Université de Montréal

Exploration génomique de la déficience intellectuelle

par
José-Mario Capo-chichi

Département de biochimie
Faculté de médecine

Thèse présentée à la Faculté des Études Supérieures
en vue de l'obtention du grade de *philosophiæ doctor* (Ph.D.)
en Biochimie
option Génomique Humaine

Août, 2014

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

Exploration génomique de la déficience intellectuelle

présentée par :

José-Mario Capo-chichi

a été évaluée par un jury composé des personnes suivantes :

Dr Muriel Aubry, président-rapporteur

Dr Jacques L. Michaud, directeur de recherche

Dr Mark E. Samuels, co-directeur de recherche

Dr Thierry Alquier, membre du jury

Dr Yannis Trakadis, examinateur externe

Dr Jacques Thibodeau, représentant du doyen

RÉSUMÉ

La déficience intellectuelle (DI) définit un groupe de conditions génétiquement hétérogènes caractérisées par l'apparition de troubles cognitifs précoces chez l'enfant. Elle affecte 1-3% de la population dans les pays industrialisés. La prévalence de la DI est beaucoup plus élevée ailleurs dans le monde, en raison de facteurs sociodémographiques comme le manque de ressources dans le système de santé, la pauvreté et la consanguinité. Des facteurs non-génétiques sont mis en cause dans l'étiologie de la DI ; on estime qu'environ 25% des cas de DI sont d'origine génétique. Traditionnellement, les bases moléculaires de la DI ont été investiguées par des analyses cytogénétiques, les approches de cartographie génétique et le séquençage de gènes candidats ; ces techniques de génétiques classiques sont encore mises à rude épreuve dans l'analyse de maladies complexes comme la DI.

La DI liée à l'X a été particulièrement étudiée, avec plus d'une centaine de gènes identifiés uniquement sur le chromosome X. Des mutations hétérozygotes composites sont mises en évidence dans la DI autosomique, dans le contexte d'unions non-consanguines. L'occurrence de ce type de mutations est rare, chez des individus non-apparentés, de sorte que les mutations dominantes *de novo* sont plus courantes. Des mutations homozygotes sont attendues dans les populations consanguines ou marquées par un effet fondateur. En fait, les bases moléculaires de la DI autosomique ont été presqu'exclusivement étudiées dans le contexte de populations avec des forts taux de consanguinité. L'origine de la DI demeure encore inconnue dans environ 60 % des cas diagnostiqués. En l'absence de facteurs environnementaux associés à la DI chez ces individus, il est possible d'envisager que des facteurs génétiques non identifiés entrent en jeu dans ces cas de DI inexpliqués.

Dans ce projet de recherche, nous voulions explorer l'origine génétique de la DI, dans vingt familles, où une transmission de la maladie selon un mode autosomique récessif est suspectée. Nous avons mis de l'avant les techniques de séquençage de nouvelle génération, afin de mettre en évidence les déterminants génétiques de la DI, à l'échelle du génome humain. En fait, nous avons priorisé la capture et le séquençage de l'exome; soient la totalité des régions codantes du génome humain et leurs sites d'épissage flanquants. Dans nos analyses, nous avons ciblé les variants qui ne sont pas rapportés trop fréquemment dans différentes bases de données d'individus contrôles, ces mutations rares cadrent mieux avec

une condition comme la DI. Nous avons porté une attention particulière aux mutations autosomiques récessives (homozygotes et hétérozygotes composites) ; nous avons confirmé que ces mutations ségréguent avec une transmission récessive dans la famille à l'étude. Nous avons identifié des mutations dans des gènes pouvant être à l'origine de la DI, dans certaines des familles analysées ; nous avons validé biologiquement l'impact fonctionnel des mutations dans ces gènes candidats, afin de confirmer leur implication dans la pathophysiologie de la DI.

Nous avons élucidé les bases moléculaires de la DI dans huit des familles analysées. Nous avons identifié le second cas de patients avec syndrome de cassure chromosomique de Varsovie, caractérisé par des dysfonctions de l'ARN hélicase DDX11. Nous avons montré qu'une perte de l'activité de TBC1D7, une des sous-unités régulatrice du complexe TSC1-TSC2, est à l'origine de la pathologie dans une famille avec DI et mégalencéphalie. Nous avons mis en évidence des mutations pathogéniques dans le gène *ASNS*, codant pour l'Asparagine synthétase, chez des patients présentant une microcéphalie congénitale et une forme progressive d'encéphalopathie. Nous avons montré que des dysfonctions dans la protéine mitochondriale MAGMAS sont mises en cause dans une condition caractérisée par un retard prononcé dans le développement associé à une forme sévère de dysplasie squelettique. Nous avons identifié une mutation tronquant dans *SPTBN2*, codant pour la protéine *spinocerebellar ataxia 5*, dans une famille avec DI et ataxie cérébelleuse. Nous avons également mis en évidence une mutation dans *PIGN,* un gène impliqué dans la voie de biosynthèse des ancres de glycosylphosphatidylinositol , pouvant être à l'origine de la maladie chez des individus avec épilepsie et hypotonie. Par ailleurs, nous avons identifié une mutation - perte de fonction dans *CLPB*, codant pour une protéine chaperonne mitochondriale, dans une famille avec encéphalopathie néonatale, hyperekplexie et acidurie 3-méthylglutaconique. Le potentiel diagnostic des techniques de séquençage de nouvelle génération est indéniable ; ces technologies vont révolutionner l'univers de la génétique moléculaire, en permettant d'explorer les bases génétiques des maladies complexes comme la DI.

Mots-clés :
Déficience intellectuelle, génétique, techniques de séquençage de nouvelle génération, séquençage d'exome, transmission autosomique récessive, *DDX11*, *TBC1D7*, *ASNS, MAGMAS, PIGN, SPTBN2, CLPB*.

ABSTRACT

Intellectual disability (ID) regroups greatly heterogeneous conditions that are characterized by early-onset cognitive impairment. ID affects about 1-3% of Western populations; but its prevalence is much higher in deprived regions of the world where socio-demographic factors like poor healthcare, lack of resources and parental consanguinity prevail. Non-genetic factors are involved in the etiology of ID; approximately 25% of ID cases are of genetic origin. Traditionally, the molecular basis of ID have been assessed through cytogenetic analyses, genetic mapping and candidate gene approaches. These classical genetic tools are still put to the test in the study of complex diseases like ID.

Until recently, X-linked ID cases were the main focus of studies on ID with more than hundred ID genes identified only on the X chromosome. Compound heterozygous mutations are identified in autosomal forms of ID, in the context of non-consanguineous unions. However, the occurrence of such mutations is rare in outbred populations, so that dominant *de novo* mutations are most common in unrelated individuals. Homozygous mutations are expected in consanguineous unions or in populations marked by a founder effect. In fact, the molecular bases of autosomic recessive ID have been almost exclusively studied in populations with high consanguinity rates. ID remains unsolved in more than 60% of patients. In the absence of environmental factors associated with ID in these individuals, it is possible to consider that unidentified genetic factors are involved in these unexplained ID cases.

In this research project, we used next generation sequencing technologies to highlight the genetic causes of ID in twenty families were an autosomal recessive mode of inheritance is expected. We prioritized the use of whole-exome sequencing, namely all coding exons in the genome of this individual. In our analyses, we filtered out variants that were too common in control individuals to describe a rare condition like ID. We focussed our attention on rare autosomic recessive varaiants (homozygous and compound heterozygous), these mutations were confirmed by Sanger re-sequencing to segregate with an autosomal recessive mode of inheritance in the family. We identified mutations in candidate genes for ID in some of the family analysed, we validated the functional impact of the mutations in these genes to confirm their involvement in the pathophysiology of ID in the family studied.

We explained the molecular basis of ID in eight of the families studied. We identified the second case of Warsaw-Breakage-Syndrome, a rare genetic disorder characterised by dysfunction of the RNA helicase DDX11. We showed that disruption in TBC1D7, a functional subunit of the TSC1-TSC2 protein complex, cause ID and megalencephaly. We demonstrated that *ASNS*, the Asparagine Synthetase gene, is defective in patients with congenital microcephaly and progressive encephalopathy. We showed that the gene coding for the mitochondrial protein MAGMAS is involved in the pathophysiology of a condition characterised by developmental delay and severe skeletal dysplasia. We identified a truncating mutation in *SPTBN2*, encoding for the spinocerebellar ataxia 5 proteins, in a family with ID and spinocerellar ataxia. We also identified a mutation in a gene involved in the biosynthetic pathway of glycosylphosphatidylinositol anchors; the mutation in *PIGN* may cause the epilepsy and hypotonia features observed in the affected individuals of that family. Finally, we identified a loss of function mutation in *CLPB,* coding for a mitochondrial chaperone, in individuals with severe encephalopathy, hypereklexia and 3-methylglutaconic aciduria. The diagnostic potential of next generation sequencing technologies is undeniable. These technologies will revolutionize the world of molecular genetics; they will help deciphering the molecular basis of complex diseases like ID.

Keywords:
Intellectual deficiency, genetics, next generation sequencing technologies, exome sequencing, autosomic recessive, *DDX11*, *TBC1D7*, *ASNS*, MAGMAS, *PIGN*, *SPTBN2*, *CLPB*.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS

ADN ; DNA: Acide désoxyribonucléique ; *Deoxyribonucleic acid*

ARN ; RNA: Acide ribonucléique; *Ribonucleic acid*

ARNm ; mRNA: ARN messager ; *messenger RNA*

ASNS: Asparagine synthétase; *Asparagine synthetase*

BSA: *Bovine serum albumin*

CCDS: *Consensus Coding Sequence*

cDNA: coding DNA

CGH: *Comparative genomic hybridization*

Chr: Chromosome

CNV: *Copy number variation*

CSF: *Cerebrospinal fluid*

DI ; ID: Déficience intellectuelle ; *Intellectual disability*

DI-A: Déficience intellectuelle autosomique

DI-AD: Déficience intellectuelle autosomique dominante

DI-AR: Déficience intellectuelle autosomique récessive

DI-NS: Déficience intellectuelle non-syndromique

DI-S: Déficience intellectuelle syndromique

DI-X: Déficience intellectuelle liée au chromosome X

DTT: *Dithiothreitol*

DPBS: *Dulbecco's Phosphate-Buffered Saline*

EEG: *Electroencephalogram*

EMSA: *Electrophoresis mobility shift assays*

FISH: *Fluorescence in situ hybridization*

G4: *G-quadruplex*

GAP: *GTPase activating protein*

GDI: *Guanine nucleotide dissociation inhbitor*

GEF: *Guanine exchange factor*

GST: *Glutathione S-transferase*

IBD: Identique par descendance *; Identical by descent*

IEM: Erreur innée du métabolisme ; *Inborn error of metabolism*

IGV: *Integrated genomics viewer*

Indels: Insertions-délétions; *Insertions-deletions*

ISF: *Interstitial fluid*

F: Coefficient de consanguinité

kb: kilo paires de bases ; *kilo base pairs*

LCL: *Lymphoblastoid cell line*

LTP: *Long term potentiation*

Mb: Méga paires de bases ; *Mega base pairs*

MCPH: *Microcephaly*

MLPA: *Multiplex ligation-dependent probe amplification*

MMC: *Mitomycin-C*

MPS: *Massive parallel sequencing*

ND: Maladie neurodéveloppementale; *neurodevelopmental disorder*

NGS: *Next generation sequencing*

NMD: *Nonsense-mediated decay*

pb ; bp: paire de bases ; base *pairs*

PBS: *Phosphate-Buffered Saline*

PCR: *Polymerase chain reaction*

QI ; IQ: Quotient intellectuel ; *Intelligence quotient*

RH; HR: Region d'homozygotie; *Homozygosity regions*

RFLP: *Restriction fragment length polymorphism*

RT-PCR: *Reverse transcription polymerase chain reaction*

SD: *Standard deviation*

SDS: *Sodium dodecyl sulfate–polyacrylamide gel electrophoresis*

SF2: *Superfamily 2*

SNP: *Single nucleotide polymorphism*

SNV: *Single-nucleotide variant*

SVZ: *Subventricular zone*

TSC: *Tuberous Sclerosis Complex*

VUS : *Variant of unknown significance*

VZ: *Ventricular zone*

WABS: *Warsaw breakage syndrome*

WES: Séquençage d'exome; *Whole-exome sequencing*

WGS: Séquençage du genome; *Whole-genome sequencing*

WT: *Wild type*

Remerciements

J'arrive au terme de cinq années mémorables. Ce fut une expérience enrichissante qui a commencé avec une maitrise et qui s'est poursuivie, tout au long de ma thèse de doctorat. Je remercie mes superviseurs de recherche Jacques Michaud et Mark Samuels pour m'avoir donné l'opportunité de réaliser cette thèse, et pour l'environnement de travail qu'ils ont créé autour de moi. Un grand merci à Fadi Hamdan, ainsi que tous ceux qui sont intervenus au cours de ce processus d'apprentissage. Certes, j'aurai acquis beaucoup de connaissances sur mon sujet de recherche ; mais j'en aurai définitivement appris sur ma personne. Si cette expérience revêt une importance particulière à mes yeux, c'est aussi grâce à toutes ces personnes que j'ai eu le plaisir de côtoyer et avec qui j'ai pu cheminer au cours des cinq dernières années. Évidemment, rien de ceci n'aurait été possible sans le support inconditionnel de ma famille et de mes proches.

CHAPITRE I: INTRODUCTION GÉNÉRALE

Les maladies neurodéveloppementales (NDs) définissent des atteintes neurologiques ayant un impact sur la formation et le développement du cerveau. Ces affections apparaissent, généralement, tôt dans l'enfance ; elles peuvent également se manifester plus tard, à l'âge adulte. Les NDs sont individuellement rares ; collectivement, elles constituent un véritable fléau socioéconomique affectant 5 à 10 % de la population. La déficience intellectuelle est un facteur de comorbidité communément associé aux NDs.

1. La déficience intellectuelle

1.1 Définition générale

La déficience intellectuelle (DI) ou retard mental est le handicap sévère le plus fréquent chez l'enfant. Au sens le plus large, la DI inclus les trois critères suivants :

- des capacités intellectuelles limitées ; quotient intellectuel (QI) < 70
- un déficit significatif du fonctionnement adaptatif dans au moins deux des sphères suivantes : la communication, l'autonomie, les aptitudes sociales et interpersonnelles, les compétences scolaires, l'usage des services publics, la prise de décision, le travail, les loisirs, la santé et la sécurité.
- l'apparition des troubles susmentionnés avant l'âge de 18 ans (American Psychiatric Association., 2000) .

L'échelle d'intelligence de Wechsler (WISC IV) est la référence internationale utilisée dans l'évaluation de l'intelligence chez l'enfant. Elle permet de déterminer le QI en se basant sur quatre indices factoriels : la compréhension verbale, le raisonnement perceptif, la mémoire de travail et la vitesse de traitement de l'information. Le test de QI est communément utilisé dans le diagnostic de la DI. Un retard global dans le développement de l'enfant, particulièrement au niveau du langage et des aptitudes psychomotrices, est généralement prédicateur d'une DI, lorsque le test de QI ne peut être effectué. Typiquement, on distingue cinq degrés de sévérité de DI, selon la performance de l'enfant au test de QI. (Tableau I.1, World Health Organisation, 1980) . La DI peut être également classifiée en deux groupes, sur la base de l'observation clinique des patients. On distingue : la DI syndromique (DI-S), caractérisée par des signes phénotypiques, morphologiques, métaboliques distinctifs ; la DI non-syndromique (DI-NS) définit plutôt une condition où la

DI est le seul trait clinique manifesté. Il n'est pas aisé de discriminer la DI-S de la DI-NS, particulièrement lorsque les traits phénotypiques ne sont pas apparents. (Ropers, 2006 ; Basel-Vanagaite, 2007) La DI-NS compte pour deux-tiers des cas de DI ; toutefois, les bases génétiques de la DI-NS n'ont pu être élucidées que dans 10-15% des cas diagnostiqués.

