnormalities [113].

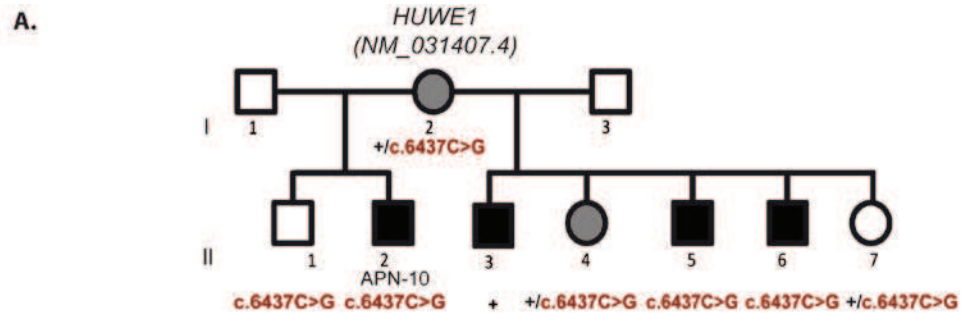
Patient APN-70 (male, born in 1988)

Preliminary genetic tests:
Fragile-X: negative; Array-CGH: dup 16q22.2 inherited from the father

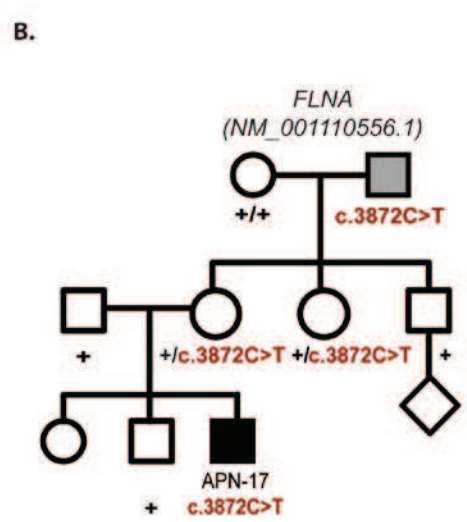
Clinical information: Moderate ID, poor autonomy (poor speech/reading/writing), communication and social interaction disorders, learning difficulties. Tip-toe walking at the beginning, autistic behavior that started at 3 years of age, stereotypic movements.

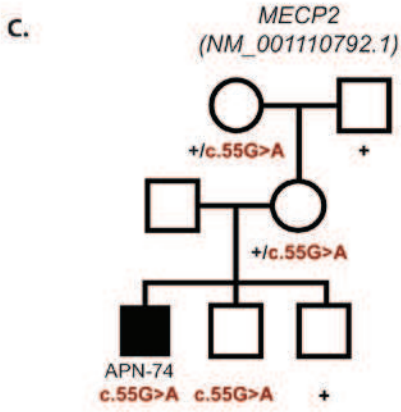
Figure S22. Other candidate variants not or ambiguously co-segregating with ID in probands' families

- APN-10: c.6437C>G (*HUWE1*, NM_031407.4), p.Thr2146Arg, hemizygous; reported in an unaffected brother
- APN-17: c.3872C>T (*FLNA*, NM_001110556.1), p.Pro1291Leu (known, rs137853319), hemizygous; maternally inherited, present in maternal grandfather with late onset psychiatric disorders
- APN-74: c.55G>A (*MECP2*, NM_001110792.1), p.Glu19Lys, hemizygous; reported in an unaffected brother
- APN-132: c.1312G>A (*FMR1*, NM_002024.5), p.Asp438Asn, hemizygous, maternally inherited

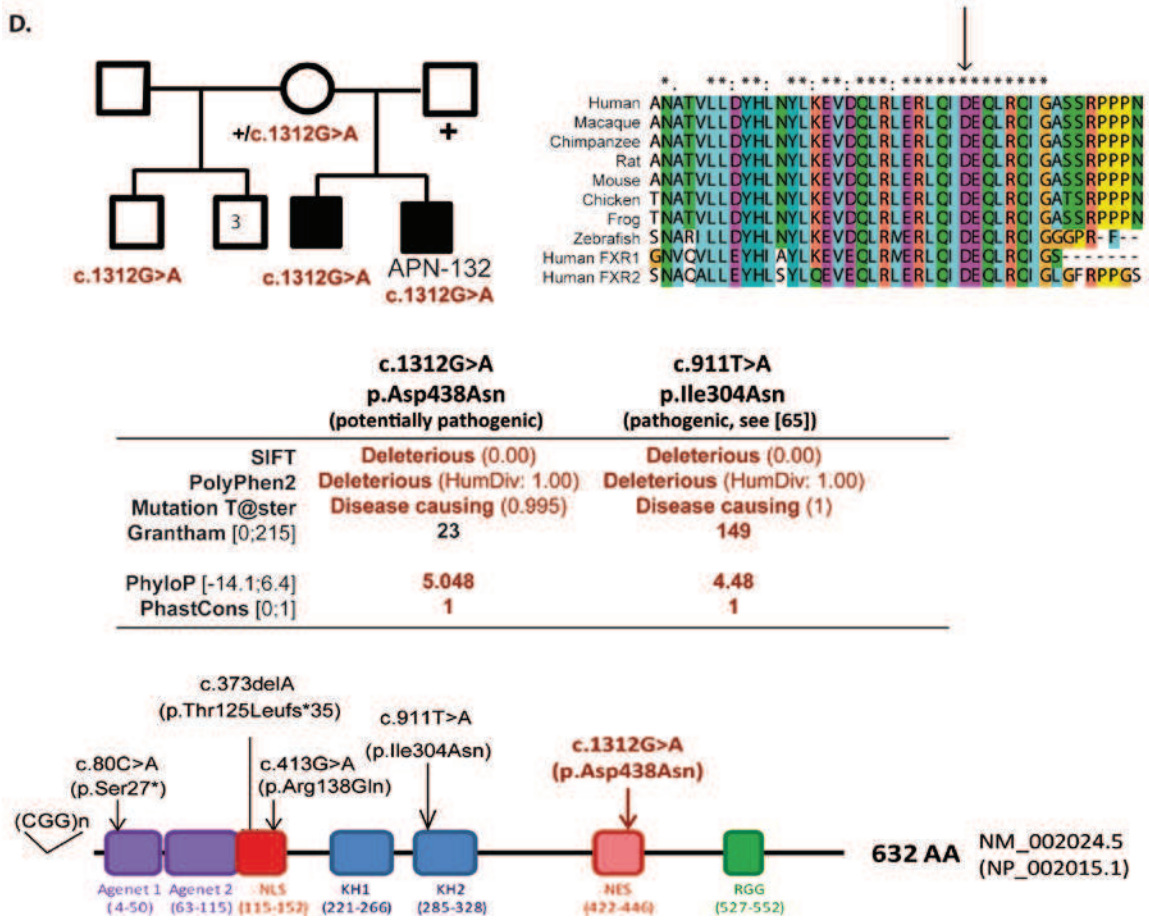


	c.6437C>G p.Thr2146Arg (benign, this study)	c.8942G>A p.Arg2981His (pathogenic, see [114])	c.12037C>T p.Arg4013Trp (pathogenic, see [114])	c.12559C>T p.Arg4187Cys (pathogenic, see [114])
SIFT	Deleterious (0.00)	Deleterious (0.00)	Deleterious (0.00)	Deleterious (0.01)
PolyPhen2	Deleterious (0.989)	Deleterious (0.999)	Deleterious (0.999)	Deleterious (1)
Mutation T@ster	Disease causing (1)	Disease causing (1)	Disease causing (1)	Disease causing (1)
Grantham [0;215]	71	29	101	180
PhyloP [-14.1;6.4]	3.595	5.610	1.980	3.110
PhastCons [0;1]	1	1	1	1





	c.55G>A p.Glu19Lys
SIFT	Deleterious (0)
PolyPhen2	Benign (HumDiv: 0.05)
Mutation T@ster	Disease causing (1)
Grantham [0;215]	56
PhyloP [-14.1;6.4]	2.22
PhastCons [0;1]	1



	c.1312G>A p.Asp438Asn (potentially pathogenic)	c.911T>A p.Ile304Asn (pathogenic, see [65])
SIFT	Deleterious (0.00)	Deleterious (0.00)
PolyPhen2	Deleterious (HumDiv: 1.00)	Deleterious (HumDiv: 1.00)
Mutation T@ster	Disease causing (0.995)	Disease causing (1)
Grantham [0;215]	23	149
PhyloP [-14.1;6.4]	5.048	4.48
PhastCons [0;1]	1	1

A. Partial pedigree of patient APN-10 carrying c.6437C>G, p.Thr2146Arg variant in *HUWE1*, large family highly suggestive of an X-linked inheritance of a disease causing mutation. The prediction scores for the missense variant are also indicated, along with those associated to previously reported missense mutations [114]; **B.** Pedigree of patient APN-17 with the *FLNA* c.3872C>T variant showing the carrier status of maternal grandfather presenting with psychiatric disorder; **C.** Pedigree of patient APN-74's family showing non-segregation of the *MECP2* c.55G>A, p.Glu19Lys missense variant, affecting the first exon of the major MeCP2_e1 transcript in brain. A p.Ala2Val missense variant was reported as disease causing in two patients [115,116]; **D.** Pedigree of patient APN-132, with protein alignment of FMR1 ortholog and paralog proteins FXR1 and FXR2, showing striking conservation of the affected Asp438 residue even in FXR1 and FXR2, associated prediction scores comparing the present non-segregating missense mutation to the well-known pathogenic missense mutation [65]; and representation of FMRP and its domains, and distribution of the few point mutations that were reported in *FMR1* in patients. The herein described missense mutation is indicated in red, located at the very end of the NES (nuclear export signal) domain.

A/ Previous implication of *HUWE1* in ID

Initially, duplications of *HUWE1* were identified in families and simplex cases with intellectual disability. More recently, three missense mutations affecting highly conserved residues and cosegregating with the ID status were identified in unrelated XLID-affected families [114,117].

Patient APN-10 (male, born in 2001)
Preliminary genetic tests:
Fragile-X test: negative; Array-CGH: negative
ARX (exon #2), *PQBP1* (exon #4), Prader-Willi: negative
Existence of an X-inactivation bias in the patient's mother

Clinical information: moderate ID, behavioral and attention disorders, myopia, facial dysmorphic traits, hyperlaxity, macrosomy. Large family with several affected males, highly suggestive of XLID.

B/ Previous implication of *FLNA* in ID

Mutations in *FLNA* can cause different disorders: loss-of-function mutations tend to cause periventricular heterotopia whereas missense mutations can lead to otopalatodigital syndrome, Melnick-Needles or fronto-metaphyseal dysplasia [118-120]. In a boy with type 2 FG syndrome, Unger *et al*. identified a hemizygous c.3872C>T, p.Pro1291Leu mutation in *FLNA* [121]. The boy presented with severe constipation, large rounded forehead, prominent ears, frontal hair upsweep, and mild delay in language acquisition. This missense is reported in one male from the Exome Variant Server, raising doubts about its pathogenicity.

Patient APN-17 (male, born in 1996)
Preliminary genetic tests:
Fragile-X test: negative; Array-CGH: negative

Clinical information: Severe ID with minimal autonomy. Absence of speech. Autistic features with hyperactivity, behavioral and sleep disorders. Epilepsy. No regression, no microcephaly.
MRI: normal (no sign of heterotopia)
Proband's maternal grandfather also carries the variant and presented later in life with schizophrenia, behavioral disorder, speech impairment, anxiety and stereotypic behaviors. He could read and worked for part of his life as a semi-skilled worker.

C/ Previous implication of *MECP2* in cognitive disorders
See **Figure S17**.

Patient APN-74 (male, born in 1988)
Preliminary genetic tests:
Fragile-X test: negative; Array-CGH: negative; Prader-Willi: negative

Clinical information: Moderate ID with partial autonomy (no reading, no writing). Epilepsy since 11 months of age with regression of motor acquisitions. Autistic features, attention disorder, polyphagy, stereotypic movements and aggressive behavior.

D/ Previous implication of FMR1 in cognitive disorders
See **Figure S13**.

Patient APN-132 (male, born in June, 2007) and his brother (male, born in April, 2006)
Preliminary genetic tests:
Fragile-X: negative; Array-CGH: negative; Karyotype: negative

Clinical information: Both boys present with severe ID, autistic features, behavioral and sleep disorders, and behavioral instability.

REFERENCES

1. Gecz J, Shoubridge C, Corbett M (2009) The genetic landscape of intellectual disability arising from chromosome X. Trends Genet 25: 308-316.
2. Lubs HA, Stevenson RE, Schwartz CE (2012) Fragile X and X-linked intellectual disability: four decades of discovery. Am J Hum Genet 90: 579-590.
3. Stevenson RE, Schwartz CE (2009) X-linked intellectual disability: unique vulnerability of the male genome. Dev Disabil Res Rev 15: 361-368.
4. Piton A, Redin C, Mandel JL (2013) XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. Am J Hum Genet 93: 368-383.
5. Kuss AW, Garshasbi M, Kahrizi K, Tzschach A, Behjati F, et al. (2011) Autosomal recessive mental retardation: homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots. Hum Genet 129: 141-148.
6. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, et al. (2011) Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 478: 57-63.
7. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, et al. (2010) A de novo paradigm for mental retardation. Nat Genet 42: 1109-1112.
8. Daoud F, Angeard N, Demerre B, Martie I, Benyaou R, et al. (2009) Analysis of Dp71 contribution in the severity of mental retardation through comparison of Duchenne and Becker patients differing by mutation consequences on Dp71 expression. Hum Mol Genet 18: 3779-3794.
9. Courcet JB, Faivre L, Malzac P, Masurel-Paulet A, Lopez E, et al. (2012) The DYRK1A gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. J Med Genet 49: 731-736.
10. Moller RS, Kubart S, Hoeltzenbein M, Heye B, Vogel I, et al. (2008) Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. Am J Hum Genet 82: 1165-1170.
11. Sitz JH, Tigges M, Baumgartel K, Khaspekov LG, Lutz B (2004) Dyrk1A potentiates steroid hormone-induced transcription via the chromatin remodeling factor Arip4. Mol Cell Biol 24: 5821-5834.
12. van Bon BW, Hoischen A, Hehir-Kwa J, de Brouwer AP, Ruivenkamp C, et al. Intragenic deletion in DYRK1A leads to mental retardation and primary microcephaly. Clin Genet 79: 296-299.
13. Yamamoto T, Shimojima K, Nishizawa T, Matsuo M, Ito M, et al. Clinical manifestations of the deletion of Down syndrome critical region including DYRK1A and KCNJ6. Am J Med Genet A 155A: 113-119.
14. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, et al. (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74: 285-299.
15. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 338: 1619-1622.
16. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485: 246-250.
17. Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, et al. (2013) De novo mutations in epileptic encephalopathies. Nature 501: 217-221.
18. Hamdan FF, Gauthier J, Araki Y, Lin DT, Yoshizawa Y, et al. Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. Am J Hum Genet 88: 306-316.
19. Parsons CG, Stoffler A, Danysz W (2007) Memantine: a NMDA receptor antagonist that improves memory by restoration of homeostasis in the glutamatergic system--too little activation is bad, too much is even worse. Neuropharmacology 53: 699-723.

20. Hung CC, Chen HY, Chen CH (2002) Systematic mutation analysis of the human glutamate receptor, ionotropic, N-methyl-D-aspartate 1 gene(GRIN1) in schizophrenic patients. Psychiatr Genet 12: 225-230.
21. Sakurai K, Toru M, Yamakawa-Kobayashi K, Arinami T (2000) Mutation analysis of the N-methyl-D-aspartate receptor NR1 subunit gene (GRIN1) in schizophrenia. Neurosci Lett 296: 168-170.
22. Tarabeux J, Kebir O, Gauthier J, Hamdan FF, Xiong L, et al. Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. Transl Psychiatry 1: e55.
23. Asadollahi R, Oneda B, Sheth F, Azzarello-Burri S, Baldinger R, et al. (2013) Dosage changes of MED13L further delineate its role in congenital heart defects and intellectual disability. Eur J Hum Genet 21: 1100-1104.
24. M MvH, G RM, Duran K, van Binsbergen E, Mpj Breur J, et al. (2014) Further confirmation of the MED13L haploinsufficiency syndrome. Eur J Hum Genet.
25. Muncke N, Jung C, Rudiger H, Ulmer H, Roeth R, et al. (2003) Missense mutations and gene interruption in PROSIT240, a novel TRAP240-like gene, in patients with congenital heart defect (transposition of the great arteries). Circulation 108: 2843-2850.
26. Elsea SH, Girirajan S (2008) Smith-Magenis syndrome. Eur J Hum Genet 16: 412-421.
27. Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH (2003) Mutations in RAI1 associated with Smith-Magenis syndrome. Nat Genet 33: 466-468.
28. Smith AC, McGavran L, Robinson J, Waldstein G, Macfarlane J, et al. (1986) Interstitial deletion of (17)(p11.2p11.2) in nine patients. Am J Med Genet 24: 393-414.
29. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, et al. (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet 39: 25-27.
30. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, et al. (2007) Contribution of SHANK3 mutations to autism spectrum disorder. Am J Hum Genet 81: 1289-1297.
31. Phelan K, McDermid HE The 22q13.3 Deletion Syndrome (Phelan-McDermid Syndrome). Mol Syndromol 2: 186-201.
32. Boccuto L, Lauri M, Sarasua SM, Skinner CD, Buccella D, et al. Prevalence of SHANK3 variants in patients with different subtypes of autism spectrum disorders. Eur J Hum Genet 21: 310-316.
33. Gauthier J, Champagne N, Lafreniere RG, Xiong L, Spiegelman D, et al. De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. Proc Natl Acad Sci U S A 107: 7863-7868.
34. Gauthier J, Spiegelman D, Piton A, Lafreniere RG, Laurent S, et al. (2009) Novel de novo SHANK3 mutation in autistic patients. Am J Med Genet B Neuropsychiatr Genet 150B: 421-424.
35. Waga C, Okamoto N, Ondo Y, Fukumura-Kato R, Goto Y, et al. Novel variants of the SHANK3 gene in Japanese autistic patients with severe delayed speech development. Psychiatr Genet 21: 208-211.
36. Bozdagi O, Sakurai T, Papapetrou D, Wang X, Dickstein DL, et al. Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. Mol Autism 1: 15.
37. Peca J, Feliciano C, Ting JT, Wang W, Wells MF, et al. Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. Nature 472: 437-442.
38. Wang X, McCoy PA, Rodriguiz RM, Pan Y, Je HS, et al. Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms of Shank3. Hum Mol Genet 20: 3093-3108.
39. Yang M, Bozdagi O, Scattoni ML, Wohr M, Roullet FI, et al. Reduced excitatory neurotransmission and mild autism-relevant phenotypes in adolescent Shank3 null mutant mice. J Neurosci 32: 6525-6541.

40. Suls A, Dedeken P, Goffin K, Van Esch H, Dupont P, et al. (2008) Paroxysmal exercise-induced dyskinesia and epilepsy is due to mutations in SLC2A1, encoding the glucose transporter GLUT1. Brain 131: 1831-1844.
41. De Giorgis V, Veggiotti P GLUT1 deficiency syndrome 2013: Current state of the art. Seizure 22: 803-811.
42. Wang D, Pascual JM, Yang H, Engelstad K, Jhung S, et al. (2005) Glut-1 deficiency syndrome: clinical, genetic, and therapeutic aspects. Ann Neurol 57: 111-118.
43. Overweg-Plandsoen WC, Groener JE, Wang D, Onkenhout W, Brouwer OF, et al. (2003) GLUT-1 deficiency without epilepsy--an exceptional case. J Inherit Metab Dis 26: 559-563.
44. Sirmaci A, Spiliopoulos M, Brancati F, Powell E, Duman D, et al. Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. Am J Hum Genet 89: 289-294.
45. Barbaric I, Perry MJ, Dear TN, Rodrigues Da Costa A, Salopek D, et al. (2008) An ENU-induced mutation in the Ankrd11 gene results in an osteopenia-like phenotype in the mouse mutant Yoda. Physiol Genomics 32: 311-321.
46. Berryer MH, Hamdan FF, Klitten LL, Moller RS, Carmant L, et al. Mutations in SYNGAP1 cause intellectual disability, autism, and a specific form of epilepsy by inducing haploinsufficiency. Hum Mutat 34: 385-394.
47. Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet 45: 825-830.
48. Hamdan FF, Daoud H, Piton A, Gauthier J, Dobrzeniecka S, et al. De novo SYNGAP1 mutations in nonsyndromic intellectual disability and autism. Biol Psychiatry 69: 898-901.
49. Hamdan FF, Gauthier J, Spiegelman D, Noreau A, Yang Y, et al. (2009) Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. N Engl J Med 360: 599-605.
50. Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, et al. (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380: 1674-1682.
51. Hamdan FF, Daoud H, Patry L, Dionne-Laporte A, Spiegelman D, et al. (2013) Parent-child exome sequencing identifies a de novo truncating mutation in TCF4 in non-syndromic intellectual disability. Clin Genet 83: 198-200.
52. Amiel J, Rio M, de Pontual L, Redon R, Malan V, et al. (2007) Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am J Hum Genet 80: 988-993.
53. Villard L, Fontes M (2002) Alpha-thalassemia/mental retardation syndrome, X-Linked (ATR-X, MIM #301040, ATR-X/XNP/XH2 gene MIM #300032). Eur J Hum Genet 10: 223-225.
54. Gibbons RJ, Picketts DJ, Villard L, Higgs DR (1995) Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell 80: 837-845.
55. Gibbons RJ, Wada T, Fisher CA, Malik N, Mitson MJ, et al. (2008) Mutations in the chromatin-associated protein ATRX. Hum Mutat 29: 796-802.
56. Abidi FE, Cardoso C, Lossi AM, Lowry RB, Depetris D, et al. (2005) Mutation in the 5' alternatively spliced region of the XNP/ATR-X gene causes Chudley-Lowry syndrome. Eur J Hum Genet 13: 176-183.
57. Guerrini R, Shanahan JL, Carrozzo R, Bonanni P, Higgs DR, et al. (2000) A nonsense mutation of the ATRX gene causing mild mental retardation and epilepsy. Ann Neurol 47: 117-121.
58. Isidor B, Pichon O, Baron S, David A, Le Caignec C (2010) Deletion of the CUL4B gene in a boy with mental retardation, minor facial anomalies, short stature, hypogonadism, and ataxia. Am J Med Genet A 152A: 175-180.

59. Cabezas DA, Slaugh R, Abidi F, Arena JF, Stevenson RE, et al. (2000) A new X linked mental retardation (XLMR) syndrome with short stature, small testes, muscle wasting, and tremor localises to Xq24-q25. J Med Genet 37: 663-668.
60. Tarpey PS, Raymond FL, O'Meara S, Edkins S, Teague J, et al. (2007) Mutations in CUL4B, which encodes a ubiquitin E3 ligase subunit, cause an X-linked mental retardation syndrome associated with aggressive outbursts, seizures, relative macrocephaly, central obesity, hypogonadism, pes cavus, and tremor. Am J Hum Genet 80: 345-352.
61. Deburgrave N, Daoud F, Llense S, Barbot JC, Recan D, et al. (2007) Protein- and mRNA-based phenotype-genotype correlations in DMD/BMD with point mutations and molecular basis for BMD with nonsense and frameshift mutations in the DMD gene. Hum Mutat 28: 183-195.
62. Lenk U, Hanke R, Thiele H, Speer A (1993) Point mutations at the carboxy terminus of the human dystrophin gene: implications for an association with mental retardation in DMD patients. Hum Mol Genet 2: 1877-1881.
63. Kerr TP, Sewry CA, Robb SA, Roberts RG (2001) Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay? Hum Genet 109: 402-407.
64. de Brouwer AP, Nabuurs SB, Verhaart IE, Oudakker AR, Hordijk R, et al. A 3-base pair deletion, c.9711_9713del, in DMD results in intellectual disability without muscular dystrophy. Eur J Hum Genet.
65. De Boulle K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, et al. (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nat Genet 3: 31-35.
66. Lugenbeel KA, Peier AM, Carson NL, Chudley AE, Nelson DL (1995) Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. Nat Genet 10: 483-485.
67. Gronskov K, Brondum-Nielsen K, Dedic A, Hjalgrim H A nonsense mutation in FMR1 causing fragile X syndrome. Eur J Hum Genet 19: 489-491.
68. Collins SC, Bray SM, Suhl JA, Cutler DJ, Coffee B, et al. Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. Am J Med Genet A 152A: 2512-2520.
69. Behnecke A, Hinderhofer K, Bartsch O, Numann A, Ipach ML, et al. (2011) Intragenic deletions of IL1RAPL1: Report of two cases and review of the literature. Am J Med Genet A 155A: 372-379.
70. Laumonnier F, Ronce N, Hamel BC, Thomas P, Lespinasse J, et al. (2002) Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. Am J Hum Genet 71: 1450-1455.
71. Lepretre F, Delannoy V, Froguel P, Vasseur F, Montpellier C (2003) Dissection of an inverted X(p21.3q27.1) chromosome associated with mental retardation. Cytogenet Genome Res 101: 124-129.
72. Bhat SS, Ladd S, Grass F, Spence JE, Brasington CK, et al. (2008) Disruption of the IL1RAPL1 gene associated with a pericentromeric inversion of the X chromosome in a patient with mental retardation and autism. Clin Genet 73: 94-96.
73. Franek KJ, Butler J, Johnson J, Simensen R, Friez MJ, et al. Deletion of the immunoglobulin domain of IL1RAPL1 results in nonsyndromic X-linked intellectual disability associated with behavioral problems and mild dysmorphism. Am J Med Genet A 155A: 1109-1114.
74. Behnecke A, Hinderhofer K, Bartsch O, Numann A, Ipach ML, et al. Intragenic deletions of IL1RAPL1: Report of two cases and review of the literature. Am J Med Genet A 155A: 372-379.
75. Nawara M, Klapecki J, Borg K, Jurek M, Moreno S, et al. (2008) Novel mutation of IL1RAPL1 gene in a nonspecific X-linked mental retardation (MRX) family. Am J Med Genet A 146A: 3167-3172.

76. Carrie A, Jun L, Bienvenu T, Vinet MC, McDonell N, et al. (1999) A new member of the IL-1 receptor family highly expressed in hippocampus and involved in X-linked mental retardation. Nat Genet 23: 25-31.
77. Piton A, Michaud JL, Peng H, Aradhya S, Gauthier J, et al. (2008) Mutations in the calcium-related gene IL1RAPL1 are associated with autism. Hum Mol Genet 17: 3965-3974.
78. Youngs EL, Henkhaus R, Hellings JA, Butler MG IL1RAPL1 gene deletion as a cause of X-linked intellectual disability and dysmorphic features. Eur J Med Genet 55: 32-36.
79. Tabolacci E, Pomponi MG, Pietrobono R, Terracciano A, Chiurazzi P, et al. (2006) A truncating mutation in the IL1RAPL1 gene is responsible for X-linked mental retardation in the MRX21 family. Am J Med Genet A 140: 482-487.
80. Shoubridge C, Tarpey PS, Abidi F, Ramsden SL, Rujirabanjerd S, et al. (2010) Mutations in the guanine nucleotide exchange factor gene IQSEC2 cause nonsyndromic intellectual disability. Nat Genet 42: 486-488.
81. Tran Mau-Them F, Willems M, Albrecht B, Sanchez E, Puechberty J, et al. (2013) Expanding the phenotype of IQSEC2 mutations: truncating mutations in severe intellectual disability. Eur J Hum Genet.
82. Jensen LR, Amende M, Gurok U, Moser B, Gimmel V, et al. (2005) Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. Am J Hum Genet 76: 227-236.
83. Ounap K, Puusepp-Benazzouz H, Peters M, Vaher U, Rein R, et al. A novel c.2T > C mutation of the KDM5C/JARID1C gene in one large family with X-linked intellectual disability. Eur J Med Genet 55: 178-184.
84. Rujirabanjerd S, Nelson J, Tarpey PS, Hackett A, Edkins S, et al. Identification and characterization of two novel JARID1C mutations: suggestion of an emerging genotype-phenotype correlation. Eur J Hum Genet 18: 330-335.
85. Santos-Reboucas CB, Fintelman-Rodrigues N, Jensen LR, Kuss AW, Ribeiro MG, et al. A novel nonsense mutation in KDM5C/JARID1C gene causing intellectual disability, short stature and speech delay. Neurosci Lett 498: 67-71.
86. Tzschach A, Lenzner S, Moser B, Reinhardt R, Chelly J, et al. (2006) Novel JARID1C/SMCX mutations in patients with X-linked mental retardation. Hum Mutat 27: 389.
87. Abidi FE, Holloway L, Moore CA, Weaver DD, Simensen RJ, et al. (2008) Mutations in JARID1C are associated with X-linked mental retardation, short stature and hyperreflexia. J Med Genet 45: 787-793.
88. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, et al. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23: 185-188.
89. Clayton-Smith J, Watson P, Ramsden S, Black GC (2000) Somatic mutation in MECP2 as a non-fatal neurodevelopmental disorder in males. Lancet 356: 830-832.
90. Meloni I, Bruttini M, Longo I, Mari F, Rizzolio F, et al. (2000) A mutation in the rett syndrome gene, MECP2, causes X-linked mental retardation and progressive spasticity in males. Am J Hum Genet 67: 982-985.
91. Orrico A, Lam C, Galli L, Dotti MT, Hayek G, et al. (2000) MECP2 mutation in male patients with non-specific X-linked mental retardation. FEBS Lett 481: 285-288.
92. Imessaoudene B, Bonnefont JP, Royer G, Cormier-Daire V, Lyonnet S, et al. (2001) MECP2 mutation in non-fatal, non-progressive encephalopathy in a male. J Med Genet 38: 171-174.
93. Zeev BB, Yaron Y, Schanen NC, Wolf H, Brandt N, et al. (2002) Rett syndrome: clinical manifestations in males with MECP2 mutations. J Child Neurol 17: 20-24.
94. Moog U, Smeets EE, van Roozendaal KE, Schoenmakers S, Herbergs J, et al. (2003) Neurodevelopmental disorders in males related to the gene causing Rett syndrome in females (MECP2). Eur J Paediatr Neurol 7: 5-12.

95. Villard L (2007) MECP2 mutations in males. J Med Genet 44: 417-423.
96. Maiwald R, Bonte A, Jung H, Bitter P, Storm Z, et al. (2002) De novo MECP2 mutation in a 46,XX male patient with Rett syndrome. Neurogenetics 4: 107-108.
97. Archer HL, Whatley SD, Evans JC, Ravine D, Huppke P, et al. (2006) Gross rearrangements of the MECP2 gene are found in both classical and atypical Rett syndrome patients. J Med Genet 43: 451-456.
98. Lebo RV, Ikuta T, Milunsky JM, Milunsky A (2001) Rett syndrome from quintuple and triple deletions within the MECP2 deletion hotspot region. Clin Genet 59: 406-417.
99. Ravn K, Nielsen JB, Skjeldal OH, Kerr A, Hulten M, et al. (2005) Large genomic rearrangements in MECP2. Hum Mutat 25: 324.
100. Laumonnier F, Holbert S, Ronce N, Faravelli F, Lenzner S, et al. (2005) Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. J Med Genet 42: 780-786.
101. Abidi F, Miano M, Murray J, Schwartz C (2007) A novel mutation in the PHF8 gene is associated with X-linked mental retardation with cleft lip/cleft palate. Clin Genet 72: 19-22.
102. Koivisto AM, Ala-Mello S, Lemmela S, Komu HA, Rautio J, et al. (2007) Screening of mutations in the PHF8 gene and identification of a novel mutation in a Finnish family with XLMR and cleft lip/cleft palate. Clin Genet 72: 145-149.
103. Friesema EC, Grueters A, Biebermann H, Krude H, von Moers A, et al. (2004) Association between mutations in a thyroid hormone transporter and severe X-linked psychomotor retardation. Lancet 364: 1435-1437.
104. Schwartz CE, May MM, Carpenter NJ, Rogers RC, Martin J, et al. (2005) Allan-Herndon-Dudley syndrome and the monocarboxylate transporter 8 (MCT8) gene. Am J Hum Genet 77: 41-53.
105. Jansen J, Friesema EC, Kester MH, Milici C, Reeser M, et al. (2007) Functional analysis of monocarboxylate transporter 8 mutations identified in patients with X-linked psychomotor retardation and elevated serum triiodothyronine. J Clin Endocrinol Metab 92: 2378-2381.
106. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, et al. (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet 34: 27-29.
107. Chih B, Afridi SK, Clark L, Scheiffele P (2004) Disorder-associated mutations lead to functional inactivation of neuroligins. Hum Mol Genet 13: 1471-1477.
108. Tabuchi K, Blundell J, Etherton MR, Hammer RE, Liu X, et al. (2007) A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. Science 318: 71-76.
109. Talebizadeh Z, Lam DY, Theodoro MF, Bittel DC, Lushington GH, et al. (2006) Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism. J Med Genet 43: e21.
110. Lubs H, Abidi FE, Echeverri R, Holloway L, Meindl A, et al. (2006) Golabi-Ito-Hall syndrome results from a missense mutation in the WW domain of the PQBP1 gene. J Med Genet 43: e30.
111. Takahashi M, Mizuguchi M, Shinoda H, Aizawa T, Demura M, et al. Polyglutamine tract-binding protein-1 binds to U5-15kD via a continuous 23-residue segment of the C-terminal domain. Biochim Biophys Acta 1804: 1500-1507.
112. Kalscheuer VM, Freude K, Musante L, Jensen LR, Yntema HG, et al. (2003) Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. Nat Genet 35: 313-315.
113. Kunde SA, Musante L, Grimme A, Fischer U, Muller E, et al. The X-chromosome-linked intellectual disability protein PQBP1 is a component of neuronal RNA granules and regulates the appearance of stress granules. Hum Mol Genet 20: 4916-4931.
114. Froyen G, Corbett M, Vandewalle J, Jarvela I, Lawrence O, et al. (2008) Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. Am J Hum Genet 82: 432-443.

115. Fichou Y, Nectoux J, Bahi-Buisson N, Rosas-Vargas H, Girard B, et al. (2009) The first missense mutation causing Rett syndrome specifically affecting the MeCP2_e1 isoform. Neurogenetics 10: 127-133.
116. Saunders CJ, Minassian BE, Chow EW, Zhao W, Vincent JB (2009) Novel exon 1 mutations in MECP2 implicate isoform MeCP2_e1 in classical Rett syndrome. Am J Med Genet A 149A: 1019-1023.
117. Froyen G, Belet S, Martinez F, Santos-Reboucas CB, Declercq M, et al. (2012) Copy-number gains of HUWE1 due to replication- and recombination-based rearrangements. Am J Hum Genet 91: 252-264.
118. Fox JW, Lamperti ED, Eksioglu YZ, Hong SE, Feng Y, et al. (1998) Mutations in filamin 1 prevent migration of cerebral cortical neurons in human periventricular heterotopia. Neuron 21: 1315-1325.
119. Robertson SP, Twigg SR, Sutherland-Smith AJ, Biancalana V, Gorlin RJ, et al. (2003) Localized mutations in the gene encoding the cytoskeletal protein filamin A cause diverse malformations in humans. Nat Genet 33: 487-491.
120. Sheen VL, Dixon PH, Fox JW, Hong SE, Kinton L, et al. (2001) Mutations in the X-linked filamin 1 gene cause periventricular nodular heterotopia in males as well as in females. Hum Mol Genet 10: 1775-1783.
121. Unger S, Mainberger A, Spitz C, Bahr A, Zeschnigk C, et al. (2007) Filamin A mutation is one cause of FG syndrome. Am J Med Genet A 143A: 1876-1879.

Paper III: 20 ans après: a second mutation in MAOA identified by targeted high-throughput sequencing in a family with altered behavior and cognition.

Piton A*, Poquet H*, **Redin C**, Masurel A, Lauer J, et al. Eur J Hum Genet. 2013 Oct.

Aim of the study: During the previous study, we identified a missense mutation in the monoamine oxydase A gene (*MAOA*), in several males affected with borderline ID, autistic and aggressive behaviors. Since no mutations had ever been reported since the initial publication in 1993, the aim of the study was to confirm the implication of *MAOA* in ID and associated neurodevelopmental disorders by highlighting a second case and functional validation of the missense mutation reported in this second family.

Contribution: Intermediary. Performed the sequencing library, analyzed and interpreted the processed data.

European Journal of Human Genetics (2013), 1–8

www.nature.com/ejhg

ARTICLE

20 ans après: a second mutation in MAOA identified by targeted high-throughput sequencing in a family with altered behavior and cognition

Amélie Piton*[,1,2,10], Hélène Poquet[3,4,10], Claire Redin[1,2], Alice Masurel[3], Julia Lauer[5], Jean Muller[1,5], Julien Thevenon[3,6], Yvan Herenger[5], Sophie Chancenotte[3,7], Marlène Bonnet[7], Jean-Michel Pinoit[4], Frédéric Huet[3], Christel Thauvin-Robinet[3,6], Anne-Sophie Jaeger[5], Stéphanie Le Gras[8], Bernard Jost[8], Bénédicte Gérard[5], Katell Peoc'h[9], Jean-Marie Launay[9], Laurence Faivre[3,6,10] and Jean-Louis Mandel*[,1,2,5,10]

Intellectual disability (ID) is characterized by an extraordinary genetic heterogeneity, with >250 genes that have been implicated in monogenic forms of ID. Because this complexity precluded systematic testing for mutations and because clinical features are often non-specific, for some of these genes only few cases or families have been unambiguously documented. It is the case of the X-linked gene encoding monoamine oxidase A (*MAOA*), for which only one nonsense mutation has been identified in Brunner syndrome, characterized in a single family by mild non-dysmorphic ID and impulsive, violent and aggressive behaviors. We have performed targeted high-throughput sequencing of 220 genes, including *MAOA*, in patients with undiagnosed ID. We identified a c.797_798delinsTT (p.C266F) missense mutation in *MAOA* in a boy with autism spectrum disorder, attention deficit and autoaggressive behavior. Two maternal uncles carry the mutation and have severe ID, with a history of maltreatment in early childhood. This novel missense mutation decreases MAOA enzymatic activity, leading to abnormal levels of urinary monoamines. The identification of this new point mutation confirms, for the first time since 1993, the monogenic implication of the *MAOA* gene in ID of various degrees, autism and behavioral disturbances. The variable expressivity of the mutation observed in male patients of this family may involve gene–environment interactions, and the identification of a perturbation in monoamine metabolism should be taken into account when prescribing psychoactive drugs in such patients.
European Journal of Human Genetics advance online publication, 30 October 2013; doi:10.1038/ejhg.2013.243

Keywords: monoamine oxidase A; X-linked intellectual deficiency; behavioral abnormalities; autism; high-throughput sequencing

INTRODUCTION

The *monoamine oxidase A* (*MAOA*) gene and its close homolog *MAOB* are located at Xp11.3 and encode enzymes crucial for the metabolic degradation of biogenic amines, and particularly neurotransmitters such as norepinephrine, dopamine and serotonin. The two enzymes share 70% of amino-acid sequence identity but differ by their expression, substrate affinities and inhibitor specificities. MAOA is mainly involved in endogenous bioamine (metanephrine (MN), normetanephrine (NMN) and serotonin (5-hydroxytryptamine or 5-HT)) degradation, whereas MAOB preferentially metabolizes exogenous bioamines such as phenylethylamine, and both enzymes are active on dopamine. Combined loss of *MAOA* and *MAOB* genes has been described in some patients with a continuous syndrome also including a deletion of the Norrie disease gene. These patients present, in addition to Norrie disease symptoms, with severe intellectual disability (ID), autistic-like behavior and seizures.[1] Severe developmental delay and hypotonia was more recently observed in a few patients with only *MAOA* and *MAOB* deletion.[2–4] In 1993, Brunner *et al*[5] described a large Dutch family with X-linked borderline ID and prominent behavioral abnormalities. The linkage study and the biochemical analyses suggested that *MAOA* could be responsible for this syndrome. This was confirmed by sequencing which revealed a nonsense c.886C>T (p.Q296*) mutation.[6] All the affected male patients in this family carried the mutation and showed, in addition to borderline ID, very characteristic abnormal behavior, in particular impaired impulse control and stress-induced aggressive and violent behavior. Shortly thereafter, aggressive behavior was described in a mouse line with an inactivated *Maoa* gene,[7] and this was later confirmed in a different mouse line carrying a spontaneous mutation mimicking the human mutation.[7–9] *Maoa*-deficient mice also present autistic-like features.[7–9] Despite the extensive attention given to these early reports, and early attempts at replication, by screening for MAOA deficiency in cohorts of patients with ID and/or abnormal behavior,[10] no other clearly pathogenic mutation in *MAOA*

[1]IGBMC, CNRS UMR 7104/INSERM U964/University of Strasbourg, Illkirch Cedex, France; [2]Chaire de Génétique Humaine, Collège de France, Paris, France; [3]Centre de génétique et Centre de Référence Anomalies du développement et Syndromes malformatifs, Hôpital d'Enfants, Dijon, France; [4]Service de Pédopsychiatrie, Hôpital d'Enfants, Dijon, France; [5]Laboratoire de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, France; [6]EA 4271 GAD, Faculté de Médecine, Université de Bourgogne, Dijon, France; [7]Centre Référent des Troubles du Langage et des Apprentissages, Hôpital d'Enfants, Dijon, France; [8]IGBMC, Microarray and Sequencing Platform, Illkirch, France; [9]Service de Biochimie et de Biologie Moléculaire, Groupe Hospitalier Universitaire Saint-Louis Lariboisière Fernand-Widal, AP-HP, Paris, France
*Correspondence: Dr A Piton or J-L Mandel, Department of Translational Medicine and Neurogenetics, IGBMC, CNRS UMR 7104/INSERM U964, 1 rue Laurent Fries, University of Strasbourg, Illkirch Cedex 67404, France. Tel: +33 3 88653417; E-mail: piton@igbmc.fr or jlmandel@igbmc.fr
[10]These authors contributed equally to this work.
Received 13 May 2013; revised 26 August 2013; accepted 29 August 2013

was reported to our knowledge in other patients in the past 20 years,[11] with the possible exception of a missense variant predicted to be damaging reported in a single patient with autism spectrum disorder (ASD).[12] Many association studies investigated the potential role of *MAOA* in risk of abnormal behaviors, focusing on a 'variable number of tandem repeats' (VNTR) polymorphism[13] located in the *MAOA* promoter region, whose alleles are associated with variations of transcriptional activity, with a 'low' and a 'high' activity frequent alleles. Notably, Caspi *et al*[14] reported that maltreated children with a genotype conferring high levels of MAOA expression were less likely to develop antisocial problems. This VNTR has also been reported to be a modifier of ASD severity, with lower intelligence quotient (IQ) and more severe behavioral problems observed in patients with the 'low activity allele'.[15,16]

Since the original publication,[6] *MAOA* is considered as an ID gene[17,18] and is included in the diagnostic panels of genes screened for X-linked ID mutations. Here we report, for the first time since 1993, a novel pathogenic mutation of the *MAOA* gene segregating in a small family with three affected male patients showing various degrees of cognitive impairment and behavioral disturbances evocative of Brunner syndrome.