Tableau I.1 Classification de la déficience intellectuelle en fonction du QI.

DI : déficience intellectuelle. QI : Quotient intellectuel.
(Modifié à partir de American Psychiatric Association, 2000)

DI	QI	Prévalence	Description sommaire
Profonde	<20	1-2%	Troubles neurologiques prononcés. Développement moteur altéré. Déficit majeur dans la communication et l'autonomie. Patients nécessitant des soins adaptés dans un environnement sous haute supervision.
Sévère	20-35	3-4%	Déficit de communication majeur (voire inaptitude à communiquer). Peut accomplir des tâches élémentaires, avec une assistance soutenue.
Modérée	35-30	10 %	Capacité à communiquer et à être autonome. Fonctionnement adaptatif du patient pouvant être amélioré avec une assistance modérée.
Faible	50-70	85 %	Déficit mineur. DI généralement diagnostiquée très tard, à l'âge scolaire.
Limite	70-85	non déterminée	Individu ne pouvant être distingué des autres enfants de son âge.

L'autisme et l'épilepsie comptent parmi les maladies neuropsychiatriques les plus fréquemment associés à la DI. Par exemple, 70% des individus présentant des troubles dans le spectre de l'autisme vont développer une forme de DI (Newschaffer et al., 2007 ; Mefford et al., 2012). L'épilepsie est également commune chez les patients avec DI, particulièrement dans les cas les plus sévères. On estime que 70-80% des enfants avec syndrome de Rett vont avoir des crises d'épilepsie entre 3 et 5 ans (Shorvon, 2000) ; des troubles épileptiques sont aussi diagnostiqués chez environ 90% des enfants avec syndrome d'Angelman (Dan et Boyd, 2003) . L'épilepsie est aussi un trait clinique récurrent dans d'autres formes

syndromiques de DI comme la trisomie 21 (syndrome de Down) et le syndrome de l'X fragile (Romano et al., 1990 ; Pueschel et al., 1991 ; Chakrabarti et Davies, 1997).

La DI est un fléau mondial affectant 1-3% de la population, toutes classes socioéconomiques confondues (Leonard et Wen, 2002) . Sa prévalence est plus élevée dans les pays non-industrialisés, particulièrement au sein de la classe moyenne (Drews et al., 1995 ; Roeleveld et al., 1997 ; Durkin et al., 1998) . On estime que la DI touche environ 30-50% plus de garçons que de filles (Stevenson et al., 2000) . Certains facteurs génétiques de susceptibilité chez les garçons ou de protection chez les filles seraient mis en cause dans ce phénomène (Jacquemont et al., 2014). L'occurrence de la DI varie en fonction de sa sévérité ; les cas de DI sévère affectent 0.3%-0.5% des populations dans les pays industrialisés (Leonard et al., 2002). La DI modérée est la forme de DI la plus fréquente; elle est mise en évidence dans environ 85% des cas diagnostiqués (American Psychiatric Association., 1994).

1.2 Causes de la déficience intellectuelle

1.2.1 Causes-non génétiques

L'exposition du nouveau-né à des facteurs environnementaux durant les périodes prénatales, périnatales ou postnatales, peut affecter le développement du cerveau de ce dernier et résulter en une forme de DI. Par exemple, dans l'hyperphénylalaninémie maternelle, un déficit intellectuel est observé chez des enfants, pourtant non phénylcétonuriques, de mères avec des concentrations élevées en phénylalanine. Le syndrome d'alcoolisme fœtal est de loin la cause environnementale la plus fréquente de DI (May et Gossage, 2001 ; Niccols, 2007) . Les infections périnatales, l'hypoxie du nouveau-né, les malformations cérébrales, les traumatismes et autres affections menant au développement prématuré du nouveau-né sont fréquemment associés à la DI. L'usage de drogues ou toxines, ainsi que l'exposition à des agents tératogènes sont aussi mis en cause dans la DI (Majnemer et Shevell, 1995 ; Curry et al., 1997) . Par ailleurs, des facteurs sociodémographiques sous-jacents comme la malnutrition, la déprivation culturelle et la consanguinité sont également prépondérants dans la DI (Seidman et al., 2000 ; Durkin, 2002 ; Musante et Ropers, 2014) . Dans l'ensemble, des facteurs non-génétiques contribuent à environ 20% de l'étiologie de la DI (Majnemer et al., 1995) . À noter que ces estimations ne

tiennent pas compte du fait que certaines causes non génétiques puissent être elles-mêmes secondaires à des lésions génétiques, comme des mutations dans des gènes exprimés au cours de l'embryogenèse ou étant requis dans le développement du cerveau du nouveau-né.

1.2.2 Causes génétiques

La DI est l'une des conditions génétiques les plus couteuses, en termes de diagnostic moléculaire; typiquement 5-10% des budgets en santé sont alloués à la prise en charge des patients avec DI, dans les pays industrialisés (Tarpey et al., 2009 ; Musante et al., 2014) . D'ailleurs, la DI est la raison la plus fréquemment évoquée dans les recours aux services génétiques (Ropers et Hamel, 2005 ; Ropers, 2010) . Les causes génétiques de la DI incluent les aberrations chromosomiques, les réarrangements submiscroscopiques, les anomalies d'empreinte génétique et les lésions monogéniques. Les approches cytogénétiques sont traditionnellement utilisées, en premier lieu, dans l'investigation des patients avec DI.

On compte plus de 450 syndromes génétiques impliquant une forme de DI (Otte et Rauch, 2013); certaines de ces maladies, en l'occurrence les erreurs innées du métabolisme (IEMs; *inborn errors of metabolism*), peuvent être diagnostiquées par des analyses de routine dans les laboratoires de génétique moléculaire. Typiquement, ces défauts sont mis en évidence dans les fluides organiques (urine, plasma) chez le nouveau-né ; ils peuvent être détectés par des tests biochimiques sur les métabolites du corps humain (acides aminés, acides organiques, glucose, cholestérol, purines, pyrimidines, substrats du cycle de Krebs) ou encore sur certaines substances excrétées (métaux et autres toxines). Dans certains cas, les IEMs peuvent être diagnostiquées en étudiant les dysfonctions au sein de certaines organelles comme les lysosomes, les peroxysomes et les mitochondries (Van Karnebeek et Stockler, 2012 ; Van Karnebeek et al., 2014) . Les IEMs peuvent se manifester par des atteintes neurologiques caractéristiques (troubles comportementaux, démence, épilepsie, ataxie, difficultés psychomotrices) ; elles peuvent également se traduire par des défauts dans d'autres systèmes (dysmorphie faciale, malformations gastro-intestinales et cardiaques, problèmes de croissance, dysfonctions rénales ou hépatiques). L'évaluation des maladies métaboliques rentre en ligne de compte dans la prise en charge des patients avec DI. Dans certains cas, des analyses moléculaires, dans une deuxième phase, sont nécessaires afin de confirmer le diagnostic. Certaines IEMs peuvent être traitées si elles sont diagnostiquées tôt;

des exemples bien documentés sont ceux de la phénylcétonurie et de la tyrosinémie.

La DI est génétiquement hétérogène; on dénombre environ 500 gènes connus de DI , il pourrait y avoir 800 à 1000 gènes de DI (Ropers, 2010) . Des facteurs génétiques sont mis en cause dans 25% de la DI (Mclaren et Bryson, 1987 ; Srivastava et Schwartz, 2014). L'origine de la DI ne peut être déterminée chez 50-60% des patients, particulièrement dans les cas de DI-NS. En l'absence de facteurs environnementaux associés à la DI chez ces patients, il est possible d'envisager que des facteurs génétiques non-identifiés seraient mis en cause dans ces cas de DI encore inexplorés (Rauch et al., 2006 ; Ropers, 2008) .

1.3 Études génétiques sur la déficience intellectuelle

1.3.1 Études cytogénétiques

Les analyses cytogénétiques ont jeté les bases des études génétiques menées sur la DI avec le caryotype. Depuis, d'autres techniques comme l'hybridation en fluorescence *in situ* (*Fluorescence in situ hybridization*, FISH), la MLPA (*Multiplex ligation-dependent probe amplification*)**,** l'hybridation génomique comparative (*Comparative genomic hybridization*, CGH) et plus récemment les techniques de puces d'ADN (*array-CGH*) ont grandement contribué à l'identification des causes génétiques de la DI. Des défauts cytogénétiques sont communément mis en évidence dans des formes syndromiques de DI incluant les syndromes de Turner, de Klinefelter, d'Angelman et de Prader-Willi (Rauch et al., 2006 ; Ropers, 2010) On détecte des variations du nombre de copies (*Copy number variations* ou CNVs chez 10-15% des patients avec DI (Zahir et Friedman, 2007). Les études cytogénétiques ont permis d'élucider les causes génétiques de la DI chez environ 15% des patients avec DI sévère, et approximativement 25% des cas de DI en général (Leonard et al., 2002 ; Ropers, 2010) . Ces approches sont toutefois limitées à la détection d'aberrations chromosomiques macroscopiques ou submicroscopiques dont la résolution varie entre 15 Mb et 100 kb.

1.3.2 Séquençage de gènes candidats

Des lésions génétiques à plus petite échelle, non détectables en cytogénétique, sont également mises en cause dans la DI. Le séquençage classique par la méthode PCR (*Polymerase chain reaction*) – Sanger est une technique de choix dans l'étude de ce type de mutations. Les approches de séquençage de gènes candidats ont été utilisées dans l'identification des gènes de DI. Par exemple, le séquençage Sanger de la totalité des régions

codantes de 718 gènes sur le chromosome X a permis d'identifier 9 nouveaux gènes de DI dans une cohorte de 208 familles avec DI liée au chromosome X (DI-X , Tarpey et al.,2009). D'autres projets d'envergure ont également contribué aux connaissances sur la DI. Par exemple, le séquençage Sanger du gène codant pour la protéine synaptique SYNGAP1 (MRD5 ; [MIM:612621]) dans une cohorte de 94 patients avec DI-NS a permis de mettre en évidence des mutations *de novo* à l'origine de la DI chez 3 patients de cette cohorte (Hamdan FF et al.,2009). Depuis, des cas additionnels avec mutations dans *SYNGAP1* ont été rapportés, faisant de *SYNGAP1* l'un des gènes les plus fréquemment impliqués dans les formes modérées ou sévères de DI-NS. Une autre étude visant le séquençage ciblé de 197 gènes codant pour des récepteurs glutamatergiques et leurs partenaires d'interaction dans une cohorte de 95 patients avec DI-NS a permis d'identifier des mutations causales *de novo* dans 6 autres gènes de DI : *CACNG2* (MRD10 ; [MIM:614256]), *EPB41L1* (MRD11; [MIM:614257]) , *GRIN1* , *KIF1A* (MRD9; [MIM:614255]) , *SHANK3* [MIM:613950] et *STXBP1* [MIM:612164] (Hamdan et al., 2011) .

Le potentiel de découverte des approches de gènes candidats est limité par nos connaissances sur certains systèmes biologiques. Ces techniques sont restreintes à l'analyse ciblée d'une fraction de gènes ; elles ne tiennent donc pas compte des autres déterminants génétiques impliqués dans la DI. Par ailleurs, les études de gènes candidats requièrent de grandes cohortes de patients, afin d'identifier de nouveaux gènes et de répliquer les résultats. Par exemple, dans l'étude de Tarpey PS, la cause de la DI-X n'a pu être confirmée que dans 25% des familles analysées. Dans plus de 70% de ces familles, une cause génétique était suspectée, mais elle n'a pu être confirmée, faute dans de cas additionnels. La caractérisation phénotypique des patients est un facteur clé, lorsqu'on veut répliquer les résultats de ce genre d'études. En effet, la cohorte de patients se doit d'être le plus homogène possible, ce qui est particulièrement difficile dans la DI-NS. La priorisation de gènes candidats s'applique davantage lorsqu' on n'a qu'une poignée de gènes à étudier. Elle se prête mieux dans des situations où une cartographie génétique a été préalablement réalisée, afin de restreindre le nombre de gènes potentiels à séquencer.

1.3.3 Cartographie génétique

Les études sur la DI-X ont été facilitées par les particularités de la transmission liée au chromosome X : la mère est obligatoirement hétérozygote pour la mutation causale et tous ses garçons hémizygotes pour l'allèle mutant manifestent nécessairement le phénotype. Dans le cas d'une transmission autosomique, il est beaucoup plus ardu de déceler le mode de transmission impliqué (dominant ou récessif), sur simple observation du pédigrée de la famille. Traditionnellement, on a plutôt recours aux approches de cartographie génétique, afin d'identifier des locus statistiquement associés au trait génétique à l'étude. Le pouvoir statistique de ces analyses dépend du devis l'étude. En fait, ces techniques nécessitent généralement de grandes cohortes de patients ou des familles avec un nombre significatif d'individus affectés pour pouvoir gagner en résolution. Elles se prêtent donc peu à l'analyse de conditions rares comme la DI.

Les connaissances sur la DI autosomique (DI-A) ont été grandement limitées par le fait que, dans les pays industrialisés, où la plupart des études sont menées, la tendance est à faire de moins en moins d'enfants. De fait, dans ces pays, les cas de DI sont sporadiques ; c'est à dire qu'ils surviennent généralement dans de petites familles avec un seul individu affecté Dans l'étude de la DI-A, l'emphase a été mise sur l'analyse de cas familiaux issus d'unions consanguines. Certaines particularités inhérentes à ce type d'unions peuvent être mises à profit dans l'exploration des causes génétiques de maladies récessives rares comme la DI (voir 2. La consanguinité).

1.4 Bases moléculaires de la déficience intellectuelle

1.4.1 Déficience intellectuelle liée au chromosome X

La DI-X (Base de données OMIM, acronyme MRX) est de loin la forme la plus explorée de DI avec plus de 215 syndromes génétiques associés (Chiurazzi et al., 2008 ; Tarpey et al., 2009). Le syndrome de l'X fragile, causé par des mutations dans le gène *FMR1* [MIM:300624], totalise pour 10-12% des cas de DI-X. Il s'agit de la forme la plus fréquente de DI-X (Mandel et Chelly, 2004 ; Kleefstra et Hamel, 2005) . Des lésions génétiques dans d'autres gènes sur le chromosome X sont également communes dans la DI-X ; c'est le cas notamment des gènes *ARX* (MRXARX ; [MIM:300419]), *ATRX* (MRXSHF1; [MIM:309580]), *CDKL5* [MIM:300672], *JARID1C* (MRXSCJ;

[MIM:300534]), *MECP2* [MIM:312750] , *OPHN1* (MRXSO; [MIM:300486]), *PQBP1* [MIM:309500] et *SLC6A8* [MIM:300352].

La DI-X est particulièrement fréquente chez les garçons , avec une prévalence de 1/600 à 1/1000 (Gecz et al., 2009). Le taux de masculinité est de 1.4 pour la DI-X sévère ; il s'élève à 1.9 pour la DI-X faible (Ropers et al., 2005). Des différences dans l'organisation des circuits neuronaux , entre garçons et filles , ou dans les patrons d'expression de gènes du chromosome X exprimés dans le cerveau , ont été avancées pour expliquer la plus grande prévalence de la DI-X chez les garçons ; ces hypothèses n'ont toutefois pas été appuyées par des études (Skuse, 2005) . La contribution génétique du chromosome X ne suffit pas pour expliquer, l'excès de la DI chez les garçons. On compte plus d'une centaine de gènes de DI-X qui ne totalisent toutefois que 10-12% de la DI chez les hommes (Mandel et al., 2004 ; Kleefstra et al., 2005). Par ailleurs, en dépit du grand nombre d'études menées sur la DI-X, il y a peu de nouveaux gènes de DI qui sont identifiés. En fait, on estime qu'une fraction de gènes explique une proportion significative de la DI-X. Par exemple, une vingtaine de gènes sur le chromosome X sont impliqués dans plus de 30% de la DI-X non-syndromique, laquelle totalise pour les deux tiers de la DI-X. La contribution de cette poignée de gènes dans l'étiologie de la DI-X pourrait être plus importante, étant donné que la majorité d'entre eux n'ont été identifiés que récemment (Ropers et al., 2005).

1.4.2 Déficience intellectuelle autosomique

Le chromosome X ne compte que 5% des gènes du génome humain, on peut donc estimer qu'il y aurait plus de 2500 gènes autosomaux pouvant causer la DI (Ropers et al., 2005 ; Musante et al., 2014) . Dans la base de données OMIM, on compte environ 300 gènes de DI-A, ce qui suggère qu'une bonne proportion des gènes de DI seraient localisés sur les autosomes. Les formes de DI les plus sévères sont généralement létales ; elles peuvent également avoir un impact important sur la fertilité et les capacités reproductives des individus affectés, de sorte qu'on estime qu'elles sont majoritairement causées par des mutations dominantes *de novo* (Rauch et al., 2006 ; Ropers, 2010 ; De Ligt et al., 2012 ; Veltman et Brunner, 2012) . Des mutations dominantes à pénétrance incomplète ou à expressivité variable peuvent être également mises jeu dans les formes de DI modérées. Une transmission autosomique récessive est attendue dans une fraction significative de la DI,

particulièrement chez les patients avec DI-NS. (Basel-Vanagaite et al., 2007 ; Ellison et al., 2013) Des mutations hétérozygotes composites sont mises en évidence dans la DI-A, dans le contexte d'unions non-consanguines ; toutefois l'occurrence de ce type de mutations est rare chez des individus non-apparentés, de sorte que le cas de figure le plus fréquent est celui des mutations dominantes. Des mutations homozygotes récessives sont davantage rencontrées dans les populations avec un fort taux de consanguinité (Figure I.1).

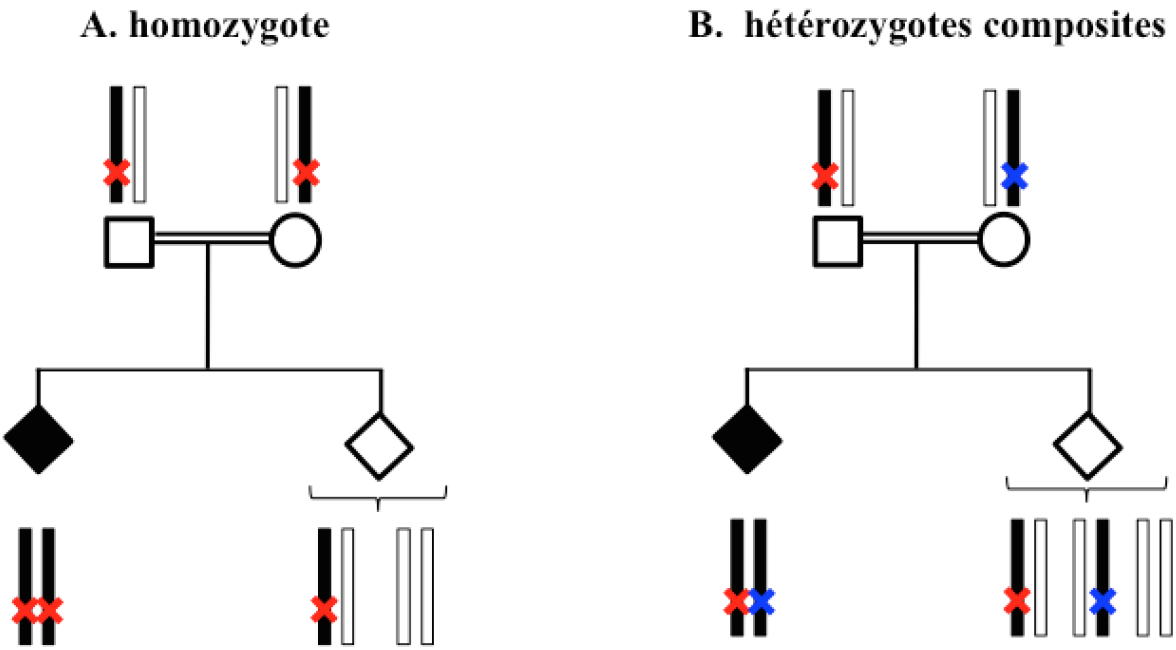


Figure I.1 Modèles de transmission d'un trait autosomique récessif.

A.Transmission homozygote B. Transmission hétérozygote composite (ou composée génétique). Les croix rouges ou bleues représentent une mutation homozygote ou hétérozygote composite, selon le cas.

2. La consanguinité

2.1 Généralités

La consanguinité traduit l'union entre deux individus apparentés. On estime qu'environ 10% de la population mondiale vit dans une union consanguine ou est issue d'un mariage consanguin (Bittles et Black, 2010) . En génétique, une union consanguine définit la reproduction sexuée entre deux individus étant, aussi peu apparentés que des cousins de deuxième génération (Figure I.2; Bittles , 2001) . Les mariages consanguins sont

traditionnellement acceptés dans certaines régions du monde, au sein d'une zone géographique délimitée à l'Est par le Pakistan et l'Afghanistan, à l'Ouest par le Maroc et au Sud par l'Inde. (Figure I.3) Plus d'un milliard d'individus vivent dans des régions où la consanguinité fait partie des mœurs, majoritairement dans des groupes socioculturels d'Afrique du Nord, du Moyen-Orient et de l'Ouest de l'Asie. Dans ces communautés, en moyenne 20-50% des mariages impliquent des membres de la même famille (Hamamy et al., 2011). Les taux de consanguinité les plus élevés ont été répertoriés dans certaines régions de l'Inde (55%) et du Pakistan (77%) (Puri et al., 1978 ; Hashmi, 1997). Culturellement, les unions entre des cousins de première génération ou de type oncle-nièce sont celles qui sont le plus fréquemment rencontrées. (Figure I.2).

La préservation de l'héritage familial et des valeurs culturelles, la facilité dans les arrangements maritaux et la réduction des couts de la dot sont autant de facteurs socioéconomiques prédisposant dans la consanguinité. Les unions entre individus apparentés sont particulièrement fréquentes au sein de certains groupes isolés caractérisés par un grand effet fondateur, et ce même dans les pays occidentaux. L'isolement de ces populations peut être d'origine géographique (Canadiens-Français du Lac Saint-Jean, Acadiens des Maritimes), religieuse (Mormons, Huttérites, Juifs Ashkénazes) ou culturelle (Gitans). Le taux de mariages consanguins est également élevé dans certains groupes d'immigrants, particulièrement lors de l'établissement de ces populations.

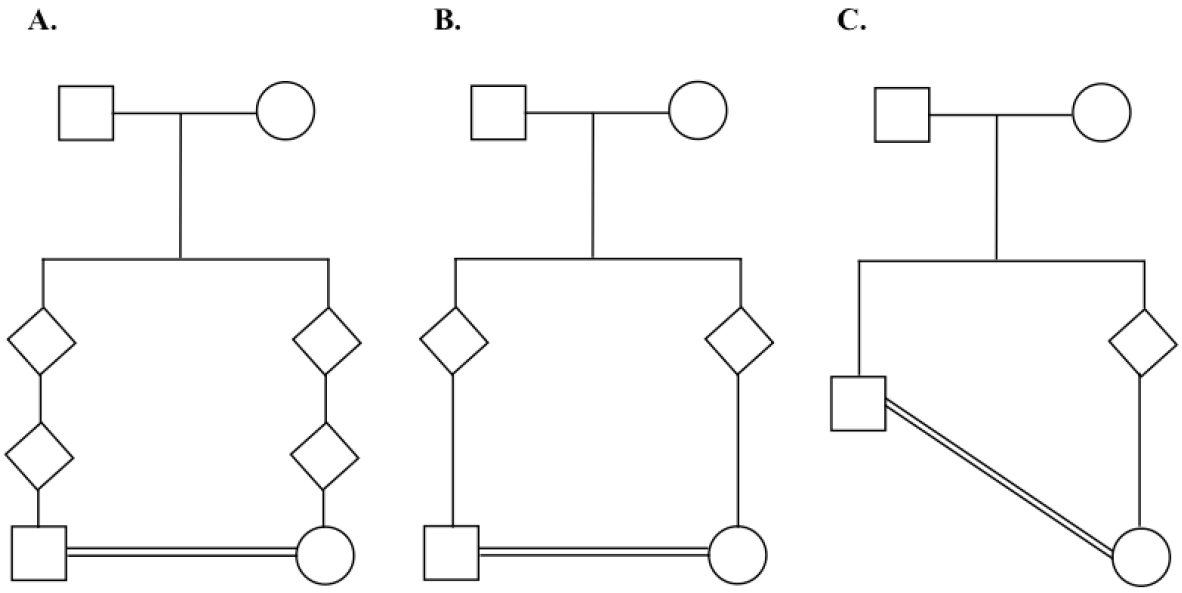


Figure I.2 Pédigrées illustrant différents degrés de consanguinité dans des familles.

A.Union entre cousins de deuxième génération. Coefficient de consanguinité (F) = 1/64. B. Union entre cousins de première génération. F = 1/16. C. Union oncle-nièce. F = 1/8. Coefficien de consanguinité : mesure de la probabilité que deux allèles d'un invidivu soient identiques par descendance.

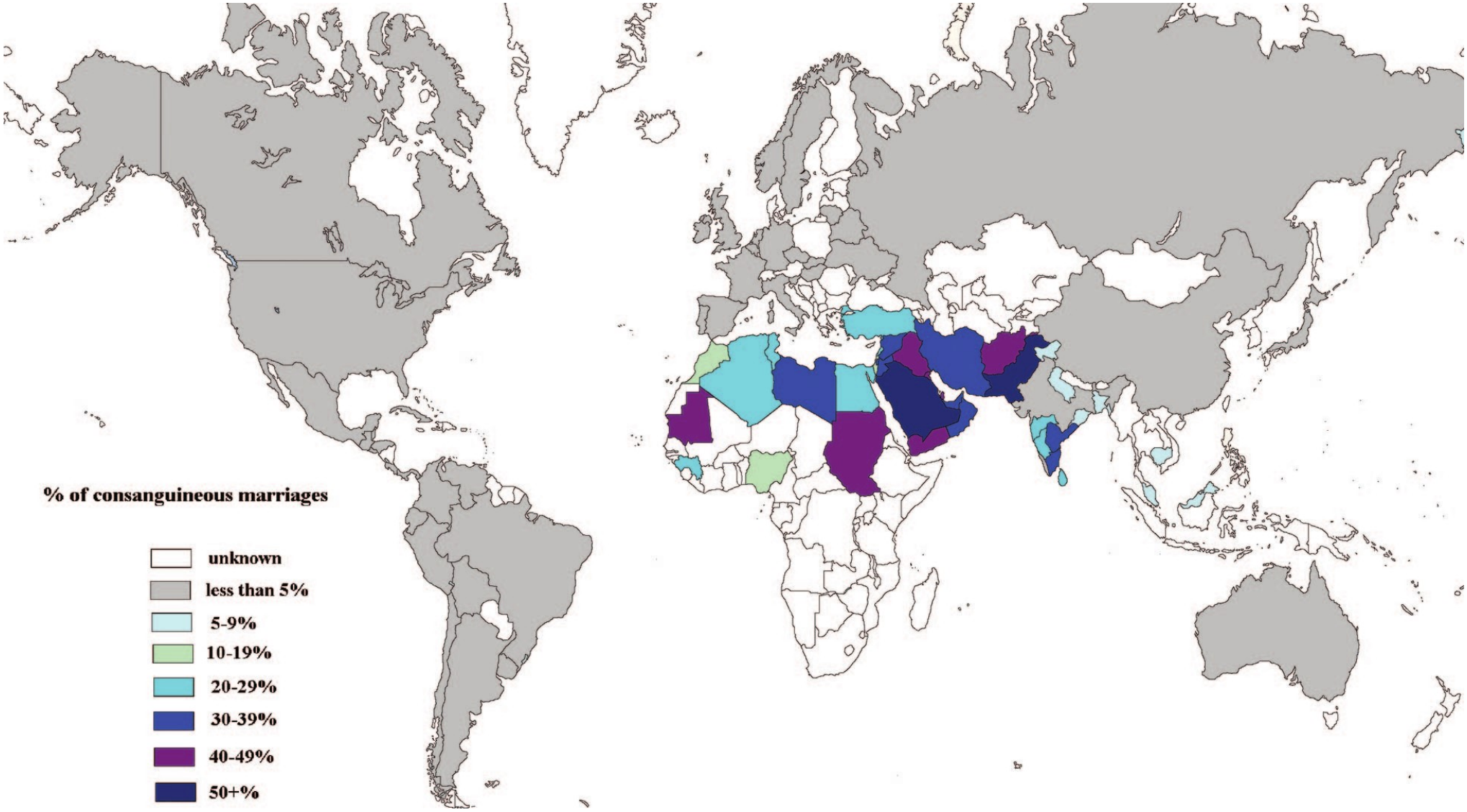


Figure I.3 Carte géographique représentant la répartition du taux de consanguinité dans le monde. (Hamamy et al.,2011)

2.2 Impact

Les unions consanguines sont caractérisées par un fort taux de natalité, en dépit du fort taux de mortalité infantile observé. Plusieurs facteurs contribuent à expliquer ce phénomène. Traditionnellement, les femmes sont amenées à la maternité beaucoup plus tôt (peu après la puberté) et vont continuer de concevoir à des âges plus avancés (proche de la ménopause). La pression sociale et les mœurs font que l'écart entre le nombre de grossesses est généralement plus court ; la période de reproduction commence peu après la cessation de l'aménorrhée lactationnelle. Le risque de récurrence des maladies génétiques rares est considérablement plus élevé dans ce type d'unions, notamment à cause du fort taux de natalité observé et de la plus grande proportion d'allèles récessifs impliqués.

Dans une union entre individus apparentés, il y a concentration du matériel génétique. De fait, des cousins de première génération, partagent environ 1/8 de leur génome, de sorte que leurs progénitures sont homozygotes pour 1/16 de leurs allèles (Hamamy et al., 2011) . Les individus issus d'une union consanguine sont donc plus à risques de manifester des traits récessifs rares. La prévalence des malformations congénitales est particulièrement élevée dans les familles consanguines (Tadmouri et al., 2009 ; Zlotogora et Shalev, 2010) . Une étude prospective menée à Birmingham sur des Britanniques d'origine pakistanaise a montré que les enfants issus de ces communautés, caractérisées par un taux de consanguinité élevé, héritent dix fois plus de malformations congénitales transmises de façon récessive. Ces maladies sont particulièrement sévères résultant généralement en une mort prématurée de l'enfant (Modell et Darr, 2002). Les auteurs de cette étude estiment aussi que le taux de morbidité infantile dans cette communauté anglo-pakistanaise pouvait être réduit de 60%, uniquement en cessant la tradition de mariages consanguins (Bundey et Alam, 1993) . Plusieurs maladies génétiques ségréguent dans les familles consanguines ; de ces maladies, les anomalies du tube neural, les malformations cardiaques congénitales, la microcéphalie, la cécité, la surdité et les troubles neurodéveloppementaux sont celles qui sont les plus fréquemment rencontrées (Mitsui et al., 1994 ; Bittles, 2001 ; Bittles et al., 2010 ; Mahmood et al., 2011) .