PATIENTS AND METHODS

Targeted HTS

DNAs from a cohort of 50 patients with ID patients (with normal caryotype and negative results from CGH array, Fragile X and ARX expansions testing) were prepared as described elsewhere.[19] They were enriched in coding sequence of 220 genes known to cause ID, including *MAOA*, by a target custom capture (SureSelect, Agilent, Santa Clara, CA, USA). These enriched libraries were tagged and pooled by 12 in one lane of a new generation sequencer (HiSeq2000, Illumina, San Diego, CA, USA) for a 100 bp paired-end run. Read mapping and variant calling were performed following standard procedures, and variants were filtered using VaRank, an in-house software which collects variant-specific information to rank them according to their predicted pathogenicity.[19,20]

Bioinformatic analyses

The potential functional effects of the amino-acid change on the protein has been assessed using several bioinformatics programs including SIFT,[21] PolyPhen2,[22] Mutation Taster[23] and KD4v.[24] Possible effects on splicing were determined by MaxEnt,[25] NNsplice,[26] GeneSplicer[27] or Human Splicing Finder[28] programs via Alamut version 2.2 (Interactive Biosoftware, Rouen, France). On the basis of the published 3D structure (2z5y), a 3D model structure of the mutated C266F MAOA was computed using the KD4v webserver. The Exome Variant Server (EVS) and dbSNP have been used to test the presence of the variation in the general population.

Sanger sequencing, RT-PCR analysis, VNTR genotyping and X-inactivation assay

Sanger sequencing was used to confirm the presence of the mutation in the proband and in the different family members and to sequence coding regions of MAOB (primers available on request). Patient III-1 mRNA was extracted from blood (PAXgene Blood RNA System, Preanalytix, Hombrechtikon, Switzerland) and was studied by RT-PCR using specific primers (5′-GTGGCCAGGAACGGAAGTTTGTA-3′ and 5′-CGGGCAAGAATGAAGCC-CATGAT-3′). VNTR genotyping of patient III-1 was performed on genomic DNA using standard primers as previously described.[29] The X chromosome inactivation assay was performed on proband's mother genomic DNA extracted from peripheral blood, as described elsewhere.[30]

Bioamine and MAOA assays

Levels of catecholamine catabolites (MN, NMN, dihydroxyphenylglycol (DHPG), homovanillic acid (HVA) and vanillymandelic acid (VMA)) were measured by HPLC with electrochemical detection.[31] Plasma 5-hydroxyindoleacetic acid (5-HIAA) was measured by HPLC with fluorimetric detection.[32] MAO (EC.1.4.3.4.)-A enzymatic activity was determined on human fibroblasts whole cell homogenates by a radioenzymatic assay using [^{14}C]-5-HT creatinine sulfate (1.96 GBq/mmol, Amersham GE Healthcare, Little Chalfont, UK, final concentration 20 μM) as substrate according to Denney *et al*[33] MAOA protein concentration was assessed by measuring the binding of [^{3}H]-Ro 41-1049 (0.31 TBq/mmol, Amersham GE Healthcare), a reversible inhibitor of MAO-A, to fibroblast membranes, as described by Cesura *et al*.[34]

Patient's evaluations

Patient 1, III-1 was evaluated using WISC IV (Wechsler Intelligence Scale for Children) and ADOS[35] scales. Considering the severity of their clinical manifestations, patients 2 and 3 (II-3 and II-4) could only be evaluated using Vineland scales.[36]

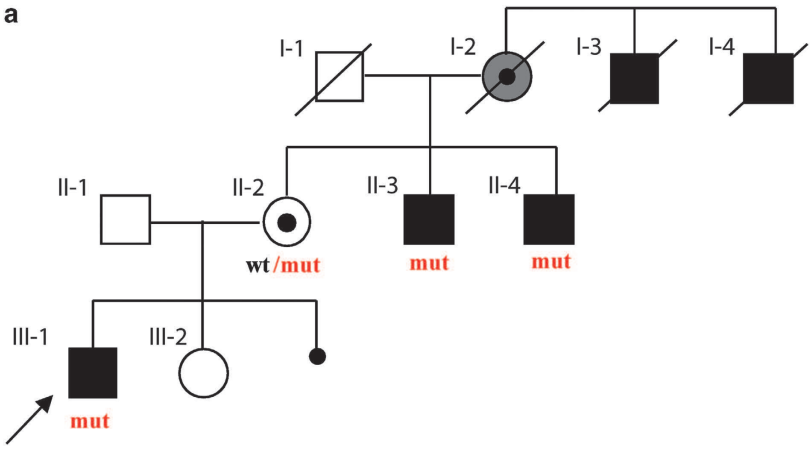
RESULTS

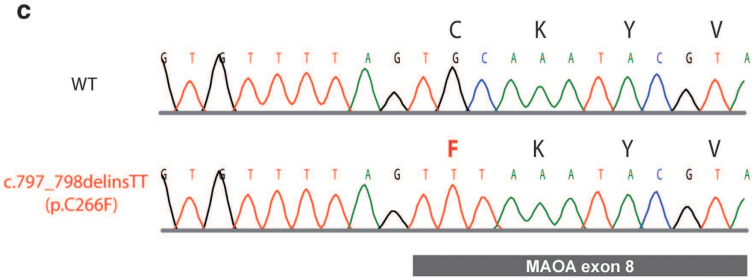
Clinical description of the family

The index case III-1 was referred at the age of 7 years for investigation of suggested ID with prominent behavioral disturbances. Family history suggested a possible X-linked inheritance, although the phenotype of two maternal uncles appeared much more severe (Figure 1). Later investigations indicated that the phenotype of II-3 and II-4 was complicated by familial neglect, maltreatment and sexual abuse during childhood with parental psychiatric disturbances and substance abuse, leading to their placement at the respective ages of 7 and 5 years.

Patient III-1. The index case is the first child of non-consanguineous healthy young parents. The patient has a healthy younger sister. Labor was induced at 35 + 3 weeks of gestation because of maternal diabetes. Apgar was 10–10 and birth measurements were normal: weight 3010 g, length 49 cm, occipital frontal circumference (OFC) 34 cm (50th percentile). Feeding difficulties were noticed at the time of food diversification with high selection of food. Sleep disorder was reported in the first year of life including difficulty in falling asleep, night terrors and frequent awakenings. Gross psychomotor acquisitions were in normal range but the infant was described as passive

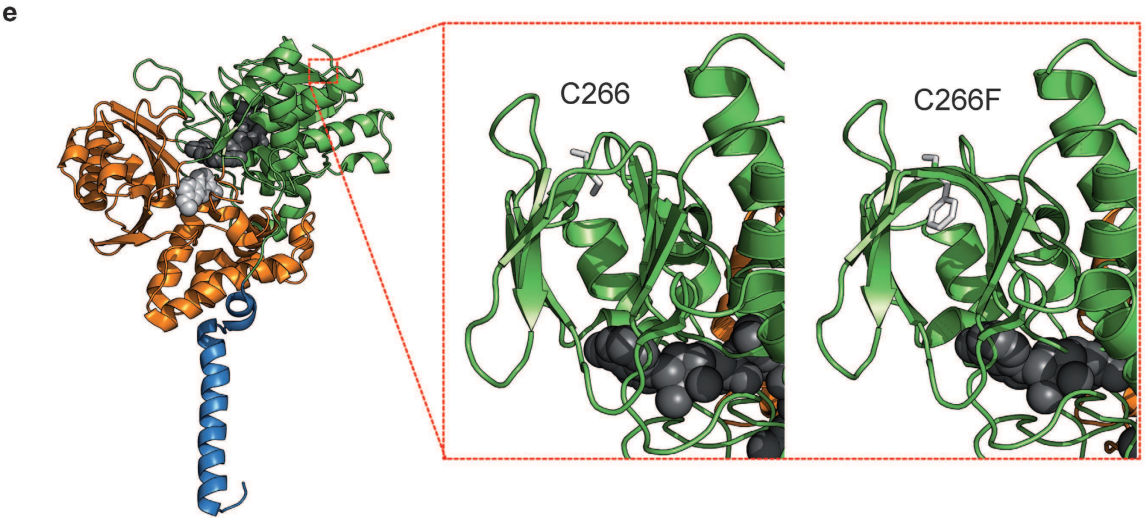
Figure 1 Identification of a c.797_798delinsTT (p.C266F) mutation in MAOA in a family with Brunner syndrome-like behavioral disturbances. (**a**) Pedigree of the family: the mutation (mut) in the proband (arrow) is inherited from his heterozygote mother and the two uncles with ID, autistic traits and aggressive behavior also carry the mutation. (**b**) Integrative Genome Viewer (IGV, Broad Institute, Cambridge, MA, USA) view of the HTS results in the *MAOA* exon 8 region: identification of a c.797_798delinsTT (p.C266F) mutation. (**c**) Sanger sequencing validation of the c.797_798delinsTT (p.C266F) mutation. (**d**) Prediction for the c.796T>A (p.C266S) mutagenesis performed by Wu *et al*[40] and for the c.797_798delinsTT (p.C266F) ID mutation identified in this paper (see Methods). The predicted effects at the protein level were calculated using SIFT, PolyPhen2, Mutation Taster and KD4v. Potential effects on splicing were determined by MaxEnt, NNsplice, GeneSplicer or Human Splicing Finder (HSF): percentage of increase or decrease of the score for the normal acceptor splice site is indicated for each mutation according to each program (**e**) MAOA structure (PDB:2Z5Y Pymol view) according to Se-Young Son *et al*: the membrane-binding domain (colored in blue) and the extra membrane domain, divided into the substrate/inhibitor subdomain (orange) and the FAD-binding subdomain (green). The MAOA cofactor FAD (dark gray) and the MAOA inhibitor Harmine (light gray) are displayed using the sphere model. A detailed view of the cysteine 266 is shown as well as the modeled structure (using KD4v) of the mutated p.C266F MAOA.





d

	c.797_798delinsTT (p.C266F) (ID mutation)	c.796T>A (p.C266S) (Wu et al.)
SIFT	deleterious (0,02)	tolerated (0,23)
PolyPhen2	possibly damaging (0,932)	benign (0,068)
Mutation taster	disease causing (0,927)	disease causing (0,779)
KD4v	deleterious	deleterious
MaxEnt splice	+1%	+5,7%
NNSplice	+12,5%	+12,5%
GeneSplicer	-4,2%	+1,4%
HSF	+0,8%	+1,6%



with low interactive skills. Sensory deficits were ruled out. Walk was acquired at 18 months and speech at 15 months. He was toilet trained at 3 years of age. He went to standard school but attended at the same time twice-weekly a treatment center for individual and group therapy, psychomotricity and school remediation. At the age of 6 years, he entered primary school with a classroom assistant. He appreciated activities such as music listening or memorizing tasks. The parents reported that he did not appreciate danger, experimenting notably with fire and autoaggressive actions (shaving his head and manipulation of household products). No treatment was introduced.

When referred to the genetics clinic at the age of 7 years, reading and writing were in a learning process. Global clinical examination was normal except few hand stereotypies and behavioral abnormalities. No specific cranio-facial dysmorphism was noticed. Both gross psychomotor skills (climb stairs and bicycle) and fine psychomotor skills (dressing and putting on shoes) were delayed. Behavioral troubles during the interview included an amimic facial expression leading to a bizarre contact, oral and motor perseverations, ideomotor slowliness, auto-mutilation and angers when frustrated or failing an exercise. He had restricted patterns of interests such as recurrent questions about water cycle or electrical circuit. The IQ was not calculable but a diagnosis of ASD was confirmed using ADOS scale (Table 1 and Supplementary data).

Patient II-3. Patient II-3 was born premature (1650 g) leading to hospitalization during his first 8 months, but no complication was reported. Psychomotor development was delayed with sitting acquired at 16 months and walking at the age of 2 years. A diagnosis of autism was proposed at the age of 3 years. He had invasive behavioral troubles that included auto and hetero aggressive bursts and very low interactive skills. Speech remained restricted to simple sentences with limited vocabulary. He started attending a school for special needs at the age of 5. Behavior worsened at the age of 6 years. Psychotropic and sedative drugs were early introduced, and stabilized the behavioral aggravation. Intellect was not tested but was severely impaired. He was evaluated at 38 years of age and measurements were as follows: weight, 72 kg; height, 1.58 m (body mass index (BMI) = 28.8); and OFC, 54 cm (−2 SD). Clinical examination revealed no extra-pyramidal sign but dystonic movements of the head and hands that were attributed to secondary effect of the treatment. The patient could neither read nor write, and he was not autonomous for daily living. Vineland Adaptative Behaviors Scale (VABS) for social worker aid was evaluated by semi-structured interview and indicated very impaired autonomy, socialization, communication and locomotion, with equivalent developmental ages between 2 and 5 years (Table 2). The clinical observation was in favor of a neurodevelopmental psycho-motor and socio-emotional very early stagnation and/or regression.

Patient II-4. Patient II-4 was born premature after 8 months of a normal pregnancy (weight: 2150 g). His mother quickly worried about encopresis and abnormal behaviors. No detailed data were available regarding early psychomotor development. He was placed in institution at the age of 5 years and followed a school for special needs. Behavioral troubles were invasive with very low interactive skills and auto-aggressive behavior. He was expressing with poor speech, echolalia and frequent perseverations. He developed stereo-typies, interests for repetitive tasks and became very intolerant to changes. Autism was diagnosed during childhood. The evolution was complicated with auto-mutilation, frequent hetero aggressive bursts and tantrum. Cognitive performances had not been evaluated but intellectual deficiency was considered as severe. He was neither able to read nor write. Psychotropic drug treatment was introduced early with anti-psychotic and sedatives drugs.

He was evaluated at 36 years of age and measurements were as follows: weight, 80 kg; height, 1.62 m (BMI = 30.8); and OFC, 54 cm (−2 SD). Clinical examination was normal except for a marked extra-pyramidal syndrome probably secondary to the high neuroleptics posology. Speech was poorly understandable. He could only execute simple orders. Daily life was evaluated in a semi-structured interview using VABS for social worker, revealing very impaired autonomy, communication, socialization and locomotion, with developmental age equivalents between about 18 months and 4 years (Table 2). This clinical observation was also in favor of a neurodevelopmental psychomotor and socio-emotional with very early stagnation and/or regression.

Other family members. Two maternal great uncles of III-1, I-3 and I-4, were reported to present encephalopathy. They were institutio-nalized all their life. No other detail could be available. The mother of III-1 (sister of II-3 and II-4) had normal scholarship, normal behavior and obtained a high school diploma. The maternal grandmother of the index case presented with a depression and psychotic disturbances.

Table 1 Neuropsychological assessments of patient III-1 using WISC IV

Wechsler Intelligence Scale For Children Fourth Edition (WISC IV)		
Factorial index	*Total score*	*Subtest's results*
Verbal comprehension index	112	Vocabulary: 11/19 Comprehension: 13/19 Similarities: 12/19
Perceptual reasoning index	Invalid calculation because of significant difference between two subtests	Cubes: 6/19[a] Concepts identification: 14/19 Matrices: 6/19[a]
Working memory index	Invalid calculation because of significant difference between two subtests	Memory of numbers: 7/19[a] Arithmetic: 2/19[b]
Processing speed index	76[b]	Code: 4/19[b] Dams: 4/19[b] Symbols: 1/19[b]

[a]Low, average low, cut-off score.
[b]Significantly lower than standard score.

Identification of a pathogenic missense mutation p.C266F in *MAOA*

Targeted high-throughput sequencing (HTS) of the coding exons of 220 'ID genes' in a cohort of 50 patients with ID (unpublished data) led to the identification of a non-synonymous mutation in the *MAOA* coding sequence in the proband III-1 (Figure 1a). This replacement of two nucleotides c.797_798delinsTT (NM_000240.3), at the very beginning of exon 8 (Figure 1b), leads to the change of a cysteine at position 266 to a phenylalanine (p.C266F). This mutation, confirmed by Sanger sequencing (Figure 1c), was never previously described, neither in dbSNP database nor in the 10 562 X chromosomes of the NHLBI Exome Sequencing project. As the mutation affects the very first bases of exon 8, we tested a possible effect on splicing but neither the different prediction programs (Figure 1d) nor RT-PCR analysis on blood mRNA (data not shown) revealed any effect. The p.C266F amino-acid change that affects drastically both the size and the chemical characteristics (hydrophobic versus polar) of the side chain of a highly conserved residue is predicted to be damaging according to all four *in silico* methods used (Figure 1d). MAOA is composed of three functional domains namely a membrane-binding domain, a substrate/inhibitor domain and the flavin adenine dinucleotide (FAD)-cofactor-binding domain (Figure 1e). Interestingly, a full-length crystal structure of the human MAOA is available (Protein Data Bank 2Z5Y, from positions 12–524 out of 527 aa).[37] The analysis of the modeled 3D structure by the prediction program KD4v indicates that the missense affects an aminoacid in a beta sheet close to the FAD-binding pocket, and the bulky aromatic ring of phenylalanine is oriented toward the FAD-binding pocket (Figure 1e), which may explain the observed effect on enzymatic activity (see below). We also genotyped the VNTR located in the *MAOA* promoter and observed that the proband III-1 carries the 3R (three repeats) allele, associated with 'low expression' of MAOA,[13] which may thus potentiate the effect of the missense mutation on the level of MAOA activity.

Sanger sequencing revealed that the missense c.797_798delinsTT (p.C266F) mutation is inherited from the proband's mother, who presents a skewed X inactivation profile (92:8), and is present in the two severely affected maternal uncles II-3 and II-4.

Reduction of MAOA level and activity in patient III-1

A reduction of MAOA activity in the patient was firstly assessed by the measurement of urinary catecholamine catabolites. The levels of urinary MN and NMN, two specific MAOA substrates, exceeded normal range, strikingly so for NMN, whereas the product VMA was at the lower threshold (Figure 2). An *in vitro* assay of MAOA activity using 5-HT, specific MAOA substrate, showed a significant reduction (0.15 ± 0.04 *vs* 0.68 ± 0.16 nmoles/mg prot/h, $P<0.0001$) in patient's III-1 fibroblasts when compared with other boys of the same age range (Figure 3a, left panel). A quantitative radioligand binding assay in fibroblast revealed a three-fold reduction of the binding to MAOA protein (0.11 ± 0.02 *vs* 0.36 ± 0.08 pmoles/mg prot, $P<0.0001$) in patient III-1 (Figure 3a, right panel), reflecting a reduction of MAOA protein level confirmed by western blot analysis (data not shown). Moreover, the plasma concentrations of 5-HIAA and 3,4-DHPG, two products of the degradation of serotonin and norepinephrine, respectively, reflecting MAOA activity *in vivo*,[38,39] were found to be 40 (4.57 ± 0.3 *vs* 177 ± 28 nM, $P<0.0001$) or 10 times (0.67 ± 0.07 *vs* 7.7 ± 0.6 nM, $P<0.0001$) reduced in the proband III-1 compared with control age-matched boys (Figure 3b).

Table 2 Vineland scales of patients II-3 and II-4

	Patient II-3		*Patient II-4*	
Sub-domains	Raw score	Age equivalent score	Raw score	Age equivalent score
Communication	47	2–1	38	1–8
Autonomy	96	5–3	76	3–11
Socialization	67	4–1	39	1–7
Locomotion	51	2–11	51	2–11

DISCUSSION

Targeted HTS of the protein-coding sequences of 220 genes involved in ID identified (after variant filtration using variant databases such as dbSNP or EVS) a non-synonymous c.797_798delinsTT (p.C266F) missense mutation in *MAOA* in a family including male patients presenting with a phenotype overlapping with the one described by Brunner *et al* in 1993[5,6] (data submitted to ClinVar database). Two other nonsynonymous variants were identified in this patient in genes associated with dominant forms of ID (c.1954A>G; p.I652V in *DOCK8* and c.2260G>A; p.V754I in *HDAC4*). However, they were both predicted to be benign and found to be inherited from one of the unaffected parents, excluding a potential pathogenic role. The missense mutation found in *MAOA* was regarded as pathogenic by all the prediction programs. In a systematic site-directed mutagenesis

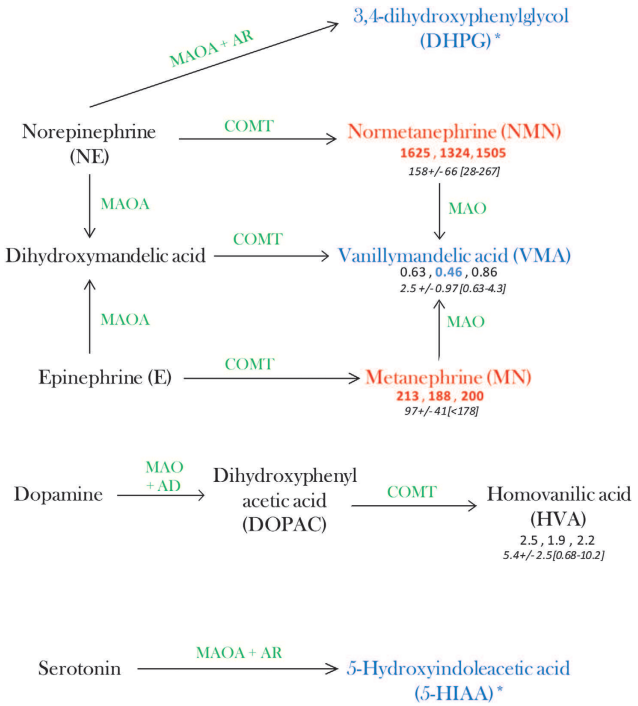


Figure 2 Schematic representation of catecholamine and serotonin metabolism and urinary metabolite dosages in patient III-1. Urinary catecholamine metabolite dosages for patient III-1 have been performed at three different days and are expressed in µmol/mmol of creatinine for VMA and HVA and in nmol/mmol of creatinin for MN and NMN. Reference values from Pussard E, 2009 are indicated in italic (into bracket the 2.5 and 97.5th percentile). In red /blue are indicated abnormal elevated/reduced values (above two standard deviations). Abbreviations: MAO, monoamine oxidase; COMT, catechol-O-methyl transferase; AD, aldehyde dehydrogenase; and AR, aldehyde reductase. *DHPG and 5-HIAA are found reduced in patient III-1 plasma (see Figure 3).

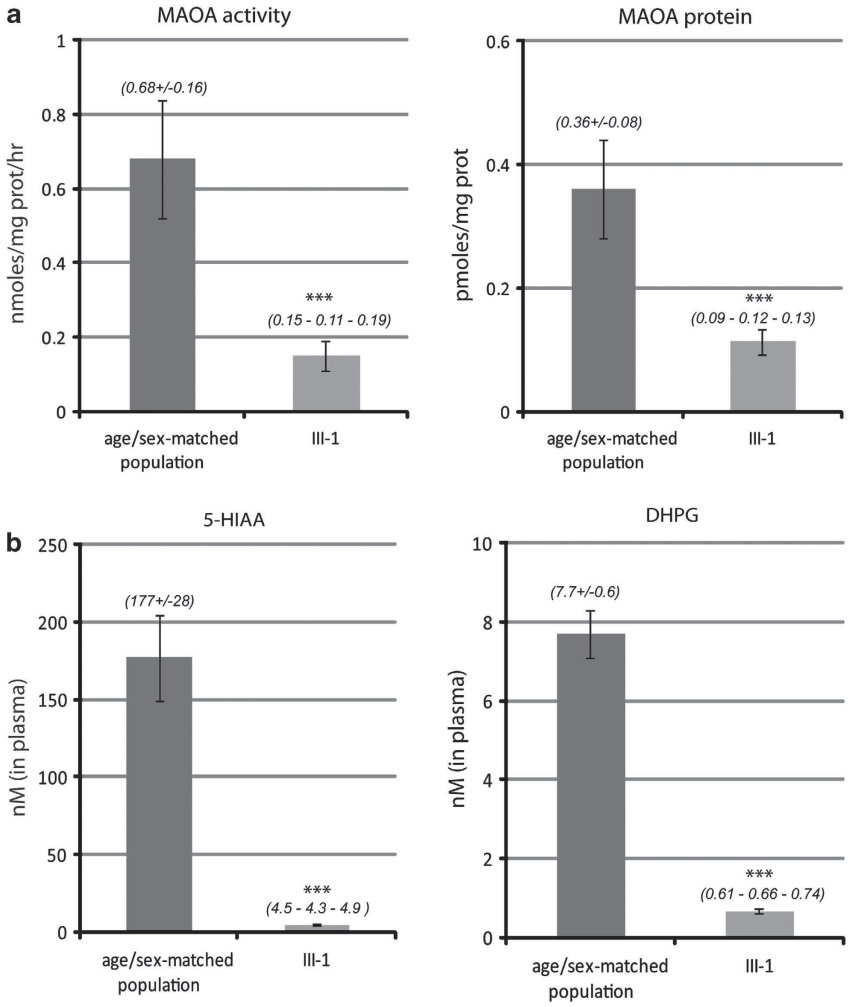


Figure 3 MAOA activity is decreased in the patient III-1 with c.797_798delinsTT (p.C266F) mutation (**a**) MAOA activity (left) and protein level (right) measures *in vitro* using fibroblasts of proband III-1 ($n=3$ measures) and normal 7-year-old boys ($n=13$ measures). The enzymatic activity assay was performed using [14C]-5-HT creatinine sulfate as MAOA substrate and the protein level was estimated with binding of a radioactive MAOA inhibitor [3H]Ro 41-1049. (**b**) Measure of the plasma 5-HT and norepinephrine metabolites 5-HIAA and 3,4-DHPG in the proband III-1 ($n=3$ measures) and in normal 7-year-old boys ($n=13$ measures) as a reflection of *in vivo* MAOA activity. ***$P<0.0001$ with Student's *t*-test.

study of the role of cysteines in the catalytic activity of human MAOA and MAOB, the change of cysteine 266 into a serine did not affect the binding of the substrate, but reduced the catalytic activity by 50%.[40] However, the change to a phenylalanine residue is predicted to be much more damaging, affecting drastically both the size and the chemical nature of the side chain, and was thus likely to have a bigger effect on MAOA activity (Figure 1d). It should be pointed out that missense changes affecting MAOA appear rare in the population, as in the EVS database, none exceeds 0.5% of minor allele frequency, and only two, c.515G>A (p.R172Q) and c.788A>C (p.H263L), predicted benign by the different programs, are present in hemizygous state in male patients (c.515G>A in five male patients; c.788A>C in only one). All the other rare missense variants are described in heterozygous female patients only. This suggests that most variants causing an amino-acid change in MAOA sequence have a phenotypic effect and are negatively selected.

The mutation c.797_798delinsTT (p.C266F) we reported here is in cis with a 'low-activity VNTR allele' in the MAOA promoter, and this might worsen its effect. Indeed, biochemical investigations of substrates and metabolites of MAOA in urine and plasma, as well as measurement of MAOA activity in fibroblasts of the proband all indicated an important perturbation in the catabolism of catecholamines (notably norepinephrine) and serotonin. In fact, *in vivo* evaluation of MAOA activity showed a more drastic effect, with a 10–40-times decrease in MAOA product (5-HIAA and DHPG) levels in the proband's serum compared with what is observed in control age-matched boys, whereas *in vitro* reduction in enzymatic activity showed only a 3–4-fold reduction. Such a difference might be explained by the fact that *in vitro* assays use optimal substrate conditions for enzyme function that could alleviate in part enzymatic dysfunction manifest under *in vivo* conditions.

The family displays variable expressivity in male patients, with systematic ASD and prominent behavioral disturbances, but variable ID. Indeed, ID could not be unambiguously diagnosed in the index case, despite some cognitive defects, while his maternal uncles had severe ID and behavioral disturbances necessitating a high level of psychotropic treatment. The history of maltreatment and sexual abuse during infancy of the two uncles in opposition to the protective familial environment of proband III-1 could explain at least in part this intrafamilial variable expression. Large-scale studies of behavioral

consequences of high *vs* low activity associated to short versus long VNTR MAOA alleles converged toward a gene/environment interaction theory.[41] It was suggested that maltreated children with the *MAOA* genotype conferring low levels of MAOA enzyme developed more often antisocial behavior in adulthood than maltreated children with a high-activity *MAOA* genotype.[42] The description of this family overlaps with the so-called Brunner syndrome, a neuropsychiatric disorder associating ASD, borderline cognitive abilities and behavioral abnormalities, caused by truncating mutation in *MAOA* segregating in a large kindred. Severe aggressiveness was a distinctive feature in this family, and was less obvious in the present family, although all three male carriers of the mutations showed auto-aggressive behavior, and the two uncles had also a history of hetero-aggressive behavior.

The absence of finding of another mutation in the past 20 years is puzzling, especially given the attention given to the initial 1993 reports. This could be attributable to a very low frequency of *MAOA* mutations. An alternative is that the low specificity of the associated clinical phenotype has precluded systematic testing for mutations in this gene in patients with borderline cognitive deficiency and mainly behavioral disorder. Indeed, apart the systematic sequencing of XLID families by Tarpey *et al*[43], there is no published report of systematic screening of mutation in patients with ID or behavioral impairments, and a few other published XLID genes similarly lack such confirmatory reports.[44] Lack of replication finding in these 20 years might also be explained by the necessity of another genetic or environmental event for the expression of the disease. An obvious candidate as modifier gene is the *MAOB* gene that is adjacent to the *MAOA* gene and carries overlapping function. We thus tested whether the present family, as well as the original Brunner-syndrome family carried some missense variants in *MAOB* that could affect enzymatic activity, and would have cosegregated with the *MAOA* nonsense mutation in the latter family. We did not identify such variant in *MAOB* coding regions in patient III-1 and in one member from the Brunner-syndrome family. This does not exclude a potential effect of some non-coding variants.

This study showed the manifest benefit of targeted HTS for genetic diagnosis of non-specific heterogeneous conditions such as ASD or ID associated with behavioral troubles. Of note, the presence of abnormal neurotransmitters in urine could be considered as a diagnostic test in patients with such phenotype, but it is likely that the use for HTS would considerably develop in the next years, and will be economically more appropriate than biochemical testing covering a wide range of potential dysfunctions.

Besides the benefit of genetic diagnosis for appropriate genetic counseling, it might also help in the patient's management. Pharmacogenetics studies on *Maoa*-deficient mice evidenced the counter indication of serotonergic drugs in these animals because of the risk of serotonergic syndrome complication.[45] MAOA agonists may be a good indication for reversing the behavioral symptoms (irritability, social withdrawal, stereotypy and repetitive speech) in the affected members, but such medication is not yet available. The role of atomoxetine, used in some countries for treatment of pervasive developmental disorders by targeting norepinephrine synaptic recapture,[46] could be questionable. Also, agonists of upstream/downstream enzyme such as the catechol-O-methyl transferase enzyme might optimize the remaining activity of the MAO enzyme, whereas use of inhibitors of MAOA activity should not be used.

To conclude, we are reporting on the second family with a non-ambiguous mutation in *MAOA*. This family presents with ASD, behavioral troubles, variable cognitive impairment and accumulation of NMN in urine associated with a phenotype overlapping with the Brunner family. This HTS result could lead to improve management by giving clues for possible therapeutic options.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Nadège Calmels, Nadine Kempf, Valérie Biancalana and all the other members of the Genetic Diagnostic Laboratory of Nouvel Hopital Civil (Strasbourg) for their help with patient DNA samples selection and preparation. We thank Serge Vicaire and Muriel Phillips for their help and technical assistance with HTS. We would also like to thank Véronique Geoffroy and Cecile Pizot for the development of VaRank. We are very grateful to Pr Han Brunner for providing us DNA for one member of the initial Brunner-syndrome family. This work was supported by Agence de BioMedecine, Fondation Jérome Lejeune, Fondation pour la Recherche Medicale and Fondation APLM. This project also benefitiated from 7th European Union Framework (grant no. 241995; GENCODYS).

WEB RESOURCES

The URLs for online tools and data presented herein are:
OMIM: http://www.omim.org/, UCSC: http://genome.ucsc.edu/, dbSNP: http://www.ncbi.nlm.nih.gov/projects/SNP/, ClinVar: http://www.ncbi.nlm.nih.gov/clinvar/, Mutation Nomenclature: http://www.hgvs.org/mutnomen/recs.html, Exome Variant Server, NHLBI Exome Sequencing Project (ESP): http://evs.gs.washington.edu/EVS/, Integrative Genomics Viewer (IGV): http://www.broadinstitute.org/igv/, PolyPhen 2: http://genetics.bwh.harvard.edu/pph2/, Sift Human Protein: http://sift.jcvi.org/www/SIFT_enst_submit.html, MutationTaster: http://www.mutationtaster.org/, KD4v: http://decrypthon.igbmc.fr/kd4v/cgi-bin/prediction, NNsplice: http://www.fruitfly.org/seq_tools/splice.html, Max Ent: MES; http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html, Human Splicing Finder: (http://www.umd.be/HSF/, GeneSplicer: http://cbcb.umd.edu/software/GeneSplicer/.

1 Collins FA, Murphy DL, Reiss AL *et al*: Clinical, biochemical, and neuropsychiatric evaluation of a patient with a contiguous gene syndrome due to a microdeletion Xp11.3 including the Norrie disease locus and monoamine oxidase (MAOA and MAOB) genes. *Am J Med Genet* 1992; **42**: 127–134.
2 Whibley A, Urquhart J, Dore J *et al*: Deletion of MAOA and MAOB in a male patient causes severe developmental delay, intermittent hypotonia and stereotypical hand movements. *Eur J Hum Genet*, **18**: 1095–1099.
3 O'Leary RE, Shih JC, Hyland K, Kramer N, Asher YJ, Graham JM Jr: *De novo* microdeletion of Xp11.3 exclusively encompassing the monoamine oxidase A and B genes in a male infant with episodic hypotonia: a genomics approach to personalized medicine. *Eur J Med Genet* 2012; **55**: 349–353.
4 Saito M, Yamagata T, Matsumoto A *et al*: MAOA/B deletion syndrome in male siblings with severe developmental delay and sudden loss of muscle tonus. *Brain Dev* 2013; **S0387-7604**: 00020–0002.
5 Brunner HG, Nelen MR, van Zandvoort P *et al*: X-linked borderline mental retardation with prominent behavioral disturbance: phenotype, genetic localization, and evidence for disturbed monoamine metabolism. *Am J Hum Genet* 1993; **52**: 1032–1039.
6 Brunner HG, Nelen M, Breakefield XO, Ropers HH, van Oost BA: Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science* 1993; **262**: 578–580.
7 Cases O, Seif I, Grimsby J *et al*: Aggressive behavior and altered amounts of brain serotonin and norepinephrine in mice lacking MAOA. *Science* 1995; **268**: 1763–1766.
8 Bortolato M, Godar SC, Alzghoul L *et al*: Monoamine oxidase A and A/B knockout mice display autistic-like features. *Int J Neuropsychopharmacol* 2013; **16**: 869–888.
9 Scott AL, Bortolato M, Chen K, Shih JC: Novel monoamine oxidase A knock out mice with human-like spontaneous mutation. *Neuroreport* 2008; **19**: 739–743.
10 Schuback DE, Mulligan EL, Sims KB *et al*: Screen for MAOA mutations in target human groups. *Am J Med Genet* 1999; **88**: 25–28.
11 Haavik J, Blau N, Thony B: Mutations in human monoamine-related neurotransmitter pathway genes. *Hum Mutat* 2008; **29**: 891–902.
12 Piton A, Gauthier J, Hamdan FF *et al*: Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol Psychiatry* 2011; **16**: 867–880.

13 Sabol SZ, Hu S, Hamer D: A functional polymorphism in the monoamine oxidase A gene promoter. *Hum Genet* 1998; **103**: 273–279.
14 Caspi A, McClay J, Moffitt TE *et al*: Role of genotype in the cycle of violence in maltreated children. *Science* 2002; **297**: 851–854.
15 Cohen IL, Liu X, Schutz C *et al*: Association of autism severity with a monoamine oxidase A functional polymorphism. *Clin Genet* 2003; **64**: 190–197.
16 May ME, Srour A, Hedges LK *et al*: Monoamine oxidase a promoter gene associated with problem behavior in adults with intellectual/developmental disabilities. *Am J Intellect Dev Disabil* 2009; **114**: 269–273.
17 Lubs HA, Stevenson RE, Schwartz CE: Fragile X and X-linked intellectual disability: four decades of discovery. *Am J Hum Genet* 2012; **90**: 579–590.
18 Ropers HH: Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet* 2010; **11**: 161–187.
19 Redin C, Le Gras S, Mhamdi O *et al*: Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alstrom syndromes. *J Med Genet* 2012; **49**: 502–512.
20 Li H, Handsaker B, Wysoker A *et al*: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
21 Ng PC, Henikoff S: SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**: 3812–3814.
22 Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
23 Schwarz JM, Rodelsperger C, Schuelke M, Seelow D: MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010; **7**: 575–576.
24 Luu TD, Rusu A, Walter V *et al*: KD4v: comprehensible knowledge discovery system for missense variant. *Nucleic Acids Res* 2012; **40**: W71–W75.
25 Yeo G, Burge CB: Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; **11**: 377–394.
26 Reese MG, Eeckman FH, Kulp D, Haussler D: Improved splice site detection in Genie. *J Comput Biol* 1997; **4**: 311–323.
27 Pertea M, Lin X, Salzberg SL: GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001; **29**: 1185–1190.
28 Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C: Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: e67.
29 Lung FW, Tzeng DS, Huang MF, Lee MB: Association of the MAOA promoter uVNTR polymorphism with suicide attempts in patients with major depressive disorder. *BMC Med Genet* 2011; **12**: 74.
30 Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW: Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet* 1992; **51**: 1229–1239.
31 Pussard E, Neveux M, Guigueno N: Reference intervals for urinary catecholamines and metabolites from birth to adulthood. *Clin Biochem* 2009; **42**: 536–539.
32 Kema IP, de Vries EG, Slooff MJ, Biesma B, Muskiet FA: Serotonin, catecholamines, histamine, and their metabolites in urine, platelets, and tumor tissue of patients with carcinoid tumors. *Clin Chem* 1994; **40**: 86–95.
33 Denney RM, Fritz RR, Patel NT, Abell CW: Human liver MAO-A and MAO-B separated by immunoaffinity chromatography with MAO-B-specific monoclonal antibody. *Science* 1982; **215**: 1400–1403.
34 Cesura AM, Bos M, Galva MD, Imhof R, Da Prada M: Characterization of the binding of [3H]Ro 41-1049 to the active site of human monoamine oxidase-A. *Mol Pharmacol* 1990; **37**: 358–366.
35 Lord C, Rutter M, Goode S *et al*: Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord* 1989; **19**: 185–212.
36 Sparrow SS, Cicchetti DV: Diagnostic uses of the Vineland Adaptive Behavior Scales. *J Pediatr Psychol* 1985; **10**: 215–225.
37 Son SY, Ma J, Kondou Y, Yoshimura M, Yamashita E, Tsukihara T: Structure of human monoamine oxidase A at 2.2-A resolution: the control of opening the entry for substrates/inhibitors. *Proc Natl Acad Sci USA* 2008; **105**: 5739–5744.
38 Celada P, Artigas F: Plasma 5-hydroxyindoleacetic acid as an indicator of monoamine oxidase-A inhibition in rat brain and peripheral tissues. *J Neurochem* 1993; **61**: 2191–2198.
39 Eisenhofer G, Finberg JP: Different metabolism of norepinephrine and epinephrine by catechol-O-methyltransferase and monoamine oxidase in rats. *J Pharmacol Exp Ther* 1994; **268**: 1242–1251.
40 Wu HF, Chen K, Shih JC: Site-directed mutagenesis of monoamine oxidase A and B: role of cysteines. *Mol Pharmacol* 1993; **43**: 888–893.
41 Fergusson DM, Boden JM, Horwood LJ, Miller AL, Kennedy MA: MAOA, abuse exposure and antisocial behaviour: 30-year longitudinal study. *Br J Psychiatry* 2011; **198**: 457–463.
42 Kim-Cohen J, Caspi A, Taylor A *et al*: MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis. *Mol Psychiatry* 2006; **11**: 903–913.
43 Tarpey PS, Smith R, Pleasance E *et al*: A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet* 2009; **41**: 535–543.
44 Piton A, Redin C, Mandel JL: XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am J Human Genet* 2013; **S0002-9297**: 00282–00286.
45 Fox MA, Panessiti MG, Moya PR *et al*: Mutations in monoamine oxidase (MAO) genes in mice lead to hypersensitivity to serotonin-enhancing drugs: implications for drug side effects in humans. *Pharmacogenomics J* 2012; e-pub ahead of print 11 September 2012; doi:10.1038/tpj.2012.35.
46 Jou RJ, Handen BL, Hardan AY: Retrospective assessment of atomoxetine in children and adolescents with pervasive developmental disorders. *J Child Adolesc Psychopharmacol* 2005; **15**: 325–330.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)

VI- Discussion

Intellectual disability (or intellectual developmental disorders, as referred by DSM-V) is a generic term regrouping an extensive number of individual conditions, that all have in common impaired intellectual and adaptative functioning. Many of them have genetic origins, and mutations in over 300 genes have been reported in patients so far. Other neurological, neurodevelopmental and psychiatric are frequent comorbid features of ID, all of them being also genetically highly heterogeneous. This complex and humongous genetic heterogeneity make the task of identifying the genetic origin of the ID in patients difficult, especially in patients presenting with no additional features but ID (non-syndromic patients). We developed an approach based on targeted high throughput sequencing of several hundred ID genes.