La consanguinité un facteur prépondérant dans les troubles neuropsychologiques (Mitsui et al., 1994) . Des études rapportent d'ailleurs une régression du QI des enfants en fonction du degré de consanguinité de leur parents (Bashi, 1977 ; Morton, 1978 ; Jensen, 1983

; Woodley, 2009). L'incidence de la DI est élevée dans les mariages consanguins, particulièrement dans le cas des formes modérées et sévères de DI. Des études menées en Suède et aux États-Unis respectivement ont montré que la DI est trois à quatre fois plus fréquente chez les enfants issus de cousins de première génération, comparativement à des enfants de parents non consanguins (Book, 1957 ; Reed et Reed, 1965). Les résultats de Madhavan et Narayan abordent dans le même sens; ils ont notamment observé que la DI est plus fréquente dans les couples consanguins ; et ce , majoritairement dans les unions de type oncle-nièce (Madhavan et al., 1991) . La régression du QI observée dans les familles consanguines corrèle avec une forte ségrégation d'allèles récessifs rares, tel qu'observé par les travaux de Morton qui a mis en évidence 325 loci récessifs ségréguant avec le QI des enfants issus d'unions entre individus apparentés (Morton, 1978) . En effet, les individus issus d'unions consanguines héritent plus fréquemment d'haplotypes récessifs que dans la population générale ; ces blocs d'haplotypes sont également plus grands que ceux observé chez des individus non-apparentés (Kirin et al., 2010).

Les familles consanguines sont caractérisées par une forte homogénéité génétique ; de fait, elles se prêtent mieux à l'étude de maladies génétiques impliquant des allèles récessifs. Le taux de natalité est particulièrement élevé dans ce type d'unions, ce qui contribue à augmenter le risque de récurrence de ces maladies récessives. Typiquement dans les unions consanguines, il y a davantage de cas familiaux (à l'opposé de cas sporadiques) pour pouvoir mener des études classique de liaisons ou d'association génétique.

2.3 Cartographie de l'autozygome

Les individus issus d'une union consanguine héritent d'une plus grande proportion d'allèles identiques par descendance (IBD, *identical by descent*). De fait, à cause de leur apparentement génétique ou coefficient de consanguinité (F), ils ont une plus grande probabilité d'être homozygotes pour un même marqueur génétique transmis par descendance. Deux allèles sont dits IBD , lorsqu'ils proviennent de la même copie d'un allèle ancestral commun (Anderson et Weir, 2007) . L'autozygome définit l'ensemble des haplotypes pour lesquels un individu est IBD (Alkuraya, 2010) . La cartographie de l'autozygome (Figure I.4) permet de mettre en évidence les allèles récessifs, en cartographiant l'ensemble des régions d'homozygotie (RH) dans le génome d'un individu (Lander et Botstein, 1987) . Une RH

définit un bloc de marqueurs génétiques pour lesquels un individu est homozygote. La puissance de la cartographie dépend de la résolution des RH, c'est à dire la distance minimale entre deux marqueurs génétiques consécutifs.

La cartographie de l'autozygome est basée sur le génotypage de polymorphismes génétiques**.** Originalement**,** les fragments de restriction (RFLP, *restriction fragment length polymorphism*), les microsatellites et les CNVs étaient les types de polymorphismes les plus couramment génotypés. Les polymorphismes nucléotidiques (SNP, *single nucleotide polymorphism*) sont désormais les marqueurs génétiques les plus utilisés ; ils représentent les polymorphismes les plus fréquents du génome humain (1 SNP / kb). Le génotypage à haut-débit a fait de la cartographie de l'autozygome une technique de choix dans l'étude des maladies qui mettent en jeu des allèles récessifs rares (Alkuraya, 2013 ; Musante et al., 2014).
La cartographie de l'autozygome présente certains avantages qui lui permettent de se démarquer des autres approches de génétique classique. Par exemple, elle peut être appliquée dans l'étude de petites familles, contrairement aux autres approches de cartographie dont le pouvoir de détection dépend grandement du nombre de cas familiaux à analyser. Par ailleurs, l'autozygome peut être mis en évidence à partir d'un seul individu, ce qui est un atout majeur dans l'étude de conditions rares qui, comme la DI, impliquent bons nombres de cas sporadiques (voir 1.3.3 cartographie génétique). Conceptuellement, la cartographie des RH n'est pas restreinte qu'à l'étude d'un type de familles en particulier. Toutefois, ce type d'approches se prête mieux à l'étude de cas issus d'unions consanguines, en raison du grand nombre d'allèles récessifs IBD transmis. D'ailleurs, à ce jour, les gènes de DI-AR ont presque tous été identifiés par cartographie des RH dans des populations consanguines. Typiquement, une transmission homozygote récessive est suspectée, dans le contexte d'unions consanguines; toutefois, il n'est pas exclu que d'autres modes de transmission soient mis en jeu dans ces familles. La cartographie de l'autozygome est limitée à l'analyse des allèles homozygotes. De fait, elle ne tient pas compte des autres types de mutations transmises ; notamment les mutations hétérozygotes composites, lesquelles peuvent être également mises en évidence dans les familles consanguines ou marquées par un effet fondateur.

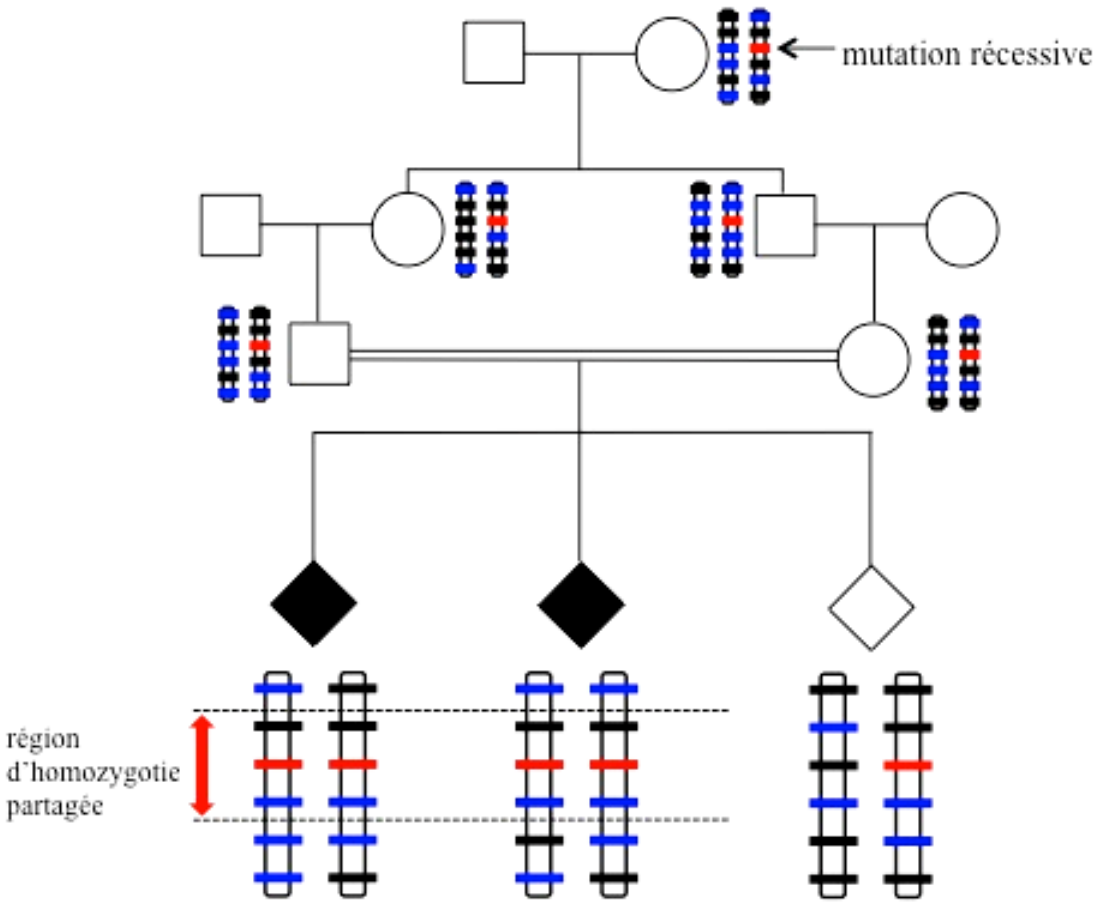


Figure I.4 Transmission d'une mutation récessive dans une union consanguine.

Une mutation récessive est transmise dans la famille, à partir d'un même ancêtre commun. D'autres marqueurs génétiques sont transmis en même temps que la mutation récessive. A chaque méiose, il y a une recombinaison des différents marqueurs génétiques. Les individus affectés issus d'unions consanguines ont plus de chance d'hériter des deux copies de la même mutation ancestrale ; ils sont homozygotes pour cette mutation. L'ensemble des haplotypes pour lesquels ces individus sont homozygotes pour les mêmes marqueurs génétiques constitue l'autozygome. La cartographie de l'autozygome permet d'identifier la mutation récessive, en ciblant les régions d'homozygotie partagées entre les individus affectés.

Typiquement, la cartographie de l'autozygome met en évidence des d'haplotypes pouvant contenir plusieurs dizaines voire des centaines de gènes candidats; des approches subséquentes de cartographie fine et de séquençage Sanger sont nécessaires, afin d'identifier la cause génétique du trait à l'étude. Une alternative, pour réduire le nombre de régions candidates à analyser, consiste à cibler les RH partagées entre plusieurs individus affectés d'une même famille. Il peut être aussi utile d'exclure les RH retrouvées chez les individus non

affectés de la famille; cette pratique peut toutefois générer des faux positifs (Abu-Safieh et al., 2013 ; Alkuraya, 2013) . Dans certains cas, les évènements de recombinaison génétique contribuent à réduire considérablement la taille de l'haplotype ancestral comportant la mutation causale; de sorte que les RH candidates sont difficilement identifiables.

3. Les techniques de séquençage de nouvelle génération

Les techniques de séquençage de nouvelle génération (NGS, *Next generation sequencing*) ont vu le jour en 2005, dans la foulée du projet de séquençage du génome humain. Typiquement on parle de NGS ou MPS (*Massive parallel sequencing*), lorsqu'on fait référence aux approches de séquençage de deuxième génération. Les approches classiques de séquençage Sanger se classent parmi les techniques de séquençage de première génération. Elles sont essentiellement basées sur la séparation électrophorétique des produits de PCR dans des réactions de séquençage individuelles (Sanger et al., 1977 ; Swerdlow et al., 1990 ; Hunkapiller et al., 1991) . Les NGS, quant à elles, permettent l'amplification clonale d'une molécule d'ADN dans plusieurs réactions en simultanée. Il existe plusieurs variantes dans les protocoles d'amplification (en émulsion ou sur phase solide) et de séquençage (terminaison réversible, en temps réel, par ligation et pyroséquençage) adoptés par les NGS (Bentley, 2006 ; Shendure et Ji, 2008 ; Metzker, 2010) . Ces approches de séquençage à haut débit promettent de révolutionner l'univers de la génétique, en rendant possible le séquençage d'importantes quantités de matériel génétique à l'échelle du génome humain, rapidement et à moindres coûts (Margulies et al., 2005 ; Metzker, 2010) . Par exemple, on estime qu'il a coûté environ 72 millions de dollars US pour séquencer le génome de Craig Venter par séquençage Sanger (Levy et al., 2007) , comparativement au million de dollars US investi pour séquencer le génome de James Watson en utilisant les NGS (Wheeler et al., 2008) . Les récents progrès technologiques ont contribués à diminuer les coûts du séquençage ; aujourd'hui il est possible de séquencer le génome humain pour aussi peu que trois mille dollars. Actuellement, nos connaissances sur la structure du génome humain et notre habileté à les interpréter sont encore limitées, de sorte que les NGS ont été principalement utilisées dans l'analyse de la portion la plus explorée du génome, soit l'exome.

3.1 Capture et séquençage d'exome

L'exome regroupe la totalité des régions codantes du génome et leurs sites introniques flanquants. Bien qu'il ne représente que 1.5% du génome humain, l'exome constitue la portion la plus informative du génome avec plus de 180 000 séquences codantes ; il totalise approximativement 85% des mutations causales associées à des maladies génétiques (Botstein et Risch, 2003 ; Pussegoda, 2010 ; Gilissen et al., 2012). De fait, le séquençage de l'exome (WES, *Whole-exome sequencing*) constitue la meilleure approche pour étudier les maladies complexes comme la DI, parce qu'il permet la recherche directe de mutations à l'échelle de l'exome ou du génome.

Plusieurs plateformes commerciales (Agilent, Illumina, Nimblegen) permettent de mettre en évidence l'exome, en procédant selon des protocoles de capture par hybridation. (Figure I.5). Brièvement, l'ADN génomique d'un individu est fragmenté ; puis il est dénaturé, afin de permettre la ligation d'adaptateurs universels pour le séquençage. L'ADN simple brin est ensuite hybridé à une librairie synthétique apprêtée de façon à reconnaitre l'exome ; la composition de la librairie peut être adaptée aux besoins de l'expérience (capture du génome, de l'exome ou de certains gènes cibles). Typiquement, la librairie de l'exome est composée des séquences codantes références CCDS (*Consensus coding sequence*) et Refseq de NCBI (Pruitt et al., 2009) . Les librairies sont régulièrement mises à jour, afin d'inclure toutes les séquences références annotées dans le génome humain. Les séquences de ces librairies synthétiques sont préalablement biotinylées ; ce qui permet la capture des séquences-cibles sur des billes de streptavidine, selon l'interaction entre la biotine et la streptavidine. Le matériel capturé est ensuite isolé, puis séquencé en faisant appel à des séquenceurs à haut-débit ; les plateformes de séquençage d'Illumina et d'ABI sont les plus communément utilisées.

3.2 Interprétation et analyse des données de séquençage

Le WES génère des milliers de séquences de 100 pb en moyenne, lesquelles doivent être alignées à un génome référence, afin de détecter les variants contenus dans l'exome. L'analyse d'une aussi importante quantité de matériel génétique requiert certaines connaissances en bioinformatique, ainsi que des ressources logistiques considérables. Plusieurs plateformes et logiciels informatiques sont utilisés pour traiter les données de séquençage. Ces outils informatiques permettent notamment d'annoter les variant identifiés ;

en recueillant autant d'informations pour les caractériser comme la position génomique du variant, le nom du gène affecté par le variant, le type de mutation engendrée, l'impact de la mutation sur protéine résultante et la fréquence allélique du variant dans différentes bases de données d'individus contrôles.

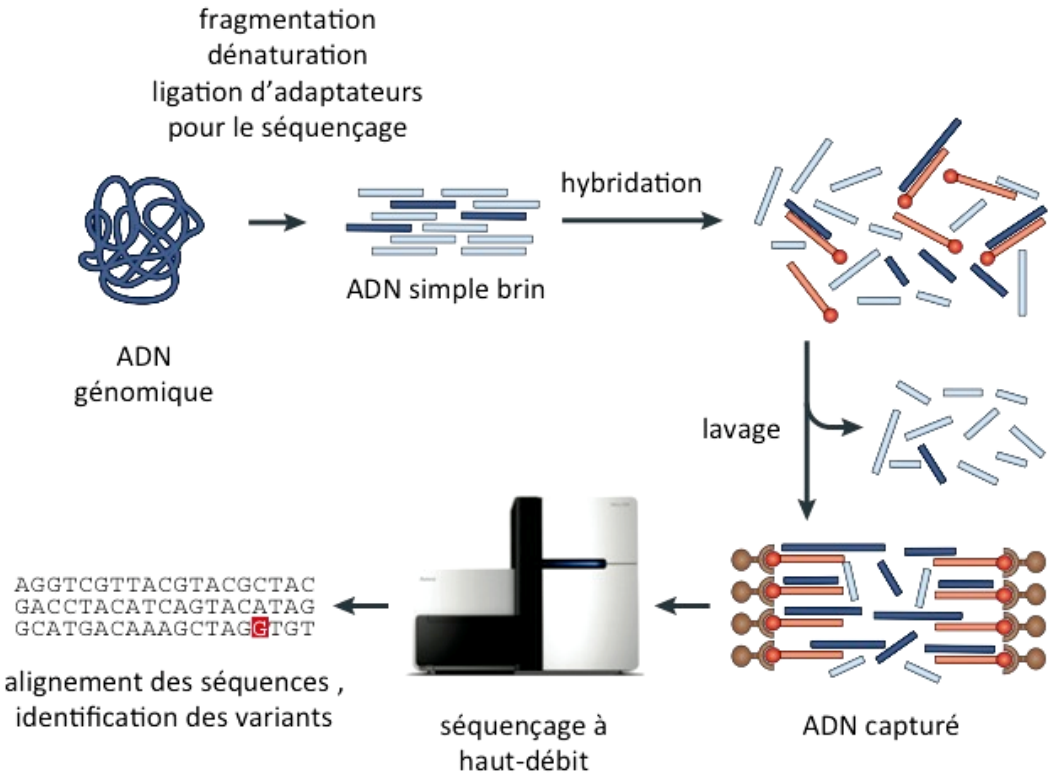


Figure I.5 Protocole de capture par hydribation.
(Modifiée à partir de Bamshad et al.,2011)

On compte en moyenne 20 000 polymorphismes (ou variants) dans l'exome d'un individu (Ng et al., 2009) , de sorte qu'il devient ardu d'identifier la mutation causale dans la maladie à l'étude. Typiquement dans l'étude d'une maladie monogénique rare, la cause génétique doit satisfaire les critères suivants :

- la mutation causale est retrouvée uniquement chez les individus affectés, c'est à dire qu'elle n'est pas également retrouvée dans le même état (homozygote ou hétérozygote) chez des individus non-affectés
- la mutation causale a un impact fonctionnel sur la protéine résultante

- des dysfonctions dans le gène muté causent un phénotype chez les individus affectés

Différents filtres doivent être appliqués dans l'analyse des résultats de séquençage à haut-débit. (Figure I.6). D'abord, il est primordial d'effectuer un contrôle minutieux des données générées, afin de se départir des séquences de mauvaise qualité. Dans le cas particulier du WES, l'analyse doit être centrée sur l'exome du patient. Dans l'étude de maladies génétiques rares, il est important de filtrer les variants rapportés trop fréquemment (fréquence allélique inférieure à 1%) dans des cohortes d'individus contrôles; ces variants peuvent être mis en évidence en criblant différentes bases de données référence comme dbSNP (*Single Nucleotide Polymorphism Database*), et *1000 Genome* (Sherry et al., 2001 ; 1000 Genomes Project Consortium et al., 2010). On estime qu'environ 95% des variants dans l'exome sont des polymorphismes (Gilissen et al., 2012). Il y a en moyenne de 300 à 500 variants rares par exome ; soit approximativement 2% des variants identifiés dans l'exome d'un individu (Bamshad et al., 2011) .

Il revient, par la suite, de sélectionner les mutations candidates; c'est à dire celles qui ségréguent avec la transmission de la maladie dans la famille. On peut par exemple, spécifier le type de mutations désirées (mutations hétérozygotes, homozygotes). Une mutation est généralement considérée comme étant hétérozygote lorsqu'elle est identifiée dans au moins 20% des séquences alignées au génome ; un minimum de 80% de couverture des lectures est requis pour désigner une mutation homozygote (Gilissen et al., 2012) . On observe en moyenne une dizaine de mutations homozygotes et une à cinq mutations *de novo* rares dans l'exome d'un individu. Il est parfois plus aisé d'avoir recours à des approches de cartographie génétique, parallèlement au WES, afin de faciliter le processus d'identification des mutations causales (Figure I.6). Les mutations candidates peuvent également être priorisées selon leur impact fonctionnel sur la protéine résultante. Des algorithmes (*Sorting Intolerant From Tolerant* - SIFT ; *Prediction of functional effects of human nsSNPs* - Polyphen) permettent de prédire le potentiel pathogénique d'un variant d'après des paramètres comme la conservation phylogénique du nucléotide ou de l'acide aminé affecté par la mutation ou encore l'impact de la mutation sur la structure de la protéine (Kumar et al., 2009b ; Adzhubei et al., 2010). Typiquement, les mutations synonymes ou celles qui sont retrouvées en dehors des sites consensus d'épissage sont filtrées, parce qu'elles ont un faible potentiel délétère.

Le WES s'est rapidement établi comme un puissant outil dans le diagnostic clinique. Uniquement entre 2009 et 2011, l'utilisation du WES aura permis d'élucider les bases moléculaires de plus de 90 maladies génétiques Gilissen et al. (2011) . Dans le cadre du projet pancanadien FORGE (*Finding of Rare Disease Genes*), l'application du WES a permis d'identifier la cause génétique de 146 maladies rares ; mettant en évidence 67 gènes qui n'avaient pas été préalablement associés à une maladie génétique (Beaulieu et al., 2014) . Les NGS ont été beaucoup utilisées dans l'exploration des causes génétiques de la DI ; particulièrement dans les formes de DI-NS .(Kaufman et al., 2010 ; Vissers et al., 2010 ; Najmabadi et al., 2011 ; De Ligt et al., 2012 ; Musante et al., 2014) . Par exemple , entre 2002 et 2011 , on ne comptait que 8 gènes de DI-AR non-syndromique; une trentaine de nouveaux gènes ont depuis été identifiés par WES (Musante et al., 2014) . Conceptuellement, les NGS peuvent être appliquées dans l'analyse des maladies monogéniques , peu importe leur modèle de transmission (autosomique dominant, *de novo*, homozygote récessif, composées génétiques ou encore lié au chromosome X) ; et ce sans avoir recours à une approche de cartographie génétique (Bamshad et al., 2011 ; Gilissen et al., 2012 ; Beaulieu et al., 2014) (Figure I.7) .

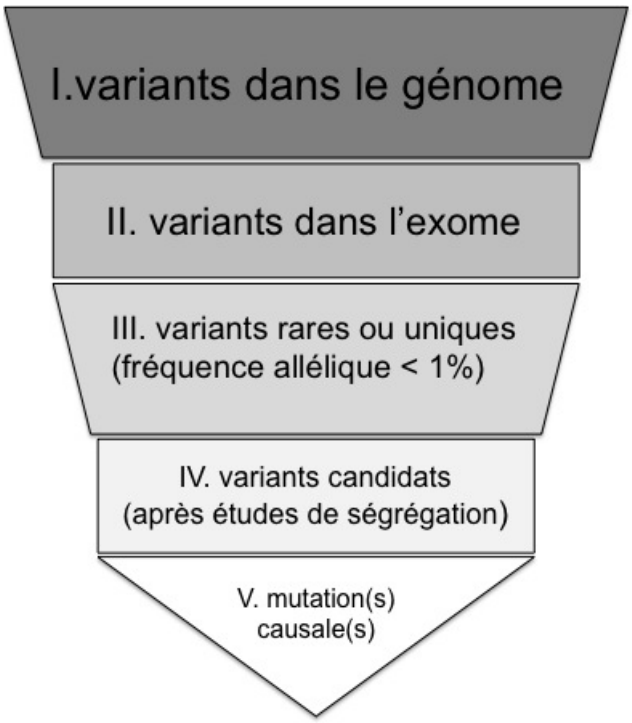


Figure I.6 Exemple d'approche de filtration de gènes.
Les filtres sont appliqués successivement.

Bien qu'elles aient grandement contribuées au diagnostic moléculaire, l'application des NGS est encore mise à rude épreuve; ces technologies sont encore en pleine effervescence , des améliorations doivent être apportées, afin de faire les progresser ces technologies et de les utiliser à leur plein potentiel.

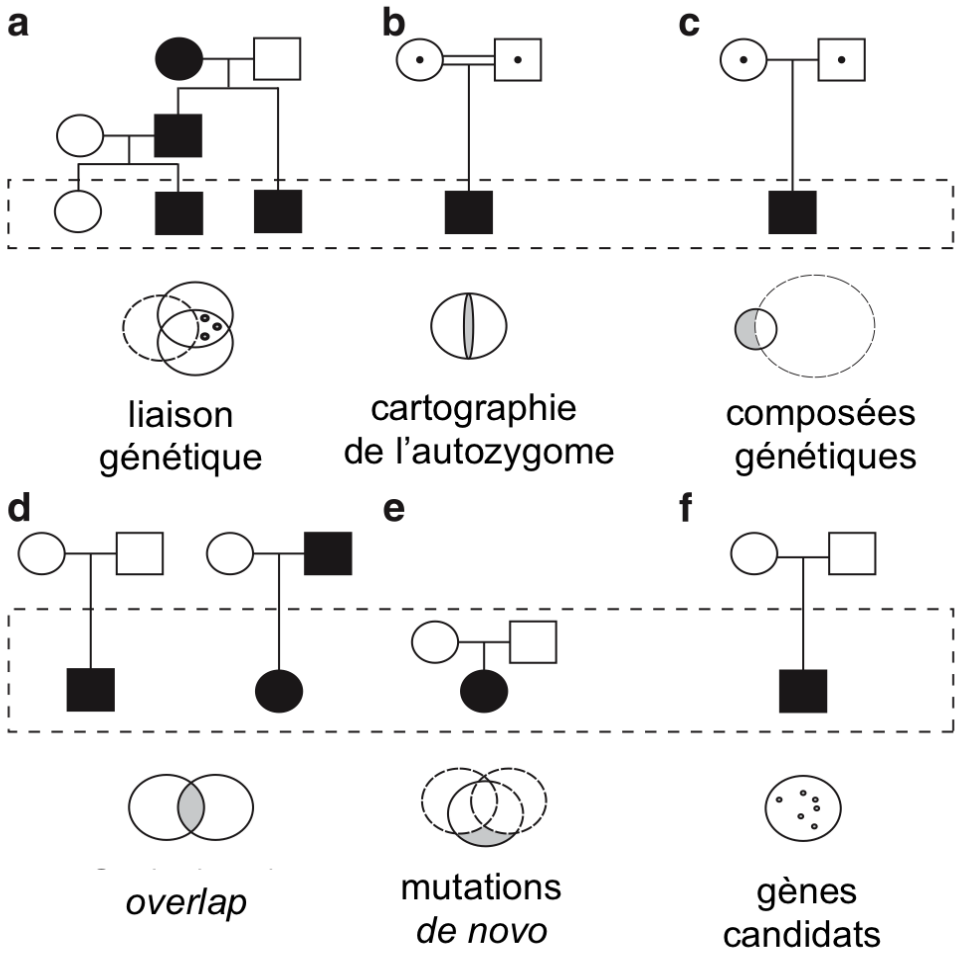


Figure I.7 Utilisation du séquençage d'exome dans le diagnostic clinique.

Modifiée à partir de Gilissen et al.,2012

Quelques stratégies utilisées pour identifier la cause génétique d'une maladie à l'aide du séquençage d'exome. **a et b.** Le séquençage d'exome est fait parallèlement à une approche de cartographie génétique : liaison génétique (**a**) ou cartographie de l'autozygome (**b**). L'exome d'un (ou de plusieurs) individu affecté est séquencé et la cause génétique de la maladie est identifiée parmi les variants contenus dans les régions candidates cartographiées. **b et c.** Stratégies utilisées pour cibler la cause

génétique de maladies récessives transmises selon un mode homozygote (**b**) ou composés génétiques (**c**). **c.** L'exome de l'individu affecté est séquencé, et l'attention est portée sur les gènes touchés par plus d'une mutation. Dans ce mode de transmission, deux mutations récessives causales (hétérozygote composites ou composées génétiques) sont attendues dans un gène, selon le modèle du *double hit*. **d et e**. Stratégies généralement utilisées pour cibler la cause génétique de maladies transmises selon un modèle dominant. **d.** L'exome d'au moins un individu affecté est séquencé ; la cause de la maladie est identifiée parmi les mutations partagées par les individus affectés séquencés. **e.** Dans cette approche, le séquençage d'exome est appliqué à l'analyse d'un trio composé d'un individu affecté et de ses parents. L'analyse est portée sur la recherche de mutations *de novo*, c'est à dire présentes uniquement chez l'individu affecté. Cette approche est généralement priorisée dans l'étude de maladies particulièrement sévères de sorte qu'elles ne peuvent être transmises aux générations futures soit parce qu'elles sont létales ou parce qu'elles affectent les capacités reproductives des individus. Les déterminants génétiques impliqués dans ce type de maladies sont généralement acquis *de novo*. **f.** L'exome de l'individu affecté est séquencé et l'analyse est axée sur une fraction de gènes candidats.

4. Les affections de la synapse

Les progrès technologiques ont grandement contribué à l'identification de nouveaux gènes de DI, et par le fait même aux connaissances sur les mécanismes impliqués dans la pathophysiologie de la DI. Globalement, les maladies neurodéveloppementales, incluant la, DI décrivent des affections de la synapse. Conceptuellement, les causes non-génétiques mises à part, l'étiologie de la DI peut être définie par des défauts survenant de manière précoce au cours du développement du cerveau et d'autres irrégularités apparaissant plus tard, après que les réseaux neuronaux aient été implantés.

4.1 Les malformations congénitales précoces du développement cortical

La DI est communément associée à des défauts dans la neurogenèse et/ou dans la mise en place du cortex cérébral. Le cortex est formé de plusieurs couches de cellules renfermant différents types de neurones et de cellules gliales. Ces cellules sont induites dans le neuroépithélium; puis elles se différencient et migrent vers leur destination finale, où elles sont disposées de manière à donner naissance au cortex cérébral proprement dit. La formation du cortex est dictée par des gradients de morphogènes et autres signaux extracellulaires exprimés par les progéniteurs neuronaux et les structures environnantes. Des dysfonctions dans le processus de formation et/ou de développement du cortex peuvent causer la DI.

4.1.1 Les défauts dans la prolifération et l'apoptose des cellules neuronales

Plusieurs syndromes associés à la DI sont causés par des anomalies dans le nombre de neurones et de cellules gliales ; elles mettent en évidence des défauts dans la prolifération

et/ou d'apoptose de ces cellules neuronales. Ces maladies peuvent être dues à une prolifération lente ou une apoptose élevée (microcéphalie), ou inversement à une prolifération accrue et une réduction dans l'apoptose (mégalencéphalie). Typiquement, on distingue :

- la **microcéphalie** causée par des défauts de la neurogenèse; typiquement des dysfonctions dans différents mécanismes du cycle cellulaire , incluant la réparation de l'ADN, la régulation de la transcription, la maturation et la duplication des centrosomes, l'activité des moteurs de dynéine et la formation du fuseau mitotique (Kumar et al., 2009a ; Thornton et Woods, 2009 ; Bilguvar et al., 2010 ; Shen et al., 2010 ; Alkuraya et al., 2011) . La microcéphalie est fréquemment observée chez les patients avec DI (Mahmood et al., 2011) .

- la **mégalencéphalie** mettant en évidence une défaillance dans les processus de croissance et de prolifération cellulaire. Des dysfonctions dans la voie de signalisation de mTOR sont mises en cause dans des cas de DI associés à une mégalencéphalie; des mutations dans des régulateurs de mTOR (PTEN [MIM:605309], (AKT3, PIK3R2, PIK3CA, [MIM:602501]) [MIM:603387] , STRADA) ont été identifiées dans des formes syndromiques et non-syndromiques de mégalencéphalie (Liaw et al., 1997 ; Puffenberger et al., 2007 ; Lee et al., 2012 ; Riviere et al., 2012).

4.1.2 Les défauts dans la migration des cellules neuronales

Les anomalies de la migration des neurones corticaux regroupent :

- l'**hétérotopie neuronale** caractérisée par la présence de neurones en dehors du cortex. Les hétérotopies peuvent être laminaires ou nodulaires selon leur aspect macroscopique et leur disposition dans la substance blanche.

- la **lissencéphalie** causée par des anomalies dans la stratification du cortex. On distingue les agyries ou les pachygryies selon qu'il y ait absence totale ou plutôt un petit nombre de circonvolutions à la surface du cortex. Le syndrome de Miller-Dieker causé par des mutations dans *LIS1* [MIM:247200] représente l'une des formes les plus fréquentes de lissencéphalie (Reiner et al., 1993 ; Lo Nigro et al., 1997) .