Although exome sequencing is a more exhaustive and less-biased approach, we opted for targeted sequencing strategy for several reasons: first for economic reasons, indeed exome-sequencing (2,000€ for a trio at 50x by BGI) is still more expensive than our targeted strategy (600€); second for throughput reasons since in a single HiSeq2500 sequencing lane you can sequence up to 50 patients compared to one single patient and his parents when doing exome sequencing; then because it generates data of better quality (higher and more reproducible depth of coverage, decreasing the number of false positives and negatives and ensuring a reliable call for structural variants). With our approach, the diagnostic yield in a cohort of 106 patients is of 25-29%, which is very similar to the ones reported in previous studies implementing the trio-exome strategy (31-47% with (Rauch *et al.*, 2012), 13-19% with (de Ligt *et al.*, 2012), or an overall diagnostic yield of 19-28% in 151 patients). When computing the supposed diagnostic yield of our method when performing the analysis on the same families from those two studies (*i.e.* considering only mutations reported in genes that are included in our panel), it would be of 24% in the study or Rauch and colleagues, which is highly similar to what we observed with our cohort. It would be of only 6% with the study of de Ligt and colleagues, but the diagnostic yield was already much lower (13%, including only known genes and validated novel genes but excluding recessive genes in which no biallelic mutation were found). *In fine*, although many ID genes are not included in our panel, we reach a similar diagnostic yield by increasing the quality of sequencing in the targeted ones. For instance, O'roak and colleagues reported that a few mutations in recurrent ASD genes had been missed by exome sequencing due to an insufficient depth of coverage, while they were detected subsequently using the MIP approach (O'Roak *et al.*, 2012a). Finally, the diagnostic yield with our targeted approach is also higher than the diagnostic yield of array-CGH and fragile-X testing altogether, which make it a relevant and efficient approach to incorporate in routine settings.

Mutations are detected evenly in males and females, and regardless of the degree of severity of the cognitive impairment (although the diagnostic yield appears higher in females, it does not reach significance). There is a trend of a higher diagnostic yield (36%) among patients presenting with microcephaly, but this is not significant either. Consequently, there does not appear to be any

key to discriminate patients with positive from negative diagnosis. Some genes that have been exclusively described in syndromic patients may as well be involved in lesser syndromic forms, but since never tested in other cohorts of patients this could have never been highlighted. One of the best illustrations is *TCF4*, which was initially implicated in Pitt-Hopkins syndrome and has been now demonstrated to be also involved in non-syndromic intellectual disability (Hamdan *et al.*, 2013).

Once again, the interrogation regarding the location of the 'missing' mutations in the remaining 75% of patients arises. Although we consider this strategy as successful, a considerable proportion of patients still remain without diagnosis. We know that we were not exhaustive in including ID genes, and especially for the autosomal and syndromic forms. We may have missed mutations because located in non-targeted regions essential for a correct expression of the gene (in enhancers, polyadenylation signal, promoter regions, deep intronic mutations creating competing splice sites, etc). Patients may also carry mutations in known ID genes that are not included in our panel and that cannot be guessed by clinical examination alone because either they present with atypical clinical features deviating from the classical phenotype, or because they present with non-syndromic ID. Mutations may also be located in novel ID genes, since we expect a large number is still left to discover. Lastly, we cannot rule out a more complex oligogenic or multifactorial scenario, where several rare variants in different genes contribute to the phenotype of the patient.

In current analyses (for both targeted sequence and exome sequencing), we only consider candidate variants under the monogenic scenario, with full penetrance. We posit that is the main weakness of current analyses. Indeed, in exome sequencing for instance, variants that are inherited from a parent are filtered out under the dominant scenario. Similarly, we discard variants located in supposedly dominant genes when also reported in unaffected relatives. This statement is also to relate with our study trying to discriminate 'true' versus false positive XLID genes (Piton *et al.*, 2013). Indeed, in this study we only assess the validity of each mutation by considering it fully penetrant and disease causing by itself. We therefore discarded all variants reported in healthy individuals, or genes for which haploinsufficiency seemed to be tolerated in controls. We do not imply that those variants cannot play a role in the observed phenotype in patients, rather that they do not fully explain the phenotype by themselves.

However, we know that many disorders associated with autosomal dominant inheritance present with incomplete penetrance, and high intrafamilial phenotypic variability. For instance some *SLC2A1* mutations, although presumably responsible for severe GLUT1 deficiency syndrome, are inherited from mildly affected to asymptomatic parents (reviewed in (Leen *et al.*, 2010)). This is also observed with CNVs, where most CNVs associated to ID are of reduced penetrance: many are (maternally) inherited from healthy parents. The question whether the phenotype in patients originates from an oligogenic model (additional variants in other genes leading to the phenotype), or a hypomorphic model (another variant in *trans* reducing the expression of the gene) is still open.

RESULTS - PART III

TARGETED SEQUENCING FOR THE DIAGNOSIS OF LEUKODYSTROPHIES

I- Introduction

1. Myelin, and the myelination process

Brain development depends on the correct sequence of several steps, among which the establishment of white matter referred to as the 'myelination' process. The myelination process is one late step, which starts around the 24th week of gestation and continues throughout childhood and adulthood. Late in the embryonic life, myelination starts with the axons of primary sensory neurons being myelinated first (necessary for vision, audition,…, myelinated during late pre-natal period and shortly after birth) while the axons of cells associated to more complex cognitive functions get myelinated later in childhood. Although myelination starts in fetal life, very little myelin is present in the brain at birth.

The myelination process starts with the proliferation, migration and maturation of oligodendrocyte progenitor cells. Those progenitors differentiate into pre-myelinating oligodendrocytes that align along target axons. Under appropriate environmental signals, mature oligodendrocytes start producing a dielectric material composed of fatty substances called myelin, which wraps around most neuronal axons and acts as an electrical insulation (**Figure 30**). The myelin sheath becomes more compacted as the number of wrapping turns, determined by the axonal diameter. Such lipid constitution ensures insulation of the axons from electrically charged atoms and molecules located in the surrounding fluid of the nervous system. Besides lipids, myelin is also composed of water and of diverse lipoproteins (mainly the proteolipid protein PLP, the myelin basic protein MBP and the myelin oligodendrocyte glycoprotein MOBP). Different cells supply such multi-layered myelin sheath from which most of the cytoplasm is extruded: Schwann cells in the PNS, oligodendrocytes in the CNS. Oligodendrocytes can provide myelin around multiple axons contrary to Schwann cells, which spiral around one single axon.

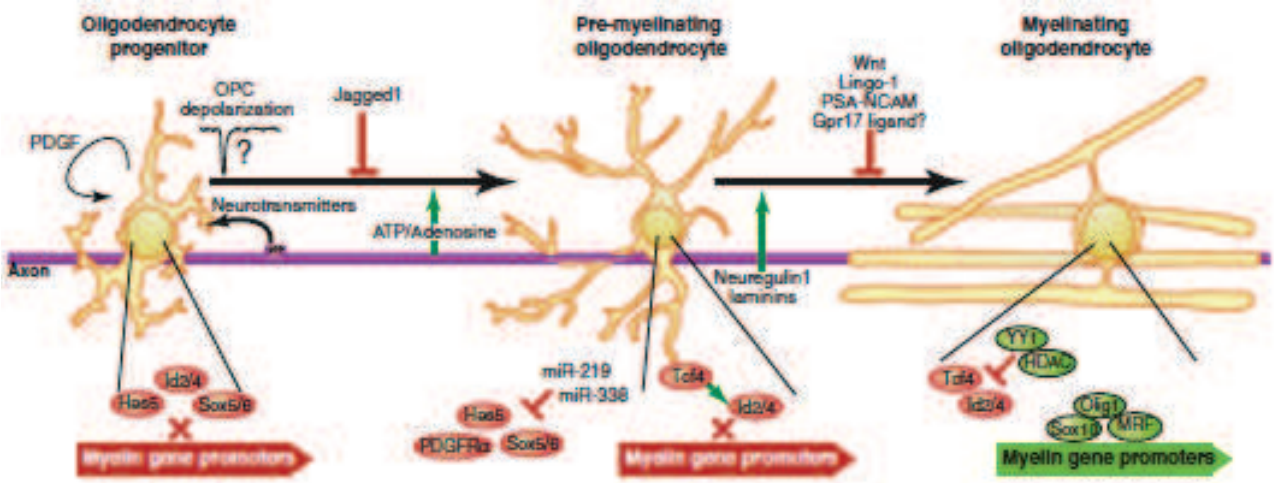


Figure 30 (from (Emery, 2010)): Schematic view of the oligodendrocytes maturation and initiation of the myelination process.

White matter itself is composed of glial cells (astrocytes, oligodendrocytes and microglia, altogether representing up to 50% of the brain volume and 90% of brain cells) and myelin (40-50% of dried weight). It is also considered as a brain compartment, connecting different functional regions through long-distance fibers (commissural, projection or associative fibers).

Myelin sheaths guarantee faster conduction of the electric nerve impulses along the axon, 10-100 times faster that would allow non-myelinated axons (the speed mostly depends on axonal diameter and thickness of the myelin sheaths). Indeed, the action potential propagates along the axon in a continuous manner when not myelinated, or by a saltatory process from one Ranvier node (or myelin sheath gaps) to the next when myelinated. By improving the speed conduction of electric messages, myelin participates in improving the overall functional efficiency of brain networks. Such tight interactions between neurons and oligodendrocytes appear not to be restricted to insulation properties, but may also influence neuronal size and axon diameter, and contribute to neural signaling (Lin and Bergles, 2004).

2. Definition of leukodystrophies

Leukodystrophies are a group of highly genetically and clinically heterogeneous heritable disorders, grouped together based on MRI findings demonstrating abnormalities of the white matter in the central nervous system. Such abnormalities reflect impairment of the production and/or maintenance of the myelin sheaths around neuronal axons, resulting in non or de-myelinated axons. They affect approximately 1 person in 8000 in the population (O Boespflug-Tanguy, personal communication).

In an attempt to discriminate leukodystrophies from leukoencephalopathies, the neuropathology of leukodystrophies is proposed to involve primarily oligodendrocytes, astrocytes or other non-neuronal cell types while that of leukoencephalopathies involves primarily neurons of the cerebral cortex or other grey matter structures (van der Knaap, in preparation). Leukodystrophies do not include acquired myelin disorders (*e.g.* multiple sclerosis), infectious damages (*e.g.* progressive multifocal leukoencephalopathy), toxic disorders, vascular insults, and inborn error of metabolism (IEMs) for which systemic illness involving another organ that the brain predominates.

In leukodystrophies, primary impairment of the white matter leads generally to regression of motor and cognitive skills with association of severe neurological symptoms. Motor defects manifest as hypotonia in early childhood and evolve to different degrees of spasticity. Peripheral nerve involvement is present in certain forms, which result in a combination of spasticity and reduced muscle reflexes. Other features such as ataxia, cognitive impairment, or global development delay are often reported in patients. Seizures are reported in some patients, but are rather unusual.

3. Genetics

The etiology of such heterogeneous disorders is thought to be mainly genetic, even though over 50% of patients with undetermined leukodystrophies remain with no molecular diagnosis because of the heterogeneity and complexity of these disorders (Schiffmann and van der Knaap, 2004).

Almost 50 genes have been reported so far as implicated in a specific form of leukodystrophy, involving genes encoding for myelin constituents such as proteolipid protein (PLP) or myelin basic protein (MBP), glial-specific proteins such as glial fibrillary acidic protein (GFAP) or connexin 47, or factors essential for protein synthesis regulating myelin formation and degradation such as EIF2B subunits or RNA polymerase III subunits (Gencic *et al.*, 1989; Brenner *et al.*, 2001; Leegwater *et al.*, 2001; Uhlenberg *et al.*, 2004; Linnankivi *et al.*, 2006; Bernard *et al.*, 2011; Saitsu *et al.*, 2011). Most genes are implicated in autosomal recessive forms.

4. Classification

MRI is the gold standard test to perform in patients with suspected leukodystrophy. Leukodystrophies are usually classified into the four hypomyelinating, de/dysmyelinating, vacuolating/cystic and undetermined subclasses depending on lesions detected on MRI profiles (**Table 17**).

Demyelinating, dysmyelinating disorders

Demyelinating and dysmyelinating disorders originate from a substantial destruction of myelin leading to unmyelinated axons.

X-linked adrenoleukodsytrophy belongs to this category and is one of the most frequent leukodystrophies with an assessed incidence of about 1:17,000 males (Bezman *et al.*, 2001). It is an X-linked disorder, caused by a mutation in *ABCD1*. Male patients can unpreferentially develop in an unpredictable way either a rapidly progressive childhood cerebral form, an adult form, or the adrenomyeloneuropathy that is much more progressive and primarily affects the peripheral nervous system. 60% of carrier females develop after 40 years of age a mild or moderate form of adrenomyeloneuropathy. *ABCD1*-related disorders can be diagnosed by a biochemical test demonstrating an accumulation in serum and tissues of very long chain fatty acids, which are an integral part of myelin.

Krabbe disease also belongs to this family. It can manifests at any onset, but the infantile form appears to be the most frequent. The prognosis is usually between 5 months and 14 years. It is an autosomal recessive disorder caused by biallelic mutations in *GALC* encoding for

galactocerebroside β galactosidase. Those mutations lead to an impaired activity of GALC and the accumulation of cerebrosides resulting in apoptosis of oligodendrocytes and demyelination. The remarkable feature of this disorder is the observation of globoid cells, which are multinucleated macrophages.

Metachromatic leukodystrophy is caused by biallelic mutations in *ARSA*, which result in deficiency in the lysosomal enzyme arylsulfatase A or in *PSAP* that is implicated in the stimulation of ARSA. In both cases, the accumulation of sulfatide leads to injury of oligodendrocytes and Schwann cells that will go under apoptosis. MRI findings highlight diffuse white matter abnormalities, sparing association fibers. Both are autosomal recessive disorders of any infantile, juvenile or adult onset.

Alexander disease is an autosomal dominant disorder characterized by heterozygous mutations in *GFAP*, coding for the glial fibrillary acidic protein that is an intermediate-filament protein specifically expressed in astrocytes. Mutations are usually gain-of-function mutations resulting in an overexpression of altered GFAP. The production of myelin is also significantly reduced. Alexander disease can be histologically diagnosed by the report of cytoplasmic GFAP inclusions in astrocytes, called Rosenthal fibers (reviewed in (Sawaishi, 2009)). Patients also present with megalencephaly.

Hypomyelinating disorders

Hypomyelination disorders are white matter disorders characterized by a significant permanent deficit in myelin originating from a defect in its development or production.

Pelizaeus-Merzbacher disease is a rare X-linked disorder caused by mutation in *PLP1* encoding for two isoforms of myelin proteolipid proteins: PLP and DM20. Altogether, both account for about half of the total protein content of the CNS. The most severe cases are observed when the mutation affects both isoforms. The entire duplication of the gene is the most frequent mutation responsible for this disorder. The consequences are deficient myelination, with central white matter being significantly reduced or even absent. Mutations can also be reported in *GJC2*, encoding connexin 47 that is specifically expressed in oligodendrocytes and whose expression is regulated in concert with other myelin genes (Menichella *et al.*, 2003). The Pelizaeus-Merzbacher-like disorder is associated with an autosomal recessive mode of inheritance. MRI analysis generally reveals symmetric confluent white matter abnormalities.

Fucosidosis is caused by biallelic mutations in *FUCA1* coding for the α-L-fucosidase (Kretz *et al.*, 1989), which is a lysosomal enzyme responsible for degrading glycoproteins and glycolipids containing fucose. Mutations result in the accumulation of α-L-fucose-rich storage products that are supposed to be neurotoxic (Galluzzi *et al.*, 2001). In MRI analysis, myelin sheaths appear disrupted, and glial cells filled with lipid granular material.

Mutations in the genes coding for the two subunits of RNA polymerase III (POLR3A and POLR3B) cause the 4H syndrome (hypomyelination, hypodontia and hypogonadotropic hypogonadism; (Bernard *et al.*, 2011; Saitsu *et al.*, 2011)). Biallelic mutations in either gene disrupt the RNA polymerase III function that is proposed to result in a global decreased expression of tRNAs during development, and resulting impaired protein synthesis. MRI analyses reveal diffuse hypomyelination.

Demyelinating/dysmyelinating disorder	**Hypomyelinating disorders**	**Vacuolating and cystic disorders**
Alexander disease	Fucosidosis	Cystical leukoencephalopathy without megalencephaly
Canavan disease	Folate receptor defect	Glycine leukoencephalopathy
Krabbe disease	Hypomyelination with atrophy of the basal ganglia and cerebellum	Leukoencephalopathy with calcifications and cysts
Leukoencephalopathy with metaphyseal chondrodysplasia	Hypomyelination and congenital cataract	Megalencephalic leukodystrophy with cysts
Metachromatic leukodystrophy	Hypomyelination, hypodontia, hypogonadotropic hypogonadism	Progressive cavitating leukoencephalopathy
Metachromatic leukodystrophy with multiple sulfatase deficiency	Hypomyelination with monocarboxylate transporter-8 deficiency	Aicardi–Goutières syndrome
X-linked adrenoleukodystrophy	Pelizaeus–Merzbacher disease and Pelizaeus–Merzbacher-like disease	Cerebrotendinous xanthomatosis
Mitochondrial disorders	Sialic acid storage disorder	
	Tremor-ataxia with central hypomyelination	

Table 17 (adapted from (Kohlschutter and Eichler, 2011)): Classification of leukodystrophies based on MRI findings

Vacuolating and cystic disorders

Vacuolating, cavitary or cystic disorders refer to a smaller group of leukodystrophies whose MRI signature is a diffuse abnormal and swollen white matter, which are associated with macrocystic or microcystic changes.

EIf2B-related disorders (notably leukoencephalopathy with vanishing white matter, VWM) are caused by biallelic mutations in the genes encoding the subunits of the eukaryotic initiation factor-2B (*eIf2B1-5*) that is a protein required for correct protein synthesis. Altered eIF2B assembly results in an apparent increased density of abnormal oligodendrocytes with vacuolated cytoplasm.

MRI analyses generally reveal progressive rarefaction and cystic degeneration of white matter, with the remaining myelin sheaths appearing thinner.

Megalencephalic leukoencephalopathy with subcortical cysts is an autosomal recessive disorder caused by biallelic mutations in *MLC1* or *MLC2/HEPACAM*. One cue for the diagnosis of patients is the presence of perinatal macrocephaly. Both proteins localize in axons and their junctions with astrocytes, and are implicated intro cell-cell contacts and brain ion-water homeostasis leading to defects in brain volume regulation (reviewed in (van der Knaap *et al.*)). MRI shows subcortical vacuolizations and cysts of the white matter and thinner myelin sheaths.

Biallelic mutations in *L2HGDH* are responsible for L2-hydroxyglutaric aciduria leading to a neurotoxic accumulation of L-2-hydroxyglutarate in the central nervous system due to impaired a L-2-hydroxyglutarate dehydrogenase activity. Characteristic MRI findings show spongiform degeneration, cystic cavitations in subcortical white matter and cerebellar atrophy (van der Knaap *et al.*, 2007).

5. Diagnosis of leukodystrophies

Diagnosing leukodystrophies is challenging. In some cases, MRI analyses alone or with additional clues (phenotypic features, biochemical markers, electrophysiological analyses…) can help defining clinical entities and directing toward the right candidate gene. About 30-40% of individuals still remain without precise clinical diagnosis despite extensive investigations and even less have a molecular diagnosis (only 1/3 in the LeukoFrance cohort; O Boespflug-Tanguy, personal communication). Precise diagnosis in adult forms appears even less efficient (Labauge and Boespflug-Tanguy, 2010). Clinicians often face with overlapping or incomplete forms of leukodystrophy making selection of the right candidate tough and of low efficiency. Other clinical forms show extensive genetic heterogeneity, with five implicated genes for Aicardi-Goutières or for CACH/VWM diseases.

As mentioned previously, the main key for the diagnosis of leukodystrophies is given by MRI findings (see **Figure 31**). However, there are some other cues that can be used.

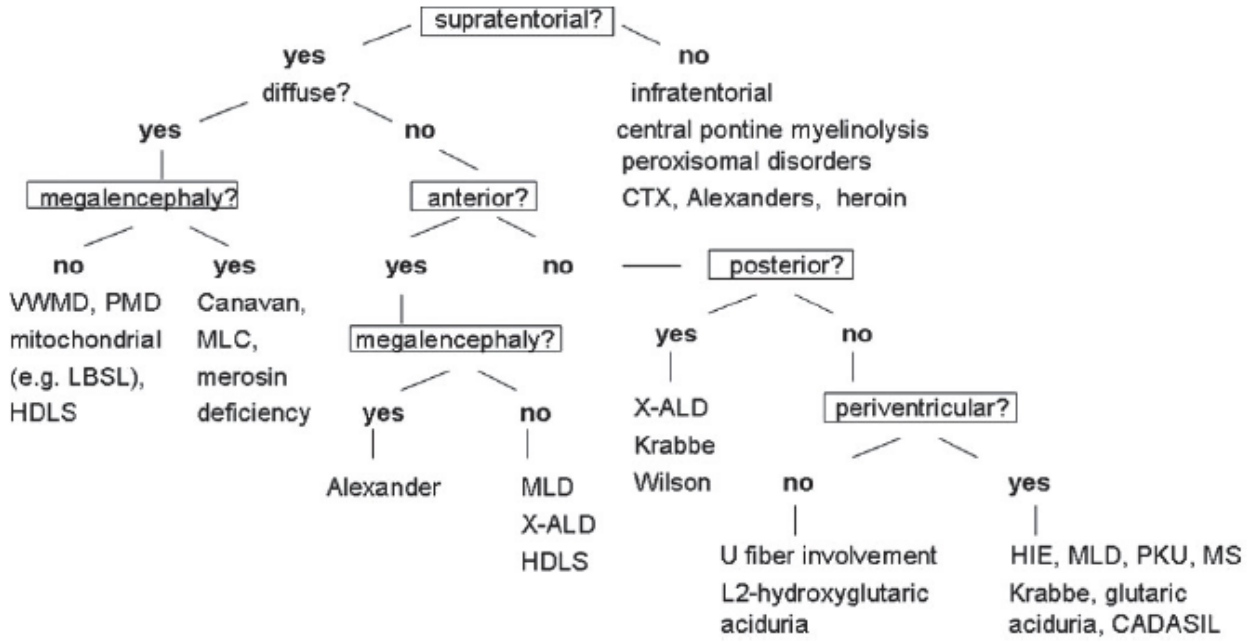


Figure 31 (from (Costello *et al.*, 2009)): Decision tree for the diagnosis of leukodystrophies using MRI patterns.

VWM: Vanishing-White matter disease, PMD: LBSL: HDLS:, MLC: CTX:, MLD:, X-ALD: X-linked adrenoleukodystrophy, HIE, PKU: phenylketonuria, MS: multiple sclerosis.

Physical examination

Age of onset of the disease, marking the apparition of the first neurological symptoms occurring in a previously healthy child or adult can bring suspicion towards leukodystrophies but hardly points out towards one particular disorder due to the phenotypic variability (in terms of age of onset and severity) reported in patients. No particular dysmorphic features are reported in patients with leukodystrophies, apart from macrocephaly/megalencephaly for some with Alexander, Canavan, megalencephalic leukodystrophy with cysts, or VWM disease.

Electrophysiological and nerve conduction tests

Some electrophysiological or nerve conduction investigations can highlight reduced velocities, which are the marker of involvement of long spinal tracts and peripheral nerves. This can be helpful in discriminating some forms of leukodystrophies from others. In some cases, the observed abnormalities are correlated with the severity of the disease in patients, such as in Krabbe disease.

Laboratory test	
Blood	
Cellular elements	Lymphocyte granulation in metachromatic leukodystrophy–multiple sulfatase deficiency Anemia in mitochondrial disorders
Very long chain fatty acids	Elevated in X-Adrenoleukodystrophy (and other peroxisomal disorders)
Lysosomal enzyme activities	Arylsulfatase A low in metachromatic leukodystrophy Galacto-cerebrosidase low in Krabbe disease Fucosidase low in fucosidosis (several other lysosomal storage disorders may show hypomyelination
Lactate	Elevated in LBSL, other mitochondrial disorders
Asialotransferrins	Elevated in congenital disorders of glycosylation
Cholestanol	Elevated in cerebrotendinous xanthomatosis
Urine	
Sulfatides	Elevated in metachromatic leukodystrophy
NNA (organic acids)	Elevated in Canavan disease
Organic acids	Organic acid disorders
Free sialic acid	Elevated in sialic acid storage disorder
Cerebrospinal fluid	
Cellular elements	Elevated in Aicardi-Goutières syndrome
Total protein	Elevated in Krabbe disease, metachromatic leukodystrophy (young patients)
Asialotransferrins	Elevated in VWMD
5-Methyltetrahydrofolate	Low in folate receptor defects (while normal in blood)
Lactate	Elevated in LBSL, other mitochondrial disorders
α-interferon	Aicardi-Goutières syndrome
Glycine (CSF:serum ratio)	Elevated in glycine leukoencephalopathy
Free sialic acid	Elevated in sialic acid storage disorder
Biopsy	
Skin	Generally useful for morphological studies and fibroblast cultivation ; Metachromatic leukodystrophy: demyelinated nerve fibers may be seen ; Krabbe disease: crystalloid inclusion bodies; Other inclusion bodies of storage disorders

Table 18 (adapted from (Kohlschutter and Eichler, 2011)): Metabolic, biochemical and other laboratory investigations that are useful to discriminate certain forms of leukodystrophies

NAA: *N*-acyl aspartate; VWMD: vanishing white matter disease; LBSL: leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation.

Biochemistry/metabolic studies

Some biochemical investigations on blood, urine or cerebrospinal fluid may be useful to diagnostic one particular form of leukodystrophies. However, there are themselves of low yield, expensive and might be invasive. Some tests are useful when performed early in the diagnostic process, but most of the time such investigations come as confirmatory tests when a prior suspicion is made based on MRI or electrophysiological findings (**Table 18**).

II- Results

1. Targeted genes

The gene selection was made in consultation with several experts in leukodystrophies (Prof. Bertini, Prof. Boespflug-Tanguy, Dr Mochel, Dr Burglen, Prof Rodriguez, Dr Sedel). Due to non-stringent size restrictions, we included as many genes as possible. Some genes were included although the clinical, imaging and molecular investigations were proposed to lead unambiguously to the specific syndrome. Indeed, we wanted to assess in an unbiased manner if there were some atypical phenotypes. This resulted in a large set of genes, implicating in leukodystrophies and leukoencephalopathies, and a few research candidate genes. In total, 68 genes were included, 47 leukodystrophy genes, and 21 leukoencephalopathy genes (see **Table 19**).

Disorder	**Official Gene Symbol**	**# exons across all isoforms**	**Gene size (bp)**	**Targeted-region size**
Leukodystrophies				
X-linked Adrenoleukodystrophy	*ABCD1*	10	19,894	2,738
Peroxisomal acyl-CoA-Oxidase Deficiency	*ACOX1*	15	37,927	8,749
Sjögren Larsen syndrome	*ALDH3A2*	11	28,845	4,301
Metachromatic leukodystrophy	*ARSA*	9	5,420	2,043
Canavan disease	*ASPA*	7	25,297	1,205
Hereditary Diffuse Leukodystrophy With Spheroids (HDLS)	*CSF1R*	22	60,082	5,221
Cerebroretinal microangiopathy with calcifications and cysts (CRMCC)	*CTC1*	23	23,275	8,545
Cerebrotendinous Xanthomatosis	*CYP27A1*	9	33,545	2,789
Leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation (LBSL)	*DARS2*	17	33,886	4,296
Leukodystrophy with Vanishing WM Disease (or CACH)	*eIF2B1*	9	12,754	2,385
	eIF2B2	8	6,683	2,000
	eIF2B3	12	136,201	2,714
	eIF2B4	13	6,106	2,681
	eIF2B5	16	10,290	3,777
Hypomyelination with congenital cataract	*FAM126A*	11	72,893	6,697
Fucosidosis	*FUCA1*	8	23,288	2,451
Globoid cell Leukodystrophy (Krabbe disease)	*GALC*	18	60,652	5,195
Polyglucosan Body Disease (PGBD)	*GBE1*	16	272,101	3,771
Alexander Disease (AxD)	*GFAP*	9	9,927	4,482
Oculodentodigital dysplasia	*GJA1*	2	14,129	3,210
Pelizaeus Merzbacher like-disease	*GJC2*	2	10,113	2,461
Megalencephalic Leukodystrophy with subcortical cysts	*HEPACAM*	7	1,7163	3,917
D-Bifunctional Protein Deficiency	*HSD17B4*	25	89,893	4,522
L2-hydroxyglutaric aciduria	*L2HGDH*	10	69,796	6,571
Autosomal Dominant Leukodystrophy with Autonomic disease (ADLD)	*LMNB1*	11	60,398	3,851

18q minus syndrome	*MBP*	10	153,986	7,546
Megalencephalic Leukodystrophy with subcortical cysts	*MLC1*	13	26,539	4,840
Peroxisome biogenesis disorders (PBD)	*PEX1*	24	41,509	5,570
	PEX6	17	15,371	4,525
	PEX26	5	13,112	4,531
	PEX10	6	7,770	2,404
	PEX12	3	3,843	2,803
	PEX2	5	20,787	4,819
Pelizaeus Merzbacher (PMD)	*PLP1*	8	16,109	3,562
4H syndrome (hypomyelination, hypodontia and hypogonadotropic hypogonadism)	*POLR3A*	31	54,392	8,084
	POLR3B	29	152,541	5,905
Metachromatic leukodystrophy due to SAP-b deficiency	*PSAP*	15	35,028	3,723
Aicardi-Goutières Syndrome	*RNASEH2A*	8	7,035	1,746
	RNASEH2B	12	60,783	2,878
	RNASEH2C	4	3,266	2,982
RNAse T2 deficient leukoencephalopathy	*RNASET2*	9	27,074	1,906
Aicardi-Goutières Syndrome	*SAMHD1*	16	60,020	4,071
SCPx deficiency	*SCP2*	17	124,389	4,626
Sialic acid storage disorders	*SLC17A5*	11	60,636	3,820
SOX10-associated disorders	*SOX10*	4	12,221	3,022
Multiple sulfatase deficiency	*SUMF1*	9	106,138	1,607
Aicardi-Goutières Syndrome	*TREX1*	1	1,816	1,856
Leukoencephalopathies				
AIMP1 related disorder	*AIMP1*	9	33,615	3,661
AMACR Deficiency	*AMACR*	5	21,130	3,552
Leukoencephalopathy with ataxia	*CLCN2*		15,508	4,353
Combined oxidative phosphorylation deficiency 12 (COXPD12)	*EARS2*	9	35,363	4,395
Fatty Acid Hydroxylase-Associated Neurodegeneration	*FA2H*	7	61,874	2,729
Muscular dystrophy-dystroglycanopathy (MDDG)	*FKRP*	4	12,530	3,748
	FKTN	11	82,989	8,038
3-Hydroxy-3-Methylglutaryl-CoA Lyase Deficiency	*HMGCL*	9	23,583	2,062
HSPD1 related disorder (or MitCHAP-60 disease)	*HSPD1 (HSP60)*	13	13,691	3,024
Cerebral Autosomal Recessive Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CARASIL)	*HTRA1*	9	53,384	2,634
Congenital Muscular Dystrophy (CMD)	*LAMA2 (MDC1A)*	65	633,425	12,717
α-dystroglycanopathies	*LARGE*	16	647,355	4,923
Mucolipidosis type IV	*MCOLN1*	14	11,400	2,916
α-dystroglycanopathies	*POMGNT1*	23	31,625	4,828
	POMT1	20	20,905	4,380
	POMT2	21	45,927	6,040
Dihydropterine reductase (DHPR) deficiency	*QDPR*	7	25,842	1,993
Allan-Herndon-Dudley syndrome	*SLC16A2*	6	112,437	4,381
Epileptic encephalopathy, early infantile	*SPTAN1*	57	81,109	10,771

Nasu Hakola Disease	*TREM2*	5	4,681	1,200
	TYROBP	5	3,909	933
	SRY	1	897	937
	Total	**897**	**4,118,102**	**281,613**

Table 19: List of targeted genes mostly implicated in leukodystrophies, but also in leukoencephalopathies.

2. Cohort Description

135 patients were included in total, with any type of leukodystrophy on basis of MRI analysis. Our cohort was rather even in terms of sex distribution (78 males, 57 females; see **Table 18**). Fifteen patients were positive controls (mostly females), *i.e.* had a molecular diagnosis prior to inclusion and served as proof of concept. For all remaining patients, inclusion criteria were based upon non-extensive prior molecular investigation and availability of DNA from both parents, and ideally of other family members. Most patients were recruited through the European Leukodystrophy consortium, and clinical features were extensively and consistently provided.

3. Proof of principle yield

Among the 15 samples with known mutations, we were able to detect mutations in 13/15. One heterozygous mutation was missed because it was located in the non-targeted 5'UTR region of *GJC2*. We were not aware of such recurrent mutation at the time of the design otherwise this region would have been targeted as well. Another heterozygous mutation in *DARS2* was missed by the variant calling pipeline due to misalignment on the reference genome. The second version of the pipeline now detects this indel correctly, which is of particular importance since it is one of the recurrent *DARS2* mutations. We also detected correctly three large structural variants: one hemizygous duplication of *PLP1*, one heterozygous triplication of *PLP1* in a female (**Figure 32**), and one heterozygous deletion encompassing both *ARSA* and *MLC1* (**Figure 33**), which was in fact a translocation. We could not map the exact breakpoints of the deletion since we did not have enough targeted genes on that chromosome. Interestingly, we also detected mutations that differed from the ones originally proposed by the clinicians. In one patient with only a single heterozygous missense mutation (c.274C>T, p.Pro92Ser) in *MLC1*, we detected a second heterozygous splice site mutation (c.177+1G>T, p.?) that had been missed by Sanger sequencing.

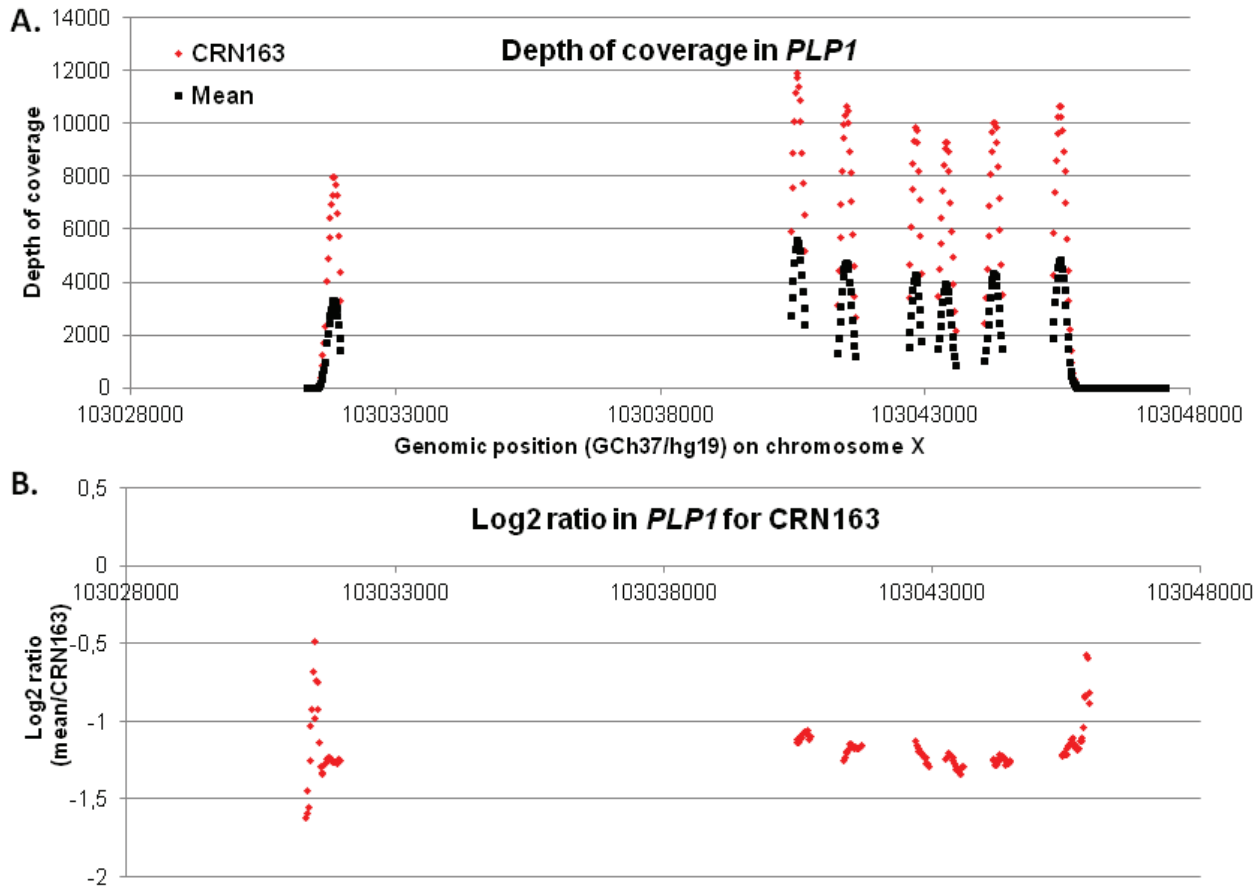


Figure 32: Illustration of a heterozygous triplication of *PLP1* in patient CRN163.

A. Comparison of the depth of coverage between patient CRN163 (red) and the mean of 8 random patients from the same sequencing run (black); **B.** Log2 ratio of the fold change (mean/CRN163) demonstrating the presence of a heterozygous triplication of *PLP1* (values close to -1).

In another patient, which was initially sent for a homozygous missense mutation (c.1151C>T, p.Pro384Leu, rs41272687) in *CYP27A1* (the diagnosis of cerebrotendinous xanthomatosis being suspected on the basis of a high cholestanol in serum), we detected a heterozygous frameshift mutation in *PLP1* (c.37_38insA, p.Gly13Glufs*26). The initially reported *CYP27A1* missense variant was not in our final list of candidates, and had been filtered out by some of our criteria. The missense mutation had already been reported in the literature in patients, but when looking at its frequency in the Exome Variant Server, it was reported in 237 carriers, and 2 individuals at the homozygous state. Its overall MAF is of 1.9%, which is far too frequent when compared to the incidence of the disease in the general population (and was bypassing our filtering criteria, requiring a MAF<0.4% in EVS). A total of 0.04% individuals would be homozygous for this variant, which is already more than the frequency of the disease itself. In sight of these new findings, the clinical data were reviewed and appeared compatible with a *PLP1* mutation in a female.

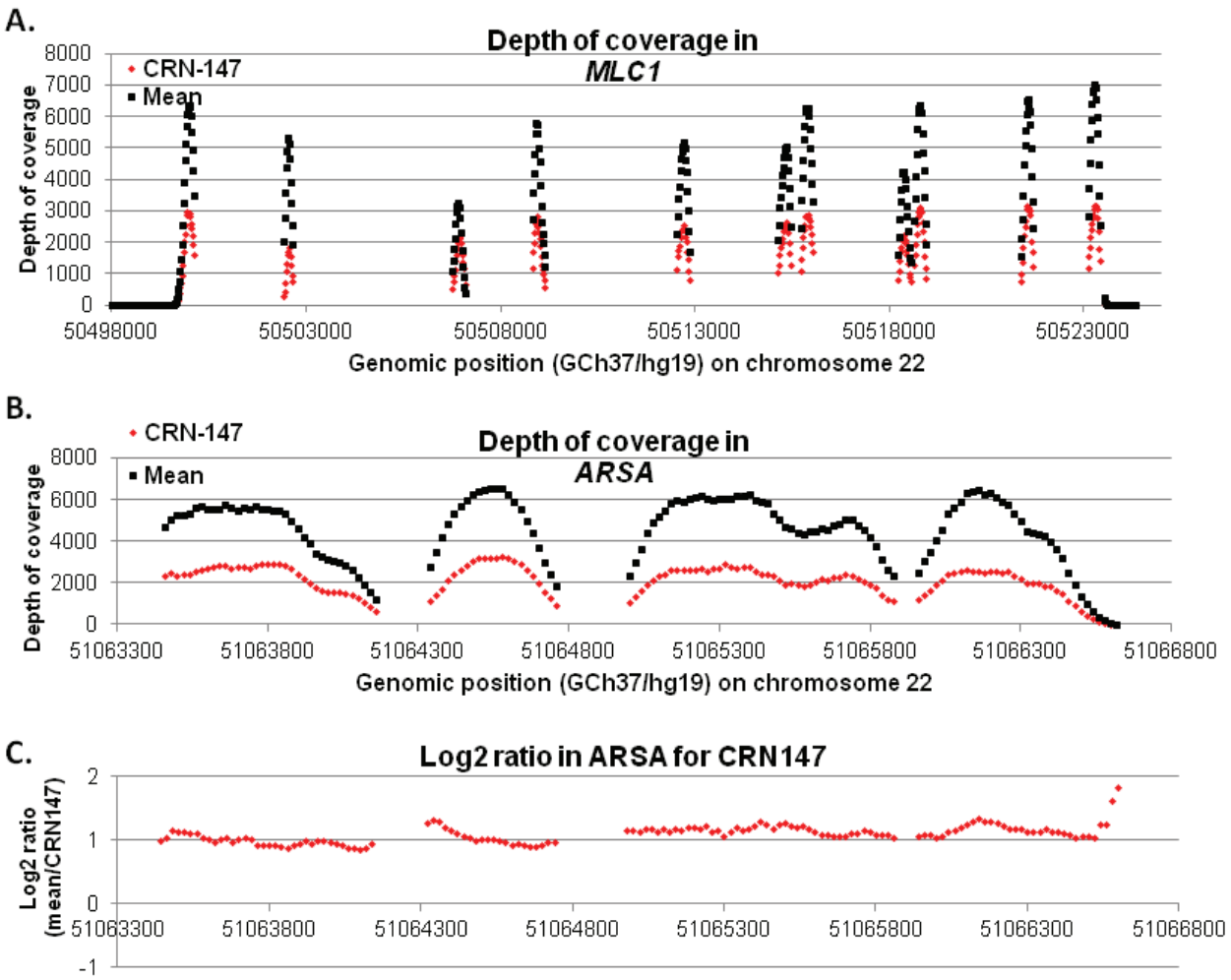


Figure 33: Illustration of a heterozygous deletion encompassing *ARSA* and *MLC1* in patient CRN147.

A-B. Comparison of the depth of coverage between patient CRN147 (red) and the mean of 8 random patients from the same sequencing run (black) in *MLC1* (**A**.) and *ARSA* (**B.**); **C.** Log2 ratio of the fold change (mean/CRN147) demonstrating the presence of a heterozygous deletion of *ARSA* (values close to 1). The corresponding graph for the log2 ratio of *MLC1* had a highly similar profile (data not shown).

4. Diagnostic yield

Among the 120 patients without known molecular diagnosis, 20 carried certainly causative mutations and 10 probably or possibly causative mutations for an overall diagnostic yield of 17-25%. The overall diagnostic yield was similar between males (12/76, 16%) and females (8/44, 18%). Most mutations were identified in autosomal recessive genes (14/20), a few in all three X-linked genes included in the design, and only a single one in the autosomal dominant gene *GFAP*. Among the five mutations in the X-linked genes, 4/5 were detected in males. Two mutations (one hemizygous entire deletion of the gene and one hemizygous full duplication of the gene) were reported in *PLP1*, two in *SLC16A2* (one hemizygous frameshift mutation, one hemizygous splice mutation) and one missense variant in *ABCD1*. Among the other genes *EIF2B2*, *POLR3A*, *POLR3B*

were hit twice in the cohort and *RNASEH2B* was hit four times, each time involving the recurrent mutation (c.529G>A, p.Ala177Thr) either at heterozygous or homozygous state.

Gender	Cohort (n=135)	Positive controls (n=13)	With positive diagnostic (n=21)				Yield (per category)
			X-linked genes	Autosomal recessive genes	Autosomal dominant genes	Total	
Male	78 (58%)	2	4	8	0	**12**	12/76 (16%)
Female	57 (42%)	13	1	6	1	**8**	8/44 (18%)
Total	*135*	*15*	*5*	*14*	*1*	***20***	*20//120 (17%)*

Table 20: Cohort description and diagnostic yield per category.

One interesting finding is the report of two variants affecting the same aminoacid residue Thr567 in *CSF1R*, implicated in hereditary diffuse leukoencephalopathy with spheroids (HDLS; (Rademakers *et al.*, 2012)). One is inherited from the father, and reported in two unaffected brothers. Nonetheless, in sight of the clinical presentation of the proband CRN206 (maniac depression, apraxia and severe gait disturbances reflecting a diffuse leukoencephalopathy with an onset of 20 years old) the implication of *CSF1R* seems highly probable. The other proband (CRN198) presents with an adrenal insufficiency and a motor neuron disease with demyelination, of onset at 30 years of age. The clinical spectrum associated with *CSF1R* mutations is very vast, from less severe symptoms of speech disorders, to cognitive impairment, depression, gait disturbances or epilepsy (Rademakers *et al.*, 2012). The age of onset of the disease is also highly variable (from 23 to 78 years old), which prevents from considering any carrier individual as definitively healthy. Also, a large intrafamilial variability is often reported. Within 6 years after the onset of the disease, patients often die of dementia. HDLS patients may be clinically diagnosed as frontotemporal dementia, Alzheimer disease, multiple sclerosis, Parkinson disease or CADASIL. MRI analyses are initially characterized by asymmetrical patchy abnormalities within the white matter, but that progress towards a confluent and symmetrical pattern. Carriers of the probably pathogenic missense that appear to be unaffected shall undergo MRI analyses, which would give better insights on the nature of this variant. The co-segregation analyses of the second missense will also shed some light in this regard.

Another intriguing finding is the full duplication of *MBP* in a patient (**Figure 34**). Indeed, *MBP* is mainly implicated in the 18q deletion syndrome, characterized by ID, cerebral abnormalities and congenital malformations (cardiac and skeletal defects, dysmorphic features, etc). Due to its function (encoding the myelin-basic proteins, second most abundant CNS myelin protein after PLP), *MBP* has been considered as a reasonable candidate to explain the dysmyelinating disorder observed in many patients with the 18q syndrome. The recurrent reports of full duplication/deletion of *PLP1* in patients with Pelizaeus-Merzbacher disease or spastic paraplegia

type 2 fostered the testing of deletions/duplications of MBP in patients with undetermined Leukodystrophies, which led to negative results (Vaurs-Barriere *et al.*, 2006).

Patient ID	Sex	Gene	Mutation	Inheritance	Causative status
X-linked genes					
CRN154	F	*PLP1*	c.37_38insA, p.Gly13Glufs*26 htz	?	Certain
CRN193	M	*PLP1*	**Full deletion**, hemz	Inherited	Certain
CRN284	M	*PLP1*	**Full duplication**, hemz	?	Certain
CRN238	M	*SLC16A2*	c.1621+2T>C, p. ?, hemz	Inherited	Certain
CRN271	M	*SLC16A2*	c.534_549del, p.Phe179Alafs*20, hemz	?	Certain
CRN299	F	*ABCD1*	c.896A>G, p.His299Arg, htz	?	Possible
Autosomal recessive genes					
CRN214	M	*ALDH3A2*	c.25_50del, p.Arg9Alafs*36, hmz	?	Certain
CRN173	F	*DARS2*	c.1313T>G, p.Val438Gly htz + c.505C>T, p.Leu169Phe, htz	Inherited from each parent	Certain
CRN211	F	*EIF2B2*	c.512C>T, p.Ser171Phe htz + c.570del, p.Ile190Metfs*2, htz	?	Certain
CRN249	M	*EIF2B2*	**c.638A>G, p.Glu213Gly**, htz + **c.599G>T, p.Gly200Val**, htz. Both variants are in trans	?	Certain
CRN296	F	*EIF2B5*	**c.318A>T, p.Leu106Phe**, htz + **c.338G>A, p.Arg113His**, htz	?	Certain
CRN175	F	*MLC1*	c.177+1G>T, p.?, hmz	?	Certain
CRN216	F	*POLR3A*	c.2434G>A, p.Gly812Ser, htz + c.1988T>C, p.Ile663Thr, htz	Inherited from each parent	Certain
CRN277	M	*POLR3A*	c.1045C>T, p.Gln349*, htz + c.2360-3C>G, p. ?, htz	?	Certain
CRN253	M	*POLR3B*	**c.1568T>A, p.Val523Glu,** htz + c.2084-6A>G, p.?, htz	?	Certain
CRN244	M	*POMGNT1*	c.385C>T, p.Arg129Trp, htz + **c.636C>T, p.=**. Both are in trans	?	Certain
CRN192	M	*RNASEH2B*	**c.529G>A, p.Ala177Thr**, hmz	?	Certain
CRN257	M	*RNASEH2B*	**c.529G>A, p.Ala177Thr**, htz + c.136del, p.Gly46Glufs*18, htz	?	Certain
CRN274	M	*RNASEH2B*	**c.529G>A, p.Ala177Thr**, hmz	?	Certain
CRN297	F	*RNASEH2B*	**c.529G>A, p.Ala177Thr**, htz +	?	Certain
CRN266	M	*POLR3A* or *POLR3B*	*POLR3B*: c.2899A>C, p.Ser967Arg, htz + c.1346T>C, p.Leu449Pro htz *POLR3A*: c.1745G>A, p.Arg582His, htz + c.2534C>T, p.Thr845Ile, htz	?	Probable
CRN172	M	*FKRP*	c.535A>G, p.Thr179Ala htz + c.1043T>G, p.Leu348Arg htz	?	Possible
CRN241	M	*LAMA2*	c.8357G>A, p.Arg2786His, htz +	?	Possible

			c.6128A>G, p.Gln2043Arg htz		
CRN285	M	*LAMA2* or *SOX10* or *SPTAN1*	*LAMA2*: c.2512G>A, p.Gly838Arg , htz + c.2462C>T, p.Thr821Met, htz *SOX10* (dominant): c.773G>T, p.Arg258Leu, htz *SPTAN1* (dominant): c.2197C>G, p.Arg733Gly, htz	?	Unlikely
Autosomal dominant genes					
CRN174	F	*GFAP*	c.235C>T, p.Arg79Cys, htz	*de novo*	Certain
CRN267	M	*SPTAN1*	c.1534T>G, p.Phe512Val, htz	?	Probable
CRN198	F	*CSF1R*	c.1699A>G, p.Thr567Ala, htz	?	Possible
CRN206	M	*CSF1R* or *GBE1*	*CSF1R*: c.1700C>T, p.Thr567Met, htz (or *GBE1*: c.1618+3_1618+6del, p.?, htz + ?)	Inherited (Pa), inherited (Ma)	Possible
CRN235	F	*SPTAN1* or *HSPD1*	*SPTAN1*: c.6728C>G, p.Ser2243Trp, htz *HSPD1*: c.185T>C, p.Val62Ala, htz	?	Possible
CRN208	M	*SOX10*	c.782G>A, p.Arg261His, htz	?	Unlikely
CRN264	M	*SPTAN1*	c.7309-12_7309-11del, p. ?, htz	?	Unlikely
Single allele in autosomal recessive genes					
CRN204	F	*GBE1*	c.691+2T>C, p.?, htz + ?	?	VOUS
CRN199	M	*MBP* or *ARSA*	*MBP*: full duplication *ARSA*: c.109_116del, p.Asp37Leufs*36, htz + ?	?	Possible
CRN200	M	*PEX1*	c.2097dup, p.Ile700Tyrfs*42 htz + ?	?	VOUS
CRN273	M	*POLR3A*	**c.3013C>T, p.Arg1005Cys**, htz + ?	?	Possible
CRN239	M	*POLR3B*	**c.1568T>A, p.Val523Glu**, htz + ?	?	Unlikely
CRN270	M	*RNASEH2B*	**c.529G>A, p.Ala177Thr**, htz + ?	?	VOUS
CRN278	M	*RNASEH2C*	c.404del, p.Glu135Glyfs*14, htz + ?	?	VOUS
CRN258	M	*SAMHD1*	c.2T>A, p.Met1?, htz, + ?	?	Unlikely

Table 21: Candidate variants detected by NGS in 122 patients with undetermined leukodystrophies.

M: male; F: female; in bold: previously reported mutations in other patients; ?: co-segregation analyses pending; VOUS: variant of unknown significance.

The report of a duplication of *MBP* appears very unusual: some 18q23 duplications have been reported but never encompassing *MBP*. No structural variants affecting the full length *MBP* are reported in the database of genomic variants (DGV) except one duplication encompassing the last three exons of the longest *MBP* isoform (Xu *et al.*, 2011). Co-segregation analyses of this CNV along with the mapping of the breakpoints will demonstrate whether gene expression may be disrupted or not and provide more insights regarding its putative impact.

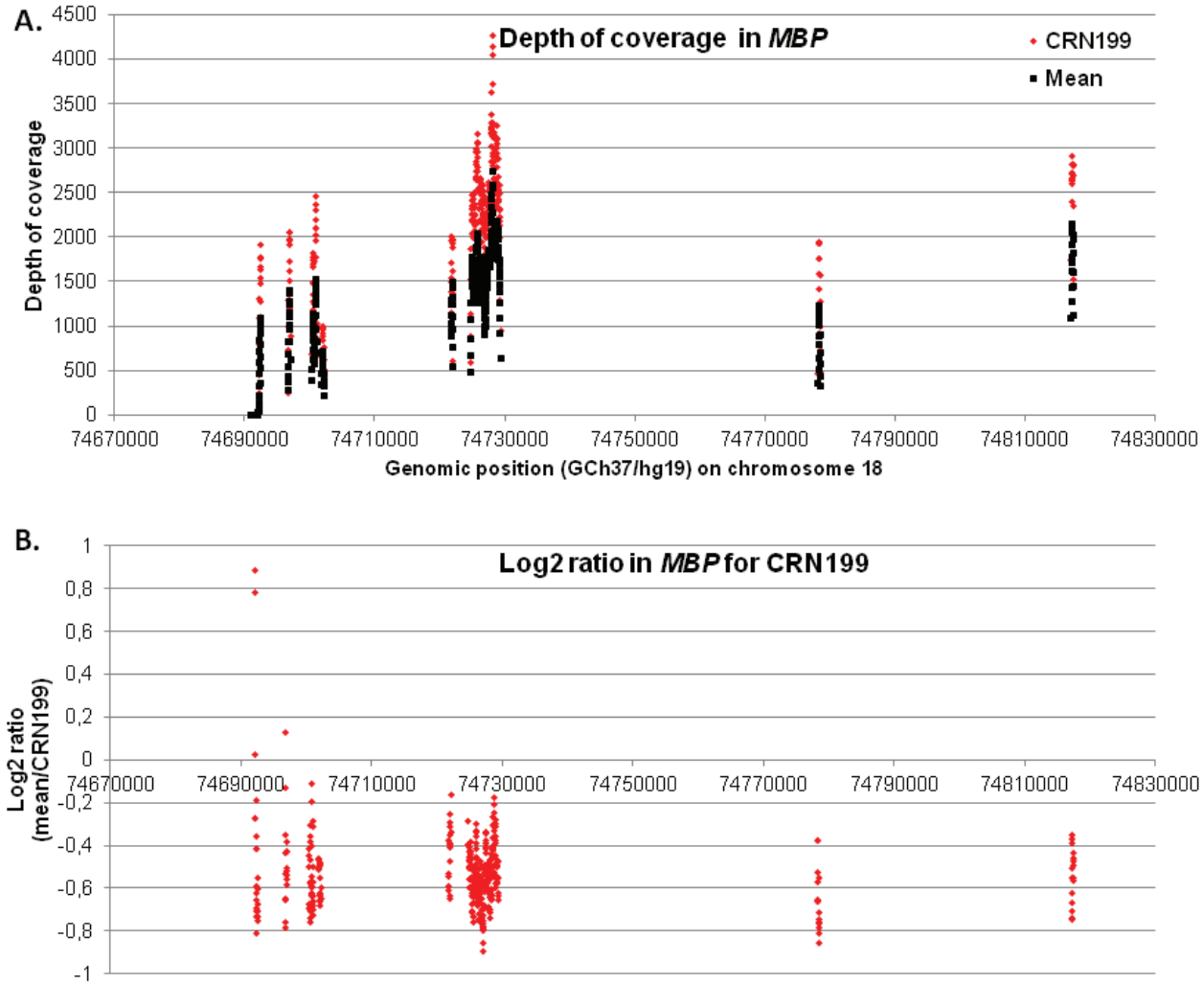


Figure 34: Illustration of the duplication of MBP in patient CRN199.

A. Comparison of the depth of coverage between patient CRN199 (red) and the mean of 8 random patients from the same sequencing run (black); **B.** Log2 ratio of the fold change (mean/ CRN199) demonstrating the presence of a heterozygous duplication (values close to -0.5).

Lastly, we also identified single truncating alleles (splice site disruptions, frameshift variations or start loss) at the heterozygous state in six patients (CRN206, CRN204, CRN199, CRN200, CRN278 and CRN258 in *GBE1*x2, *ARSA*, *PEX1*, *RNASEH2C* and *SAMHD1* respectively) in autosomal recessive genes. We posit that either we missed the second mutation (located in a non targeted region), either there might be a polymorphic allele in trans suggesting a hypomorphic mechanism such as in TAR syndrome (Albers *et al.*, 2012), or that there might exist some dominant forms associated to some of these genes (which does not seem plausible at least for *ARSA*, due to the frequency of the pseudodeficiency allele). The co-segregation analyses will permit us to conclude at least regarding the last hypothesis.

5. Clinical findings

PLP1

CRN193 carries a full hemizygous deletion of *PLP1* and is now 24 years old. He presents with a general psychomotor retardation and regression with a spastic tetrapyramidal syndrome. He presented delayed seating and standing acquisitions, and never acquired walking. In early infancy, he was operated for vesicoureteral reflux. Since 15 years old, he cannot reach the vertical position. He also had many urinary tract infections. He now moves with an electric wheelchair, and is less and less interacting socially. His status seems to progressively deteriorate. MRI analyses show diffuse bilateral white matter abnormalities, sparing the basal ganglia and association fibers. He undertook an extensive metabolic balance that appeared normal (arysulfatase, hexosaminidase, alphagalactosidases, galactosylceramides, very long chain fatty acid; sialic acid, folate and neurotransmitters levels in CSF). The only abnormal finding was high sulfatides level in urine. He also had multiple nevi requiring dermatological monitoring. He has three healthy brothers and sisters. His sister's son presents at 15 months old some neurological symptomes, with developmental delay and convulsions. He will hence be tested for the deletion of *PLP1*.

CRN-284 is a male of 25 years-old, who was first hospitalized at 6 months of age for a severe hypotonia, developmental delay and nystagmus. An MRI analysis at 16 months of age demonstrated a diffuse abnormality of the white matter, evocative and compatible with Pelizaeus-Merzbacher disease.

SLC16A2

CRN-238 is a male of 24 years-old presenting with major hypotonia and speech delay. MRI analysis demonstrates a delayed myelinisation without cerebellar syndrome. Intellectual disability is not mentioned. One brother presents with similar clinical features. The mutation is maternally inherited, and detected in both affected brothers.

CRN-271 is a male of 20 years-old, with an early-onset major hypotonia, epilepsy and nystagmus. Intellectual disability is not mentioned. A maternal uncle died of unknown cause at 15 months of age. MRI analysis in the proband demonstrates hypomyelination.

RNASEH2B

CRN-192 carries the recurrent mutation (c.529G>A, p.Ala177Thr) at the homozygous state, although the parents are not supposedly consanguineous. Both are Moroccan but from distinct regional origin. There is no other familial history of neurodevelopmental disorders. The patient acquired walking at 13 months, so without particular delay. At the same age, he developed an episode of intense fever without apparent infectious origin that persisted for a month. The first

neurologic symptoms appeared at 22 months-old, when he developed gait disturbances, with recurrent falls, and speech delay. MRI analyses showed hyperintensity of the periventricular white matter with a discrete enlargement of the lateral ventricles and thinning of the corpus callosum. No other abnormalities were noted among all other tests (electromyography, electroencephalogram, auditory and visual tests, CPKs, metabolic tests). At 33 months old, IRM showed no evolution, with similar abnormalities of white matter. At 5 years old, the patient presents with a stable pyramidal syndrome of lower limbs. His legs are a little stiff when walking, resulting in still a few falls. There are no apparent learning disabilities and correct speech capacities despite of the initial delay. The patient presents now with a phenotype of stable spastic paraplegia.

III- Discussion

Leukodystrophies are a large group of genetically heterogeneous disorders with poor diagnostic yield in patients when no evident imaging, metabolic or morphological clue is detected. The proposed method of targeted sequencing of 68 genes implicated in leukodystrophies or leukoencephalopathies in a cohort of 120 patients without molecular diagnosis led to a diagnostic yield of 17-25%. Most of the detected mutations are located in autosomal recessive genes, but most reported genes associated to leukodystrophies and leukoencephalopathies are associated with an autosomal recessive mode of inheritance. With only 3 X-linked genes included in our design, we detected mutations in patients in all three of them highlighting their importance. A single causative mutation (and two putative, in *CSF1R*) was detected in autosomal dominant genes.

Interestingly, several genes were hit more than once, with a particular recurrence of mutations in the Aicardi-Goutières syndrome gene *RNASEH2B*, reported in 4/120 patients (*i.e.* 3%, 95% confidence interval: 0.6-6%). In all patients, *RNASEH2B* mutations invariably involve the recurrent mutation p.Ala177Thr at either heterozygous or homozygous state. Other genes such as the X-linked genes *PLP1*, *SLC16A2* or the autosomal genes *EIF2B2* and *POLR3A* were also hit at least twice. Such genes might be more recurrent leukodystrophy genes, which will be confirmed by studying larger cohorts of patients. Alternatively, the clinical spectrum associated to some genes may have been incomplete, and the patients of our cohort may represent one end of the spectrum with atypical clinical features that were not considered initially as related to mutations in those genes.

Regarding the relevance of the targeted sequencing approach, this design is almost exclusive concerning the reported leukodystrophy genes. In sight of the substantial proportion of patients without mutations in those genes, we can issue several assumptions. First, many leukodystrophy genes remain to be discovered, which explains why we do not find mutations in many patients. Second, the presumed white matter disorders primary involving glial cells or myelin abnormalities reported in those patients might be hardly distinguishable from those where the white matter defects are secondary and result from neuronal involvement or other systemic manifestations. Indeed, many genetic and mitochondrial leukoencephalopathy genes were not included in our design. Although most patients with positive diagnosis carry mutations in leukodystrophy genes, two of them carry mutations in *SLC16A2* responsible for Allan-Herndon-Dudley syndrome, which was initially identified as gene associated to severe ID and later associated to white matter defects. Both patients were still addressed for a leukodystrophy diagnosis, based on MRI findings. We thus posit that this rather low (or insufficient) diagnostic yield may be due to the non-specific clinical feature of white-matter defects, as detected by MRI. Although a few leukodystrophy disorders demonstrate highly specific MRI patterns (*e.g.* butterfly pattern in metachromatic leukodystrophy, characteristic pattern in Krabbe disease), the others might be confounded with non primary white matter defects, reported in patients with spastic paraplegia, epileptic syndromes, or ID. It would be interesting for instance

to screen some of the negative patients presenting with additional cognitive defects with our targeted ID panel.

The diagnostic yield of this approach is still significant. The main interrogation is: what would be the diagnostic yield in cohorts of leukodystrophy patients if performing exome sequencing instead? The implementation of exome sequencing on negative patients from our cohort would answer to both our hypotheses, either many patients carry mutations in novel leukodystrophy genes or many patients carry mutations in known genes associated to secondary white matter defects. If the first hypothesis proved to be correct, another updated targeted design could be tested in order to compare both diagnostic yield, and estimate the proportion of novel genes left to discover. In any case, the results of such large-scale exome study will shed much light on the genetics of leukodystrophies, and delineating precisely the boundary of primary white matter disorders.

GENERAL DISCUSSION

During my PhD, I developed targeted next-generation sequencing to test its efficiency as an alternative method for the molecular diagnosis of three different disorders (or group of disorders): Bardet-Biedl and Alström syndromes, intellectual disability, and leukodystrophies. I analyzed over 400 patients with this method (164 patients with BBS or other related ciliopathies, 106 patients with ID and 135 in Leukodystrophies). A positive diagnostic identifying the genetic etiology of the disease could be reported via the diagnostic laboratories of Strasbourg University Hospital or of Robert Debré in 125 patients (79 patients with Bardet-Biedl or Alström syndrome, 26 patients with ID, 20 patients with leukodystrophy). Obtaining a molecular diagnosis is primordial for patients and families since it marks the end of a laborious diagnostic search and of a long phase of misunderstanding and guilt regarding the disease of a close relative. It can also have major consequences on the medical prognosis and care of patients, and can even indicate therapeutic options. At the broader level of the family, it is the first step towards genetic counseling, allowing the calculation of recurrence risks.

The overall diagnostic yield reached with targeted NGS in all three disorders is nonetheless variable. For a disorder of relative genetic heterogeneity and specific clinical manifestations like Bardet-Biedl syndrome, the diagnostic yield is as high as 81% in patients that comply with the diagnostic criteria. For such disorders, targeted NGS seems a highly efficient method to implement in routine diagnostic settings. For disorders of much more extensive genetic heterogeneity, and with larger clinical variability such as intellectual disability or leukodystrophies, the diagnostic yield reached with targeted NGS is much lower (25% and 17% respectively). To assess whether the implementation of targeted NGS is relevant in such disorders, a comparison with the efficiency of alternative diagnostic methods is required. While for the molecular diagnosis of intellectual disability targeted-NGS still seems to be a reasonable approach, for the one of leukodystrophies it still needs to be compared to the efficiency of other strategies such as exome or full genome sequencing.

As illustrated throughout this manuscript, the emergence of next-generation technologies has tremendously improved our understanding of the genetic causes of many genetically heterogeneous disorders. Soon, multifactorial disorders will also benefit from this technology. Whether NGS technologies are ‘the’ ultimate solution or not will be proven within the next few years. Targeted sequencing seems to be a relatively efficient approach for disorders of limited genetic heterogeneity or of homogeneous clinical entity. As long as it remains of much higher throughput, higher quality, and lower costs than exome sequencing, targeted sequencing will remain the optimal choice for the diagnosis of those disorders. However, the more exome sequencing will tend to perfection (*i.e.* with an even and sufficient depth of coverage across all exons and with reduced costs) the more targeted sequencing will become obsolete. Among other advantages, it will allow a universal implementation whatever the condition (chromosomal and monogenic disorders, cancer genomics, etc.), which will simplify considerably the task for diagnostic laboratories. The diverging point will be at the bioinformatic pipeline step, requiring various strategies to handle the data depending on

whether single nucleotide variants, indels, structural variants, mosaic or multiple events are sought. Deciphering the putative oligogenic model of neurodevelopmental disorders for instance will depend on our capacity to develop accurate and pertinent programs to model the oligogenic scenario and then process and analyze the data.

The same requirements apply for full genome sequencing. At the very emergence of next-generation sequencing technologies, the throughput of sequencers was considerably restricting their application for full genome sequencing. Using the initial NGS machines, the first human genome was sequenced in two months at a depth of coverage of less than 10X (Wheeler *et al.*, 2008). Now, a full genome can be sequenced in a single run in as little as a day using the ultra high throughput HiSeq2500 sequencer from Illumina. The limitations now lie in our capacity to interpret the generated data. For now, we are only able to interpret mostly variants located within coding regions, which means that in a full-genome we are only able to analyze 1.5% of the data corresponding to… the exome. This is the reason of the slow inertia associated to full genome projects. However, some teams do advocate full-genome sequencing instead of exome-sequencing for prenatal diagnosis, particularly since reducing the sample processing time (*i.e*. removing the enrichment step) delivering the results more rapidly, and leading to a more even representation of exons than in exome sequencing by eliminating the bias introduced with the enrichment step (Saunders *et al.*, 2012). One of the major challenges within the next decade in human genetics lies in our capacity to understand and model the impact of variations, and the application of this through the development of bioinformatics programs capable of analyzing and interpreting complex data from well-defined patients with full genome sequencing.

A prior requirement will nonetheless be to gather and share accurate and extensive clinical information for each patient analyzed. Indeed, it is currently one of the obstacles the community is facing with the development of large public databases overflowing with variants detected by massive high-throughput sequencing studies, where returning to the phenotype and clinic of the person carrying a variant is impossible. This precludes accurate interpretation of the impact of a particular variant on an individual's phenotype. The development of a database of SNVs and indels on the model of the DGV (database of genomic variants) database is necessary, with even more extensively detailed clinical manifestations. For instance, in ID where there are over 300 genes identified, we can consider that there are approximately the same number of rare disorders or syndromes. The rarity of each disorder in terms of affected patients complicates considerably the studies of genotype-to-phenotype correlations, or the evaluation of efficacy of specific pre or post-symptomatic treatments under development. An extensive clinical evaluation and report would be therefore crucial for both, but are often not available because of the limited time available for clinicians to this task for each patient. One alternative that we are currently trying to develop and we feel would be much more efficient, is 'patients-oriented' databases. In such databases, the clinical features would be provided by patients' families who presumably would find those highly valuable and may be highly motivated in providing such information. The approach was already

proven successful in GWAS and clinical discovery for Parkinson's disease, type 2 diabetes and other medical phenotypes (Do *et al.*, 2011; Tung *et al.*, 2011; Wicks *et al.*, 2011). In clinically heterogeneous disorders, such databases would allow refining clinical features associated with a particular rare gene defect. In genetically heterogeneous or multifactorial disorders such as ID or ASD, that information would be of high value to recruit patients for a global treatment proposed for ID or ASDs that are often consider as a single clinical entity, and then to follow on a longitudinal manner the putative benefits or negative effects of the treatments according to each particular gene defect. Indeed, one treatment might be highly efficient for many gene defects, but inefficient in another. The development of such complete 'genotype + phenotype' public databases appears essential and would open the way towards personalized medicine.

Assuming we overcome these huge challenges leading to a comprehensive understanding of genetic variants and their implication in disease, what comes next? In which way will societies incorporate NGS technologies in daily lives? In the same line as the *Gattaca* movie (which was rather visionary for 1997) one of the obvious short-term outcomes is the routine implementation of prenatal exome (or full genome) sequencing to select against genetic disorders. This will hopefully be tightly controlled (in some countries) by ethical laws and jurisdictions and will only allow terminating pregnancies when the impact on the fetus' genotype is associated with high probability to a severe and incurable disorder. Another outcome that might soon be proposed by private companies would be performing exome sequencing in couples to determinate whether they are 'genetically compatible' or not. For instance, if both carry a heterozygous mutation in a same gene associated to an autosomal recessive monogenic disorder they may be advised to monitor the pregnancy for the apparition of suspicious signs or simply perform Sanger (or exome) sequencing to screen the gene of concern. Pre-conception screening is already systematically done for high prevalence diseases in some countries (*e.g.*: β-thalassemia in Cyprus). I believe that, as time will pass, our way of selecting against disorders will become more and more stringent (since our understanding will increase more and more regarding the impact of each genetic variation). The null risk for genetic disorders will nonetheless probably never exist, since there will always be *de novo* mutations or structural rearrangements occurring or epigenetic modifications that will lead to disorders that we will not have yet fully understood.

In parallel, as we unravel the genetics of many disorders, developments on therapeutic research evolve rapidly, highlighting therapeutic targets or molecular therapeutic tools, but still rarely leads to effective treatments. For instance, in Fragile-X syndrome, the better understanding of FMRP function and its mRNA targets led to the identification of several therapeutic targets currently under clinical trials: the use of some mGluR5 antagonists suggested modest (and uneven) clinical benefits, and some are under phase III clinical trials (Berry-Kravis *et al.*, 2009; Jacquemont *et al.*, 2011); the use of $GABA_a$ and $GABA_b$ agonists although proved efficient in FXS animal models still remained to be proved efficient in human, and was under a large-scale phase III clinical study that was stopped (Berry-Kravis *et al.*, 2012); matrix metalloproteinase agonists were also

tested in human and suggested to efficiently improve cognitive, language and behavioral defects (Utari *et al.*, 2010). Nonetheless, the evaluation criteria in clinical trials mostly involve behavioral or physiological parameters, and never evaluate their impact on cognitive function.

Another neurodevelopmental disorder, X-linked adrenoleukodystrophy (ALD), was a pioneer disease for testing and proving the efficiency of bone-marrow transplantation (Aubourg *et al.*, 1990) and later of gene therapy using lentiviral vectors (Cartier *et al.*, 2009). Treatments with other compounds such as valproic acid usually used as an antiepileptic have also been proposed (Fourcade *et al.*, 2010; Salsano *et al.*, 2012). Nonetheless, forty years after its first suggestion (Friedmann and Roblin, 1972), apart for ALD very few gene therapies have been approved and proposed for routine treatments of genetic disorders (Leber congenital amaurosis, and severe combined immunodeficiencies; (Maguire *et al.*, 2008; Fischer *et al.*, 2010)). One recent breakthrough theoretically opening new therapeutic hopes for Down syndrome patients is the success of inactivation of one extra-copy of chromosome 21 in patients' stem cells using genome editing (Jiang *et al.*, 2013). This inactivation of the supernumerary chromosome already demonstrated efficiency in *in vitro* studies and opens new therapeutic avenues using 'chromosome therapy', if used at embryonic stages. Nonetheless, in child or adult patients, *i.e.* at postnatal stages, this option does not appear possible since every cell of the patient (especially the ones from the CNS) would need to inactivate their supernumerary chromosome to stop the disease.

Other research seem to move towards molecular therapeutics and personalized medicine using stop-codon read-through, exon-skipping or genome editing approaches. Some are under preclinical or clinical trials. Stop codon read through and exon skipping strategies aim at forcing the production of a protein whose encoding transcripts are otherwise rapidly degraded due to nonsense-mediated mRNA decay in patients' stem cells. Both approaches are under clinical trial for Duchenne muscular dystrophy, but appear deceptive at first sight (Bhagavati, 2014). The genome editing approaches (using CRISPR, zinc finger nucleases or TALEN) aim at correcting a point mutation, by replacing a nucleotide to change a premature stop codon into another amino acid residue. Such genome editing was recently proven to be efficient on primate embryos (Niu *et al.*, 2014). Nonetheless, this approach only seems to be applicable at embryonic stages, or for genetic disorders in which a specific type of cells is affected (*e.g.* autoimmune diseases). Altogether, the development of efficient therapeutic strategies for genetic disorders thus appears to be of low yield (*e.g.* Lesch-Nyhan syndrome, characterized in 1968, first gene therapy trials in animals in 1984, still no treatments available 30 years late ; (Kelley, 1968; Miller *et al.*, 1984)). We can posit that the future implementation of routine exome/genome prenatal screening might outstrip the usefulness of these therapies, making them obsoletes.

Supposing a therapy is found for one particular disorder, for which prenatal diagnosis was previously accepted, how do the laws evolve in regard of this particular disorder? Are prenatal diagnoses still authorized or not? The requirements for prenatal diagnosis in France, according to

the French law are the following: '*there should be a high probability that the unborn child suffers from a disorder of particular gravity recognized as incurable at the moment of diagnosis*.' Severe disorders for which therapeutic options exist should not therefore be accepted as the object of prenatal diagnoses' demands. Nonetheless, when consulting the figures related to the disorders for which prenatal diagnoses were performed in France over the last few years, some disorders such as hemophilia or Sickle-cell disease appear (15 and 22 medical abortions respectively in 2011; Agence de Biomédecine source), confirming that the answer remains not so evident. Also, as prenatal diagnoses are controlled in France by regional pluridisciplinary expert commissions, the acceptation of prenatal diagnosis demands may vary from one to another center, leading to what we could call 'clinical tourism' (*i.e.* families presenting demands successively to several regional commissions). I feel that there should be homogeneous guidelines providing with a clearer definition regarding what is a 'severe' disorder, in order to provide a homogeneous and less subjective consideration for prenatal diagnoses in France.

To conclude, next-generation sequencing is a fantastic tool that opened the way towards novel diagnostic approaches. For patients, molecular diagnosis is the first step towards genetic counseling but also personalized medicine and therapeutic options, whether they are gene therapies, molecular therapeutics or pharmacologic treatments. Nonetheless, the generalization of the application of NGS as a diagnostic or preventive tool raises many ethical issues.

Appendix 1: *DPY19L2* deletion as a major cause of globozoospermia.

Koscinski I*, Elinati E*, Fossard C, **Redin C**, Muller J, Velez de la Calle J, Schmitt F, Ben Khelifa M, Ray PF, Kilani Z, Barratt CL, Viville S.

Aim of the study: Reporting *DPY19L2* deletions in four consanguineous families as responsible for the globozoospermia phenotype in affected males, and dissecting the underlying mechanism responsible for those recurrent duplications.

Contributions: Minor. Performed bioinformatic mining to understand the molecular mechanism responsible for the *DPY19L2* deletions recurrently observed in patients. Analyzed and identified two Alu elements within flanking LCRs, suggesting a nonallelic homologous recombination (NAHR) mechanism. Wrote the findings accordingly.

REPORT

DPY19L2 Deletion as a Major Cause of Globozoospermia

Isabelle Koscinski,[1,2,11] Elias ElInati,[2,11] Camille Fossard,[2] Claire Redin,[2] Jean Muller,[2,3] Juan Velez de la Calle,[4] Françoise Schmitt,[5] Mariem Ben Khelifa,[6] Pierre Ray,[6,7,8] Zaid Kilani,[9] Christopher L.R. Barratt,[10] and Stéphane Viville[1,2,*]

Globozoospermia, characterized by round-headed spermatozoa, is a rare (< 0.1% in male infertile patients) and severe teratozoospermia consisting primarily of spermatozoa lacking an acrosome. Studying a Jordanian consanguineous family in which five brothers were diagnosed with complete globozoospermia, we showed that the four out of five analyzed infertile brothers carried a homozygous deletion of 200 kb on chromosome 12 encompassing only *DPY19L2*. Very similar deletions were found in three additional unrelated patients, suggesting that *DPY19L2* deletion is a major cause of globozoospermia, given that 19% (4 of 21) of the analyzed patients had such deletion. The deletion is most probably due to a nonallelic homologous recombination (NAHR), because the gene is surrounded by two low copy repeats (LCRs). We found *DPY19L2* deletion in patients from three different origins and two different breakpoints, strongly suggesting that the deletion results from recurrent events linked to the specific architectural feature of this locus rather than from a founder effect, without fully excluding a recent founder effect. *DPY19L2* is associated with a complete form of globozoospermia, as is the case for the first two genes found to be associated with globozoospermia, *SPATA16* or *PICK1*. However, in contrast to *SPATA16*, for which no pregnancy was reported, pregnancies were achieved, via intracytoplasmic sperm injection, for two patients with *DPY19L2* deletion, who then fathered three children.

In a previous work, we described the identification of a mutation in *SPATA16* (MIM 609856) responsible for a complete form of globozoospermia (MIM 102530) in three affected brothers of an Ashkenazi Jewish family.[1] We report here the involvement of *DPY19L2* on chromosome 12 in cases of complete forms of globozoospermia. Our study started with the analysis of a Jordanian family of ten siblings, including two sisters and three brothers who were naturally fertile (all of them having 3–7 children) and five affected brothers with a complete form of globozoospermia (Figure 1A). Intracytoplasmic sperm injection (ICSI) was performed for the five brothers at the Farah Hospital, Amman, Jordan, and despite a total of 20 cycles there was only one single-term pregnancy, as well as two miscarriages.[2] Because of the consanguineous marriage of the grandparents, who were first-degree cousins, we postulated that the genetic abnormality was transmitted as an autosomal-recessive disorder. Four of the five affected individuals, Globo1 to Globo4 (IV-1 to IV-4 in Figure 1A), and all three fertile individuals, brothers B1 to B3 (IV-5 to IV-7 in Figure 1A), consented to give blood samples for research (Figure 1A).

We performed a genome-wide scan analysis of all seven brothers, using 10K SNP arrays (Affymetrix Genechip). Regions of homozygosity were defined by the presence of at least 25 consecutive homozygous SNPs. We identified a unique region of 30 homozygous SNPs shared by the four affected brothers; the fertile brothers were heterozygous for this region (Figure 1B). This region of about 6.4 Mb on chromosome 12 (positions 63060074 to 69409722, on the GRCh37/hg19 version of the human genome, corresponding to SNPs rs345945 and rs2172989; see Table S1 available online) contains 101 genes, 40 of which are expressed in the testis, according to the UCSC Genome Browser[3] (Table S2), and thus are potentially involved in spermatogenesis. We selected *DPY19L2* as the most plausible candidate gene because of its predominant testis expression and its potential involvement in cellular polarization.[4] The spermatozoon is a highly polarized cell, and the initial stages of spermiogenesis involve polarization of the round spermatid. Functional studies of DPY-19 (an ortholog of DPY19L2) in *C. elegans* highlighted its role in the establishment of polarity in the migrating neuroblasts.[4] *DPY19L2* is composed of 22 exons and belongs to the *DPY19L* family[5] coding for proteins harboring 9–11 predicted transmembrane domains. This gene family, encompassing four genes and six pseudogenes, has a complex evolutionary history involving several duplications and pseudogenizations.[5] In particular, *DPY19L2* stems from an initial duplication of *DPY19L1* on chromosome 7, followed by a recent relocalization on chromosome 12 in humans, leaving the initial copy on chromosome 7 as a pseudogene (*DPY19L2P1*). This evolutionary event may rely on the presence of low copy repeats

[1]Service de Biologie de la Reproduction, Centre Hospitalier Universitaire, Strasbourg, F-67000 France; [2]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de Santé et de Recherche Médicale (INSERM) U964/Centre National de Recherche Scientifique (CNRS) UMR 1704/Université de Strasbourg, 67404 Illkirch, France; [3]Laboratoire de Diagnostic Génétique, CHU Strasbourg, Nouvel Hôpital Civil, 67000 Strasbourg, France; [4]Unité FIV, Clinique Pasteur, 29200 Brest, France; [5]Laboratoire de Biologie de la Reproduction, Centre hospitalier de Mulhouse 68100 Mulhouse, France; [6]Faculté de Médecine-Pharmacie, Domaine de la Merci, Université Joseph Fourier, Cedex 9, Grenoble, France; [7]TIMC-IMAG, UMR CNRS 5525, Team AGIM, 38710 La Tronche, France; [8]Biochimie et Génétique Moléculaire, CHU Grenoble, 38700 Grenoble, France; [9]Farah Hospital, Amman 11183, Jordan; [10]Reproductive and Developmental Biology, Maternal and Child Health Science Laboratories, Centre for Oncology and Molecular Medicine, Ninewells Hospital, University of Dundee, Dundee, DD1 9SY, Scotland
[11]These authors contributed equally to this work
*Correspondence: viville@igbmc.fr
DOI 10.1016/j.ajhg.2011.01.018.

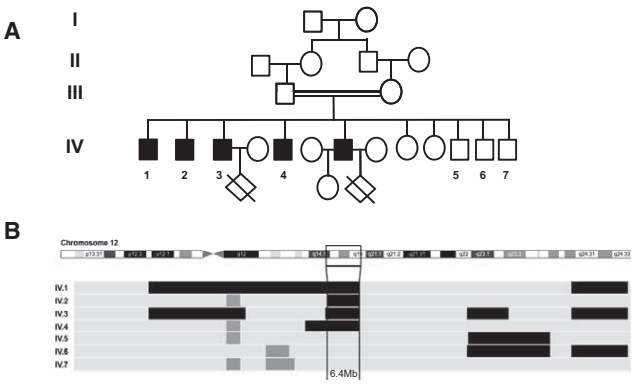
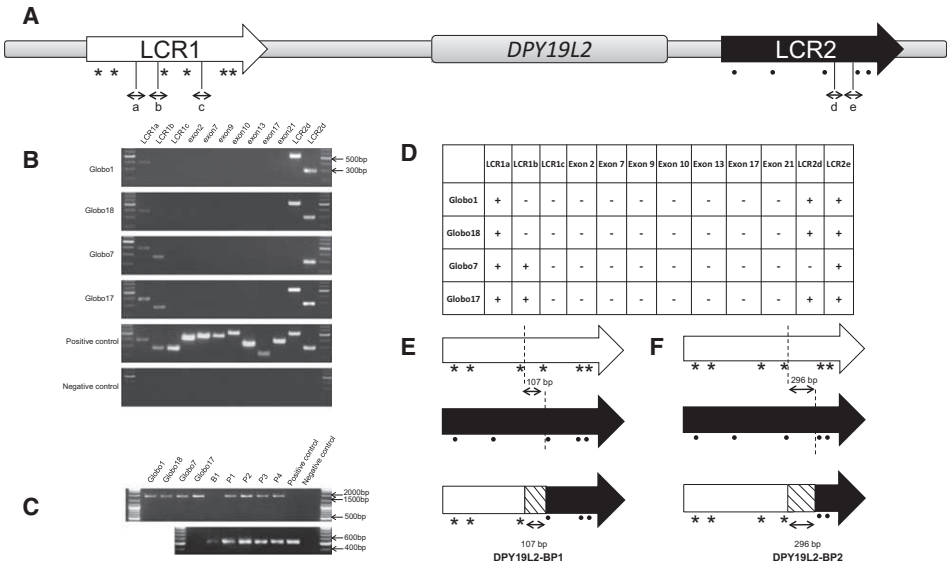


Figure 1. Pedigree and Linkage Analysis of the Jordanian Family
(A) Jordanian pedigree showing second-degree consanguinity.
(B) SNP array results of the four infertile and three fertile brothers for the region of chromosome12. Shared regions of homozygosity are visualized by the HomoSNP software, which displays one patient per line. The areas of homozygosity with 25 or more SNPs are black, whereas homozygosity regions defined by 15–25 consecutive SNPs are gray. Regions of heterozygosity are light gray. The four affected siblings, Globo1 to Globo4 (IV-1 to IV-4), share a region of homozygosity of 6.4 Mb on chromosome 12, which is not shared by the three fertile brothers, B1 to B3 (IV-5 to IV-7).