4.1.3 Les défauts dans l'organisation du cortex

D'autres maladies sont causées par des défauts survenant après la migration des cellules neuronales, lors de la mise en place du cortex. Typiquement, on distingue :

- la **micropolygyrie** caractérisée par une désorganisation de la disposition en couches du cortex, avec des lésions dans les couches corticales les plus profondes. On observe également

un repliement excessif des couches corticales, probablement causé par une fusion anormale des structures gyrales (Barkovich et al., 1999) . Les syndromes de Zellweger et d'Adams-Oliver décrivent des conditions avec une micropolygirie associée à la DI.

- la **schizencéphalie** marquée par une malformation dans le manteau cérébral ; il se crée une fente linéaire de matière grise (ou *porus*) à partir de l'épendyme des ventricules cérébraux jusqu'à la pie-mère (Sarnat et Curatolo, 2008) . Des cas familiaux de schizencéphalie ont été rapportés chez des patients avec des mutations germinales dans le facteur de transcription *EMX2* [MIM:269160] (Brunelli et al., 1996) .

- **les dysplasies corticales focales** associées à des anomalies dans la structure et l'organisation du cortex : neurones géants, cellules gliales ballonisées, épaississement du manteau cortical (Taylor et al., 1971) .

4.2 Les affections tardives de la synapse

Ces maladies de la synapse apparaissant après la neurogenèse et l'implantation des circuits neuronaux, en lien avec des problèmes au niveau de la survie et du fonctionnement des cellules neuronales. Une défaillance métabolique est généralement mise en cause dans ces affections tardives de la synapse. De fait, des dysfonctions au sein des « centrales métaboliques » de la cellule sont au cœur de ces maladies ; plusieurs maladies mitochondriales et peroxysomales sont mises en évidence dans les affections tardives de la synapse. Les mitochondries, par le biais de la production d'ATP, contrôlent différents processus indispensables au fonctionnement des neurones , notamment le trafic des vésicules synaptiques, la génération du potentiel d'action, l'assemblage du cytosquelette d'actine, la formation du cône de croissance, l'axonogenèse et la plasticité neuronale (Valenti et al., 2014). Les peroxysomes sont bien connus , entre autres, pour leur rôle dans la détoxification des neurones, la formation et le maintien des fibres de myélines (Lizard et al., 2012) . Les maladies mitochondriales et peroxysomales sont particulièrement sévères ; elles se manifestent généralement par des atteintes neurologiques majeures pouvant résulter en une forme de DI comme dans les syndrome de MELAS et d'Aicardi-Goutières. Les maladies mitochondriales traduisent des défauts dans la chaine respiratoire mitochondriale, tels que mis en évidence dans les encéphalopathies néonatales. Des mutations dans des gènes impliqués dans la biogenèse des peroxysomes (membres de la famille PEX) causent des maladies peroxysomales

associées à la DI comme le syndrome de Zellweger et la chondrodysplasie ponctuée rhizomélique de type 1 (Braverman et al., 1997 ; Crane, 2014) .

5. Voies de signalisation impliquées dans la déficience intellectuelle

Les NDs impliquent plusieurs centaines de gènes avec des fonctions biologiques diversifiées. Bien qu'il n'existe pas de gènes-type de NDs; certaines voies de signalisation cellulaire sont communément défectueuses dans la DI, l'autisme et l'épilepsie. Typiquement, ces voies mécanistiques mettent en évidence des dysfonctions au niveau des protéines synaptiques, leurs effecteurs et autres partenaires d'interaction (Figure I.8). Les notions acquises sur ces mécanismes cellulaires ont grandement contribué à l'élucidation des bases moléculaires des affections de la synapse. Parmi les mécanismes les plus communément mis en défaut dans la pathophysiologie de la DI ; les plus étudiés sont ceux qui impliquent :

5.1 La famille des petites GTPases

Les GTPases sont des protéines G caractérisées par leur habileté à lier et hydrolyser le GTP en GDP ; elles sont dans un état actif ou inactif, selon qu'elles sont liées au GTP ou au GDP. L'activité des GTPases est modulée par leurs protéines activatrices (GEFs - *Guanine Exchange Factor*) et inhibitrices (GAPs - *GTPase Activating Proteins* ; GDIs - *Guanine nucleotide Dissociation Inhbitors*) associées. Les GTPases et leurs modulateurs participent au fonctionnement du neurone, en intervenant notamment dans l'organisation du cytosquelette d'actine et l'enternalisation et le trafic de récepteurs membranaires ; des dysfonctions dans la voie de signalisation des GTPases causent la DI.

5.1.1 Les Rho GTPases

Des mutations dans *OPHN1* (MRXSO; [MIM:300486]) ont été identifiées chez des patients avec DI; ce gène code pour l'Oligophrenin-1, une **RhoA GTPase** jouant un rôle dans la maturation des synapses et la plasticité neuronale via la voie de signalisation des récepteurs AMPA (Billuart et al., 1998 ; Govek et al., 2004 ; Nadif Kasri et al., 2009).

La voie des **Rac/Cdc42/PAKs GTPases** est également associée à la pathophysiologie de la DI; des dysfonctions dans les GEF ARHGEF6 (MRX46; [MIM:300436]) et FGD1 [MIM:305400], des régulateurs de Cdc42, ont été observées chez des patients avec DI (Kutsche et al., 2000 ; Schwartz et al., 2000) . PAK3 (MRX30) ; [MIM:300558]) et LIMK, des effecteurs de la voie Rac/Cdc42/PAKs, sont également mis en cause dans la DI (Osborne

et al., 1996 ; Allen et al., 1998) . L'activité des Rac/Cdc42/PAKs est notamment requise dans la dépolymérisation des filaments d'actine médiée par la cofiline ; des défauts dans cette voie de signalisation altèrent notamment la morphologie des épines dendritiques et les processus de mémoire à long terme [LTP, *Long Term Potentiation*; Meng et al.,2002 ; Boda et al., 2004 ; Rex et al.,2009]. Typiquement, la voie Rac/Cdc42 promeut la croissance et la stabilisation des épines dendritiques ; à l'opposé la voie des RhoA GTPases réprimerait ces différents processus (Ramakers, 2002 ; Tashiro et Yuste, 2004) .

5.1.2 Les Ras GTPases

Des mutations dans SYNGAP1, une Ras/Rap GTPase, ont été identifiées chez des patients avec DI-NS modérée à sévère (Hamdan et al., 2009a) . SYNGAP1 est une composante importante de la densité post-synaptique, où elle régule l'activité des récepteurs ionotropiques activés par le glutamate (AMPA et NMDA). SYNGAP1 inhibe l'activation des récepteurs NMDA par la voie ERK; elle intervient également dans la LTP en modulant le trafic des récepteurs AMPA. SYNGAP1 prévient l'insertion des récepteurs AMPA dans la membrane synaptique, en inhibant la voie de Ras/ERK (Kim et al., 2005 ; Qin et al., 2005 ; Ligeti et al., 2012) . Des dysfonctions dans RSK2 (MRX19), un effecteur de la voie Ras/ERK, ont été identifiées chez des patients avec syndrome de Coffin-Lowry [MIM:303600], une forme syndromique de DI-X (Trivier et al., 1996 ; Merienne et al., 1999) .

5.1.3 Les Rab et les ARF GTPases

Des défaillances dans la voie des Rab GTPases sont également mises en cause dans la pathophysiologie de la DI. Par exemple, des mutations dans Rab39 et des régulateurs des Rab, comme la Rab-GDI GDI1 et la Rab-GAP TBCD124 , ont été identifiées chez des patients avec DI (D'adamo et al., 1998 ; Corbett et al., 2010 ; Giannandrea et al., 2010) . À l'instar de Rab39, les Rab sont au cœur du fonctionnement de la synapse, elles interviennent notamment dans le trafic des vésicules synaptiques et la migration neuronale (Kawauchi et al., 2010 ; Pavlowsky et al., 2012) . D'autres GTPases, participent également dans des processus neuronaux; ARF6, une ARF GTPase, intervient entre autres dans la différenciation neuronale, la formation des axones et la morphologie des épines dendritiques (Jaworski, 2007) . Des mutations dans IQSEC2 (MRX1; [MIM:309530]), une GEF de ARF6 ont été identifiées chez des patients avec DI-NS (Shoubridge et al., 2010) .

5.2 Les récepteurs de glutamates et leurs protéines associées

Les récepteurs ionotropiques activés par le glutamate et leurs partenaires d'interaction sont couramment impliqués dans la pathophysiologie de la DI. Par exemple, des mutations dans les sous-unités des récepteurs AMPA (GluR3; MRX94; [MIM:300699]), NMDA (NR2A, [MIM:245570]; NR2B (MRD6, [MIM:613970]) ou du kaïnate (GluR6, MRT6, [MIM:611092]) ont été mis en évidence chez des patients avec DI-NS (Motazacker et al., 2007 ; Wu et al., 2007 ; Endele et al., 2010). Les partenaires d'interaction de ces protéines sont également mutés dans la DI. C'est le cas notamment de DLG3 (MRX90 ; [MIM:300850]) et CASK ([MIM:300749]) qui sont des membres de la famille MAGUK (*Membrane Associated Guanylate Kinase*), impliquées dans l'ancrage des récepteurs ionotropiques du glutamate dans la fente post-synaptique des neurones excitateurs (Tarpey et al., 2004 ; Najm et al., 2008) .

5.3 Les molécules d'adhésion cellulaire

Les molécules d'adhésion cellulaire sont indispensables dans le maintien de l'intégrité du neurone, la formation des synapses et la plasticité synaptique. Différents types de molécules d'adhésion sont associées à la pathophysiologie de la DI ; c'est le cas notamment des neurexines (NRXN1, NRXN2), des neurégulines (NLGN3, [MIM:300425]; NLGN4, [MIM:300495]) et des cadhérines (CDH15 ,MRD3, [MIM:612580]); PCDH19, [MIM:300088]) (Jamain et al., 2003 ; Wu et al., 2007 ; Gauthier et al., 2011 ; Vincent et al., 2012 ; Yangngam et al., 2014) .

5.4 Le trafic des vésicules synaptiques

Les neurones communiquent par le biais de neurotransmetteurs, lesquels sont emmagasinés dans des vésicules synaptiques, avant d'être relâchés suite à un influx nerveux. Des dysfonctions dans des protéines impliquées dans le trafic et l'exocytose de ces vésicules sont mises en cause la DI. Par exemple, des mutations dans des membres de la famille des SNARE (*Soluble N-éthylmaleimide-sensitive-factor Attachment protein Receptor*), comme SNAP25 et STXBP1 [MIM:612164], ont été identifiées chez des individus avec DI et épilepsie (Hamdan et al., 2009b ; Rohena et al., 2013) .

5.5 Les régulateurs de la transcription génique

Des défaillances dans les mécanismes de régulation de transcription génique sont mises en évidence dans les NDs. Typiquement, ces voies de signalisation cellulaires impliquent:

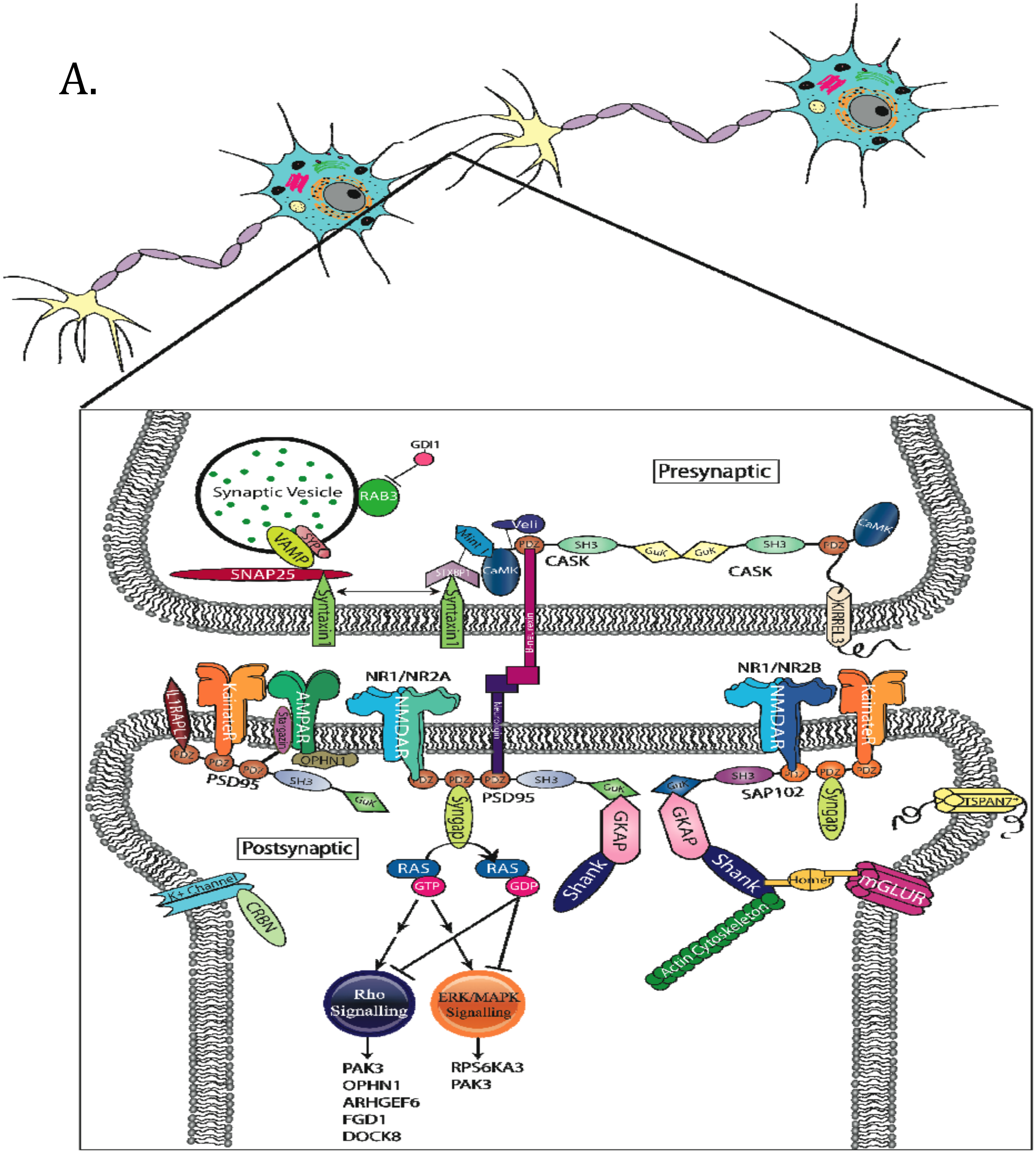
- les facteurs de transcription comme le gène *ARX* , un facteur de transcription *homeobox* , mis en cause dans des formes de DI sévère associées à l'épilepsie incluant les syndromes de West [MIM : 308350] et de Partington [MIM : 309510] (Stromme et al., 2002a ; Stromme et al., 2002b) . Les facteurs de transcription en structure doigt de zinc ou *zinc finger* (ZNF) sont également impliqués dans la pathophysiologie de la DI; des mutations dans ZNF41, ZNF81 (de la famille des KRAB - *Kruppel-associated box*) et ZNF764 ont été identifiées chez des patients avec DI (Shoichet et al., 2003 ; Kleefstra et al., 2004 ; Lugtenberg et al., 2006 ; Alesi et al., 2012) . Des dysfonctions dans d'autres types de facteurs de transcription peuvent également causer la DI; c'est le cas notamment de CC2D1A (MRT3 ; [MIM:608443]) , un répresseur des récepteurs de sérotonine et de dopamine impliqué dans l'activation de la voie NF-κB ; et de TRAPPC9 (MRT13; [MIM:613192]) également un activateur de la voie de signalisation de NF-κB (Basel-Vanagaite et al., 2006 ; Mochida et al., 2009) .

- les régulateurs épigénétiques, des protéines qui contrôlent l'expression génique, en modulant la structure de la chromatine. Des dysfonctions dans les mécanismes de régulation épigénitique ont été mises en évidence dans plusieurs syndromes associés à la DI; ces défaillances affectent différentes classes de protéines dont les histones acétyltransférases/déacétylases (CBP [MIM:600140], HDAC4 [MIM:600430]), les histones méthyltransférases/déméthylases (NSD1 [MIM:117550], JARID1C [MIM:300534]), les ADN méthyltransférases/déméthylases (DNMT3A [MIM:615879], DNMT3B [MIM:242860], les facteurs de liaison aux groupes méthyles de l'ADN (MECP2), les facteurs de remodelage de la chromatine (ATRX [MIM:600140],CHD7 [MIM:214800],) et autres protéines liant les histones (BRWD3 [MIM:300659],BCOR) (Xu et al., 1999 ; Petrij et al., 2000 ; Villard et al., 2000 ; Watson et al., 2001 ; Kurotaki et al., 2002 ; Winnepenninckx et al., 2002 ; Kalscheuer et al., 2003 ; Mnatzakanian et al., 2004 ; Ng et al., 2004 ; Vissers et al., 2004 ; Mari et al., 2005 ; Field et al., 2007 ; Williams et al., 2010 ; Tatton-Brown et al., 2014) .

5.6 Les régulateurs de la traduction de l'ARNm et de la synthèse protéique

La voie de signalisation de mTOR contrôle des processus essentiels à l'homéostasie de la cellule ; elle régule la transcription de l'ARNm, la biogenèse des ribosomes et la synthèse protéique. L'activité de la voie de mTOR est indispensable au fonctionnement des neurones ; elle module notamment la migration axonale, la formation des dendrites et les différentes formes de plasticité synaptique (Tang et al., 2002 ; Jaworski et al., 2005 ; Kumar et al., 2005) . Des mutations dans les régulateurs de la voie mTOR (TSC1 [MIM:191100]; TSC2 [MIM:613254]) causent la sclérose tubéreuse (TSC, *Tuberous Sclerosis*), une maladie multisystémique associée à la DI, l'autisme et l'épilepsie (European Chromosome 16 Tuberous Sclerosis, 1993 ; Van Slegtenhorst et al., 1997) . Le syndrome de l'X Fragile, causé par des mutations dans le gène *FMR1*, est l'une des causes les plus fréquente de DI ; il est caractérisé par des irrégularités dans les mécanismes de répression de la traduction de l'ARNm, en lien avec des mutations dans la protéine FMRP (Siomi et al., 1993) .

A.



GDI1
Synaptic Vesicle
RAB3
Presynaptic
VAMP
SNAP25
Syntaxin1
STXBP1
Mint 1
Veli
CaMK
PDZ
SH3
GuK
CASK
KIRREL3
NR1/NR2A
NR1/NR2B
L1RAPL1
KainateR
AMPAR
Stargazin
OPHN1
NMDAR
PSD95
SAP102
Syngap
GKAP
Shank
Homer
mGLUR
TSPAN7
Postsynaptic
K+ Channel
CRBN
RAS
GTP
GDP
Actin Cytoskeleton
Rho Signalling
ERK/MAPK Signalling
PAK3
OPHN1
ARHGEF6
FGD1
DOCK8
RPS6KA3
PAK3

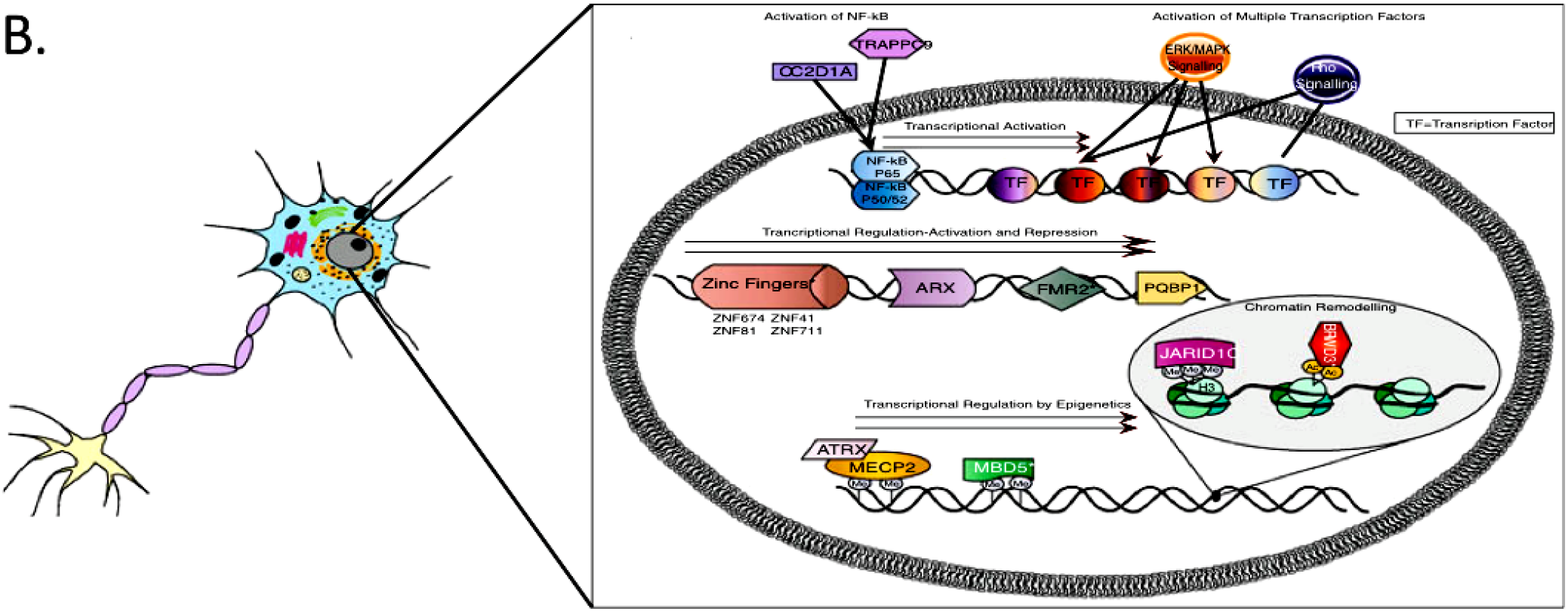


Figure I.8 Voies de signalisation impliquées dans les maladies de la synapse.

Modifiée à partir de Kaufman et al., 2010

A. Représentation schématique d'une synapse excitatrice mettant en évidence les récepteurs ionotropiques activés par le glutamate, leurs partenaires d'interactions et protéines effectrices associées. B. Mécanismes cellulaires régulant la transcription génique dans le noyau d'un neurone.

HYPOTHÈSE ET OBJECTIFS

Les bases génétiques de la DI sont difficilement approchables ; l'origine de la DI est inconnue dans 50-60% des cas diagnostiqués. En l'absence de facteurs environnementaux associés à la DI chez ces patients, il est possible d'envisager que des facteurs génétiques non-identifiés sont mis en cause dans ces cas de DI non élucidés. Les techniques de génétique classique (études de liaison et d'association génétique, approches de gènes candidats et séquençage PCR-Sanger) sont mises à rude épreuve dans l'étude de la DI. Les NGS, en l'occurrence le WES, offrent la meilleure alternative, afin d'explorer les bases génétiques de maladies complexes comme la DI. L'hypothèse de ce projet de recherche est que, en rendant possible la recherche directe de mutations dans les régions codantes du génome humain, le WES va permettre d'explorer les bases génétiques de la DI dans de petites familles. Ce projet porte sur l'étude de vingt familles avec DI syndromique ou non-syndromique ; nous avons concentré nos analyses sur des familles où une transmission récessive de la maladie est suspectée. Les individus affectés analysés ont subi des analyses préliminaires, afin d'exclure la possibilité que la DI soit causée par des anomalies chromosomiques détectables en cytogénétique. Les résultats de cette étude sont présentés dans les chapitres suivants.

CHAPITRE II: Identification and biochemical characterization of a novel mutation in DDX11 causing Warsaw Breakage Syndrome

Publié dans: *Human Mutation*, 34(1): 103-7, 2013

Identification and biochemical characterization of a novel mutation in DDX11 causing Warsaw Breakage Syndrome.

José-Mario Capo-Chichi,[1] Sanjay Kumar Bharti,[2] Joshua A. Sommers,[2] Tony Yammine,[3] Eliane Chouery,[3] Lysanne Patry,[1] Guy A. Rouleau,[4] Mark E. Samuels,[1,5] Fadi F. Hamdan,[1] Jacques L. Michaud,[1] Robert M. Brosh Jr,[2*] André Mégarbane,[3,6] and Zoha Kibar[1,7 *]

[1] Center of Excellence in Neuroscience of Université de Montréal, Centre de Recherche du CHU Sainte-Justine, Montréal, Canada
[2] Laboratory of Molecular Gerontology, National Institute on Aging, NIH Biomedical Research Center, Baltimore, Maryland
[3] Unité de Génétique Médicale et laboratoire associé Inserm UMR_S910, Université Saint Joseph, Beirut, Lebanon
[4] CHUM Notre-Dame Hospital Research Center and the Department of Medicine, Center of Excellence in Neuroscience of Université de Montréal, Montreal, Canada
[5] Department of Medicine, University of Montreal, Montreal, Canada
[6] Institut Jérôme Lejeune, Paris, France; [7]Department of Obstetrics and Gynecology, Université de Montréal, Montreal, Canada

***Correspondence to**:

1) Zoha Kibar,
CHU Sainte Justine Research Center,
3175 Côte-Ste-Catherine, Room A711,
Montreal, QC H3T 1C5, Canada.

2) Robert M. Brosh Jr,
Laboratory of Molecular Gerontology, National Institute on Aging, NIH,
NIH Biomedical Research Center,
251 Bayview Blvd, Baltimore, MD 21224, USA.

Conflict of interest: none declared

Key Words: *DDX11*; Warsaw breakage syndrome; WABS; helicase

Contribution:

J'ai effectué les analyses génétiques ayant permis d'identifier la mutation causale (c.G788A :p.R263Q dans le gène *DDX11*) dans cette famille. J'ai également validé génétiquement l'implication de cette mutation dans la pathologie de la famille. Bharti SK, Sommers JA ont validé biologiquement l'impact de la mutation p.R263Q sur l'activité de DDX11. Yammine T et Chouery E ont réalisé les expériences montrant la cassure chromosomique dans les lymphocytes des patients. Mégarbane A a caractérisé le phénotype des patients de la famille à l'étude. Lysanne P, Rouleau GA, Samuels ME, Hamdan FF, Michaud JL et Kibar Z ont aidé dans le *design* des expériences et la supervision de la partie génétique des travaux. J'ai rédigé la première version du manuscrit, Hamdan FF, Michaud JL, Brosh Jr RM, Mégarbane A et Kibar Z ont également contribué à la rédaction et à la révision du manuscrit.

ABSTRACT

Mutations in the gene encoding the iron–sulfur-containing DNA helicase DDX11 (ChlR1) were recently identified as a cause of a new recessive cohesinopathy, Warsaw breakage syndrome (WABS), in a single patient with severe microcephaly, pre- and postnatal growth retardation, and abnormal skin pigmentation. Here, using homozygosity mapping in a Lebanese consanguineous family followed by exome sequencing, we identified a novel homozygous mutation (c.788G>A [p.R263Q]) in *DDX11* in three affected siblings with severe intellectual disability and many of the congenital abnormalities reported in the WABS original case. Cultured lymphocytes from the patients showed increased Mitomycin-C induced chromosomal breakage, as found in WABS. Biochemical studies of purified recombinant DDX11 indicated that the p.R263Q mutation impaired DDX11 helicase activity by perturbing its DNA binding and DNA-dependent ATP hydrolysis. Our findings thus confirm the involvement of *DDX11* in WABS, describe its phenotypical spectrum, and provide novel insight into the structural requirement for DDX11 activity**.**

INTRODUCTION

Warsaw breakage syndrome (WABS; MIM# 613398) is a new form of cohesinopathy showing defects in sister chromatid cohesion and hypersensitivity to chemicals that induce replication stress, thus combining distinct cytogenetic features seen in Roberts syndrome (MIM# 268300) and Fanconi anemia (MIM# 227650), respectively (Van Der Lelij et al., 2010b)). WABS has been reported in a single individual with several congenital abnormalities, including severe pre- and postnatal growth retardation, microcephaly, facial dysmorphy, deafness, ventricular septal defects, mild intellectual disability (ID), and abnormal skin pigmentation. This patient carried compound heterozygous mutations in *DDX11* (MIM# 601150; NM_030653.3), including a splice site mutation (c.2271+2T>C, previously reported as IVS22+2T>C) and a 3-bp in-frame C-terminal deletion (c.2689_2691del [p.K897del]) that was recently shown to abrogate the DDX11 helicase activity(Van Der Lelij et al., 2010a ; Wu et al., 2012) So far, no other patients with *DDX11* mutations have been identified.

DDX11 (ChlR1), an orthologue of the yeast Chl1, is a member of the superfamily 2 (SF2) of ATP-dependent DEAH-box DNA helicases (Hirota et Lahti, 2000 ; Skibbens, 2004). DDX11 shares sequence homology with the related SF2 DNA helicases FANCJ (MIM# 609054), ERCC2 (XPD; MIM# 126340), and RTEL1 (MIM#608833), which all contain an iron–sulfur (Fe–S) motif between helicase domains IA and II (Rudolf et al., 2006 ; Wu et al., 2009) FANCJ and ERCC2 (also known as XPD) are also implicated with genetic instability disorders in humans, and RTEL1 had been suggested to play a role in the maintenance of telomere length and genome stability in mice (Lehmann, 2001 ; Ding et al., 2004 ; Levitus et al., 2005) In human cells, DDX11 was shown to interact with components of the cohesin complex and play a role in sister chromatid cohesion (Parish et al., 2006).

RESULTS

Here, we describe the identification and biochemical characterization of a novel deleterious homozygous mutation in *DDX11* in a consanguineous Lebanese family showing many of the symptoms associated with WABS. This family, which was recruited for this study with the approval of our institutional ethics committee, consists of two healthy first cousins once removed and their three affected children (Fig. II.1). We first performed whole-genome SNP genotyping in two affected siblings (patients V-1 and V-3) by using the Illumina Human 610 Genotyping BeadChip panel, which interrogates 620,901 SNPs, and we used PLINK (v.1.06) to search for homozygosity regions (HR) containing >30 consecutive SNPs and extending over >1 MB. We identified 10 candidates HR shared by the two siblings (Supp. Table SII.1). To find potential causative mutations in these regions, we performed exome capture (Agilent SureSelect 50 Mb kit) and sequencing (paired-end SOLiD4) on one of the affected siblings (patient V-1) and obtained an average per target base coverage of 25×, with at least 85% of the target region being covered at >= 3×. Exome capture, read mapping as well as variant calling and annotation were performed as previously described (Daoud et al., 2012) as noted in the Supporting Information. In total, 148 homozygous non synonymous coding and splicing variants were identified in the HR, of which only two were not found in our in-house set of 198 control exomes (from 30 healthy individuals and from 168 patients with other rare diseases, including amyotrophic lateral sclerosis, essential tremor, aneurysm, and hereditary spastic paraplegia) or not reported at frequencies > 0.5% in public SNP databases (dbSNP135, 1000 genomes and EVS Exome Variant Server). Sanger sequencing confirmed that only one of these two variants, a missense mutation in *DDX11* (c.788G>A [p.R263Q]; NM_030653.3) present in the largest HR (chr12: 33.6 Mb), was heterozygous in the parents and homozygous in the three affected siblings (Fig. II.1; Supp.Table SII.2). This variant was submitted to a gene-specific database (http://www.lovd.nl/DDX11). We also genotyped 150 individuals from the Middle East, including 100 Lebanese and 50 Palestinians, and did not identify a single carrier of the *DDX11* p.R263Q mutation, indicating that it is rare. This mutation, predicted to be damaging by various in silico algorithms (SIFT, Polyphen-2, Mutation Taster),

affects a conserved residue in the Fe–S domain, which is also conserved in the other related SF2 helicases (FANCJ, ERCC2 [XPD], and RTEL) (Rudolf et al., 2006 ; Wu et al., 2009) (Fig. II.1E). Interestingly, the Fe–S domain was shown to be essential for the helicase activity of FANCJ and XPD (Rudolf et al., 2006 ; Wu et al., 2010). Moreover, a mutation of the homologous arginine residue in ERCC2 (XPD) (c.335G>A, p.R112H; NM_000400.3), which causes trichothiodystrophy (TTD; MIM# 601675), results in a loss of the helicase activity, a concomitant defect in nucleotide excision repair, and a reduction in the repair/transcription factor TFIIH (GTF2H2; MIM# 601748), all indicating that this arginine residue (R263 in DDX11; R112 in XPD) is functionally important (Botta et al., 2002 ; Dubaele et al., 2003) .