(LCRs) surrounding the DPY19L2 locus (Figure 2A). Because of the high level of conservation between *DPY19L2* and *DPY19L2P1*,[5] the design of *DPY19L2*-specific oligonucleotides was complex. PCR conditions were optimized for exons 2, 7, 9, 10, 13, 17, and 21, resulting in specific amplifications in controls but never in the patients, suggesting a large deletion of the whole gene (Figure 2B). In order to better characterize the two described LCRs surrounding *DPY19L2*, we performed a Blastn analysis[6] using a 100 kb region up- and downstream of the gene. This approach revealed a highly similar duplicated region (96.5% identity), corresponding to the two identified LCRs (Figure 2B). The telomeric LCR (named here LCR1) is 26,998 bp long and is located upstream of exon 1 (64,119,249–64,146,247), and the centromeric LCR (named here LCR2) is 28,200 bp long and starts ~1000 bp downstream of exon 22 (63,923,419–63,951,619). In order to define as precisely as possible the breakpoints within the recombined LCRs, specific oligonucleotides were designed to walk on both sides of the deletion (Figures 2B–2D and Table S4). This allowed us to define a first target segment of 107 bp, which we named DPY19L2-BP1. Further refinement is not possible because of the high percentage of identity shared by the two LCRs (Figure 2E and Figure S1). The exact size of the deletion is 195,150 bp. All three healthy brothers B1, B2 and B3 are homozygous wild-type.

The *DPY19L2* deletion was then screened in 24 globozoospermic patients from 20 independent families, originating from Algeria, France, Iran, Italy, Lybia, Morocco, and Tunisia. Twelve of them had a complete form of globozoospermia, and six had a partial form (defined as between 20% and 90% of the spermatozoa having round heads[7]). For the remaining two, we did not have any detailed information (Table 1). We found a similar deletion of *DPY19L2* in three unrelated cases presenting with a complete form of globozoospermia (Figures 2C and 2D), two cases originating from France (Globo17 and Globo18) and one familial case from Algeria comprising two brothers (Globo7 and Globo8). One of the French patients (Globo17) had two children, born after ICSI treatment (Table 1). We localized the breakpoints for all three additional patients. For Globo18, the deletion breakpoints are located, as suggested by the PCR walk (Figure 2D), in the same 107 bp LCR area (DPY19L2-BP1) as that of the Jordanian family. Both of his parents are heterozygous for the deletion, as demonstrated by the possibility of amplifying both the recombined region (i.e., absence of *DPY19L2*) and exon 10 (i.e., at least one copy of *DPY19L2*) (Figure 2C). The analysis of Globo7 and Globo8 allowed us to identify a different breakpoint, which we named DPY19L2-BP2, (Figure 2D), in an area of 296 bp, positioned 169 bp 3′ to the 107 bp area. Interestingly, Globo17 is a heterozygous composite, because the two different breakpoints (DPY19L2-BP1 and DPY19L2-BP2) could be observed (Figure 2D). The analysis of his parents showed their heterozygous status for the deletion and allowed us to detect the two different breakpoints on each allele. The maternal allele carries DPY19L2-BP2 and the paternal allele carries DPY19L2-BP1. For both parents we could amplify exon 10, demonstrating the presence of at least one copy of *DPY19L2* (Figure 2C). The deletion of *DPY19L2* was not found in the homozygous state in 105 European-descent males and 101 fertile Jordanian males.

Several studies have reported copy-number variants (CNVs) at the genome-scale level that are fully available and searchable in databases such as the Database of Genomic Variants (DGV).[8] DGV reports, in its latest version (November 2010), 66,741 CNVs from 42 studies. At the DPY19L2 locus, DGV references 29 CNVs that are described by 12 different studies encompassing different ethnic groups (Figure 3 and Table S3). Considering only the nonredundant CNVs identified in non-BAC (bacterial artificial chromosome) studies and those that are fully overlapping *DPY19L2*, only five studies can be considered.[8–13] The detailed inventory of the CNVs revealed 64 duplications and 22 heterozygous deletions out of 4886 patients studied, whereas no homozygous deletion was reported (Table 2). Interestingly, about three times more duplications than deletions can be observed at this locus. This is surprising because in most of the cases studied, deletions are more frequent than duplications.[14] We do not have a clear explanation for this, but it would be of interest to determine whether the heterozygous status is affecting male fertility, which would explain the selection against deletion. The frequency of heterozygous deletion in the surrounding region of *DPY19L2* is 1/222, implying an overall disease risk of ~1/200,000 (see Table 2). Examining the localization of the CNVs, one can observe that all deletion breakpoints



	LCR1a	LCR1b	LCR1c	Exon 2	Exon 7	Exon 9	Exon 10	Exon 13	Exon 17	Exon 21	LCR2d	LCR2e
Globo1	+	-	-	-	-	-	-	-	-	-	+	+
Globo18	+	-	-	-	-	-	-	-	-	-	+	+
Globo7	+	+	-	-	-	-	-	-	-	-	-	+
Globo17	+	+	-	-	-	-	-	-	-	-	+	+

Figure 2. Analysis of *DPY19L2*

(A) Schematic representation of *DPY19L2* flanked by two LCRs, LCR1 and LCR2. LCR1 is located upstream of the gene and LCR2 downstream. Different sequences were amplified on both sides of the gene in order to determine the breakpoints: "a," "b," and "c" sequences are localized in LCR1, whereas "d" and "e" sequences are localized in LCR2. Mismatches between the two LCRs sequences are represented by (*) in the 5′ side and by (.) in the 3′ side.

(B) PCR results of *DPY19L2* exons and flanking regions. Patients Globo1, Globo7, Globo17, and Globo18 are deleted for exons 2, 7, 9, 10, 13, 17, and 21, whereas the positive control showed an amplification of all of these exons. All patients showed specific amplifications for region "a" in LCR1 and region "e" in LCR2. Globo1 and Globo18 are missing region "b" in LCR1, which is not the case for Globo7 and Globo17. Globo7 is the only patient missing region "d" in LCR2. A DNA ladder was included on both sides of each gel for accurate determination of the PCR band size.

(C) PCR of *DPY19L2* breakpoints. This PCR was performed with the use of the "LCR1a forward" and "DPY19L2-BP reverse" primers (Table S4). The four globozoospermic patients showed a fragment of 1700 bp. The fertile brother (B1, IV-5 in Figure 1A) is homozygous wild-type, because no fragment was amplified, whereas parents of Globo18 (P1 and P2) and Globo17 (P3 and P4) are heterozygous. A fertile man (positive control) was used as control of the PCR specificity. Amplification of exon 10 (lower panel) was used to control the presence or absence of *DPY19L2*.

(D) Table summarizing the PCR results of the four patients: (+) indicates presence of a PCR band at the corresponding size, and (-) indicates absence of a PCR band.

(E) Schematic representation of the breakpoint area determined for Globo1, Globo18, and the paternal allele of Globo17. The breakpoint area (DPY19L2-BP1) identified by PCR and sequencing is defined between the two discontinued lines and represented by a hatched box in the recombined LCR.

(F) Schematic representation of the homologous recombination that occurred in Globo7 and the maternal allele of Globo17. The breakpoint area (DPY19L2-BP2) identified by PCR and sequencing is defined between the two discontinued lines and represented by a hatched box in the recombined LCR.

recorded in DGV fall into the two LCRs, and this is also true for most of the duplications (Figure 3).

Considering the presence of highly conserved LCRs surrounding *DPY19L2*, the most probable mechanism to explain the multiple occurrence of this deletion would be NAHR between highly similar sequences. Indeed, such a mechanism has been reported for a large number of disorders involving deletions, duplications, inversions, or gene fusions[15–17] (reviewed in[14]). It is estimated that this is one of the most common mechanisms responsible for genomic disorders. So far, no obvious founder effect has been described for genetic disorders involving NAHR.[14] That we found the *DPY19L2* deletion in patients from three different origins (the two French cases are not known to be related), that two different breakpoints were observed, and that Globo18 is not sharing the same haplotype of the Jordanian brothers (Table S1) strongly suggest that the deletion results from recurrent events linked to the specific architectural feature of this locus rather than from a founder effect, without fully excluding a recent founder effect. It is worth noting that the two identified breakpoints are located near an Alu repeat (AluSq2), one being fully included in this repeated element (Figure S1).

Interestingly, the patients with *DPY19L2* deletion present a complete globozoospermia yet have normal or, in one case (Globo8), near-normal concentration of sperm (Table 1). This contrasts with the case in which *SPATA16* is mutated, suggesting that DPY19L2 may disrupt only spermiogenesis and not germ cell proliferation and meiosis. It should also be noted that, despite the testicular expression of the other three members of the *DPY19L* family,[5] it does not appear that there is functional redundancy. No homozygosity was found for the loci of the other functional DPY19L genes.

We selected *DPY19L2* as the primary candidate linked to globozoospermia in our Jordanian family because of its

Table 1. List of Globozoospermic Patients

Patient	Origin	SNP Screening	Sperm Concentration (Million/mL)	Progressive Motility (%)	Acrosome less Morphology (%)	Fertilization Attempts and Results
Complete Globozoospermia						
Globo1	Jordan	Yes	52	22	100	6 ICSI failed
Globo2	Jordan	Yes	223	4	100	3 ICSI failed
Globo3	Jordan	Yes	118	4	100	4 ICSI failed
Globo4	Jordan	Yes	22	30	100	3 ICSI failed (1 miscarriage)
Globo5	France	No	ND	ND	100	ND
Globo6	France	No	ND	ND	100	ND
Globo7	Algeria	No	90	10	100	ND
Globo8	Algeria	No	13.2	2	100	ND
Globo9	France	No	0.35	ND	100	1 ICSI failed
Globo10	France	Yes	38	ND	100	1 ICSI succeeded
Globo11	France	Yes	109	40	100	2 ICSI failed
Globo12	ND	No	ND	ND	100	DSI
Globo13	Iran	No	10.5	ND	100	ND
Globo14	Tunisia	No	52.6	25	100	ND
Globo15	Tunisia	No	ND	ND	100	1 ICSI failed
Globo16	Libya	No	4	8	100	3 ICSI failed
Globo17	France	Yes	78	45	100	IMSI succeeded twice
Globo18	France	Yes	71	30	100	ND
Partial Globozoospermia						
Globo19	Italy	No	50	ND	ND	1 ICSI failed
Globo20	France	Yes	0.6	ND	84	1 IUI + 3 ICSI failed
Globo21	France	Yes	2	ND	ND	2 ICSI failed
Globo22	Morocco	No	ND	ND	ND	ND
Globo23	France	Yes	ND	ND	95	1 ICSI succeeded
Globo24	France	Yes	ND	ND	98	1 ICSI succeeded
Globo25	Tunisia	No	59	50	88	ND
Globo26	Tunisia	No	100	49	99	ND
Globo27	ND	No	ND	ND	ND	ND
Globo28	ND	No	ND	ND	ND	ND

Abbreviations are as follows: ND, not determined; ICSI, intracytoplasmic sperm injection; IUI, intrauterine insemination; IMSI, intracytoplasmic morphologically selected sperm injection; DSI, donor sperm insemination.

testicular expression and function of the DPY-19 ortholog in *C. elegans*, involved in the polarization of neuroblasts.[4] This would suggest that DPY19L2 is not directly involved in the biogenesis of the acrosome, as for SPATA16[1] or PICK1 (MIM 605926),[18,19] but rather in preceding steps associated with, for example, the polarity of the spermatid, therefore allowing the correct positioning of the different organelles such as the acrosome or the flagella. However, this hypothesis is not supported by the examination of the immature germ cells in the semen of three globozoospermic patients (Globo7, Globo17, and Globo18). For Globo7, all stages of spermatogenesis could be observed (Figure 4A) and there was no obvious abnormality in the earlier stages of germ cells. In contrast, no proacrosomal vesicle was observed in the round spermatids, yet they appeared correctly polarized. Indeed, the polarization of the spermatid was indicated by the correct positioning of the chromatoid body (in the area of the attachment of the flagellum to the spermatid nucleus) observed in elongating spermatids. In Globo17 and Globo18 semen samples, we observed elongating spermatids as well as spermatozoa that are comparable to that of Globo7

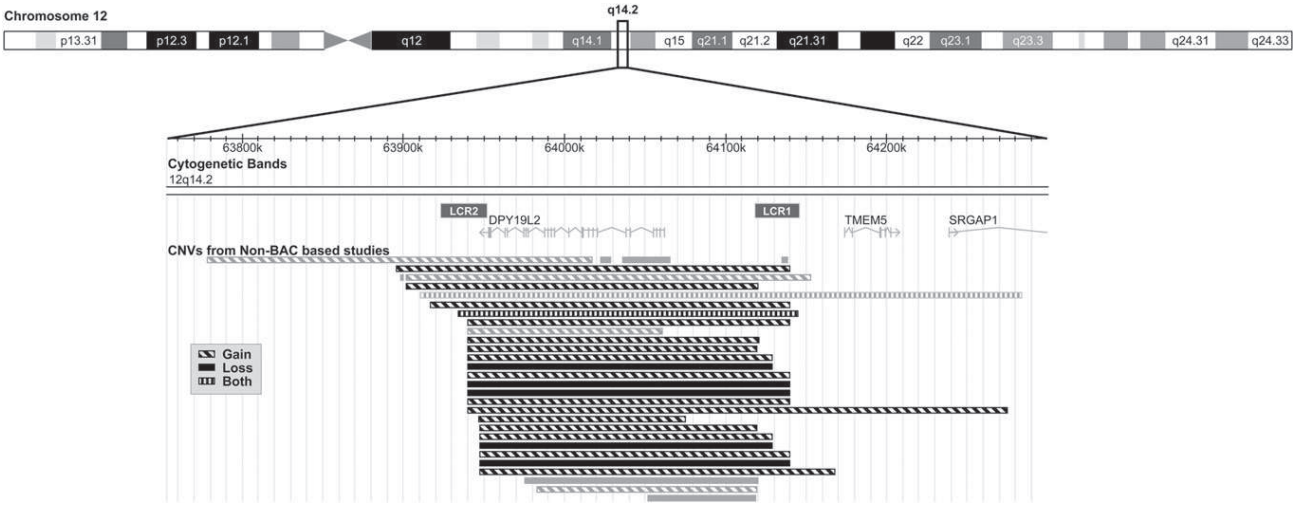


Figure 3. CNVs at the DPY19L2 Locus
CNVs are displayed as rectangles according to the following scheme: gains are drawn with hatched lines, losses with plain rectangles, and vertical lines are used when both gains and losses are observed. The CNVs excluded from the count are colored in gray. LCR1 and LCR2 are displayed as dark gray rectangles on top of the gene representation.

(Figures 4C–4E). In contrast to what is observed in *C. elegans*, the function of DPY19L2 does not seems to be involved in the polarization of the spermatozoa but, rather, in the formation of the acrosome.

DPY19L2 is the third gene identified as being associated with globozoospermia[1,19] and, according to our data, the most frequently mutated in this phenotype. Indeed, in our series of 21 unrelated cases, we identified four individual cases, meaning that the *DPY19L2* deletion is found in 19% of the globozoospermic cases. Looking at the frequency of heterozygous deletion in control populations, we would expect an incidence of 1/200,000 for homozygous individuals, whereas the incidence of globozoospermia is estimated at less than 0.1% in infertile males.[7] Although we cannot fully exclude a positioning effect, we do not believe that this could explain the observed phenotype, because the four genes surrounding the DPY19L2 locus (two centromeric: *AVPR1A* and *PPM1H*; two telomeric: *TMEM5* and *SRGAP1*) are not expressed in the testis and their function has no relation with cell polarity or acrosome formation.

Considering that round-headed spermatozoa do not have a higher incidence of chromosomal abnormalities[20–22] and that pregnancy can be obtained through ICSI, albeit at a low frequency,[2,23–26] it is reasonable to propose ICSI treatment for all globozoospermic patients. It will be interesting to enlarge the cohort of patients and establish which conditions are able to give rise to live births. Indeed, so far, no pregnancies have been obtained with patients carrying a *SPATA16* mutation[1] (data not shown), in contrast to patients with *DPY19L2* deletion. Such an analysis may suggest that a genetic test would assist in determining the best option for treatment.

Our strategy has proven its interest and should be extended to other countries, given that family clustering of male infertility cases has been found in other studies and appears to be related to patterns of consanguineous marriage over generations.[27,28]

Supplemental Data

Supplemental Data include Supplemental Material and Methods, four tables, and one figure and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

We are very grateful to our colleagues Manuel Mark, for his help in interpreting the gametogenesis of our patients, and Michel Koenig and Regen Drouin, for their critical reading of this manuscript. We would also like to thanks all of the clinicians—C. Jimenez (Dijon), F. Carré-Pigeon (Reims), J.M. Grillot (Marseille), F. Brugnon (Clermont Ferrand), D. De Briel (Colmar), and S. Declève-Paulhac

Table 2. CNV Frequency of Individuals at the *DPY19L2* Locus and Expected Prevalence of Homozygotes

	Pinto et al.[11]	de Smith et al.[10]	Conrad et al.[9]	Itsara et al.[13]	Shaikh et al.[12]	Total	Frequency	Expected Homozygous Deletion Frequency
Gain	2	1	4	22	35	64	1/76	
Loss (Htz)	0	3	0	5	14	22	1/222	1/197,136
Population	506	50	450	1854	2026	4886		

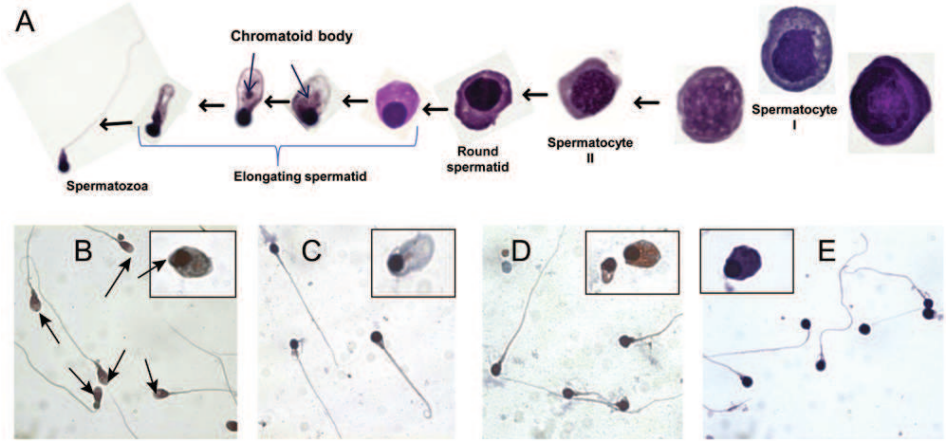


Figure 4. Seminal Germ Cells in Patients with *DPY19L2* Deletion
(A) Detailed seminal germ cell of Globo7. There is no evidence of defective spermatogenesis in immature germ cells; the spermatid polarization and elongation is consistent with controls: e.g., the correct positioning of the chromatid body (blue arrows), except for the absence of the acrosomal vesicle.
(B) Spermatozoa and spermatids of fertile controls. The acrosome is clearly observed in spermatozoa and elongating spermatid (arrows).
(C–E) Spermatozoa and spermatid of Globo17, Globo18, and Globo7, respectively. The spermatozoa present the typical round head and an absence of acrosome. No evident polarization abnormality in seminal spermatids was observed in the globozoospermic patients.

(Limoges)—who have sent us samples from globozoospermic patients and for whom we still expect to find a mutation. We would like to recognize the intellectual input of Jean-Louis Mandel, Christelle Thibault-Carpentier, and Mirna Assoum and the technical help of Anne-Sophie Jaeger and Nathalie Drouot. We would like to thank Frederic Plewniak (plewniak@igbmc.fr) from the Bioinformatics Platform of Strasbourg for the development of the HomoSNP software. We are grateful to the Institute of Genetics and Molecular and Cellular Biology (IGBMC) services and the Department of Human Genetics for their invaluable assistance. This work was supported by the French Centre National de la Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM), the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche, the University of Strasbourg, the University Hospital of Strasbourg, and the Agence de la BioMédecine.

Received: November 22, 2010
Revised: January 6, 2011
Accepted: January 27, 2011
Published online: March 10, 2011

Web Resources

The URLs for data presented herein are as follows:

Database of Genomic Variants, http://projects.tcag.ca/variation/
UCSC Genome Browser Gateway, http://genome.ucsc.edu/cgi-bin/hgGateway
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim

References

1. Dam, A.H., Koscinski, I., Kremer, J.A., Moutou, C., Jaeger, A.S., Oudakker, A.R., Tournaye, H., Charlet, N., Lagier-Tourenne, C., van Bokhoven, H., and Viville, S. (2007). Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. Am. J. Hum. Genet. *81*, 813–820.
2. Kilani, Z., Ismail, R., Ghunaim, S., Mohamed, H., Hughes, D., Brewis, I., and Barratt, C.L. (2004). Evaluation and treatment of familial globozoospermia in five brothers. Fertil. Steril. *82*, 1436–1439.
3. Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., et al. (2010). The UCSC Genome Browser database: update 2010. Nucleic Acids Res. *38* (Database issue), D613–D619.
4. Honigberg, L., and Kenyon, C. (2000). Establishment of left/right asymmetry in neuroblast migration by UNC-40/DCC, UNC-73/Trio and DPY-19 proteins in C. elegans. Development *127*, 4655–4668.
5. Carson, A.R., Cheung, J., and Scherer, S.W. (2006). Duplication and relocation of the functional DPY19L2 gene within low copy repeats. BMC Genomics *7*, 45.
6. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
7. Dam, A.H., Feenstra, I., Westphal, J.R., Ramos, L., van Golde, R.J., and Kremer, J.A. (2007). Globozoospermia revisited. Hum. Reprod. Update *13*, 63–75.
8. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nat. Genet. *36*, 949–951.
9. Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al; Wellcome Trust Case Control Consortium. (2010). Origins and functional impact of copy number variation in the human genome. Nature *464*, 704–712.
10. de Smith, A.J., Tsalenko, A., Sampas, N., Scheffer, A., Yamada, N.A., Tsang, P., Ben-Dor, A., Yakhini, Z., Ellis, R.J., Bruhn, L., et al. (2007). Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. Hum. Mol. Genet. *16*, 2783–2794.

11. Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., et al. (2009). Population analysis of large copy number variants and hotspots of human genetic disease. Am. J. Hum. Genet. *84*, 148–161.
12. Pinto, D., Marshall, C., Feuk, L., and Scherer, S.W. (2007). Copy-number variation in control population cohorts. Hum. Mol. Genet. *16* (Spec. No. 2), R168–R173.
13. Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'Arcy, M., et al. (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res. *19*, 1682–1690.
14. Gu, W., Zhang, F., and Lupski, J.R. (2008). Mechanisms for human genomic rearrangements. Pathogenetics *1*, 4.
15. Tézenas Du Montcel, S., Mendizabai, H., Aymé, S., Lévy, A., and Philip, N. (1996). Prevalence of 22q11 microdeletion. J. Med. Genet. *33*, 719.
16. Hoogendijk, J.E., Hensels, G.W., Gabreëls-Festen, A.A., Gabreëls, F.J., Janssen, E.A., de Jonghe, P., Martin, J.J., van Broeckhoven, C., Valentijn, L.J., Baas, F., et al. (1992). De-novo mutation in hereditary motor and sensory neuropathy type I. Lancet *339*, 1081–1082.
17. Lakich, D., Kazazian, H.H., Jr., Antonarakis, S.E., and Gitschier, J. (1993). Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat. Genet. *5*, 236–241.
18. Xiao, N., Kam, C., Shen, C., Jin, W., Wang, J., Lee, K.M., Jiang, L., and Xia, J. (2009). PICK1 deficiency causes male infertility in mice by disrupting acrosome formation. J. Clin. Invest. *119*, 802–812.
19. Liu, G., Shi, Q.W., and Lu, G.X. (2010). A newly discovered mutation in PICK1 in a human with globozoospermia. Asian J. Androl. *12*, 556–560.
20. Carrell, D.T., Emery, B.R., and Liu, L. (1999). Characterization of aneuploidy rates, protamine levels, ultrastructure, and functional ability of round-headed sperm from two siblings and implications for intracytoplasmic sperm injection. Fertil. Steril. *71*, 511–516.
21. Machev, N., Gosset, P., and Viville, S. (2005). Chromosome abnormalities in sperm from infertile men with normal somatic karyotypes: teratozoospermia. Cytogenet. Genome Res. *111*, 352–357.
22. Viville, S., Mollard, R., Bach, M.L., Falquet, C., Gerlinger, P., and Warter, S. (2000). Do morphological anomalies reflect chromosomal aneuploidies?: case report. Hum. Reprod. *15*, 2563–2566.
23. Banker, M.R., Patel, P.M., Joshi, B.V., Shah, P.B., and Goyal, R. (2009). Successful pregnancies and a live birth after intracytoplasmic sperm injection in globozoospermia. J. Hum. Reprod. Sci. *2*, 81–82.
24. Bechoua, S., Chiron, A., Delcleve-Paulhac, S., Sagot, P., and Jimenez, C. (2009). Fertilisation and pregnancy outcome after ICSI in globozoospermic patients without assisted oocyte activation. Andrologia *41*, 55–58.
25. Coetzee, K., Windt, M.L., Menkveld, R., Kruger, T.F., and Kitshoff, M. (2001). An intracytoplasmic sperm injection pregnancy with a globozoospermic male. J. Assist. Reprod. Genet. *18*, 311–313.
26. Dirican, E.K., Isik, A., Vicdan, K., Sozen, E., and Suludere, Z. (2008). Clinical pregnancies and livebirths achieved by intracytoplasmic injection of round headed acrosomeless spermatozoa with and without oocyte activation in familial globozoospermia: case report. Asian J. Androl. *10*, 332–336.
27. Inhorn, M.C., Kobeissi, L., Nassar, Z., Lakkis, D., and Fakih, M.H. (2009). Consanguinity and family clustering of male factor infertility in Lebanon. Fertil. Steril. *91*, 1104–1109.
28. Inhorn, M.C., Kobeissi, L., Nassar, Z., Lakkis, D., and Fakih, M.H. (2009). Consanguinity and Male Infertility in Lebanon: The Power of Ethnographic Epidemiology. In Anthropology and Public Health: Bridging Differences in Culture and Society, R.A. Hahn and M.C. Inhorn, eds. (New York: Oxford University Press), pp. 165–195.

Appendix 2: Globozoospermia is mainly due to *DPY19L2* deletion via non-allelic homologous recombination involving two recombination hotspots.

Elinati E*, Kuentz P, **Redin C**, Jaber S, Vanden Meerschaut F, Makarian J, Koscinski I, Nasr-Esfahani MH, Demirol A, Gurgan T, Louanjli N, Iqbal N, Bisharah M, Pigeon FC, Gourabi H, De Briel D, Brugnon F, Gitlin SA, Grillo JM, Ghaedi K, Deemeh MR, Tanhaei S, Modarres P, Heindryckx B, Benkhalifa M, Nikiforaki D, Oehninger SC, De Sutter P, Muller J, Viville S. Hum Mol Genet. 2012 Aug.

Aim of the study: Refining the frequency of the implication of *DPY19L2* deletions and point muations in patients with globozoospermia. Characterizing the several breakpoint clusters from *DPY19L2* deletions allowing to precise the underlying molecular mechanism.

Contributions: Minor. Performed bioinformatic analysis, mined and characterized the location of the breakpoints that appeared to coincide with the location of proposed recombination hotspots. Wrote the findings accordingly.

Human Molecular Genetics, 2012, Vol. 21, No. 16 **3695–3702**
doi:10.1093/hmg/dds200
Advance Access published on May 31, 2012

Globozoospermia is mainly due to *DPY19L2* deletion via non-allelic homologous recombination involving two recombination hotspots

Elias Ellnati[1,†], Paul Kuentz[1,2,†], Claire Redin[1], Sara Jaber[1,3], Frauke Vanden Meerschaut[4], Joelle Makarian[1], Isabelle Koscinski[1,5], Mohammad H. Nasr-Esfahani[6], Aygul Demirol[7], Timur Gurgan[7], Noureddine Louanjli[8], Naeem Iqbal[9], Mazen Bisharah[9], Frédérique Carré Pigeon[10], H. Gourabi[11], Dominique De Briel[12], Florence Brugnon[13], Susan A. Gitlin[14], Jean-Marc Grillo[15], Kamran Ghaedi[6], Mohammad R. Deemeh[6], Somayeh Tanhaei[6], Parastoo Modarres[6], Björn Heindryckx[4], Moncef Benkhalifa[16], Dimitra Nikiforaki[4], Sergio C. Oehninger[14], Petra De Sutter[4], Jean Muller[1,17] and Stéphane Viville[1,18,*]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de Santé et de Recherche Médicale (INSERM) U964, Centre National de Recherche Scientifique (CNRS) UMR1704, Université de Strasbourg, Illkirch 67404, France, [2]Centre Hospitalier Universitaire de Besançon, Besançon CEDEX 25030, France, [3]Institut Curie, 26 rue d'UIM, Paris Cedex 75248, France, [4]Department of Reproductive Medicine, University Hospital, De Pintelaan 185, Gent B-9000, Belgium, [5]Service de Biologie de la Reproduction, Centre Hospitalier Universitaire, Strasbourg 67000, France, [6]Department of Reproduction and Development, Reproductive Biomedicine Center, Royan Institute for Animal Biotechnology, ACECR, Isfahan, Iran, [7]Clinic Women Health, Infertility and IVF Center, Ankara, Turkey, [8]Laboratoire LABOMAC, Casablanca 20000, Morocco, [9]King Faisal Specialist Hospital and Research Center, Jeddah 21499, Kingdom of Saudi Arabia, [10]Service de Gynécologie-Obstétrique, CHU Reims, Institut mère-enfant Alix de Champagne, 45 rue Cognacq-Jay, Reims F-51092, France, [11]Royan Institute Reproductive Biomedicine and Stem Cell Research Center, Tehran, Iran, [12]Service de Microbiologie, Centre Hospitalier de Colmar, 39, avenue de la Liberté, Colmar 68024, France, [13]Laboratoire de Biologie de la Reproduction, Université Clermont 1, UFR Médecine, EA 975, Clermont Ferrand Cedex 1 F-63001, France, [14]Department of Obstetrics and Gynecology, The Jones Institute for Reproductive Medicine, Eastern Virginia Medical School, Norfolk, VA, USA, [15]Département d'Urologie, AP-HM, Salvatore Marseille, France, [16]ATL R&D Laboratory, 78320 la verrière et Unilabs, Paris 75116, France, [17]Laboratoire de Diagnostic Génétique, CHU Strasbourg, Nouvel Hôpital Civil, Strasbourg 67000, France and [18]Centre Hospitalier Universitaire, Strasbourg F-67000, France

Received March 28, 2012; Revised and Accepted May 22, 2012

To date, mutations in two genes, *SPATA16* and *DPY19L2*, have been identified as responsible for a severe teratozoospermia, namely globozoospermia. The two initial descriptions of the *DPY19L2* deletion lead to a very different rate of occurrence of this mutation among globospermic patients. In order to better estimate the contribution of *DPY19L2* in globozoospermia, we screened a larger cohort including 64 globozoospermic patients. Twenty of the new patients were homozygous for the *DPY19L2* deletion, and 7 were compound heterozygous for both this deletion and a point mutation. We also identified four additional mutated patients. The final mutation load in our cohort is 66.7% (36 out of 54). Out of 36 mutated patients, 69.4% are homozygous deleted, 19.4% heterozygous composite and 11.1% showed a homozygous point mutation. The mechanism underlying the deletion is a non-allelic homologous recombination (NAHR) between the flanking

*To whom correspondence should be addressed. Tel: +33 388653322; Fax: +33 388653201; Email: viville@igbmc.fr
†These authors contributed equally to this work.

low-copy repeats. Here, we characterized a total of nine breakpoints for the *DPY19L2* NAHR-driven deletion that clustered in two recombination hotspots, both containing direct repeat elements (*Alu*Sq2 in hotspot 1, THE1B in hotspot 2). Globozoospermia can be considered as a new genomic disorder. This study confirms that *DPY19L2* is the major gene responsible for globozoospermia and enlarges the spectrum of possible mutations in the gene. This is a major finding and should contribute to the development of an efficient molecular diagnosis strategy for globozoospermia.

INTRODUCTION

Globozoospermia is a rare and severe teratozoospermia characterized by round-headed spermatozoa lacking an acrosome. The acrosome plays a crucial role during fertilization, allowing the spermatozoa to penetrate the zona pellucida and to reach the oocyte cytoplasmic membrane (1). Therefore, patients suffering from a complete form of globozoospermia are infertile. Round-headed spermatozoa do not present chromosomal abnormalities (2–4) and pregnancy can be obtained through ICSI, although at a low frequency (5–10). Previous analysis of globozoospermia families allowed us to identify mutations in two genes, *SPATA16* and *DPY19L2* (11,12). The deletion of exon 4 in *SPATA16* was found in an Ashkenazi Jewish family with three affected brothers. No other mutations were identified in a screen of 21 patients. A large deletion of ~200 kb encompassing the entire *DPY19L2* locus was detected in a consanguineous Jordanian family and in three additional unrelated patients (12). The gene, located on 12q14.2, has 22 exons encoding for a 9 transmembrane domain protein and is flanked by two low-copy repeats (LCRs) sharing 96.5% identity. The mechanism underlying the deletion is most probably a non-allelic homologous recombination (NAHR) between the flanking LCRs (13). Indeed, sequences with high nucleotide similarity (usually >95%) can serve as substrates for NAHR or ectopic recombination (14–16). Recently, a study of recurrent rearrangements associated with Smith–Magenis syndrome (SMS, MIM 182290) showed that NAHR crossover frequencies are correlated with the flanking LCR length and are inversely influenced by the distance between the LCRs (17). Homologous recombination on the Y chromosome is also known to impair fertility (18,19). However, so far, the deletion of *DPY19L2* is the first example where CNVs on autosomes are reported to be a causative infertility factor (20).

In our initial study, 19% of our patients (4 out of 21) were observed with such a deletion (12). In contrast, Harbuz *et al.* (21) detected a higher rate of *DPY19L2*-deleted patients (75%) in a cohort mainly composed of Tunisian patients. The difference is probably due to the broader geographic distribution of our patients (i.e. Algeria, France, Iran, Italy, Libya, Morocco and Tunisia).

In order to better estimate the contribution of *DPY19L2* in globozoospermia, we screened the largest cohort of globozoospermic patients to date, including 64 patients (from 13 different countries) and corresponding to 54 genetically independent individuals for all types of mutations. Twenty of the new patients were homozygous for the *DPY19L2* deletion, and seven were compound heterozygous for this deletion and a point mutation. We also identified four additional mutated patients. The final mutation load in our cohort is 66.7% (36 out of 54). Out of 36 mutated patients, 69.4% are homozygous deleted, 19.4% heterozygous composite and 11.1% showed a homozygous point mutation.

Recombination hotspots have been associated with genomic disorders and, when analyzing the sequence in the breakpoint (BP) cluster region, a common degenerate 13mer motif (CCNCCNTNNCCNC) was found (22), which has recently been elucidated as the *PRDM9* recognition binding motif (23–25). Here, we characterized nine new BPs for the *DPY19L2* NAHR-driven deletion defining two recombination hotspots, one containing the *PRDM9* recognition motif.

In this study, we confirmed that *DPY19L2* is the major gene responsible for globozoospermia, we enlarged the spectrum of possible mutations in the gene (deletion of the whole locus, nonsense, missense, splicing mutations and partial deletion) and we present globozoospermia as a new genomic disorder. This is a major finding and should contribute to the development of an efficient molecular diagnosis strategy for globozoospermia.

RESULTS

Deletion screening

In addition to our initial cohort of 21 patients, we recruited 33 new globozoospermic patients. As a first step, we screened for the deletion of *DPY19L2* in all 33 patients by performing a PCR of exon 10 and a PCR encompassing the previously described deletion BPs. Of these 33 new patients, 13 (Globo 29–31, 36, 39, 45, 48, 51, 54, 56, 57, 59, 61) did not show any amplification of exon 10, whereas the PCR of *DPY19L2* BPs revealed a fragment of 1700 bp, suggesting a homozygous deletion and a BP located within the 1700 bp fragment (Fig. 1A). Eight patients (Globo 28, 32, 33, 47, 50, 53, 58, 60) did not show any amplification at all, suggesting a homozygous deletion, but with a BP situated outside the tested region. We confirmed the deletion of most of the gene by testing for the presence of exons 4 and 16, which could not be amplified. Using specific oligonucleotides to walk on both sides of the deletion, we were able to amplify a common region of 1600 bp for two patients (Globo 28, 32). Sequencing of the 1600 bp fragment identified two BPs within an area of 117 bp (BP8 and BP9) (Fig. 2A). This second region of recombination is localized 9450 bp 3′ of the first one (Fig. 2B). For a third patient (Globo 33), we were able to amplify a fragment of 1500 bp. Sequencing of this fragment revealed another BP within a region of 265 bp (BP7) (Fig. 2A). This third region of recombination is situated 693 bp 3′ of the first one (Fig. 2B). For four patients (Globo

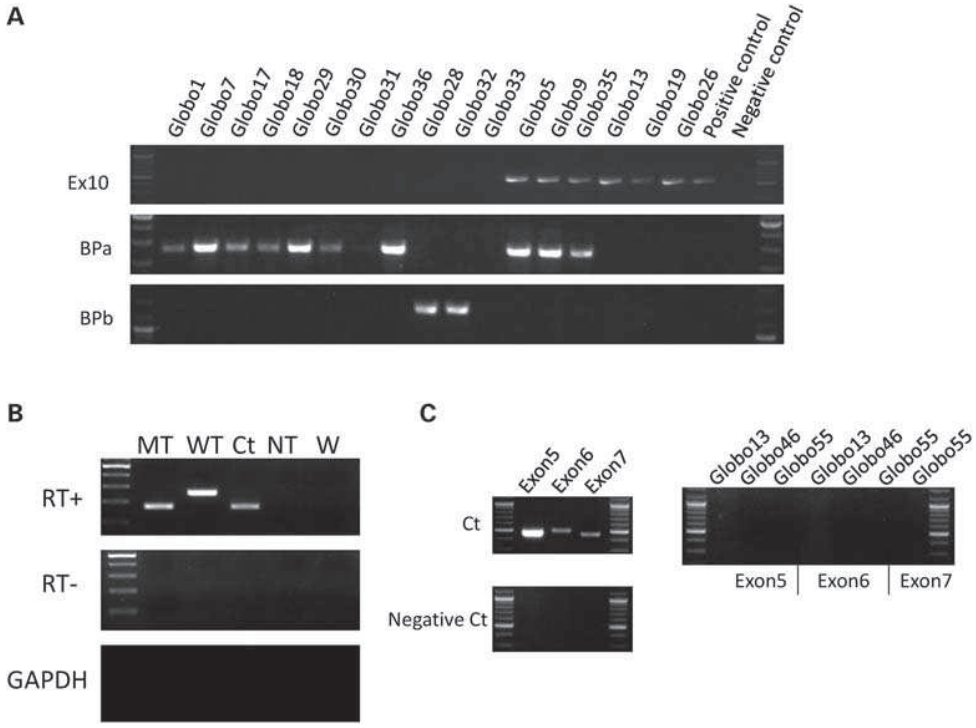


Figure 1. (**A**) PCR results of *DPY19L2* exon 10 and the BPs. Patients Globo1, Globo7, Globo17, Globo18, Globo29, Globo30, Globo31, Globo36, Globo28, Globo32, Globo33 are deleted for exon 10, whereas Gobo5, Gobo9, Gobo35, Gobo13, Gobo19, Gobo26 and the positive control showed an amplification of all of this exon. Globo1, Globo7, Globo17, Globo18, Globo29, Globo30, Globo31, Globo36, Globo28, Gobo5, Gobo9, Gobo35 showed an amplification for BP a, whereas Globo29 and Globo33 showed an amplification for BP b. For Globo39, 40, 42, 43, 45–48, 50, 51, 53, 54, 56–61, data are not shown. (**B**) Minigene constructs used to test the splicing of exon 11. Mt, mutant form of exon 11; WT, wild-type form of exon 11; Ct, construct without any cloned exon; NT, non-transfected cells; W, water. (**C**) PCR results for *DPY19L2* exons 5, 6 and 7. Globo13 and Globo46 are deleted for exons 5 and 6, whereas Globo55 is deleted for exons 5, 6 and 7. The positive control showed amplification of all these exons.

50, 53, 58, 60), we were not able to identify the BP. Five patients (Globo 35, 40, 42, 43, 46) showed an amplification of both exon 10 and the 1700 bp BP fragment, suggesting that these patients are heterozygous for the *DPY19L2* deletion (Fig. 1A). For the seven (Globo 34, 37, 38, 44, 49, 62, 63) remaining patients, exon 10 was detected, but no deletion could be found. Since we found one patient heterozygous for the deletion, we checked all remaining patients from our previous cohort. We identified two patients (Globo 5, 9) who were heterozygous for the *DPY19L2* deletion (Fig. 3A).

Point mutation screening

In a second step, in conjunction with screening for the *DPY19L2* deletion, we sequenced all coding exons and intron boundaries in the 7 compound heterozygous patients for the deletion and the 23 non-deleted patients (the exon localization of the mutations are shown in Fig. 2A). Concerning the heterozygous patients, sequence analysis of Globo5 revealed a variation in exon 8 (c.869G>A), leading to an amino acid change of a highly conserved residue (p.R290H) that was predicted to be deleterious by two programs: PolyPhen and SIFT (Supplementary Material, Table S1). Globo9 exhibited a variation in exon 9 (c.1033C>T), introducing a premature stop codon (p.Q345X). Globo35 presented a variation in exon 15 (c.1478C>G), leading to a non-synonymous mutation (p.T493R). This mutation is predicted to be deleterious by PolyPhen and tolerated by SIFT. Globo40 showed a variation in exon 21 (c.2038A>T), introducing a premature stop codon (p.K680X). Globo42 and 43, two unrelated patients, showed the same nucleotide deletion in exon 11 (c.1183delT), introducing a premature stop codon (p.S395LfsX7). Globo46 presented a deletion of exons 5 and 6. An aberrant splicing between exons 4 and 7 would give rise to a frame shift, introducing a premature stop codon (Supplementary Material, Table S1).

Among non-deleted patients, four, descending from consanguineous families, were found homozygous for a mutation. Globo26 showed a homozygous variation in exon 8 (c.892C>T), leading to a non-synonymous mutation (p.R298C), located at a strictly conserved amino acid position. This mutation is predicted to be deleterious by both PolyPhen and SIFT. Globo19 presented a homozygous donor splice-site mutation in intron 11 (c.1218+1G>A). Splice-site models (see Materials and Methods) predicted that the mutation disrupts the 5′ splice site of intron 11.

Unfortunately, the DPY19L2 protein presents a testis-restricted expression, and we were not allowed to use fresh sperm cells or to perform a testicular biopsy in these patients, in order to verify the predicted aberrant splicing *in vivo*. In order to test the prediction, minigene constructs were made, including either the wild-type (WT) or mutated form of exon 11

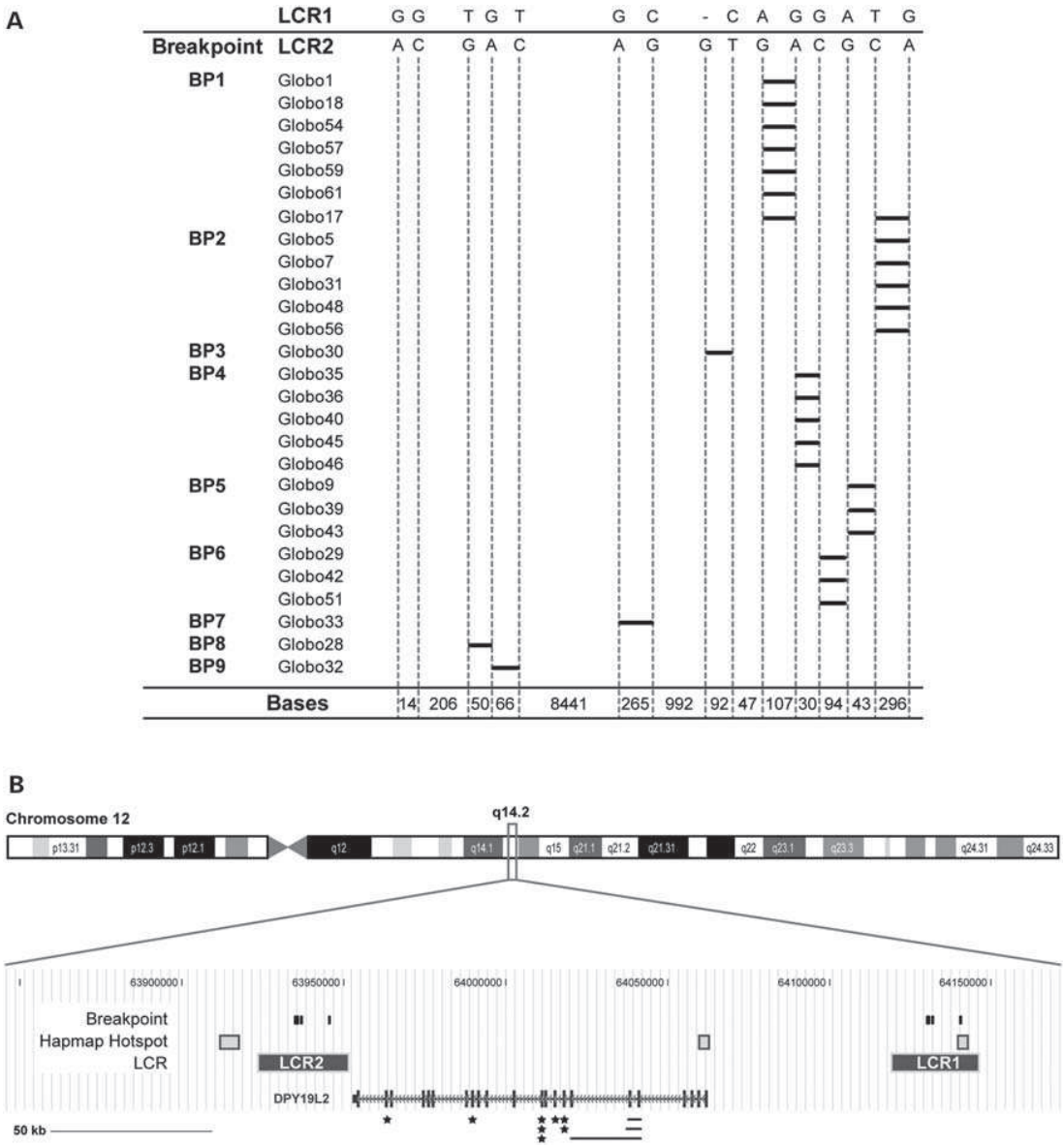


Figure 2. (**A**) For each LCR, the specific nucleotide differences within the hotspot regions are shown at the top. The length and distance between the different nucleotides are displayed at the bottom. The recombination region is represented for each patient by a black bar. The patients are ordered according to the nine BPs. (**B**) The *DPY19L2* locus is represented together with the two LCR, the HapMap recombination hotspots, the nine different BPs and the identified mutations (missenses are marked with a star and larger deletions are represented using a black horizontal line).

and flanking intronic sequences. These minigene constructs were transfected into COS1 and HeLa cells, and transcripts were analyzed by reverse transcription-PCR (RT-PCR) 48 h after the transfection. As shown in Figure 1B, WT exon 11 is invariably included in the final mRNA, as confirmed by the sequencing of the PCR product. In contrast, the mutated exon gives rise to one aberrant splicing form, leading to exon 11 skipping. An aberrant splicing of exon 11 could eventually give rise to a new splicing between exons 10 and 12 and would produce a protein with a deletion of 28 amino acids from position 378 to 406, corresponding to the seventh transmembrane helix (position 372–394).

We identified a homozygous deletion of exons 5 and 6 in Globo13. In order to pinpoint the two BP zones, subsequent amplifications on both sides of the exons were performed. Amplification across the deletion gave rise to a PCR fragment of 3500 bp for Globo13, whereas no amplification was possible from a control subject known to be fertile (Fig. 1C). Analysis of the sequence allowed the determination of the BPs and showed an insertion of 73 bp that corresponds to a part of a LINE sequence (Supplementary Material, Fig. S1). The deletion encompasses a 15.7 kb region with one BP located 8.35 kb from the 5′ side of exon 5 and the second one 4.36 kb from the 3′ side of exon 6. An aberrant splicing

A

Breakpoint zone	Point mutation	Patient	Origin
BP1	-	Globo1	Jordan
BP1	-	Globo18	France
BP1	-	Globo54	Iran
BP1	-	Globo57	Iran
BP1	-	Globo59	Iran
BP1	-	Globo61	Iran
BP2	p.R290H	Globo5	France
BP2	-	Globo7	Algeria
BP2	-	Globo31	Turkey
BP2	-	Globo48	Iran
BP2	-	Globo56	Iran
BP1/BP2	-	Globo17	France
BP3	-	Globo30	Saoudi Arabia
BP4	p.T493R	Globo35	Turkey
BP4	-	Globo36	Morocco
BP4	p.K680X	Globo40	Belgium
BP4	-	Globo45	Turkey
BP4	Ex5_6del	Globo46	USA
BP5	p.Q345X	Globo9	France
BP5	-	Globo39	Belgium
BP5	p.S395LfsX7	Globo43	Belgium
BP6	-	Globo29	Morocco
BP6	p.S395LfsX7	Globo42	Belgium
BP6	-	Globo51	Iran
BP7	-	Globo33	Turkey
BP8	-	Globo28	Morocco
BP9	-	Globo32	Turkey
-	Ex5_6del	Globo13	Iran
-	c.1218+1G>A	Globo19	Italy
-	p.R298C	Globo26	Morocco
-	Ex5_7del	Globo55	Iran
UBP	-	Globo47	Iran
UBP	-	Globo50	Iran
UBP	-	Globo53	Iran
UBP	-	Globo58	Iran
UBP	-	Globo60	Iran

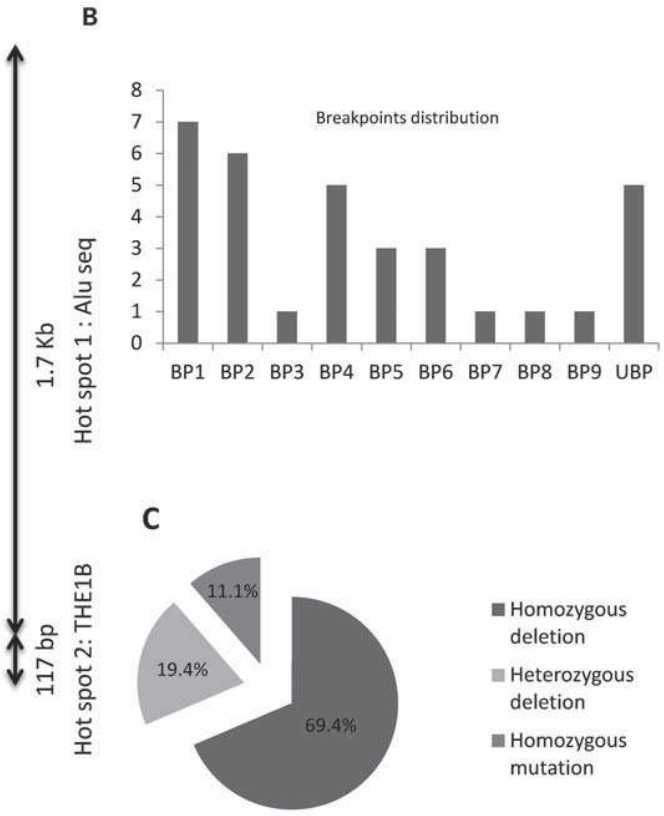


Figure 3. Distribution of BPs and mutations among patients. (**A**) The table presenting the type of BPs, mutations and patient origins. Vertical arrows delimit the two hotspots and their size. Hotspot 1 and hotspot 2 contain direct repeat elements, *Alu*Sq2 and THE1B, respectively. (**B**) The chart showing the distribution of the BPs among the patients totally deleted for *DPY19L2* (UBP, undetermined BP). (**C**) Pie chart showing the percentage of mutation types among the patients deleted and/or mutated for *DPY19L2*. Among these patients: 69.4% have a complete deletion of *DPY19L2* in the homozygous state; 19.4% have a complete deletion associated with a point mutation or a partial deletion; 11.1% have a point mutation or a partial deletion in the homozygous state.

between exons 4 and 7, if not degraded, would give rise to a frame shift, producing a truncated protein of 226 amino acids due to a premature stop codon at position 227. Since no repetitive sequence could be detected nearby, the deletion could be explained by a non-homologous end joining, which is often associated with the insertion of a DNA fragment at the BP (14). Globo46, who is compound heterozygous, also presented a deletion of exons 5 and 6, but with different BPs. Finally, Globo55 presented a deletion of exons 5, 6 and 7. An aberrant splicing between exons 4 and 8 would give rise to a protein with a deletion of 91 amino acids from position 196 to 287, corresponding to transmembrane domains 3, 4 and 5, at positions 195–215, 243–265 and 269–286, respectively (Supplementary Material, Table S1).

None of the variations described here were found, either in dbSNP v134 or when testing at least 188 control chromosomes.

BP analysis

We mapped linkage disequilibrium (LD) patterns and recombination rates in HapMap2 on chromosome 12. We observed that both LCR positions correlate with strongly suggested recombination spots, thus confirming the NAHR, which is the mechanism hypothesized to be responsible for the disease (26). Many other disorders involving NAHR between LCRs have been identified, such as DiGeorge syndrome/velocardiofacial syndrome, Williams–Beuren syndrome, Prader–Willi syndrome or Charcot–Marie–Tooth disease type 1A, which belong to the group of genomic disorders (27).

Previously, we identified two BPs (BP 1 and 2) localized close together in a region of 296 bp containing an *Alu* repetitive element (12). Sequence analysis of the amplification product across the deletions identified seven new BPs (BP3–9). Interestingly, seven of the BPs (BP1–7) are located in a small region of 1.7 kb, whereas the other two (BP8–9) are located in an area of 117 bp, 9.5 kb away from the first one (Fig. 2B). The underlying sequences of the BPs, as well as their genomic positions, are given in Supplementary Material, Figure S1. We can therefore suggest the presence of two recombination hotspots (hotspot 1 and 2, see Fig. 3A), comprising BP1–7 and BP8–9, respectively, both containing direct repeat elements (*Alu*Sq2 in hotspot 1, THE1B in hotspot 2). Interestingly, it was shown that AHR/NAHR hotspots

usually cluster within small regions of 1–2 kb of almost perfect identity, with repeat elements, transposons or minisatellites located nearby that act as substrates for double-strand breaks (22). Analysis of these hotspots in NAHR disorders identified a 13mer CCNCCNTNNCCNC motif frequently present in the flanking LCRs (22). We analyzed the sequences surrounding both identified recombination hotspots involved in the *DPY19L2* deletion and indeed found the PRDM9 13mer recognition motif (which is part of the THE1B repeat element), but only in hotspot 2 (Fig. 3A). Interestingly, hotspot 2 also coincides perfectly with a previously identified recombination hotspot (28) of chromosome 12 (Fig. 2B).

In total, we identified 32 patients with at least one deleted allele, among which BP1 occurs seven times, BP2 six times, BP4 five times, BP5 three times, BP6 three times and BP3, 7, 8, 9 were found only once (Fig. 3B). All recurrent BPs were found among patients originating from different regions, and patients from the same country showed different BPs (Fig. 3A). Out of 36 mutated patients, 69.4% were homozygous deleted, 19.4% heterozygous composite and 11.1% showed a homozygous point mutation (Fig. 3C).

DISCUSSION

The *DPY19L2* deletion was identified as being the major cause of globozoospermia in two different studies (12,21). It was shown by both groups that the most probable mechanism explaining this deletion is NAHR mediated by two LCRs surrounding *DPY19L2*. Thus, globozoospermia can be considered a new genomic disorder (27,29). Although both reports described a deletion rate of high frequency, a large difference was observed. Indeed, 19% (4 out of 21) of globozoospermic patients were found deleted in our study, whereas the other reported a frequency of 75% (15 out of 20). This difference could be due to a bias either in patient recruitment or in the limited number of patients analyzed. The higher frequency observed in Harbuz *et al*. (21) could be explained by the fact that most of the patients were Tunisians, whereas in our study, the patients were from seven different countries. Unfortunately, Harbuz *et al*. (21) were unable to localize the BPs, making it impossible to exclude a local founder effect in this cohort of globozoospermic patients. We present here the analysis of a larger cohort of globozoospermic patients, which allowed us, first, to refine the frequency rate of the deletion in our cohort, and, second, to enlarge the mutation spectrum. Indeed, in the context of this study and our previous work, we identified 24 homozygous deleted patients, 7 patients heterozygous for the deletion and presenting a point mutation in the remaining allele and 4 homozygous patients presenting a non-sense, a splicing mutation or deletions of exons 5 and 6 or exons 5 to 7 (Fig. 3A). In total, we identified nine BP zones, of which seven are clustered in a region of 1.7 kb and two in an area of 117 bp situated ~9 kb away from the first spot of recombination. Both regions seem to define two ectopic recombination hotspots for *DPY19L2* deletion (Supplementary Material, Fig. S1), each being <2 kb (in LCRs of ~27 kb) and containing a repeat element. Hotspot 1 accounts for almost all NAHR-driven *DPY19L2* deletions. In spite of its lower frequency, hotspot 2 contains the PRDM9 recognition binding motif, further supporting its involvement in the recombination mechanism (23,30). In addition, the fact that hotspot 2 coincides perfectly with a previously identified recombination hotspot supports the observation that NAHR and AHR share common features, including association with identical hotspot motifs.

We did not find any predominating BP. The fact that the same BPs are shared by patients from completely different regions and that patients from the same country can show different BPs tends to exclude any founder effect, even a recent one, and strongly suggests that the deletion results from recurrent events linked to the specific genomic architectural feature of this locus.

Our study allows us to calculate a more accurate prevalence of *DPY19L2* involvement in globozoospermia as a new genomic disorder. We analyzed a cohort of 54 globozoospermic patients and found that 36 of them were mutated for *DPY19L2* (66.7%). Despite the identification of subtle mutations, the most frequent alteration remains the deletion of the whole gene. Indeed, more than two-thirds (69%) of our patients are homozygously deleted, which also suggests that in one out of three of the cases, a search for point mutations is justified. We still have 18 patients with no identified mutation in *DPY19L2*, including 3 pairs of brothers, suggesting that new genes remain to be identified.

Since they do not present any other symptoms, globozoospermic patients are always recruited via *in vitro* fertilization centers. Considering the low frequencies of fertilization and birth obtained so far, one can wonder whether it is justified to propose that these couples go through the burden of such a heavy technology. A recent report, describing one globozoospermic patient, suggested that globozoospermia could be associated with the absence of phospholipase C zeta (PLCζ). This might explain the absence of fertilization, since PLCζ is the main physiological actor responsible for oocyte activation (31). An artificial activation, via a treatment with calcium ionophore, could improve fertilization rates and help to obtain embryos, thus increasing the chance of pregnancy (32). Such calcium ionophore activation treatment has been previously used successfully, considerably improving the fertilization and pregnancy rates in patients deficient for oocyte activation, including globozoospermic patients (33,34). It would be of great interest to see, in a larger cohort of globozoospermic patients, whether calcium ionophore activation would allow a better pregnancy rate and whether there is any correlation between the presence of a mutation in either *DPY19L2* or *SPATA16* and the pregnancy outcome. We are currently collecting DNA from globozoospermic patients for whom ICSI attempts have been done with or without oocyte activation.

MATERIALS AND METHODS

Patient recruitment and DNA preparation

Patients were selected by *in vitro* fertilization centers through a semen analysis implemented according to World Health Organization recommendations. Globozoospermia was diagnosed using a spermocytogram after a Harris–Shorr coloration.

Genomic DNA was extracted either from peripheral blood leukocytes using QIAamp DNA Blood Midi Kit (QIAGEN, Germany) or from saliva using Oragene DNA Self-Collection Kit (DNAgenotech, Ottawa, Canada), according to the manufacturer's instructions. This study was approved by the local Ethical Committee (Comité de protection de la personne, CPP) of Strasbourg University Hospital. For each case analyzed, informed written consent was obtained according to CPP recommendations.

PCR analysis

PCR covering all the exons of DPY19L2 and their exon–intron boundaries or the previously identified deletion (previously identified BPs BP1 and BP2) was performed using genomic DNA, and amplicons were sequenced by GATC (Konstanz, Germany). For the newly identified deletions, 5′ and 3′ walks were then carried out to identify the BPs. All primer sequences and PCR conditions are available in Supplementary Material, Table S2. Because of the high conservation level of duplicated regions, special care was taken to choose specific oligonucleotides with a unique sequence specifying a single location.

Plasmids and transfection

To study the effect of the c.1218+1G>A splice mutation, we constructed WT and mutated hybrid minigenes, using the pDUP33 vector (35). The genomic DNA region from Globo26 or from a control containing exon 11 (87 bp) and intronic flanking sequences (654 bp upstream from the 5′ exon end and 329 bp downstream from the 3′ exon end) were PCR-amplified using a forward primer (DPY19L2e11BamHI; 5′-GGGCCCtacatatggtattggtgatatcca-3′) and a reverse primer (DPY19L2e11BglII; 5′-AGATCTcactgcaatgagtacttaacc-3′). The 1063 bp PCR products were purified from agarose gel, using the Gel Band Purification Kit (Amersham Biosciences), and were sequentially digested by *Bam*HI and *Bgl*II. The insert was directionally cloned into the first intron of β-globin in the Dup33 plasmid. Recombinant plasmids were sequenced to confirm the presence of the mutated or the WT sequence. Transfection of plasmid DNA (WT, mutant and control pDUP33) in Cos and HeLa cells and RT-PCR were performed as previously described (11).

Bioinformatic analysis

Single-nucleotide variations (SNVs) were analyzed using the Alamut software (Interactive BioSoftware), which systematically gives access to several prediction algorithms. In particular, missense variations were tested with SIFT (36) and PolyPhen v2 (37). Missense and intronic variations were analyzed for splicing effects, using MaxEntScan (38), NNSPLICE (39) and HSF (40).

Allelic crossover hotspots [build 37 (36)], recombination rates from HapMap Phase II [build 36 (28)] and LD patterns from the 1000 Genomes project [build 37 (41)] were mapped onto the *DPY19L2* locus using the UCSC browser (42).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are very grateful to James R Lupski, Claudia Carvalho and our colleagues Jean-Louis Mandel and Julie Thompson for their critical reading of this manuscript. We are also grateful for the services of the Institute of Genetics and Molecular and Cellular Biology (IGBMC).

Conflict of Interest statement. None declared.

AUTHORS' ROLES

Contributors E.E., P.K. and S.V. designed the study; E.E., P.K., S.J., J.M. and I.K. performed genetic analysis and mutation screening; F.V.M., M.H.N.-E., A.D., T.G., N.L., N.I., M.B., F.C.P., H.G., D.D., F.B., S.A.G., J.-M.G., S.C.O., P.D. recruited patients and collected clinical data; C.R. and J.M. contributed to the bioinformatics analysis; E.E., P.K., C.R., J.M. and S.V. contributed to analysis and interpretation of data; E.E., C.R., J.M. and S.V. wrote the manuscript.

FUNDING

This work was supported by the French Centre National de la Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM), the Ministère de l'Education Nationale, the Enseignement Supérieur et de la Recherche, the University of Strasbourg, the University Hospital of Strasbourg and the Agence de la BioMédecine.

REFERENCES

1. Ikawa, M., Inoue, N., Benham, A.M. and Okabe, M. (2010) Fertilization: a sperm's journey to and interaction with the oocyte. *J. Clin. Invest.*, **120**, 984–994.
2. Carrell, D.T., Emery, B.R. and Liu, L. (1999) Characterization of aneuploidy rates, protamine levels, ultrastructure, and functional ability of round-headed sperm from two siblings and implications for intracytoplasmic sperm injection. *Fertil. Steril.*, **71**, 511–516.
3. Machev, N., Gosset, P. and Viville, S. (2005) Chromosome abnormalities in sperm from infertile men with normal somatic karyotypes: teratozoospermia. *Cytogenet. Genome Res.*, **111**, 352–357.
4. Viville, S., Mollard, R., Bach, M.L., Falquet, C., Gerlinger, P. and Warter, S. (2000) Do morphological anomalies reflect chromosomal aneuploidies? Case report. *Hum. Reprod.*, **15**, 2563–2566.
5. Banker, M.R., Patel, P.M., Joshi, B.V., Shah, P.B. and Goyal, R. (2009) Successful pregnancies and a live birth after intracytoplasmic sperm injection in globozoospermia. *J. Hum. Reprod. Sci.*, **2**, 81–82.
6. Bechoua, S., Chiron, A., Delcleve-Paulhac, S., Sagot, P. and Jimenez, C. (2009) Fertilisation and pregnancy outcome after ICSI in globozoospermic patients without assisted oocyte activation. *Andrologia*, **41**, 55–58.
7. Coetzee, K., Windt, M.L., Menkveld, R., Kruger, T.F. and Kitshoff, M. (2001) An intracytoplasmic sperm injection pregnancy with a globozoospermic male. *J. Assist. Reprod. Genet.*, **18**, 311–313.
8. Dirican, E.K., Isik, A., Vicdan, K., Sozen, E. and Suludere, Z. (2008) Clinical pregnancies and livebirths achieved by intracytoplasmic injection of round headed acrosomeless spermatozoa with and without oocyte activation in familial globozoospermia: case report. *Asian J. Androl.*, **10**, 332–336.

9. Kilani, Z., Ismail, R., Ghunaim, S., Mohamed, H., Hughes, D., Brewis, I. and Barratt, C.L. (2004) Evaluation and treatment of familial globozoospermia in five brothers. *Fertil. Steril.*, **82**, 1436–1439.
10. Sermondade, N., Hafhouf, E., Dupont, C., Bechoua, S., Palacios, C., Eustache, F., Poncelet, C., Benzacken, B., Levy, R. and Sifer, C. (2011) Successful childbirth after intracytoplasmic morphologically selected sperm injection without assisted oocyte activation in a patient with globozoospermia. *Hum. Reprod.*, **26**, 2944–2949.
11. Dam, A.H., Koscinski, I., Kremer, J.A., Moutou, C., Jaeger, A.S., Oudakker, A.R., Tournaye, H., Charlet, N., Lagier-Tourenne, C., van Bokhoven, H. *et al.* (2007) Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. *Am. J. Hum. Genet.*, **81**, 813–820.
12. Koscinski, I., Elinati, E., Fossard, C., Redin, C., Muller, J., Velez de la Calle, J., Schmitt, F., Ben Khelifa, M., Ray, P.F., Kilani, Z. *et al.* (2011) DPY19L2 deletion as a major cause of globozoospermia. *Am. J. Hum. Genet.*, **88**, 344–350.
13. Stankiewicz, P. and Lupski, J.R. (2002) Genome architecture, rearrangements and genomic disorders. *Trends Genet.*, **18**, 74–82.
14. Gu, W., Zhang, F. and Lupski, J.R. (2008) Mechanisms for human genomic rearrangements. *Pathogenetics*, **1**, 4.
15. Inoue, K. and Lupski, J.R. (2002) Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.*, **3**, 199–242.
16. Lupski, J.R. (2004) Hotspots of homologous recombination in the human genome: not all homologous sequences are equal. *Genome Biol.*, **5**, 242.
17. Liu, P., Lacaria, M., Zhang, F., Withers, M., Hastings, P.J. and Lupski, J.R. (2011) Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. *Am. J. Hum. Genet.*, **89**, 580–588.
18. Blanco, P., Shlumukova, M., Sargent, C.A., Jobling, M.A., Affara, N. and Hurles, M.E. (2000) Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J. Med. Genet.*, **37**, 752–758.
19. McLachlan, R.I. and O'Bryan, M.K. (2010) Clinical review: state of the art for genetic testing of infertile men. *J. Clin. Endocrinol. Metab.*, **95**, 1013–1024.
20. Carvalho, C.M., Zhang, F. and Lupski, J.R. (2011) Structural variation of the human genome: mechanisms, assays, and role in male infertility. *Syst. Biol. Reprod. Med.*, **57**, 3–16.
21. Harbuz, R., Zouari, R., Pierre, V., Ben Khelifa, M., Kharouf, M., Coutton, C., Merdassi, G., Abada, F., Escoffier, J., Nikas, Y. *et al.* (2011) A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. *Am. J. Hum. Genet.*, **88**, 351–361.
22. Myers, S., Freeman, C., Auton, A., Donnelly, P. and McVean, G. (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.*, **40**, 1124–1129.
23. Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G. and de Massy, B. (2011) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, **327**, 836–840.
24. McVean, G. and Myers, S. (2010) PRDM9 marks the spot. *Nat. Genet.*, **42**, 821–822.
25. Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G. and Donnelly, P. (2010) Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, **327**, 876–879.
26. Lindsay, S.J., Khajavi, M., Lupski, J.R. and Hurles, M.E. (2006) A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am. J. Hum. Genet.*, **79**, 890–902.
27. Lupski, J.R. (2009) Genomic disorders ten years on. *Genome Med.*, **1**, 42.
28. International HapMap Consortium (2003) The International HapMap Project. *Nature*, **426**, 789–796.
29. Lupski, J.R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.*, **14**, 417–422.
30. Berg, I.L., Neumann, R., Lam, K.W., Sarbajna, S., Odenthal-Hesse, L., May, C.A. and Jeffreys, A.J. (2011) PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat. Genet.*, **42**, 859–863.
31. Kashir, J., Heindryckx, B., Jones, C., De Sutter, P., Parrington, J. and Coward, K. (2010) Oocyte activation, phospholipase C zeta and human infertility. *Hum. Reprod. Update*, **16**, 690–703.
32. Taylor, S.L., Yoon, S.Y., Morshedi, M.S., Lacey, D.R., Jellerette, T., Fissore, R.A. and Oehninger, S. (2010) Complete globozoospermia associated with PLCzeta deficiency treated with calcium ionophore and ICSI results in pregnancy. *Reprod. Biomed. Online*, **20**, 559–564.
33. Heindryckx, B., De Gheselle, S., Gerris, J., Dhont, M. and De Sutter, P. (2008) Efficiency of assisted oocyte activation as a solution for failed intracytoplasmic sperm injection. *Reprod. Biomed. Online*, **17**, 662–668.
34. Heindryckx, B., Van der Elst, J., De Sutter, P. and Dhont, M. (2005) Treatment option for sperm- or oocyte-related fertilization failure: assisted oocyte activation following diagnostic heterologous ICSI. *Hum. Reprod.*, **20**, 2237–2241.
35. Dominski, Z. and Kole, R. (1992) Cooperation of pre-mRNA sequence elements in splice site selection. *Mol. Cell. Biol.*, **12**, 2108–2114.
36. Kumar, P., Henikoff, S. and Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.
37. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
38. Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
39. Reese, M.G., Eeckman, F.H., Kulp, D. and Haussler, D. (1997) Improved splice site detection in Genie. *J. Comput. Biol.*, **4**, 311–323.
40. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M. and Beroud, C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.*, **37**, e67.
41. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
42. Karolchik, D., Hinrichs, A.S. and Kent, W.J. (2011) The UCSC Genome Browser. *Curr. Protoc. Hum. Genet.*, **18**, 18.6.1–18.6.33.

Appendix 3: A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, Drosophila, and humans.

Bricker DK*, Taylor EB, Schell JC, Orsak T, Boutron A, Chen YC, Cox JE, Cardon CM, Van Vranken JG, Dephoure N, **Redin C**, Boudina S, Gygi SP, Brivet M, Thummel CS, Rutter J. Science. 2012, Jul.

Aim of the study: Reporting on the identification of the pyruvate carrier in yeast, drosophila, and humans.

Contributions: Minor. Performed the linkage analyses on three families with apparent deficient pyruvate transporter, and provided with one candidate linkage interval containing a dozen of genes among which the final human gene for the mitonchondrial pyruvate carrier. Wrote the findings accordingly.

Sciencexpress Reports

A Mitochondrial Pyruvate Carrier Required for Pyruvate Uptake in Yeast, *Drosophila*, and Humans

Daniel K. Bricker,[1]* Eric B. Taylor,[2]* John C. Schell,[2]* Thomas Orsak,[2]* Audrey Boutron,[3] Yu-Chan Chen,[2] James E. Cox,[4] Caleb M. Cardon,[2] Jonathan G. Van Vranken,[2] Noah Dephoure,[5] Claire Redin,[6] Sihem Boudina,[7] Steven P. Gygi,[5] Michèle Brivet,[3] Carl S. Thummel,[1] Jared Rutter[2]†

[1]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112, USA.

[2]Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84112, USA.

[3]Laboratoire de Biochimie, AP-HP Hôpital de Bicêtre, Le Kremlin Bicêtre, France.

[4]Metabolomics Core Research Facility, University of Utah School of Medicine, Salt Lake City, UT 84112, USA.

[5]Department of Cell Biology, Harvard Medical School, Boston, MA, USA.

[6]Institut de Genetique et de Biologie Moleculaire et Cellulaire (IGBMC), Strasbourg, France.

[7]Department of Medicine, University of Utah School of Medicine, Salt Lake City, UT 84112, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: rutter@biochem.utah.edu

Pyruvate constitutes a critical branch point in cellular carbon metabolism. We have identified two proteins, Mpc1 and Mpc2, as essential for mitochondrial pyruvate transport in yeast, *Drosophila*, and humans. Mpc1 and Mpc2 associate to form an ~150 kilodalton complex in the inner mitochondrial membrane. Yeast and *Drosophila* mutants lacking *MPC1* display impaired pyruvate metabolism, with an accumulation of upstream metabolites and a depletion of tricarboxylic acid cycle intermediates. Loss of yeast Mpc1 results in defective mitochondrial pyruvate uptake, while silencing of *MPC1* or *MPC2* in mammalian cells impairs pyruvate oxidation. A point mutation in *MPC1* provides resistance to a known inhibitor of the mitochondrial pyruvate carrier. Human genetic studies of three families with children suffering from lactic acidosis and hyperpyruvatemia revealed a causal locus that mapped to *MPC1*, changing single amino acids that are conserved throughout eukaryotes. These data demonstrate that Mpc1 and Mpc2 form an essential part of the mitochondrial pyruvate carrier.

Pyruvate occupies a pivotal node in the regulation of carbon metabolism as it is the end product of glycolysis and a major substrate for the tricarboxylic acid (TCA) cycle in mitochondria. Pyruvate lies at the intersection of these catabolic pathways with anabolic pathways for lipid synthesis, amino acid biosynthesis, and gluconeogenesis. As a result, the failure to correctly partition carbon between these fates lies at the heart of the altered metabolism evident in diabetes, obesity and cancer (*1*, *2*). Due to the fundamental importance of pyruvate, the mitochondrial pyruvate carrier (MPC) has been studied extensively (*3*, *4*). This included the discovery that α-cyanocinnamate analogs, such as UK-5099, act as specific and potent inhibitors of carrier activity (*5*). In spite of this characterization, however, the gene or genes that encode the mitochondrial pyruvate carrier remain unknown (*6*, *7*).

As part of an ongoing effort to characterize mitochondrial proteins that are conserved through evolution, we initiated studies of the Mitochondrial Pyruvate Carrier (MPC) protein family (originally designated BRP44 and BRP44L in humans) (*8*). This family contains three members in *S. cerevisiae*, encoded by *YGL080W*, *YHR162W* and *YGR243W*, hereafter referred to as *MPC1*, *MPC2* and *MPC3*, respectively. Mpc2 and Mpc3 are 79% identical in amino acid sequence and appear to be the product of a recent gene duplication event. Mpc1, Mpc2 and Mpc3 co-localize with mitochondria (Fig. 1A and fig. S2A), consistent with published mitochondrial proteomic studies (*9*, *10*). The mitochondrial localization of Mpc1 and Mpc2 was confirmed by biochemical fractionation (Fig. 1B). Mpc1, Mpc2, and Mpc3 were enriched in mitochondrial membranes (fig. S2B), consistent with the presence of predicted transmembrane domains in their sequences (fig. S1). Mpc1 and Mpc2 were resistant to protease treatment unless the mitochondrial outer membrane was ruptured (Fig. 1B and fig. S2C), implying that they are embedded in the mitochondrial inner membrane. Chromatographic purification of tagged variants of Mpc1 and Mpc2, followed by mass spectrometry, revealed that Mpc2 and Mpc3 were among the major interacting proteins of Mpc1, and Mpc1 and Mpc3 were among the major interacting proteins of Mpc2 (table S1). Consistent with this, immunoprecipitation of tagged Mpc1 co-purified Mpc2 and vice versa (Fig. 1C, lanes 3,4). In addition, Mpc2 can interact with itself (Fig. 1C, lane 8), while an Mpc1 homotypic interaction was not detected (Fig. 1C, lane 7). Blue native-polyacrylamide gel electrophoresis (BN-PAGE) showed that both Mpc1 and Mpc2 migrated as part of an ~150 kilodalton (kDa) complex (fig. S2D). Loss of Mpc2 prevents Mpc1 from migrating in this complex, while an *mpc1*Δ strain showed elevated Mpc2 complex formation (fig. S2E). We conclude that Mpc1 and Mpc2 form a multimeric complex embedded in the mitochondrial inner membrane, with Mpc2 likely being the major structural subunit.

Mutant yeast strains were subjected to a variety of growth conditions. The *mpc1*Δ and *mpc2*Δ cells displayed mild growth defects on non-fermentable carbon sources like glycerol, with greater effects on glucose medium (fig. S3) and a strong growth defect in the absence of leucine (Fig. 1D). In contrast, *mpc3*Δ mutant displayed no apparent growth phenotypes. Yeast, *Drosophila* or human *MPC1* orthologs, but not human *MPC2*, could rescue the *mpc1*Δ growth phenotype (Fig. 1E), indicating that Mpc1 function is conserved through evolution.

To analyze the physiological function of MPCs in a multicellular animal, we extended our studies to the *Drosophila* ortholog of MPC1 (dMPC1; encoded by *CG14290*), which also localized to mitochondria (fig. S4). Analogous to yeast *mpc1*Δ mutants, *dMPC1* mutants (fig. S5) were viable on standard food, but sensitive to a carbohydrate-only diet, with rapid lethality after transfer to a sucrose medium (Fig. 2A). While ATP was reduced in *dMPC1* mutants (Fig. 2C), along with TAG and protein (fig. S6B,C), carbohydrates were elevated, including the circulating sugar trehalose (Fig. 2D), glucose (Fig. 2E), fructose, and glycogen (fig. S6A,D). These results suggest that *dMPC1* mutants are defective in carbohydrate metabolism and may consume stored fat and protein for

energy. Consistent with this, the lethality of *dMPC1* mutants on the sugar diet was rescued by expression of the wild-type gene in tissues that depend heavily on glucose metabolism: the fat body, muscle, and neurons (Fig. 2B).

Metabolomic analyses revealed that pyruvate was highly elevated, while TCA cycle intermediates were significantly depleted in *dMPC1* mutants on the sugar diet (Fig. 2F). Similarly, glycine and serine, which can interconvert with glycolytic intermediates, were elevated in the mutants on the sugar diet (fig. S6E), while glutamate, aspartate, and proline, which can interconvert with TCA cycle intermediates, were depleted under these conditions (fig. S6F). Consistent with this, metabolomic analysis of *mpc1*Δ and *mpc2*Δ yeast mutants revealed elevated pyruvate levels (Fig. 3A), depletion of malate (fig. S7), depleted acetyl-CoA, and elevated CoA levels (Fig. 3B). Taken together, these results suggest that *MPC1* mutants are unable to efficiently convert cytosolic pyruvate to mitochondrial acetyl-CoA to drive the TCA cycle and ATP production.

These phenotypes could arise from either a defect in mitochondrial pyruvate uptake or the conversion of mitochondrial pyruvate into acetyl-CoA by the pyruvate dehydrogenase (PDH) complex. Yeast lacking Mpc1, however, had nearly wild-type PDH activity, unlike the strong decrease seen in *pda1*Δ mutants (Fig. 3C), which lack PDH function (*11*). A decrease in PDH activity also does not explain the growth defect of *mpc1*Δ mutants, which is more severe than that of the *pda1*Δ mutant (fig. S8). However, combining the *mpc1*Δ allele with a deletion for *mae1*, which encodes a malic enzyme that converts malate to pyruvate in the mitochondrial matrix (*12*), revealed a profound growth defect on glucose medium that was completely rescued by plasmid expression of either *MAE1* or *MPC1* (Fig. 3D). Importantly, mitochondria from the *mpc1*Δ mutant displayed almost no uptake of ^{14}C-pyruvate, which could be fully rescued by plasmid expression of wild-type *MPC1* (Fig. 3E). Moreover, Mpc1 appears to be a key target for UK-5099, which is an inhibitor of the mitochondrial pyruvate carrier (*5*). The *mae1*Δ *mpc1*Δ double mutant displayed reduced growth on glucose medium lacking leucine, and this phenotype could be effectively rescued by transgenic expression of wild-type *MPC1* in the absence, but not the presence, of UK-5099 (Fig. 3F). By screening for *MPC1* mutants that could grow in the presence of UK-5099 we recovered a D118G substitution in Mpc1 that conferred UK-5099 resistance (Fig. 3F). Moreover, while ^{14}C-pyruvate uptake into mitochondria expressing wild-type *MPC1* was almost completely inhibited by UK-5099, efficient pyruvate uptake that is resistant to UK-5099 was recovered upon expression of *MPC1-D118G* (Fig. 3G). We conclude that Mpc1 is a key component of the mitochondrial pyruvate carrier that corresponds to the activity studied for decades by Halestrap and others (*5*, *13*).

Depletion of *MPC1* in mouse embryonic fibroblasts (fig. S9) caused a modest decrease in pyruvate-driven oxygen consumption under basal conditions, and a stronger reduction in the presence of carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone (FCCP), which stimulates maximal respiration (Fig. 4A). Similar results were also seen upon silencing *MPC2* (Fig. 4B, fig. S9). This suppression of pyruvate oxidation, which occurred without affecting components of the oxidative phosphorylation machinery (fig. S9B,C), suggests that mammalian Mpc1 and Mpc2 mediate mitochondrial pyruvate uptake in a manner similar to that seen in yeast and *Drosophila*.

We have previously described a French-Algerian family with two offspring that exhibited a devastating defect in mitochondrial pyruvate oxidation (*14*) (Fig. 4C, Family 1). We subsequently discovered two additional families, each with one affected child who displayed a similar, but less severe, phenotype (Fig. 4C, Families 2 and 3). Linkage analysis and homozygosity mapping allowed us to focus on one candidate region on chromosome 6 (163,607,637-166,842,083, GRCh37/hg19). This interval contained 10 potential candidate genes: *PACRG, QKI, C6orf118, PDE10A, SDIM1, T, PRR18, SFT2D1, RPS6KA2*, and *BRP44L*, which is the human *MPC1*. DNA sequencing of the exons and intron/exon boundaries of the *MPC1* gene in fibroblasts from the affected patients in families 2 and 3 revealed the same molecular lesion, c.236T>A, causing a predicted p.Leu79His alteration (Fig. 4D). Analysis of DNA from family 1 revealed a distinct sequence change, c.289C>T, which resulted in a predicted p.Arg97Trp mutation (Fig. 4D). Both of the affected residues are conserved through evolution between *MPC1* orthologs, and Arg97 is conserved amongst both *MPC1* and *MPC2* orthologs (fig. S1).

Cells from the affected individuals in families 1 and 2 exhibited impaired basal and FCCP-stimulated pyruvate oxidation (Fig. 4E), while glutamine-driven oxygen consumption was normal or elevated, demonstrating that they have not acquired a generalized impairment of mitochondrial respiration (Fig. 4E). As expected, expression of wild-type human *MPC1* in the cells from family 2 (Fig. 4F) or family 1 (Fig. 4G) either completely or partially rescued the defect in FCCP-induced pyruvate oxidation. Moreover, expression of the *MPC1-Leu79His* allele was less effective at suppressing the yeast *mpc1*Δ growth defect relative to wild-type human *MPC1* (Fig. 4H), while the stronger *MPC1-Arg97Trp* allele was essentially inactive (Fig. 4H), suggesting that *MPC1* function is evolutionarily conserved from yeast to man.

In summary, the data presented herein demonstrate that the Mpc1/Mpc2 complex is an essential component of the mitochondrial pyruvate carrier in yeast, flies and mammals. This is consistent with experiments performed in rat liver, heart, and castor beans, which implicated proteins of 12-15 kDa in mitochondrial pyruvate uptake (*15*) – similar to the molecular weights of Mpc1 (15kDa), Mpc2 (14kDa) and Mpc3 (16kDa). Although these individual sizes are relatively small, Mpc1 and Mpc2 form a complex of ~150kDa, suggesting that an oligomeric structure mediates pyruvate transport. Definitive proof that the Mpc1/Mpc2 complex is the mitochondrial pyruvate carrier, however, awaits the reconstitution of pyruvate transport in a heterologous system [Herzig, *et al*., accompanying manuscript (*16*)]. Finally, it is important to note that the degree to which carbohydrates are imported into mitochondria and converted into acetyl-CoA is a critical step in normal glucose oxidation as well as the onset of diabetes, obesity and cancer. Thus, like PDH, which is controlled by allostery and post-translational modification (*17*), the mitochondrial import of pyruvate is likely to be a precisely regulated process (*18*, *19*). The identification of Mpc1 and Mpc2 as critical for mitochondrial pyruvate transport provides a new framework for understanding this level of metabolic control as well as new directions for potential therapeutic intervention.

References and Notes

1. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646 (2011). doi:10.1016/j.cell.2011.02.013 Medline
2. S. E. Kahn, R. L. Hull, K. M. Utzschneider, Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* **444**, 840 (2006). doi:10.1038/nature05482 Medline
3. A. P. Halestrap, R. M. Denton, Specific inhibition of pyruvate transport in rat liver mitochondria and human erythrocytes by alpha-cyano-4-hydroxycinnamate. *Biochem. J.* **138**, 313 (1974). Medline
4. A. P. Halestrap, Pyruvate and ketone-body transport across the mitochondrial membrane. Exchange properties, pH-dependence and mechanism of the carrier. *Biochem. J.* **172**, 377 (1978). Medline
5. A. P. Halestrap, The mitochondrial pyruvate carrier. Kinetics and specificity for substrates and inhibitors. *Biochem. J.* **148**, 85 (1975). Medline
6. S. Todisco, G. Agrimi, A. Castegna, F. Palmieri, Identification of the mitochondrial NAD+ transporter in Saccharomyces cerevisiae. *J. Biol. Chem.* **281**, 1524 (2006). doi:10.1074/jbc.M510425200 Medline
7. J. C. Hildyard, A. P. Halestrap, Identification of the mitochondrial pyruvate carrier in Saccharomyces cerevisiae. *Biochem. J.* **374**, 607 (2003). doi:10.1042/BJ20030995 Medline
8. M. Jiang *et al*., The molecular characterization of the brain protein 44-like (Brp44l) gene of Gekko japonicus and its expression changes in spinal cord after tail amputation. *Mol. Biol. Rep.* **36**, 215 (2009). doi:10.1007/s11033-

007-9169-0 Medline
9. D. J. Pagliarini *et al*., A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112 (2008). doi:10.1016/j.cell.2008.06.016 Medline
10. A. Sickmann *et al*., The proteome of Saccharomyces cerevisiae mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13207 (2003). doi:10.1073/pnas.2135385100 Medline
11. H. Y. Steensma, L. Holterman, I. Dekker, C. A. van Sluis, T. J. Wenzel, Molecular cloning of the gene for the E1 alpha subunit of the pyruvate dehydrogenase complex from Saccharomyces cerevisiae. *Eur. J. Biochem.* **191**, 769 (1990). doi:10.1111/j.1432-1033.1990.tb19186.x Medline
12. E. Boles, P. de Jong-Gubbels, J. T. Pronk, Identification and characterization of MAE1, the Saccharomyces cerevisiae structural gene encoding mitochondrial malic enzyme. *J. Bacteriol.* **180**, 2875 (1998). Medline
13. S. Papa, G. Paradies, On the mechanism of translocation of pyruvate and other monocarboxylic acids in rat-liver mitochondria. *Eur. J. Biochem.* **49**, 265 (1974). doi:10.1111/j.1432-1033.1974.tb03831.x Medline
14. M. Brivet *et al*., Impaired mitochondrial pyruvate importation in a patient and a fetus at risk. *Mol. Genet. Metab.* **78**, 186 (2003). doi:10.1016/S1096-7192(03)00016-7 Medline
15. A. P. Thomas, A. P. Halestrap, Identification of the protein responsible for pyruvate transport into rat liver and heart mitochondria by specific labelling with [3H]N-phenylmaleimide. *Biochem. J.* **196**, 471 (1981). Medline
16. S. Herzig *et al*., Identification and functional expression of the mitochondrial pyruvate carrier. *Science* 10.1126/science.1218530; published online 24 May 2012.
17. R. A. Harris, M. M. Bowker-Kinley, B. Huang, P. Wu, Regulation of the activity of the pyruvate dehydrogenase complex. *Adv. Enzyme Regul.* **42**, 249 (2002). doi:10.1016/S0065-2571(01)00061-9 Medline
18. F. M. Zwiebel, U. Schwabe, M. S. Olson, R. Scholz, Role of pyruvate transporter in the regulation of the pyruvate dehydrogenase multienzyme complex in perfused rat liver. *Biochemistry* **21**, 346 (1982). doi:10.1021/bi00531a023 Medline
19. R. Rognstad, The role of mitochondrial pyruvate transport in the control of lactate gluconeogenesis. *Int. J. Biochem.* **15**, 1417 (1983). doi:10.1016/0020-711X(83)90073-3 Medline
20. Materials and methods are available as Supporting Online Material on *Science* Online.

Acknowledgements: We thank members of the Rutter, Thummel, Winge, Stillman, Shaw and Metzstein laboratories for helpful discussions. We thank the Shaw and Winge labs for the anti-Fzo1, anti-Cyb2 and anti-Mge1 antibodies and mito-RFP constructs. We thank Pr. Saudubray, Dr. Burglen and Dr. Tevissen for referring patients; and to C Thibault and Pr. JL Mandel (Institut de Genetique et de Biologie Moleculaire et Cellulaire (IGBMC, Strasbourg, France) for assistance in SNP array hybridization. This research was supported by NIH grants RO1GM083746 (JR) and GM0RC1DK086426 (CST) and R24DK092784 (JR and CST) and a pilot grant from P30DK072437 (JR). DKB and CMC were supported by the NIH Genetics Predoctoral Training Grant T32GM007464. EBT was supported by NIH Pathway to Independence award K99AR059190 (EBT). DKB, TO, CST and JR are inventors on a patent application by the University of Utah covering the discovery of the MPC complex.

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1218099/DC1
Materials and Methods
Figs. S1 to S10
Table S1
References

19 December 2011; accepted 9 May 2012

Published online 24 May 2012
10.1126/science.1218099

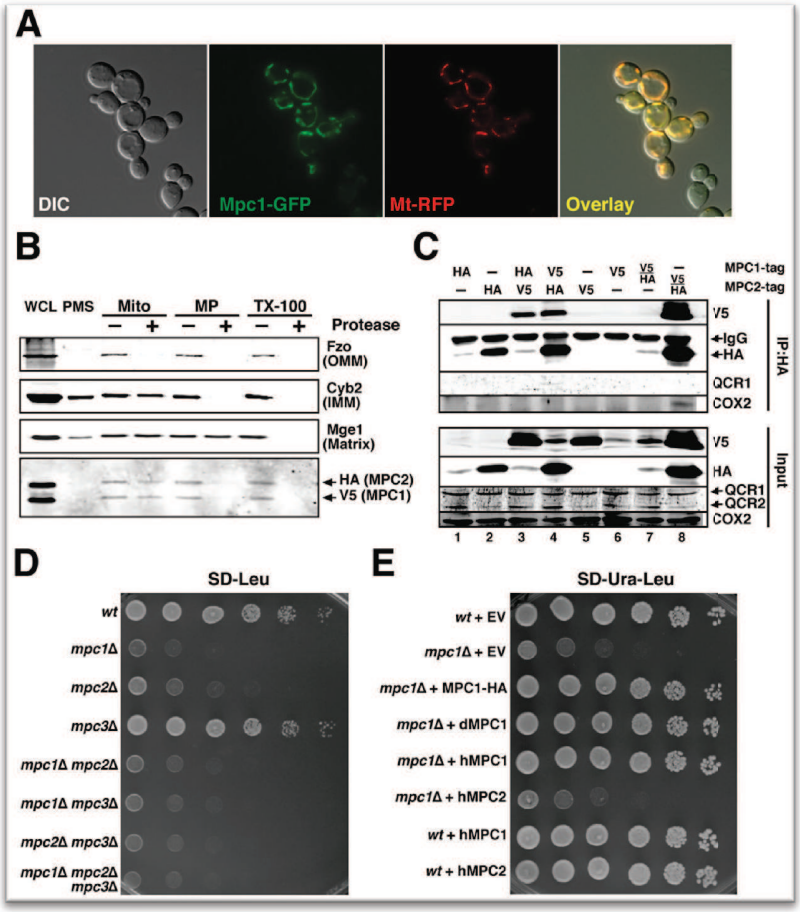


Fig. 1. Mpc1 and Mpc2 are evolutionarily-conserved mitochondrial inner membrane proteins. (A) Mpc1-GFP and mitochondrial targeted RFP (mtRFP) coexpressed in yeast cells. (**B**) Intact mitochondria, hypotonic-swollen mitoplasts, and TritonX-100–solubilized mitochondria from a strain expressing Mpc1-V5 and Mpc2-His_6/HA_2 with (+) or without (-) Proteinase K incubation. An immunoblot of extracts using the indicated antibodies with the whole cell lysate (WCL) and post mitochondrial supernatant (PMS) is shown. Mge1, Cyb2, and Fzo1 are matrix, intermembrane space, and outer membrane proteins, respectively. (**C**) Immunoprecipitations from mitochondrial extracts from *mpc1*Δ *mpc2*Δ cells expressing Mpc1 and Mpc2 tagged as indicated. Immunoblot of either immunoprecipitate (IP:HA) or input is shown. QCR1 and 2 (ubiquinol-cytochrome-c reductase complex core protein 1 and 2) along with Cox II (cytochrome c oxidase subunit 2) are controls for the specificity of the immunoprecipitation. (**D**) Serial dilutions of the indicated yeast strains spotted on synthetic media lacking leucine and grown at 30°C for 24 hours. (**E**) Serial dilutions of indicated strains spotted on synthetic media lacking leucine and grown at 30°C for 48 hours. *wt*=wild-type; *EV*=empty vector.

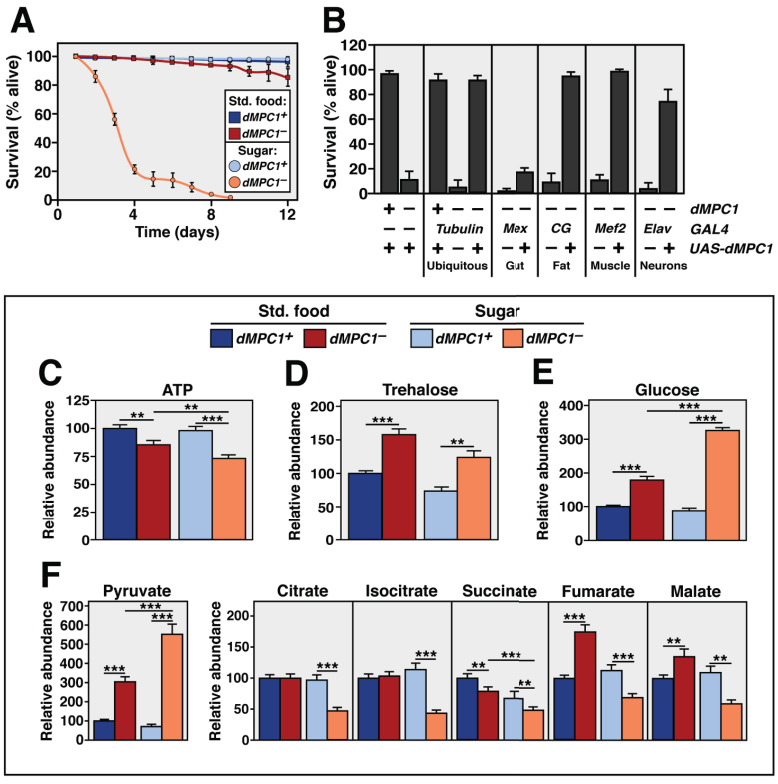


Fig. 2. *dMPC1* is required for pyruvate metabolism in *Drosophila.* (**A**) Percentage of living control (*dMPC1*$^+$) or *dMPC1* mutant (*dMPC*$^-$) flies after transfer to standard laboratory medium (std. food) or to media containing only sugar. (**B**) Percentage of living *dMPC1*$^+$ or *dMPC*$^-$ flies carrying the indicated GAL4 and UAS transgenes on sugar media after eight days. (**C-E**) Relative concentration of ATP (**C**), trehalose (**D**) and glucose (**E**) in extracts from *dMPC1*$^+$ or *dMPC*$^-$ flies on the indicated diet after either two (**D-E**) or three (**C**) days. (**F**) Relative abundance of pyruvate and TCA cycle intermediates in *dMPC1*$^+$ or *dMPC*$^-$ flies after two days on the indicated diet measured by GC/MS. *p<0.05, **p<0.01 and ***p<0.001 (Student's *t* test). Mean±SEM.

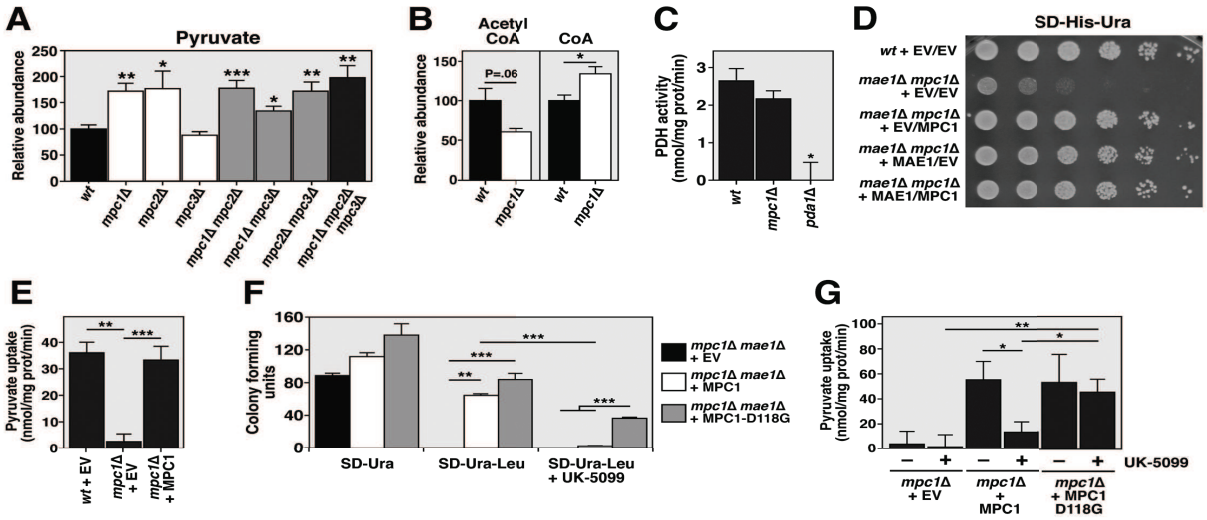


Fig. 3. *MPC1* is required for mitochondrial pyruvate uptake. (**A**) Relative abundance of pyruvate in the indicated strains. p-values relative to *wt*. (**B**) Relative abundance of acetyl-CoA and CoA in the indicated strains. (**C**) Mitochondrial pyruvate dehydrogenase activity in the indicated strains. p value relative to *wt* and *mpc1*Δ. (**D**) Serial dilutions of the indicated strain on glucose medium grown at 30°C for 48 hours. (**E**) Uptake of ^{14}C-pyruvate into mitochondria purified from either *wt* or *mpc1*Δ cells containing the indicated plasmid. p value relative to *wt + EV* and *mpc1*Δ *+ MPC1*. (**F**) *Mae1Δ mpc1Δ* cells transformed with the indicated plasmid and plated on media containing or lacking combinations of leucine or UK-5099. (**G**) Uptake of ^{14}C-pyruvate into mitochondria isolated from the *mpc1*Δ strain containing the indicated plasmid in the presence or absence of UK-5099. *** p< 0.001, ** p< 0.01, * p< 0.05, NS=not significant (Student's *t* test). Mean±SEM.

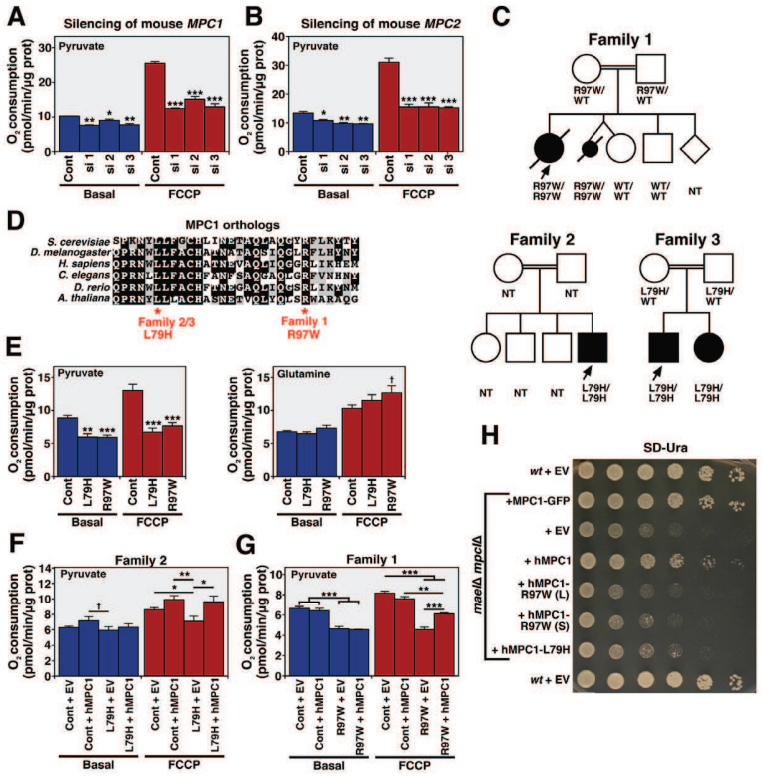


Fig. 4. Mammalian *MPC1* and *MPC2* are required for normal pyruvate metabolism. (**A-B**) Pyruvate-driven respiration in mouse embryonic fibroblasts (MEFs) under basal and FCCP-stimulated conditions in cells transfected with either control (cont.) siRNAs or three different siRNAs (si1-3) targeted to either *MPC1* (**A**) or *MPC2*. (**B**) p-values relative to control. (**C**) Pedigrees of families 1, 2, and 3. Circles indicate females, squares males, and diamonds unknown sex. Black indicates deceased and white living. Arrows mark individuals from whom fibroblasts were obtained. (**D**) The protein region of MPC1 containing the predicted amino acid substitutions from all three families aligned by ClustalW. (**E**) Pyruvate- (left) and glutamine- (right) supported respiration of fibroblasts harboring either L79H or R97W *MPC1* mutations. (**F**) Pyruvate-supported respiration of either a control or L79H patient cell line after transduction with the indicated vector. (**G**) Pyruvate-supported respiration of either a control or R97W patient cell line after transduction with the indicated vector. (**H**) Serial dilutions of *wt* or *mae1*Δ *mpc1*Δ yeast strains carrying the indicated plasmid grown on medium lacking uracil for plasmid selection at 30°C for 40 hours. Both long (L) and short (S) forms of R97W were used (with or without exon 4). *** P < 0.001, ** P < 0.01, * P < 0.05, † P < 0.10 and NS=not significant (Student's *t* test). Mean±SEM.

References

Abdollahi, M. R., E. Morrison, et al. (2009). "Mutation of the variant alpha-tubulin TUBA8 results in polymicrogyria with optic nerve hypoplasia." Am J Hum Genet **85**(5): 737-44.

Abu-Safieh, L., S. Al-Anazi, et al. (2012). "In search of triallelism in Bardet-Biedl syndrome." Eur J Hum Genet **20**(4): 420-7.

Adzhubei, I. A., S. Schmidt, et al. (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-9.

Afzelius, B. A. (1976). "A human syndrome caused by immotile cilia." Science **193**(4250): 317-9.

Albers, C. A., D. S. Paul, et al. (2012). "Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome." Nat Genet **44**(4): 435-9, S1-2.

Albert, T. J., M. N. Molla, et al. (2007). "Direct selection of human genomic loci by microarray hybridization." Nat Methods **4**(11): 903-5.

Aldahmesh, M. A., Y. Li, et al. (2014). "IFT27, encoding a small GTPase component of IFT particles, is mutated in a consanguineous family with Bardet-Biedl syndrome." Hum Mol Genet.

Amir, R. E., I. B. Van den Veyver, et al. (1999). "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2." Nat Genet **23**(2): 185-8.

Ansley, S. J., J. L. Badano, et al. (2003). "Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome." Nature **425**(6958): 628-33.

Aubourg, P., S. Blanche, et al. (1990). "Reversal of early neurologic and neuroradiologic manifestations of X-linked adrenoleukodystrophy by bone marrow transplantation." N Engl J Med **322**(26): 1860-6.

Avidor-Reiss, T., A. M. Maer, et al. (2004). "Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis." Cell **117**(4): 527-39.

Baala, L., S. Audollent, et al. (2007). "Pleiotropic effects of CEP290 (NPHP6) mutations extend to Meckel syndrome." Am J Hum Genet **81**(1): 170-9.

Badano, J. L., S. J. Ansley, et al. (2003a). "Identification of a novel Bardet-Biedl syndrome protein, BBS7, that shares structural features with BBS1 and BBS2." Am J Hum Genet **72**(3): 650-8.

Badano, J. L., J. C. Kim, et al. (2003b). "Heterozygous mutations in BBS1, BBS2 and BBS6 have a potential epistatic effect on Bardet-Biedl patients with two mutations at a second BBS locus." Hum Mol Genet **12**(14): 1651-9.

Badano, J. L., C. C. Leitch, et al. (2006). "Dissection of epistasis in oligogenic Bardet-Biedl syndrome." Nature **439**(7074): 326-30.

Bakircioglu, M., O. P. Carvalho, et al. (2011). "The essential role of centrosomal NDE1 in human cerebral cortex neurogenesis." Am J Hum Genet **88**(5): 523-35.

Bardet, G. (1920). "Sur un syndrome d'obesite congenitale avec polydactylie et retinite pigmentaire (contribution a l'etude des formes cliniques de l'obesite hypophysaire)." Univ. de Paris. Thesis no. 470. Amedee LeGrand. Paris, France.

Baroff, G. S. and J. G. Olley (1999). Mental Retardation: Nature, Cause and Management.

Basel-Vanagaite, L., R. Attia, et al. (2006). "The CC2D1A, a member of a new gene family with C2 domains, is involved in autosomal recessive non-syndromic mental retardation." J Med Genet **43**(3): 203-10.

Bates, S. H., W. H. Stearns, et al. (2003). "STAT3 signalling is required for leptin regulation of energy balance but not reproduction." Nature **421**(6925): 856-9.

Beales, P. L., J. L. Badano, et al. (2003). "Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome." Am J Hum Genet **72**(5): 1187-99.

Beales, P. L., N. Elcioglu, et al. (1999). "New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey." J Med Genet **36**(6): 437-46.

Beirne-Smith, M., P. J.R., et al. (2006a). Mental retardation: An introduction to intellectual disabilities, Upper Saddle River (NJ): Pearson Merrill Prentice Hall.

Beirne-Smith, M., J. R. Patton , et al. (2006b). Mental retardation: An introduction to intellectual disabilities.

Benazet, J. D. and R. Zeller (2009). "Vertebrate limb development: moving from classical morphogen gradients to an integrated 4-dimensional patterning system." Cold Spring Harb Perspect Biol **1**(4): a001339.
Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) **57**(1): 289-300.
Bernal, A., U. Ear, et al. (2001). "Genomes OnLine Database (GOLD): a monitor of genome projects world-wide." Nucleic Acids Res **29**(1): 126-7.
Bernard, G., E. Chouery, et al. (2011). "Mutations of POLR3A encoding a catalytic subunit of RNA polymerase Pol III cause a recessive hypomyelinating leukodystrophy." Am J Hum Genet **89**(3): 415-23.
Berry-Kravis, E., D. Hessl, et al. (2009). "A pilot open label, single dose trial of fenobam in adults with fragile X syndrome." J Med Genet **46**(4): 266-71.
Berry-Kravis, E. M., D. Hessl, et al. (2012). "Effects of STX209 (arbaclofen) on neurobehavioral function in children and adults with fragile X syndrome: a randomized, controlled, phase 2 trial." Sci Transl Med **4**(152): 152ra127.
Bezman, L., A. B. Moser, et al. (2001). "Adrenoleukodystrophy: incidence, new mutation rate, and results of extended family screening." Ann Neurol **49**(4): 512-7.
Bhagavati, S. (2014). "Questions about efficacy of exon-skipping therapy for duchenne muscular dystrophy." Ann Neurol **75**(2): 326-9.
Biedl, A. (1922). "Ein Geschwisterpaar mit adiposo-genitaler Dystrophie." Dtsch. Med. Wochenschr. **48:1630.**
Bienvenu, T., K. Poirier, et al. (2002). "ARX, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation." Hum Mol Genet **11**(8): 981-91.
Billingsley, G., A. Vincent, et al. (2012). "Mutational analysis of SDCCAG8 in Bardet-Biedl syndrome patients with renal involvement and absent polydactyly." Ophthalmic Genet **33**(3): 150-4.
Bisgrove, B. W. and H. J. Yost (2006). "The roles of cilia in developmental disorders and disease." Development **133**(21): 4131-43.
Bowen, M. E., E. D. Boyden, et al. (2011). "Loss-of-function mutations in PTPN11 cause metachondromatosis, but not Ollier disease or Maffucci syndrome." PLoS Genet **7**(4): e1002050.
Braslavsky, I., B. Hebert, et al. (2003). "Sequence information can be obtained from single DNA molecules." Proc Natl Acad Sci U S A **100**(7): 3960-4.
Brenner, M., A. B. Johnson, et al. (2001). "Mutations in GFAP, encoding glial fibrillary acidic protein, are associated with Alexander disease." Nat Genet **27**(1): 117-20.
Bricker, D. K., E. B. Taylor, et al. (2012). "A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, Drosophila, and humans." Science **337**(6090): 96-100.
Brownstein, Z., L. M. Friedman, et al. (2011). "Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in Middle Eastern families." Genome Biol **12**(9): R89.
Brumm, V. L., D. Bilder, et al. (2010). "Psychiatric symptoms and disorders in phenylketonuria." Mol Genet Metab **99 Suppl 1**: S59-63.
Brunelli, S., A. Faiella, et al. (1996). "Germline mutations in the homeobox gene EMX2 in patients with severe schizencephaly." Nat Genet **12**(1): 94-6.
Brunner, H. G., M. Nelen, et al. (1993). "Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A." Science **262**(5133): 578-80.
Cabezas, D. A., R. Slaugh, et al. (2000). "A new X linked mental retardation (XLMR) syndrome with short stature, small testes, muscle wasting, and tremor localises to Xq24-q25." J Med Genet **37**(9): 663-8.
Cacciagli, P., J. Sutera-Sardo, et al. (2013). "Mutations in BCAP31 cause a severe X-linked phenotype with deafness, dystonia, and central hypomyelination and disorganize the Golgi apparatus." Am J Hum Genet **93**(3): 579-86.
Cappello, S., M. J. Gray, et al. (2013). "Mutations in genes encoding the cadherin receptor-ligand pair DCHS1 and FAT4 disrupt cerebral cortical development." Nat Genet **45**(11): 1300-8.
Cartier, N., S. Hacein-Bey-Abina, et al. (2009). "Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy." Science **326**(5954): 818-23.
Carvalho, C. M., F. Zhang, et al. (2009). "Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching." Hum Mol Genet **18**(12): 2188-203.

Cavalier-Smith, T. (2002). "The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa." Int J Syst Evol Microbiol **52**(Pt 2): 297-354.
Chahrour, M. H., T. W. Yu, et al. (2012). "Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism." PLoS Genet **8**(4): e1002635.
Chapman, M., P. Iddon, et al. (2011). "The misdiagnosis of epilepsy in people with intellectual disabilities: a systematic review." Seizure **20**(2): 101-6.
Chen, C. P., Y. N. Su, et al. (2011). "De novo duplication of Xq22.1-->q24 with a disruption of the NXF gene cluster in a mentally retarded woman with short stature and premature ovarian failure." Taiwan J Obstet Gynecol **50**(3): 339-44.
Chiang, A. P., J. S. Beck, et al. (2006). "Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11)." Proc Natl Acad Sci U S A **103**(16): 6287-92.
Chiang, A. P., D. Nishimura, et al. (2004). "Comparative genomic analysis identifies an ADP-ribosylation factor-like gene as the cause of Bardet-Biedl syndrome (BBS3)." Am J Hum Genet **75**(3): 475-84.
Chizhikov, V. V., J. Davenport, et al. (2007). "Cilia proteins control cerebellar morphogenesis by promoting expansion of the granule progenitor pool." J Neurosci **27**(36): 9780-9.
Claes, L., J. Del-Favero, et al. (2001). "De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy." Am J Hum Genet **68**(6): 1327-32.
Clarke, J., H. C. Wu, et al. (2009). "Continuous base identification for single-molecule nanopore DNA sequencing." Nat Nanotechnol **4**(4): 265-70.
Coffee, B., K. Keith, et al. (2009). "Incidence of fragile X syndrome by newborn screening for methylated FMR1 DNA." Am J Hum Genet **85**(4): 503-14.
Cohen, A. S., D. R. Najarian, et al. (1988). "Rapid separation and purification of oligonucleotides by high-performance capillary gel electrophoresis." Proc Natl Acad Sci U S A **85**(24): 9660-3.
Collins, F. S., E. D. Green, et al. (2003). "A vision for the future of genomics research." Nature **422**(6934): 835-47.
Consortium, E., E. M. Consortium, et al. (2012). "Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32." Hum Mol Genet **21**(24): 5359-72.
Consortium, E. P., E. Birney, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.
Conti, V., A. Carabalona, et al. (2013). "Periventricular heterotopia in 6q terminal deletion syndrome: role of the C6orf70 gene." Brain **136**(Pt 11): 3378-94.
Cooper, G. M., B. P. Coe, et al. (2011). "A copy number variation morbidity map of developmental delay." Nat Genet **43**(9): 838-46.
Copp, A. J., P. Stanier, et al. (2013). "Neural tube defects: recent advances, unsolved questions, and controversies." Lancet Neurol **12**(8): 799-810.
Coppieters, F., S. Lefever, et al. (2010). "CEP290, a gene with many faces: mutation overview and presentation of CEP290base." Hum Mutat **31**(10): 1097-108.
Corbit, K. C., P. Aanstad, et al. (2005). "Vertebrate Smoothened functions at the primary cilium." Nature **437**(7061): 1018-21.
Costello, D. J., A. F. Eichler, et al. (2009). "Leukodystrophies: classification, diagnosis, and treatment." Neurologist **15**(6): 319-28.
Cremers, F. P., R. A. Pfeiffer, et al. (1987). "An interstitial duplication of the X chromosome in a male allows physical fine mapping of probes from the Xq13-q22 region." Hum Genet **77**(1): 23-7.
Cristino, A. S., S. M. Williams, et al. (2014). "Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system." Mol Psychiatry **19**(3): 294-301.
Curry, C. J., R. E. Stevenson, et al. (1997). "Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics." Am J Med Genet **72**(4): 468-77.
Daniels, A. B., M. A. Sandberg, et al. (2012). "Genotype-phenotype correlations in Bardet-Biedl syndrome." Arch Ophthalmol **130**(7): 901-7.
de Ligt, J., M. H. Willemsen, et al. (2012). "Diagnostic exome sequencing in persons with severe intellectual disability." N Engl J Med **367**(20): 1921-9.
Deardorff, M. A., M. Bando, et al. (2012). "HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle." Nature **489**(7415): 313-7.
Deary, I. J., L. Penke, et al. (2010). "The neuroscience of human intelligence differences." Nat Rev Neurosci **11**(3): 201-11.

den Hollander, A. I., R. K. Koenekoop, et al. (2006). "Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis." Am J Hum Genet **79**(3): 556-61.
Derrien, T., R. Johnson, et al. (2012). "The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression." Genome Res **22**(9): 1775-89.
des Portes, V., M. O. Livet, et al. (2002). "[A practical diagnostic approach to mental deficiency in 2002]." Arch Pediatr **9**(7): 709-25.
des Portes, V., J. M. Pinard, et al. (1998). "A novel CNS gene required for neuronal migration and involved in X-linked subcortical laminar heterotopia and lissencephaly syndrome." Cell **92**(1): 51-61.
Desmet, F. O., D. Hamroun, et al. (2009). "Human Splicing Finder: an online bioinformatics tool to predict splicing signals." Nucleic Acids Res **37**(9): e67.
Deveault, C., G. Billingsley, et al. (2011). "BBS genotype-phenotype assessment of a multiethnic patient cohort calls for a revision of the disease definition." Hum Mutat **32**(6): 610-9.
Djebali, S., C. A. Davis, et al. (2012). "Landscape of transcription in human cells." Nature **489**(7414): 101-8.
Do, C. B., J. Y. Tung, et al. (2011). "Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease." PLoS Genet **7**(6): e1002141.
Dumitrescu, A. M., X. H. Liao, et al. (2004). "A novel syndrome combining thyroid and neurological abnormalities is associated with mutations in a monocarboxylate transporter gene." Am J Hum Genet **74**(1): 168-75.
Durkin, M. (2002). "The epidemiology of developmental disabilities in low-income countries." Ment Retard Dev Disabil Res Rev **8**(3): 206-11.
El Zein, L., H. Omran, et al. (2003). "Lateralization defects and ciliary dyskinesia: lessons from algae." Trends Genet **19**(3): 162-7.
Eley, L., L. M. Yates, et al. (2005). "Cilia and disease." Curr Opin Genet Dev **15**(3): 308-14.
Eliasson, R., B. Mossberg, et al. (1977). "The immotile-cilia syndrome. A congenital ciliary abnormality as an etiologic factor in chronic airway infections and male sterility." N Engl J Med **297**(1): 1-6.
Elinati, E., P. Kuentz, et al. (2012). "Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots." Hum Mol Genet **21**(16): 3695-702.
Emerson, E. and C. Hatton (2007). "Contribution of socioeconomic position to health inequalities of British children and adolescents with intellectual disabilities." Am J Ment Retard **112**(2): 140-50.
Emery, B. (2010). "Regulation of oligodendrocyte differentiation and myelination." Science **330**(6005): 779-82.
Engbers, H. M., R. Berger, et al. (2008). "Yield of additional metabolic studies in neurodevelopmental disorders." Ann Neurol **64**(2): 212-7.
Evans, J. E., J. J. Snow, et al. (2006). "Functional modulation of IFT kinesins extends the sensory repertoire of ciliated neurons in Caenorhabditis elegans." J Cell Biol **172**(5): 663-9.
Fan, Y., M. A. Esmail, et al. (2004). "Mutations in a member of the Ras superfamily of small GTP-binding proteins causes Bardet-Biedl syndrome." Nat Genet **36**(9): 989-93.
Fischer, A., S. Hacein-Bey-Abina, et al. (2010). "20 years of gene therapy for SCID." Nat Immunol **11**(6): 457-60.
Fliegauf, M., T. Benzing, et al. (2007). "When cilia go bad: cilia defects and ciliopathies." Nat Rev Mol Cell Biol **8**(11): 880-93.
Fombonne, E. (2009). "Epidemiology of pervasive developmental disorders." Pediatr Res **65**(6): 591-8.
Fourcade, S., M. Ruiz, et al. (2010). "Valproic acid induces antioxidant effects in X-linked adrenoleukodystrophy." Hum Mol Genet **19**(10): 2005-14.
Fox, J. W., E. D. Lamperti, et al. (1998). "Mutations in filamin 1 prevent migration of cerebral cortical neurons in human periventricular heterotopia." Neuron **21**(6): 1315-25.
Francis, F., G. Meyer, et al. (2006). "Human disorders of cortical development: from past to present." Eur J Neurosci **23**(4): 877-93.
Frazer, K. A. (2012). "Decoding the human genome." Genome Res **22**(9): 1599-601.
Friedmann, T. and R. Roblin (1972). "Gene therapy for human genetic disease?" Science **175**(4025): 949-55.

Frosk, P., T. Weiler, et al. (2002). "Limb-girdle muscular dystrophy type 2H associated with mutation in TRIM32, a putative E3-ubiquitin-ligase gene." Am J Hum Genet **70**(3): 663-72.
Galluzzi, P., A. Rufa, et al. (2001). "MR brain imaging of fucosidosis type I." AJNR Am J Neuroradiol **22**(4): 777-80.
Gecz, J., C. Shoubridge, et al. (2009). "The genetic landscape of intellectual disability arising from chromosome X." Trends Genet **25**(7): 308-16.
Gencic, S., D. Abuelo, et al. (1989). "Pelizaeus-Merzbacher disease: an X-linked neurologic disorder of myelin metabolism with a novel mutation in the gene encoding proteolipid protein." Am J Hum Genet **45**(3): 435-42.
Gherman, A., E. E. Davis, et al. (2006). "The ciliary proteome database: an integrated community resource for the genetic and functional dissection of cilia." Nat Genet **38**(9): 961-2.
Girirajan, S., Z. Brkanac, et al. (2011). "Relative burden of large CNVs on a range of neurodevelopmental phenotypes." PLoS Genet **7**(11): e1002334.
Gleeson, J. G., K. M. Allen, et al. (1998). "Doublecortin, a brain-specific gene mutated in human X-linked lissencephaly and double cortex syndrome, encodes a putative signaling protein." Cell **92**(1): 63-72.
Grayton, H. M., C. Fernandes, et al. (2012). "Copy number variations in neurodevelopmental disorders." Prog Neurobiol **99**(1): 81-91.
Guilmatre, A., C. Dubourg, et al. (2009). "Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation." Arch Gen Psychiatry **66**(9): 947-56.
Haack, T. B., P. Hogarth, et al. (2012). "Exome sequencing reveals de novo WDR45 mutations causing a phenotypically distinct, X-linked dominant form of NBIA." Am J Hum Genet **91**(6): 1144-9.
Habas, R., I. B. Dawid, et al. (2003). "Coactivation of Rac and Rho by Wnt/Frizzled signaling is required for vertebrate gastrulation." Genes Dev **17**(2): 295-309.
Hagberg, B., J. Aicardi, et al. (1983). "A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases." Ann Neurol **14**(4): 471-9.
Hamdan, F. F., H. Daoud, et al. (2013). "Parent-child exome sequencing identifies a de novo truncating mutation in TCF4 in non-syndromic intellectual disability." Clin Genet **83**(2): 198-200.
Hamdan, F. F., J. Gauthier, et al. (2011). "Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability." Am J Hum Genet **88**(3): 306-16.
Hamdan, F. F., J. Gauthier, et al. (2009a). "Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation." N Engl J Med **360**(6): 599-605.
Hamdan, F. F., A. Piton, et al. (2009b). "De novo STXBP1 mutations in mental retardation and nonsyndromic epilepsy." Ann Neurol **65**(6): 748-53.
Hammonds, R. G., Jr. (1987). "Protein sequence of DMD gene is related to actin-binding domain of alpha-actinin." Cell **51**(1): 1.
Han, Y. G., H. J. Kim, et al. (2009). "Dual and opposing roles of primary cilia in medulloblastoma development." Nat Med **15**(9): 1062-5.
Haraguchi, R., R. Mo, et al. (2001). "Unique functions of Sonic hedgehog signaling during external genitalia development." Development **128**(21): 4241-50.
Harakalova, M., M. J. van den Boogaard, et al. (2012). "X-exome sequencing identifies a HDAC8 variant in a large pedigree with X-linked intellectual disability, truncal obesity, gynaecomastia, hypogonadism and unusual face." J Med Genet **49**(8): 539-43.
Harbour, C. K. and P. K. Maulik (2010). History of Intellectual Disability. International Encyclopedia of Rehabilitation.
Harris, J. C. (2006). "Intellectual disability: understanding its development, causes, classification, evaluation, and treatment." New York: Oxford University Press: 42-98.
Harrow, J., A. Frankish, et al. (2012). "GENCODE: the reference human genome annotation for The ENCODE Project." Genome Res **22**(9): 1760-74.
Haycraft, C. J., B. Banizs, et al. (2005). "Gli2 and Gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function." PLoS Genet **1**(4): e53.
Hehr, U., D. E. Pineda-Alvarez, et al. (2010). "Heterozygous mutations in SIX3 and SHH are associated with schizencephaly and further expand the clinical spectrum of holoprosencephaly." Hum Genet **127**(5): 555-61.

Heinzen, E. L., C. Depondt, et al. (2012). "Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy." Am J Hum Genet **91**(2): 293-302.
Hichri, H., C. Stoetzel, et al. (2005). "Testing for triallelism: analysis of six BBS genes in a Bardet-Biedl syndrome family cohort." Eur J Hum Genet **13**(5): 607-16.
Higgins, J. J., J. Pucilowska, et al. (2004). "A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation." Neurology **63**(10): 1927-31.
Hildebrandt, F., T. Benzing, et al. (2011). "Ciliopathies." N Engl J Med **364**(16): 1533-43.
Hirata, H., I. Nanda, et al. (2013). "ZC4H2 mutations are associated with arthrogryposis multiplex congenita and intellectual disability through impairment of central and peripheral synaptic plasticity." Am J Hum Genet **92**(5): 681-95.
Hoffman, E. P., R. H. Brown, Jr., et al. (1987). "Dystrophin: the protein product of the Duchenne muscular dystrophy locus." Cell **51**(6): 919-28.
Hong, S. E., Y. Y. Shugart, et al. (2000). "Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human RELN mutations." Nat Genet **26**(1): 93-6.
Houdayer, C., V. Caux-Moncoutier, et al. (2012). "Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants." Hum Mutat **33**(8): 1228-38.
Howald, C., A. Tanzer, et al. (2012). "Combining RT-PCR-seq and RNA-seq to catalog all genic elements encoded in the human genome." Genome Res **22**(9): 1698-710.
Hu, J. and P. C. Ng (2012). "Predicting the effects of frameshifting indels." Genome Biol **13**(2): R9.
Huang, X. C., M. A. Quesada, et al. (1992). "DNA sequencing using capillary array electrophoresis." Anal Chem **64**(18): 2149-54.
Huangfu, D., A. Liu, et al. (2003). "Hedgehog signalling in the mouse requires intraflagellar transport proteins." Nature **426**(6962): 83-7.
Hui, C. C. and S. Angers (2011). "Gli proteins in development and disease." Annu Rev Cell Dev Biol **27**: 513-37.
Huntley, C. C. and R. E. Stevenson (1969). "Maternal phenylketonuria. Course of two pregnancies." Obstet Gynecol **34**(5): 694-700.
Imessaoudene, B., J. P. Bonnefont, et al. (2001). "MECP2 mutation in non-fatal, non-progressive encephalopathy in a male." J Med Genet **38**(3): 171-4.
Imhoff, O., V. Marion, et al. (2011). "Bardet-Biedl syndrome: a study of the renal and cardiovascular phenotypes in a French cohort." Clin J Am Soc Nephrol **6**(1): 22-9.
Inglis, P. N., K. A. Boroevich, et al. (2006). "Piecing together a ciliome." Trends Genet **22**(9): 491-500.
International Human Genome Sequencing, C. (2004). "Finishing the euchromatic sequence of the human genome." Nature **431**(7011): 931-45.
Iossifov, I., M. Ronemus, et al. (2012). "De novo gene disruptions in children on the autistic spectrum." Neuron **74**(2): 285-99.
Ishikawa, H. and W. F. Marshall (2011). "Ciliogenesis: building the cell's antenna." Nat Rev Mol Cell Biol **12**(4): 222-34.
Jacquemont, S., A. Curie, et al. (2011). "Epigenetic modification of the FMR1 gene in fragile X syndrome is associated with differential response to the mGluR5 antagonist AFQ056." Sci Transl Med **3**(64): 64ra1.
Jaglin, X. H., K. Poirier, et al. (2009). "Mutations in the beta-tubulin gene TUBB2B result in asymmetrical polymicrogyria." Nat Genet **41**(6): 746-52.
Jiang, J. and C. C. Hui (2008). "Hedgehog signaling in development and cancer." Dev Cell **15**(6): 801-12.
Jiang, J., Y. Jing, et al. (2013). "Translating dosage compensation to trisomy 21." Nature **500**(7462): 296-300.
Kajiwara, K., E. L. Berson, et al. (1994). "Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci." Science **264**(5165): 1604-8.
Kalscheuer, V. M., J. Tao, et al. (2003). "Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation." Am J Hum Genet **72**(6): 1401-11.
Kamiya, K., M. Kaneda, et al. (2004). "A nonsense mutation of the sodium channel gene SCN2A in a patient with intractable epilepsy and mental decline." J Neurosci **24**(11): 2690-8.
Karger, B. L. and A. Guttman (2009). "DNA sequencing by CE." Electrophoresis **30 Suppl 1**: S196-202.
Kasperaviciute, D., C. B. Catarino, et al. (2010). "Common genetic variation and susceptibility to partial epilepsies: a genome-wide association study." Brain **133**(Pt 7): 2136-47.

Katsanis, N. (2004). "The oligogenic properties of Bardet-Biedl syndrome." Hum Mol Genet **13 Spec No 1**: R65-71.
Katsanis, N., S. J. Ansley, et al. (2001). "Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder." Science **293**(5538): 2256-9.
Katsanis, N., P. L. Beales, et al. (2000). "Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome." Nat Genet **26**(1): 67-70.
Katsanis, N., E. R. Eichers, et al. (2002). "BBS4 is a minor contributor to Bardet-Biedl syndrome and may also participate in triallelic inheritance." Am J Hum Genet **71**(1): 22-9.
Katz, G. and E. Lazcano-Ponce (2008). "Intellectual disability: definition, etiological factors, classification, diagnosis, treatment and prognosis." Salud Publica Mex **50 Suppl 2**: s132-41.
Kaufman, L., M. Ayub, et al. (2010). "The genetic basis of non-syndromic intellectual disability: a review." J Neurodev Disord **2**(4): 182-209.
Keays, D. A., G. Tian, et al. (2007). "Mutations in alpha-tubulin cause abnormal neuronal migration in mice and lissencephaly in humans." Cell **128**(1): 45-57.
Kelley, W. N. (1968). "Hypoxanthine-guanine phosphoribosyltransferase deficiency in the Lesch-Nyhan syndrome and gout." Fed Proc **27**(4): 1047-52.
Kheradmand Kia, S., E. Verbeek, et al. (2012). "RTTN mutations link primary cilia function to organization of the human cerebral cortex." Am J Hum Genet **91**(3): 533-40.
Kim, J., Y. D. Chung, et al. (2003). "A TRPV family ion channel required for hearing in Drosophila." Nature **424**(6944): 81-4.
Kim, S. K., A. Shindo, et al. (2010). "Planar cell polarity acts through septins to control collective cell movement and ciliogenesis." Science **329**(5997): 1337-40.
King, B. H. (2009). Comprehensive textbook of psychiatry
Kitamura, K., M. Yanazawa, et al. (2002). "Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans." Nat Genet **32**(3): 359-69.
Koboldt, D. C., K. M. Steinberg, et al. (2013). "The next-generation sequencing revolution and its impact on genomics." Cell **155**(1): 27-38.
Kodera, H., K. Nakamura, et al. (2013). "De novo mutations in SLC35A2 encoding a UDP-galactose transporter cause early-onset epileptic encephalopathy." Hum Mutat **34**(12): 1708-14.
Kohlschutter, A. and F. Eichler (2011). "Childhood leukodystrophies: a clinical perspective." Expert Rev Neurother **11**(10): 1485-96.
Koscinski, I., E. Elinati, et al. (2011). "DPY19L2 deletion as a major cause of globozoospermia." Am J Hum Genet **88**(3): 344-50.
Kretz, K. A., J. K. Darby, et al. (1989). "Characterization of EcoRI mutation in fucosidosis patients: a stop codon in the open reading frame." J Mol Neurosci **1**(3): 177-80.
Kulaga, H. M., C. C. Leitch, et al. (2004). "Loss of BBS proteins causes anosmia in humans and defects in olfactory cilia structure and function in the mouse." Nat Genet **36**(9): 994-8.
Kumamoto, N., Y. Gu, et al. (2012). "A role for primary cilia in glutamatergic synaptic integration of adult-born neurons." Nat Neurosci **15**(3): 399-405, S1.
Kumar, P., S. Henikoff, et al. (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-81.
Kuss, A. W., M. Garshasbi, et al. (2011). "Autosomal recessive mental retardation: homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots." Hum Genet **129**(2): 141-8.
Kyrpides, N. C. (1999). "Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide." Bioinformatics **15**(9): 773-4.
Kyttala, M., J. Tallila, et al. (2006). "MKS1, encoding a component of the flagellar apparatus basal body proteome, is mutated in Meckel syndrome." Nat Genet **38**(2): 155-7.
Labauge, P. and O. Boespflug-Tanguy (2010). "Maladies démyélinisantes d'origine génétique." La Presse Médicale **39**(3): 363-370.
Laing, N. G. (2012). "Genetics of neuromuscular disorders." Crit Rev Clin Lab Sci **49**(2): 33-48.
Lancaster, M. A., J. Schroth, et al. (2011). "Subcellular spatial regulation of canonical Wnt signalling at the primary cilium." Nat Cell Biol **13**(6): 700-7.
Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Landsman, L., A. Parent, et al. (2011). "Elevated Hedgehog/Gli signaling causes beta-cell dedifferentiation in mice." Proc Natl Acad Sci U S A **108**(41): 17010-5.

Laurence, J. Z. and R. C. Moon (1995). "Four cases of "retinitis pigmentosa" occurring in the same family, and accompanied by general imperfections of development. 1866." Obes Res **3**(4): 400-3.
Laurier, V., C. Stoetzel, et al. (2006). "Pitfalls of homozygosity mapping: an extended consanguineous Bardet-Biedl syndrome family with two mutant genes (BBS2, BBS10), three mutations, but no triallelism." Eur J Hum Genet **14**(11): 1195-203.
Leblond, C. S., J. Heinrich, et al. (2012). "Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders." PLoS Genet **8**(2): e1002521.
Lederer, D., B. Grisart, et al. (2012). "Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome." Am J Hum Genet **90**(1): 119-24.
Lee, J. H., M. Huynh, et al. (2012). "De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly." Nat Genet **44**(8): 941-5.
Leegwater, P. A., G. Vermeulen, et al. (2001). "Subunits of the translation initiation factor eIF2B are mutant in leukoencephalopathy with vanishing white matter." Nat Genet **29**(4): 383-8.
Leen, W. G., J. Klepper, et al. (2010). "Glucose transporter-1 deficiency syndrome: the expanding clinical and genetic spectrum of a treatable disorder." Brain **133**(Pt 3): 655-70.
Leitch, C. C., N. A. Zaghloul, et al. (2008). "Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome." Nat Genet **40**(4): 443-8.
Lejeune, J., R. Turpin, et al. (1959). "[Chromosomic diagnosis of mongolism]." Arch Fr Pediatr **16**: 962-3.
Leonard, H. and X. Wen (2002). "The epidemiology of mental retardation: challenges and opportunities in the new millennium." Ment Retard Dev Disabil Res Rev **8**(3): 117-34.
Levy, D., M. Ronemus, et al. (2011). "Rare de novo and transmitted copy-number variation in autistic spectrum disorders." Neuron **70**(5): 886-97.
Li, J. B., J. M. Gerdes, et al. (2004). "Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene." Cell **117**(4): 541-52.
Li, T., W. K. Snyder, et al. (1996). "Transgenic mice carrying the dominant rhodopsin mutation P347S: evidence for defective vectorial transport of rhodopsin to the outer segments." Proc Natl Acad Sci U S A **93**(24): 14176-81.
Lin, S. C. and D. E. Bergles (2004). "Synaptic signaling between neurons and glia." Glia **47**(3): 290-8.
Linehan, C., M. P. Kerr, et al. (2010). "Examining the prevalence of epilepsy and delivery of epilepsy care in Ireland." Epilepsia **51**(5): 845-52.
Linnankivi, T., P. Tienari, et al. (2006). "18q deletions: clinical, molecular, and brain MRI findings of 14 individuals." Am J Med Genet A **140**(4): 331-9.
Liolios, K., I. M. Chen, et al. (2010). "The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata." Nucleic Acids Res **38**(Database issue): D346-54.
Liolios, K., N. Tavernarakis, et al. (2006). "The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide." Nucleic Acids Res **34**(Database issue): D332-4.
Lionel, A. C., K. Tammimies, et al. (2014). "Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes." Hum Mol Genet.
Liu, A., B. Wang, et al. (2005). "Mouse intraflagellar transport proteins regulate both the activator and repressor functions of Gli transcription factors." Development **132**(13): 3103-11.
Liu, Q., J. Zuo, et al. (2004). "The retinitis pigmentosa 1 protein is a photoreceptor microtubule-associated protein." J Neurosci **24**(29): 6427-36.
Lo Nigro, C., C. S. Chong, et al. (1997). "Point mutations and an intragenic deletion in LIS1, the lissencephaly causative gene in isolated lissencephaly sequence and Miller-Dieker syndrome." Hum Mol Genet **6**(2): 157-64.
Lohmueller, K. E., T. Sparso, et al. (2013). "Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes." Am J Hum Genet **93**(6): 1072-86.
Lubs, H. A., R. E. Stevenson, et al. (2012). "Fragile X and X-linked intellectual disability: four decades of discovery." Am J Hum Genet **90**(4): 579-90.
Maguire, A. M., F. Simonelli, et al. (2008). "Safety and efficacy of gene transfer for Leber's congenital amaurosis." N Engl J Med **358**(21): 2240-8.
Mak, C. M., H. C. Lee, et al. (2013). "Inborn errors of metabolism and expanded newborn screening: review and update." Crit Rev Clin Lab Sci **50**(6): 142-62.
Margulies, M., M. Egholm, et al. (2005). "Genome sequencing in microfabricated high-density picolitre reactors." Nature **437**(7057): 376-80.

Marion, V., F. Stutzmann, et al. (2012). "Exome sequencing identifies mutations in LZTFL1, a BBSome and smoothened trafficking regulator, in a family with Bardet--Biedl syndrome with situs inversus and insertional polydactyly." J Med Genet **49**(5): 317-21.
Marshall, C. R., A. Noor, et al. (2008). "Structural variation of chromosomes in autism spectrum disorder." Am J Hum Genet **82**(2): 477-88.
Marshall, J. D., E. G. Hinman, et al. (2007). "Spectrum of ALMS1 variants and evaluation of genotype-phenotype correlations in Alstrom syndrome." Hum Mutat **28**(11): 1114-23.
Marshall, J. D., P. Maffei, et al. (2013). "Clinical utility gene card for: Alstrom Syndrome - update 2013." Eur J Hum Genet **21**(11).
Masyuk, A. I., T. V. Masyuk, et al. (2008). "Cholangiocyte primary cilia in liver health and disease." Dev Dyn **237**(8): 2007-12.
Mattei, M. G., P. M. Alliel, et al. (1986). "The gene encoding for the major brain proteolipid (PLP) maps on the q-22 band of the human X chromosome." Hum Genet **72**(4): 352-3.
Maulik, P. K., M. N. Mascarenhas, et al. (2011). "Prevalence of intellectual disability: a meta-analysis of population-based studies." Res Dev Disabil **32**(2): 419-36.
McCulloch, S. D. and T. A. Kunkel (2008). "The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases." Cell Res **18**(1): 148-61.
Mei, R., P. C. Galipeau, et al. (2000). "Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays." Genome Res **10**(8): 1126-37.
Melom, J. E. and J. T. Littleton (2011). "Synapse development in health and disease." Curr Opin Genet Dev **21**(3): 256-61.
Menichella, D. M., D. A. Goodenough, et al. (2003). "Connexins are critical for normal myelination in the CNS." J Neurosci **23**(13): 5963-73.
Mertes, F., A. Elsharawy, et al. (2011). "Targeted enrichment of genomic DNA regions for next-generation sequencing." Brief Funct Genomics **10**(6): 374-86.
Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nat Rev Genet **11**(1): 31-46.
Michaelson, J. J., Y. Shi, et al. (2012). "Whole-genome sequencing in autism identifies hot spots for de novo germline mutation." Cell **151**(7): 1431-42.
Miller, A. D., R. J. Eckner, et al. (1984). "Expression of a retrovirus encoding human HPRT in mice." Science **225**(4662): 630-2.
Mockel, A., Y. Perdomo, et al. (2011). "Retinal dystrophy in Bardet-Biedl syndrome and related syndromic ciliopathies." Prog Retin Eye Res **30**(4): 258-74.
Molinari, F., M. Rio, et al. (2002). "Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation." Science **298**(5599): 1779-81.
Morrow, E. M. (2010). "Genomic copy number variation in disorders of cognitive development." J Am Acad Child Adolesc Psychiatry **49**(11): 1091-104.
Muller, J., C. Stoetzel, et al. (2010). "Identification of 28 novel mutations in the Bardet-Biedl syndrome genes: the burden of private mutations in an extensively heterogeneous disease." Hum Genet **127**(5): 583-93.
Musante, L. and H. H. Ropers (2014). "Genetics of recessive cognitive disorders." Trends Genet **30**(1): 32-9.
Musselman, K. E., C. T. Stoyanov, et al. (2014). "Prevalence of ataxia in children: a systematic review." Neurology **82**(1): 80-9.
Mykytyn, K., T. Braun, et al. (2001). "Identification of the gene that, when mutated, causes the human obesity syndrome BBS4." Nat Genet **28**(2): 188-91.
Mykytyn, K., D. Y. Nishimura, et al. (2003). "Evaluation of complex inheritance involving the most common Bardet-Biedl syndrome locus (BBS1)." Am J Hum Genet **72**(2): 429-37.
Mykytyn, K., D. Y. Nishimura, et al. (2002). "Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome." Nat Genet **31**(4): 435-8.
Nachury, M. V., A. V. Loktev, et al. (2007). "A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis." Cell **129**(6): 1201-13.
Najmabadi, H., H. Hu, et al. (2011). "Deep sequencing reveals 50 novel genes for recessive cognitive disorders." Nature **478**(7367): 57-63.
Nauli, S. M., F. J. Alenghat, et al. (2003). "Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells." Nat Genet **33**(2): 129-37.
Neale, B. M., Y. Kou, et al. (2012). "Patterns and rates of exonic de novo mutations in autism spectrum disorders." Nature **485**(7397): 242-5.

Newschaffer, C. J., L. A. Croen, et al. (2007). "The epidemiology of autism spectrum disorders." Annu Rev Public Health **28**: 235-58.
Ng, B. G., K. J. Buckingham, et al. (2013). "Mosaicism of the UDP-galactose transporter SLC35A2 causes a congenital disorder of glycosylation." Am J Hum Genet **92**(4): 632-6.
Ng, P. C. and S. Henikoff (2003). "SIFT: Predicting amino acid changes that affect protein function." Nucleic Acids Res **31**(13): 3812-4.
Ng, S. B., K. J. Buckingham, et al. (2010). "Exome sequencing identifies the cause of a mendelian disorder." Nat Genet **42**(1): 30-5.
Ng, S. B., E. H. Turner, et al. (2009). "Targeted capture and massively parallel sequencing of 12 human exomes." Nature **461**(7261): 272-6.
Nicholson, P., H. Yepiskoposyan, et al. (2010). "Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors." Cell Mol Life Sci **67**(5): 677-700.
Nishimura, D. Y., M. Fath, et al. (2004). "Bbs2-null mice have neurosensory deficits, a defect in social dominance, and retinopathy associated with mislocalization of rhodopsin." Proc Natl Acad Sci U S A **101**(47): 16588-93.
Nishimura, D. Y., C. C. Searby, et al. (2001). "Positional cloning of a novel gene on chromosome 16q causing Bardet-Biedl syndrome (BBS2)." Hum Mol Genet **10**(8): 865-74.
Nishimura, D. Y., R. E. Swiderski, et al. (2005). "Comparative genomics and gene expression analysis identifies BBS9, a new Bardet-Biedl syndrome gene." Am J Hum Genet **77**(6): 1021-33.
Niu, Y., B. Shen, et al. (2014). "Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos." Cell **156**(4): 836-43.
Nonaka, S., Y. Tanaka, et al. (1998). "Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein." Cell **95**(6): 829-37.
O'Roak, B. J., P. Deriziotis, et al. (2011). "Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations." Nat Genet **43**(6): 585-9.
O'Roak, B. J., L. Vives, et al. (2012a). "Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders." Science **338**(6114): 1619-22.
O'Roak, B. J., L. Vives, et al. (2012b). "Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations." Nature **485**(7397): 246-50.
Oeseburg, B., G. J. Dijkstra, et al. (2011). "Prevalence of chronic health conditions in children with intellectual disability: a systematic literature review." Intellect Dev Disabil **49**(2): 59-85.
Okou, D. T., K. M. Steinberg, et al. (2007). "Microarray-based genomic selection for high-throughput resequencing." Nat Methods **4**(11): 907-9.
Otto, E. A., T. W. Hurd, et al. (2010). "Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy." Nat Genet **42**(10): 840-50.
Ou, G., O. E. Blacque, et al. (2005). "Functional coordination of intraflagellar transport motors." Nature **436**(7050): 583-7.
Papavasiliou, A. S., H. Bazigou, et al. (2000). "Neurometabolic testing in developmental delay." J Child Neurol **15**(9): 620-2.
Park, T. J., S. L. Haigo, et al. (2006). "Ciliogenesis defects in embryos lacking inturned or fuzzy function are associated with failure of planar cell polarity and Hedgehog signaling." Nat Genet **38**(3): 303-11.
Pei, B., C. Sisu, et al. (2012). "The GENCODE pseudogene resource." Genome Biol **13**(9): R51.
Perche, O., F. Laumonnier, et al. (2010). "[Autism, genetics and synaptic function alterations]." Pathol Biol (Paris) **58**(5): 381-6.
Pereiro, I., B. E. Hoskins, et al. (2011). "Arrayed primer extension technology simplifies mutation detection in Bardet-Biedl and Alstrom syndrome." Eur J Hum Genet **19**(4): 485-8.
Petterson, B., J. Bourke, et al. (2007). "Co-occurrence of birth defects and intellectual disability." Paediatr Perinat Epidemiol **21**(1): 65-75.
Piao, X., R. S. Hill, et al. (2004). "G protein-coupled receptor-dependent development of human frontal cortex." Science **303**(5666): 2033-6.
Pilz, D. T., J. Kuc, et al. (1999). "Subcortical band heterotopia in rare affected males can be caused by missense mutations in DCX (XLIS) or LIS1." Hum Mol Genet **8**(9): 1757-60.
Pinard, J. M., J. Motte, et al. (1994). "Subcortical laminar heterotopia and lissencephaly in two families: a single X linked dominant gene." J Neurol Neurosurg Psychiatry **57**(8): 914-20.
Pinto, D., A. T. Pagnamenta, et al. (2010). "Functional impact of global rare copy number variation in autism spectrum disorders." Nature **466**(7304): 368-72.

Piton, A., C. Redin, et al. (2013). "XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing." Am J Hum Genet **93**(2): 368-83.
Poirier, K., D. Lacombe, et al. (2006). "Screening of ARX in mental retardation families: Consequences for the strategy of molecular diagnosis." Neurogenetics **7**(1): 39-46.
Pollard, K. S., M. J. Hubisz, et al. (2010). "Detection of nonneutral substitution rates on mammalian phylogenies." Genome Res **20**(1): 110-21.
Porreca, G. J., K. Zhang, et al. (2007). "Multiplex amplification of large sets of human exons." Nat Methods **4**(11): 931-6.
Quet, F., F. Rafael, et al. (2011). "Investigating epilepsy in Africa: 10 years of data collection using a standardized questionnaire in 2,269 peoples with epilepsy." Epilepsia **52**(10): 1868-76.
Rademakers, R., M. Baker, et al. (2012). "Mutations in the colony stimulating factor 1 receptor (CSF1R) gene cause hereditary diffuse leukoencephalopathy with spheroids." Nat Genet **44**(2): 200-5.
Radford, J. P. (1991). "Sterilization versus segregation: Control of the 'feebleminded', 1900-1938." Social Science and Medicine **33**: 449-458.
Radmanesh, F., A. O. Caglayan, et al. (2013). "Mutations in LAMB1 cause cobblestone brain malformation without muscular or ocular abnormalities." Am J Hum Genet **92**(3): 468-74.
Rahmouni, K., M. A. Fath, et al. (2008). "Leptin resistance contributes to obesity and hypertension in mouse models of Bardet-Biedl syndrome." J Clin Invest **118**(4): 1458-67.
Rainer, J. D. (1971). "Chronological parameters, twin studies, and mental diseases." Acta Genet Med Gemellol (Roma) **20**(4): 359-72.
Rauch, A., J. Hoyer, et al. (2006). "Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation." Am J Med Genet A **140**(19): 2063-74.
Rauch, A., D. Wieczorek, et al. (2012). "Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study." Lancet **380**(9854): 1674-82.
Reese, M. G., F. H. Eeckman, et al. (1997). "Improved splice site detection in Genie." J Comput Biol **4**(3): 311-23.
Rehman, A. U., R. J. Morell, et al. (2010). "Targeted capture and next-generation sequencing identifies C9orf75, encoding taperin, as the mutated gene in nonsyndromic deafness DFNB79." Am J Hum Genet **86**(3): 378-88.
Riviere, J. B., G. M. Mirzaa, et al. (2012). "De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes." Nat Genet **44**(8): 934-40.
Roach, J. C., G. Glusman, et al. (2010). "Analysis of genetic inheritance in a family quartet by whole-genome sequencing." Science **328**(5978): 636-9.
Robin, N. H., C. J. Taylor, et al. (2006). "Polymicrogyria and deletion 22q11.2 syndrome: window to the etiology of a common cortical malformation." Am J Med Genet A **140**(22): 2416-25.
Roessler, E., E. Belloni, et al. (1996). "Mutations in the human Sonic Hedgehog gene cause holoprosencephaly." Nat Genet **14**(3): 357-60.
Ronan, J. L., W. Wu, et al. (2013). "From neural development to cognition: unexpected roles for chromatin." Nat Rev Genet **14**(5): 347-59.
Ropers, H. H. (2010). "Genetics of early onset cognitive impairment." Annu Rev Genomics Hum Genet **11**: 161-87.
Rosenbaum, J. L. and G. B. Witman (2002). "Intraflagellar transport." Nat Rev Mol Cell Biol **3**(11): 813-25.
Ross, A. J., H. May-Simera, et al. (2005). "Disruption of Bardet-Biedl syndrome ciliary proteins perturbs planar cell polarity in vertebrates." Nat Genet **37**(10): 1135-40.
Ruiz-Martinez, M. C., J. Berka, et al. (1993). "DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection." Anal Chem **65**(20): 2851-8.
Saccone, V., M. Palmieri, et al. (2008). "Mutations that impair interaction properties of TRIM32 associated with limb-girdle muscular dystrophy 2H." Hum Mutat **29**(2): 240-7.
Saitsu, H., M. Kato, et al. (2008). "De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy." Nat Genet **40**(6): 782-8.
Saitsu, H., H. Osaka, et al. (2011). "Mutations in POLR3A and POLR3B encoding RNA Polymerase III subunits cause an autosomal-recessive hypomyelinating leukoencephalopathy." Am J Hum Genet **89**(5): 644-51.

Salinas, P. C. (2012). "Wnt signaling in the vertebrate central nervous system: from axon guidance to synaptic function." Cold Spring Harb Perspect Biol **4**(2).
Salsano, E., O. Gambini, et al. (2012). "Effectiveness of valproate for the treatment of manic-like behavior in X-linked adrenoleukodystrophy." Neurol Sci **33**(5): 1197-9.
Sanders, S. J., A. G. Ercan-Sencicek, et al. (2011). "Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism." Neuron **70**(5): 863-85.
Sanders, S. J., M. T. Murtha, et al. (2012). "De novo mutations revealed by whole-exome sequencing are strongly associated with autism." Nature **485**(7397): 237-41.
Saunders, C. J., N. A. Miller, et al. (2012). "Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units." Sci Transl Med **4**(154): 154ra135.
Sawaishi, Y. (2009). "Review of Alexander disease: beyond the classical concept of leukodystrophy." Brain Dev **31**(7): 493-8.
Sayer, J. A., E. A. Otto, et al. (2006). "The centrosomal protein nephrocystin-6 is mutated in Joubert syndrome and activates transcription factor ATF4." Nat Genet **38**(6): 674-81.
Schaaf, C. P., A. Sabo, et al. (2011). "Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders." Hum Mol Genet **20**(17): 3366-75.
Schaaf, C. P. and H. Y. Zoghbi (2011). "Solving the autism puzzle a few pieces at a time." Neuron **70**(5): 806-8.
Schaefer, E., J. Lauer, et al. (2013). "Mesoaxial polydactyly is a major feature in Bardet-Biedl syndrome patients with LZTFL1 (BBS17) mutations." Clin Genet.
Schaefer, E., A. Zaloszyc, et al. (2011). "Mutations in SDCCAG8/NPHP10 Cause Bardet-Biedl Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly." Mol Syndromol **1**(6): 273-281.
Scheidecker, S., C. Etard, et al. (2014). "Exome sequencing of Bardet-Biedl syndrome patient identifies a null mutation in the BBSome subunit BBIP1 (BBS18)." J Med Genet **51**(2): 132-6.
Schell-Apacik, C. C., B. Ertl-Wagner, et al. (2009). "Maternally inherited heterozygous sequence change in the sonic hedgehog gene in a male patient with bilateral closed-lip schizencephaly and partial absence of the corpus callosum." Am J Med Genet A **149A**(7): 1592-4.
Schiffmann, R. and M. S. van der Knaap (2004). "The latest on leukodystrophies." Curr Opin Neurol **17**(2): 187-92.
Schwarz, J. M., C. Rodelsperger, et al. (2010). "MutationTaster evaluates disease-causing potential of sequence alterations." Nat Methods **7**(8): 575-6.
Sebat, J., B. Lakshmi, et al. (2007). "Strong association of de novo copy number mutations with autism." Science **316**(5823): 445-449.
Seo, S., L. M. Baye, et al. (2010). "BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly." Proc Natl Acad Sci U S A **107**(4): 1488-93.
Sertie, A. L., V. Sossi, et al. (2000). "Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome)." Hum Mol Genet **9**(13): 2051-8.
Shaheen, R., H. E. Shamseldin, et al. (2014). "Mutations in CSPP1, encoding a core centrosomal protein, cause a range of ciliopathy phenotypes in humans." Am J Hum Genet **94**(1): 73-9.
Sheen, V. L., V. S. Ganesh, et al. (2004). "Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex." Nat Genet **36**(1): 69-76.
Shen, Y., K. A. Dies, et al. (2010). "Clinical genetic testing for patients with autism spectrum disorders." Pediatrics **125**(4): e727-35.
Shendure, J., G. J. Porreca, et al. (2005). "Accurate multiplex polony sequencing of an evolved bacterial genome." Science **309**(5741): 1728-32.
Shimojima, K., T. Mano, et al. (2012). "Pelizaeus-Merzbacher disease caused by a duplication-inverted triplication-duplication in chromosomal segments including the PLP1 region." Eur J Med Genet **55**(6-7): 400-3.
Siepel, A., G. Bejerano, et al. (2005). "Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes." Genome Res **15**(8): 1034-50.
Simon, D., B. Laloo, et al. (2010). "A mutation in the 3'-UTR of the HDAC6 gene abolishing the post-transcriptional regulation mediated by hsa-miR-433 is linked to a new form of dominant X-linked chondrodysplasia." Hum Mol Genet **19**(10): 2015-27.
Simons, M., J. Gloy, et al. (2005). "Inversin, the gene product mutated in nephronophthisis type II, functions as a molecular switch between Wnt signaling pathways." Nat Genet **37**(5): 537-43.

Simpson, D. A., G. R. Clark, et al. (2011). "Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa." J Med Genet **48**(3): 145-51.
Singla, V. and J. F. Reiter (2006). "The primary cilium as the cell's antenna: signaling at a sensory organelle." Science **313**(5787): 629-33.
Slavotinek, A. M., E. M. Stone, et al. (2000). "Mutations in MKKS cause Bardet-Biedl syndrome." Nat Genet **26**(1): 15-6.
Sleigh, M. A. (1974). "Cilia and Flagella." **London: Academic Press**.
Solis-Cohen, S., Weiss, E. (1925). "Dystrophia adiposogenitalis with atypical retinitis pigmentosa and mental deficiency." Am. J. Med. Sci. **169**: 489-505.
Spassky, N., Y. G. Han, et al. (2008). "Primary cilia are required for cerebellar development and Shh-dependent expansion of progenitor pool." Dev Biol **317**(1): 246-59.
Stenson, P. D., M. Mort, et al. (2014). "The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine." Hum Genet **133**(1): 1-9.
Stevenson, R. E., K. R. Holden, et al. (2012). "Seizures and X-linked intellectual disability." Eur J Med Genet **55**(5): 307-12.
Stoetzel, C., V. Laurier, et al. (2006). "BBS10 encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus." Nat Genet **38**(5): 521-4.
Stoetzel, C., J. Muller, et al. (2007). "Identification of a novel BBS gene (BBS12) highlights the major role of a vertebrate-specific branch of chaperonin-related proteins in Bardet-Biedl syndrome." Am J Hum Genet **80**(1): 1-11.
Stone, D. L., A. Slavotinek, et al. (2000). "Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome." Nat Genet **25**(1): 79-82.
Stromme, P., M. E. Mangelsdorf, et al. (2002). "Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy." Nat Genet **30**(4): 441-5.
Takano, K., D. Liu, et al. (2012). "An X-linked channelopathy with cardiomegaly due to a CLIC2 mutation enhancing ryanodine receptor channel activity." Hum Mol Genet **21**(20): 4497-507.
Tarjan, G., H. F. Dingman, et al. (1960). "Statistical expectations of selected handicaps in the mentally retarded." Am J Ment Defic **65**: 335-41.
Tarpey, P. S., F. L. Raymond, et al. (2007). "Mutations in CUL4B, which encodes a ubiquitin E3 ligase subunit, cause an X-linked mental retardation syndrome associated with aggressive outbursts, seizures, relative macrocephaly, central obesity, hypogonadism, pes cavus, and tremor." Am J Hum Genet **80**(2): 345-52.
Tarpey, P. S., R. Smith, et al. (2009). "A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation." Nat Genet **41**(5): 535-43.
Tavtigian, S. V., M. S. Greenblatt, et al. (2008). "In silico analysis of missense substitutions using sequence-alignment based methods." Hum Mutat **29**(11): 1327-36.
Tewhey, R., J. B. Warner, et al. (2009). "Microdroplet-based PCR enrichment for large-scale targeted sequencing." Nat Biotechnol **27**(11): 1025-31.
Timal, S., A. Hoischen, et al. (2012). "Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing." Hum Mol Genet **21**(19): 4151-61.
Tung, J. Y., C. B. Do, et al. (2011). "Efficient replication of over 180 genetic associations with self-reported medical data." PLoS One **6**(8): e23473.
Tuz, K., R. Bachmann-Gagescu, et al. (2014). "Mutations in CSPP1 cause primary cilia abnormalities and Joubert syndrome with or without Jeune asphyxiating thoracic dystrophy." Am J Hum Genet **94**(1): 62-72.
Twigg, S. R., S. L. Versnel, et al. (2009). "Frontorhiny, a distinctive presentation of frontonasal dysplasia caused by recessive mutations in the ALX3 homeobox gene." Am J Hum Genet **84**(5): 698-705.
Uhlenberg, B., M. Schuelke, et al. (2004). "Mutations in the gene encoding gap junction protein alpha 12 (connexin 46.6) cause Pelizaeus-Merzbacher-like disease." Am J Hum Genet **75**(2): 251-60.
Utari, A., W. Chonchaiya, et al. (2010). "Side effects of minocycline treatment in patients with fragile X syndrome and exploration of outcome measures." Am J Intellect Dev Disabil **115**(5): 433-43.
Valente, E. M., J. L. Silhavy, et al. (2006). "Mutations in CEP290, which encodes a centrosomal protein, cause pleiotropic forms of Joubert syndrome." Nat Genet **38**(6): 623-5.

van der Knaap, M. S., I. Boor, et al. (2012). "Megalencephalic leukoencephalopathy with subcortical cysts: chronic white matter oedema due to a defect in brain ion and water homoeostasis." Lancet Neurol **11**(11): 973-85.
van der Knaap, M. S., T. Linnankivi, et al. (2007). "Hypomyelination with atrophy of the basal ganglia and cerebellum: follow-up and pathology." Neurology **69**(2): 166-71.
van Karnebeek, C. D., M. C. Jansweijer, et al. (2005). "Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness." Eur J Hum Genet **13**(1): 6-25.
van Karnebeek, C. D. and S. Stockler (2012). "Treatable inborn errors of metabolism causing intellectual disability: a systematic literature review." Mol Genet Metab **105**(3): 368-81.
Vaurs-Barriere, C., M. N. Bonnet-Dupeyron, et al. (2006). "Golli-MBP copy number analysis by FISH, QMPSF and MAPH in 195 patients with hypomyelinating leukodystrophies." Ann Hum Genet **70**(Pt 1): 66-77.
Veeramah, K. R., J. E. O'Brien, et al. (2012). "De novo pathogenic SCN8A mutation identified by whole-genome sequencing of a family quartet affected by infantile epileptic encephalopathy and SUDEP." Am J Hum Genet **90**(3): 502-10.
Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.
Verpelli, C., C. Montani, et al. (2013). "Mutations of the synapse genes and intellectual disability syndromes." Eur J Pharmacol **719**(1-3): 112-6.
Vissers, L. E., J. de Ligt, et al. (2010). "A de novo paradigm for mental retardation." Nat Genet **42**(12): 1109-12.
Volpi, L., G. Roversi, et al. (2010). "Targeted next-generation sequencing appoints c16orf57 as clericuzio-type poikiloderma with neutropenia gene." Am J Hum Genet **86**(1): 72-6.
von Bohlen Und Halbach, O. (2010). "Dendritic spine abnormalities in mental retardation." Cell Tissue Res **342**(3): 317-23.
Watson, P., G. Black, et al. (2001). "Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein." J Med Genet **38**(4): 224-8.
Weese-Mayer, D. E., S. P. Lieske, et al. (2006). "Autonomic nervous system dysregulation: breathing and heart rate perturbation during wakefulness in young girls with Rett syndrome." Pediatr Res **60**(4): 443-9.
Wei, X. and K. K. Sulik (1993). "Pathogenesis of craniofacial and body wall malformations induced by ochratoxin A in mice." Am J Med Genet **47**(6): 862-71.
Weijerman, M. E. and J. P. de Winter (2010). "Clinical practice. The care of children with Down syndrome." Eur J Pediatr **169**(12): 1445-52.
Weinstock, G. M. (2007). "ENCODE: more genomic empowerment." Genome Res **17**(6): 667-8.
Wheatley, D. N., A. M. Wang, et al. (1996). "Expression of primary cilia in mammalian cells." Cell Biol Int **20**(1): 73-81.
Wheeler, D. A., M. Srinivasan, et al. (2008). "The complete genome of an individual by massively parallel DNA sequencing." Nature **452**(7189): 872-6.
Wicks, P., T. E. Vaughan, et al. (2011). "Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm." Nat Biotechnol **29**(5): 411-4.
Willard, H. F. and J. R. Riordan (1985). "Assignment of the gene for myelin proteolipid protein to the X chromosome: implications for X-linked myelin disorders." Science **230**(4728): 940-2.
Williamson, R. E. (2006). "This Month in the Journal." Am J Hum Genet **79**(4): i–ii.
Wong, S. Y., A. D. Seol, et al. (2009). "Primary cilia can both mediate and suppress Hedgehog pathway-dependent tumorigenesis." Nat Med **15**(9): 1055-61.
Woo, S. L., A. S. Lidsky, et al. (1983). "Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria." Nature **306**(5939): 151-5.
Xu, H., W. T. Poh, et al. (2011). "SgD-CNV, a database for common and rare copy number variants in three Asian populations." Hum Mutat **32**(12): 1341-9.
Yeo, G. and C. B. Burge (2004). "Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals." J Comput Biol **11**(2-3): 377-94.
Yost, H. J. (2003). "Left-right asymmetry: nodal cilia make and catch a wave." Curr Biol **13**(20): R808-9.
Zaghloul, N. A., Y. Liu, et al. (2010). "Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome." Proc Natl Acad Sci U S A **107**(23): 10602-7.

Zhang, Q., S. Seo, et al. (2012). "BBS proteins interact genetically with the IFT pathway to influence SHH-related phenotypes." Hum Mol Genet **21**(9): 1945-53.
Zimmerman, K. W. (1898). "Beitrage zur Kenntniss einiger Drusen und epithelien." Arch. Mikrosk. Anat **52**(52): 552–706.

Claire Redin

NGS-based approaches for the diagnosis of intellectual disability and other genetically heterogeneous developmental disorders

Résumé

Certaines maladies héréditaires monogéniques sont caractérisées par une grande hétérogénéité génétique. Chez des individus présentant un phénotype clinique similaire, les mutations causales peuvent être retrouvées dans un des gènes parmi un sous-ensemble décrits comme impliqués dans la maladie. Cette hétérogénéité génétique limite considérablement les offres diagnostiques pour les patients, et une majorité reste sans diagnostic moléculaire. Nous avons développé une approche diagnostique alternative par séquençage à haut débit ciblé (ciblant spécifiquement les régions codantes des gènes d'intérêt par capture d'exons), au travers de trois pathologies génétiquement hétérogènes: le syndrome de Bardet-Biedl (19 gènes décrits), les leucodystrophies (50 gènes), et la déficience intellectuelle (>400 gènes). Au vu de son efficacité dans le syndrome de Bardet-Biedl et la déficience intellectuelle (80% et 25% de mutations détectées respectivement, soit des taux nettement supérieurs à ceux des méthodes précédentes), elle est depuis appliquée en routine diagnostique. Au-delà du diagnostic, cette approche permet de manière non biaisée de revoir la contribution de chacun des gènes dans la pathologie et donc d'identifier les gènes récurrents, et d'établir de nouvelles corrélations génotype/phénotype.

Summary

Some monogenic disorders are characterized by a vast genetic heterogeneity. In individuals with similar clinical phenotype, causative mutations can be found in one gene from a subset described as implicated in the disease. Such genetic heterogeneity limits considerably the diagnostic offer for the patients, and a majority is left without molecular diagnosis. We developed an alternative diagnostic approach by targeted high throughput sequencing (specific to the coding regions of genes of interest by a technique of exon capture) through three genetically heterogeneous disorders: Bardet-Biedl syndrome (19 genes reported), leukodystrophies (50 genes), and intellectual disability (>400 genes). In light of its efficiency, this approach has since been implemented in diagnostic routine for Bardet-Biedl syndrome and intellectual disability (80% and 25% of diagnostic yields respectively, significantly higher than those of previous methods). Beyond diagnosis, this approach allows unbiased means to assess the contribution of each gene in the disease and highlight recurrent genes, and establish new correlations genotype to phenotype, overall providing much insight in the genetics of a particular disease.