Metaphase cells from lymphocytes of the two affected siblings (patients V-1 and V-3) showed an increase in chromosomal breakage (86% and 68%, respectively, compared to 10% for the healthy control) after treatment with the DNA cross-linking agent Mitomycin - C (MMC; see Supporting Information for methods) (Fig.II.1G). Of these cells with MMC-induced chromosomal breaks, 46.5% and 44% in patients V-1 and V-3, respectively, showed centromeric heterochromatin repulsion (“railroads”) and premature chromatid separation. These data are consistent with observations in similarly treated cells from the original WABS patient (Van Der Lelij et al., 2010a). Patient V-2 could not be studied at the cytogenetic level.

We next investigated whether the p.R263Q mutation affects DDX11 activity. For this, we purified recombinant human hexa-histidine-tagged wild-type (WT) and mutant DDX11 (DDX11-R263Q) from transfected HEK293T and assayed the proteins activities (see Supporting Information for detailed methods) (Fig. II.2). We first examined DNA unwinding activity catalyzed by DDX11-WT and DDX11-R263Q mutant. Using a forked 19-bp duplex DNA substrate, DDX11-WT unwound the substrate in a protein concentration dependent manner, as was previously shown (Wu et al., 2012) (Fig. II.2B). In comparison, DDX11-R263Q unwound the forked duplex substrate much less efficiently (Fig. II.2C). Quantitative analysis of the helicase data showed that 0.06 nM DDX11-WT unwound approximately 50% of the substrate whereas DDX11-R263Q displayed no detectable helicase activity. At 0.47 nM protein concentration, DDX1-R263Q unwound only approximately 10% of the substrate, whereas DDX11-WT unwound the substrate to

near completion (Fig. II.2D). We also tested the ability of the mutant DDX11 protein to unwind a two-stranded antiparallel G-quadruplex (G4) DNA substrate (OX-1-G2'), as we have previously performed for DDX11-WT (Wu et al., 2012) . As shown in Supp. Figure S1, DDX11-R263Q unwound the G4 substrate much less efficiently than DDX11-WT. On the basis of these results, we conclude that the R263Q mutation negatively affects DDX11 helicase activity on both duplex and G4DNA substrates.

The reduced helicase activity of DDX11-R263Q might reflect an impaired ability to bind DNA. To address this issue, we performed electrophoresis mobility shift assays (EMSA) with the mutant and WT proteins in the absence of ATP using the same radiolabeled forked duplex DNA used for the helicase assays. DDX11-WT bound the forked duplex in a protein concentration dependent manner with approximately 75% of the substrate bound at a protein concentration of 1.2 nM (Fig. 2E). In contrast, little or no detectable DNA binding was observed at 1.2 nM DDX11-R263Q (Fig. II.2F). Increasing the protein concentration to 2.4 nM resulted in nearly complete binding of the forked duplex by the WT DDX11, whereas only approximately 3% of the substrate was bound by DDX11-R263Q at the 2.4 nM protein concentration. From these results, we conclude that the p.R263Q mutation negatively affects the ability of DDX11 to bind DNA.

We next examined the DNA-dependent ATPase activity of DDX11-R263Q compared with DDX11-WT in the presence of M13ss DNA circle as the DNA effector (Supp. Table S3). The *K*m for ATP binding by DDX11-R263Q was approximately three fold lower than that of DDX11-WT (0.19 ± 0.05 mM vs. 0.66 ± 0.04 mM, respectively), suggesting that the mutant protein bound ATP more favorably. From the Lineweaver–Burke analysis, however, it was evident that the DDX11-R263Q mutant showed a *V*max that was approximately 8.5-fold lower than that of DDX11-WT (0.066 ± 0.008 nmol min^{-1} vs. 0.56 ± 0.05 nmol min^{-1}, respectively). A kinetic analysis of ATPase activity demonstrated that the *k*cat value for DDX11-R263Q was 18-fold lower than that of DDX11-WT (12 ±4min–1 vs.220 ± 40 min–1, respectively), confirming that the mutant protein was defective in its ability to hydrolyze ATP. These results suggest that the R263Q mutation negatively affected the ability of DDX11 to perform DNA-dependent ATP hydrolysis, a result that is consistent with its DNA-binding defect and poor ability to unwind DNA substrates efficiently unwound by WT DDX11. The modestly reduced *K*m

value for DDX11-R263Q compared to WT DDX11 may reflect a reduced ability of the mutant enzyme to turnover ATP. Collectively, these biochemical analyses of mutant DDX11 indicate that the p.R263Q mutation impairs the helicase activity of DDX11 by perturbing its DNA binding and DNA-dependent ATP hydrolysis. The reduced ability of the DDX11-R263Q mutant protein to bind DNA may be consistent with structural and biochemical data that indicate a role of the Fe–S cluster in the sequence-related XPD helicase in the interaction of the protein with DNA (Kuper et al., 2012 ; Pugh et al., 2012).

Members of the consanguineous family described here originate from North Lebanon. There was no history of birth defects, or increased miscarriage rate reported by this family and known relatives. All sibs were born at 32 weeks of gestation. No known toxic, medical exposure, or unusual events were reported during the gestations.

Patient V-1 was first seen when she was 20 months old. At birth, her weight was 1,500*g*, her length was 35 cm, and her head circumference was 27cm (all values far below the third percentile). Fontanelles were closed at the age of 3 months, and she walked unhelped at the age of 14 months. Upon examination, her weight was 7.5 kg, height was 74 cm, and head circumference was 36.5 cm (all values below the third percentile). She showed severe ID. She had a small and receding forehead, short nose, small nares, short neck, clinodactyly of the fifth fingers, and a single palmar crease on both hands. Neurological examination revealed a hypotonic girl. Abdominal ultrasonography and echocardiography did not show any abnormalities. CT imaging of the brain revealed rather small and rounded cochlea without visible cochlear turns or spirals. Bilateral sensorineural deafness was diagnosed by auditory evoked potential.

Patient V-2 was first evaluated by us at the age of 1 year for severe delayed milestones. At birth, his weight was 1,500*g*, length was 38 cm, and head circumference was 25 cm (all values below the third percentile). An echocardiogram evidenced a tetralogy of Fallot. Bilateral sensorineural deafness was diagnosed by auditory evoked potential at 6 months. His fontanelles were closed at the age of 7 months. At examination, his weight was 6.2 kg, height was 64 cm, and head circumference was 31 cm (all values below the third percentile). Physical examination showed the same dysmorphic features as those found in his sister with, in addition, prominent cheeks, flat philtrum, microretrognathism, and strabismus. He also showed severe ID. Abdominal

ultrasonography did not show any abnormalities. He died at the age of 4 years because of heart failure. Patient V-3 showed the same dysmorphic features as her siblings, as well as sensorineural deafness and severe ID but no cardiac malformations.

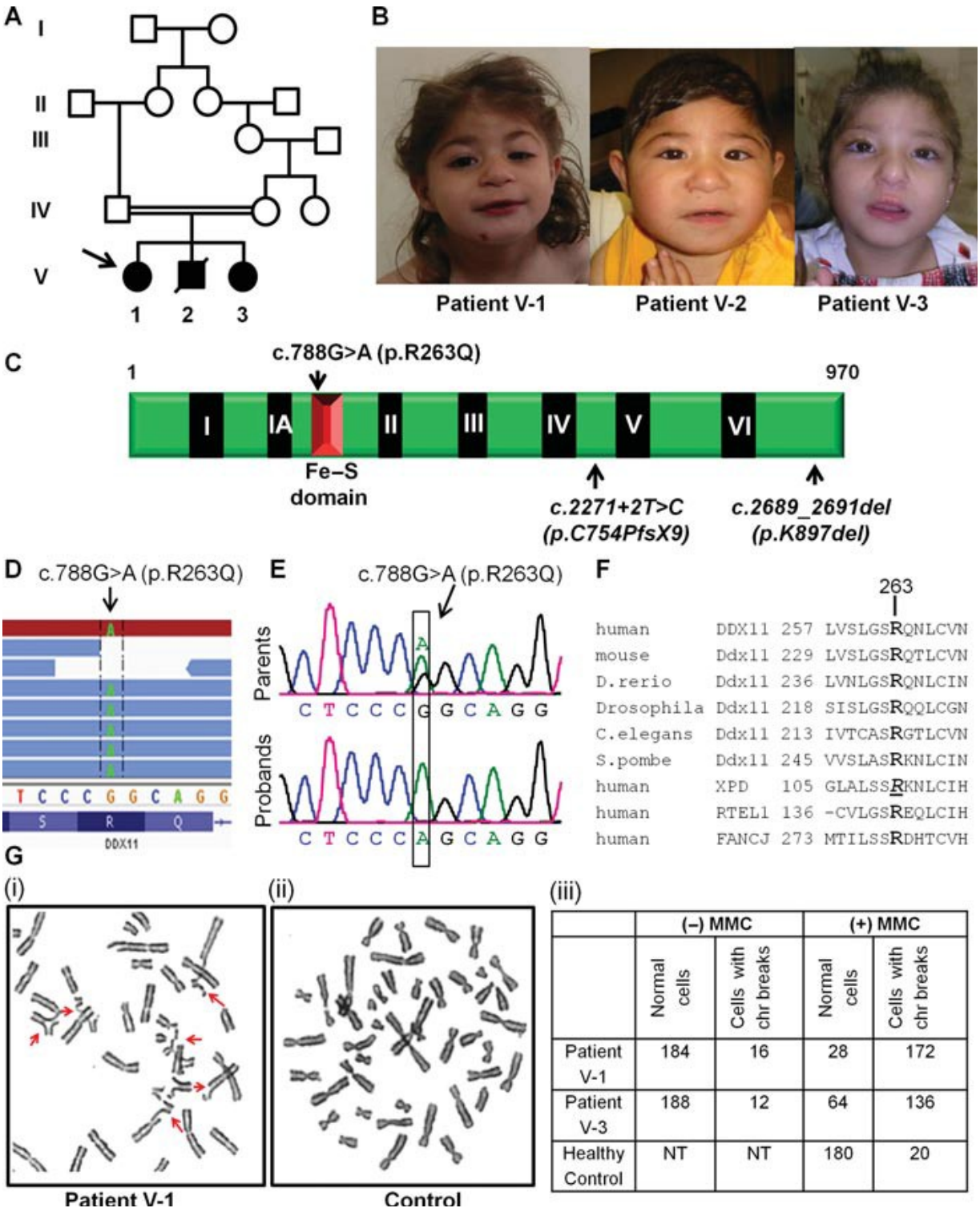
DISCUSSION

The phenotypic profiles of our patients with the mutation p.R263Q in DDX11 overlap in presentation and severity with that of the previously described WABS patient who carried compound heterozygous mutations in *DDX11* (Supp. Table SII.4) (Van Der Lelij et al., 2010a). They showed ID growth retardation, and severe congenital abnormalities including microcephaly, facial dysmorphism, deafness due to cochlear abnormalities (in two of the sibs), and cardiac malformations (in one of the sibs). Unlike the previously reported WABS patient who had mild ID and skin pigmentation abnormality, the affected sibs reported herein had severe ID and no skin pigmentation abnormality, differences that may be attributed to the type of DDX11 mutations, and differences in genetic background. It is of interest that the DDX11 protein with the p.R263Q substitution in its Fe–S domain retained partial (albeit severely compromised) helicase activity at higher protein concentrations, whereas the previously characterized DDX11-K897del protein lacking a single lysine in the C-terminal end was completely inactive in DNA unwinding (Wu et al., 2012). Our current findings demonstrate that the integrity of the Fe–S domain is critical for ATPase-driven DNA unwinding by DDX11. In this regard, a clinically relevant missense mutation in the conserved Fe–S domain of FANCJ interfered with its DNA helicase activity as well (Wu et al., 2010). Future studies of clinically relevant mutations in Fe–S helicases should help elucidate their molecular and cellular defects.

In conclusion, we describe here the second WABS family with mutations in *DDX11*, and the first with multiple affected individuals who all carry a homozygous missense mutation in the Fe–S domain of DDX11 that impairs the protein's activity. Our data confirm that *DDX11* mutations cause WABS and provide additional insights into the clinical phenotypes and functional consequences associated with pathogenic *DDX11* mutations.

Acknowledgments

J.L.M. is a National Scientist of the Fonds de Recherche en Santé du Québec. J.M.C. holds a salary award from the Réseau de Médecine Génétique Appliquée du Québec (RMGA). We thank the patients and their family for participating in this study. We also thank the members of the RMGA bioinformatics team (Alexandre Dionne-Laporte, Dan Spiegelman, Edouard Henrion, and Ousmane Diallo) for the primary analysis of the exome sequencing data. *Disclosure Statement*: The authors confirm that they have no conflicts of interest.



	(–) MMC		(+) MMC	
	Normal cells	Cells with chr breaks	Normal cells	Cells with chr breaks
Patient V-1	184	16	28	172
Patient V-3	188	12	64	136
Healthy Control	NT	NT	180	20

Figure II.1

Figure II.1 A pathogenic mutation identified in DDX11 in a family affected with WABS.

A, **B:** Pedigree of the consanguineous Lebanese family studied here and their facial features. The proband is indicated by an arrow. Patient V-2 was deceased hence certain analyses could not be extended to biological sampling of this child.
C: Schematic representation of DDX11 showing the helicase motifs (I–VI) as well as the Fe–S motif of DDX11. p.R263Q identified herein as well as the two previously published DDX11 mutations (c.2271+2T>C and p.K897del) are shown.
D: Integrated Genomics Viewer (IGV) tracks of exome sequencing reads showing the presence of the homozygous DDX11 c.788G>A (p.R263Q) mutation in patientV-1.
E: Sanger sequencing validation of c.788G>A (p.R263Q) DDX11 mutation. Representative chromatograms of the parents (both heterozygous forc.788G>A) and the three probands (all homozygous for c.788G>A) are shown. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon.
F: Amino acid alignment of the beginning of Fe–S domain region of DDX11 in different species as well as the corresponding regions in the human-related DNA helicases: XPD (ERCC2), FANCJ, and RTEL1. R263 in human DDX11 is well conserved (in bold). R112 residue in human XPD whose mutation (c.335G>A, p.R112H) causes trichothiodystrophy is italicized and underlined.
G: Mitomycin C (MMC)-induced chromosomal breakage. Representative metaphase from blood cultures treated with MMC showed increased chromosomal defects (arrows), including breakage, radial formations, centromeric heterochromatin repulsion (“railroads”) from patient V-1 (i), in contrast to the metaphases prepared from a similarly treated blood cultures prepared from a healthy control individual (ii). Table showing the representation of cells with chromosomal (chr) breaks in presence or absence of MMC. (NT, not tested.) A total of 200 cells (metaphases) were screened for each individual (iii).

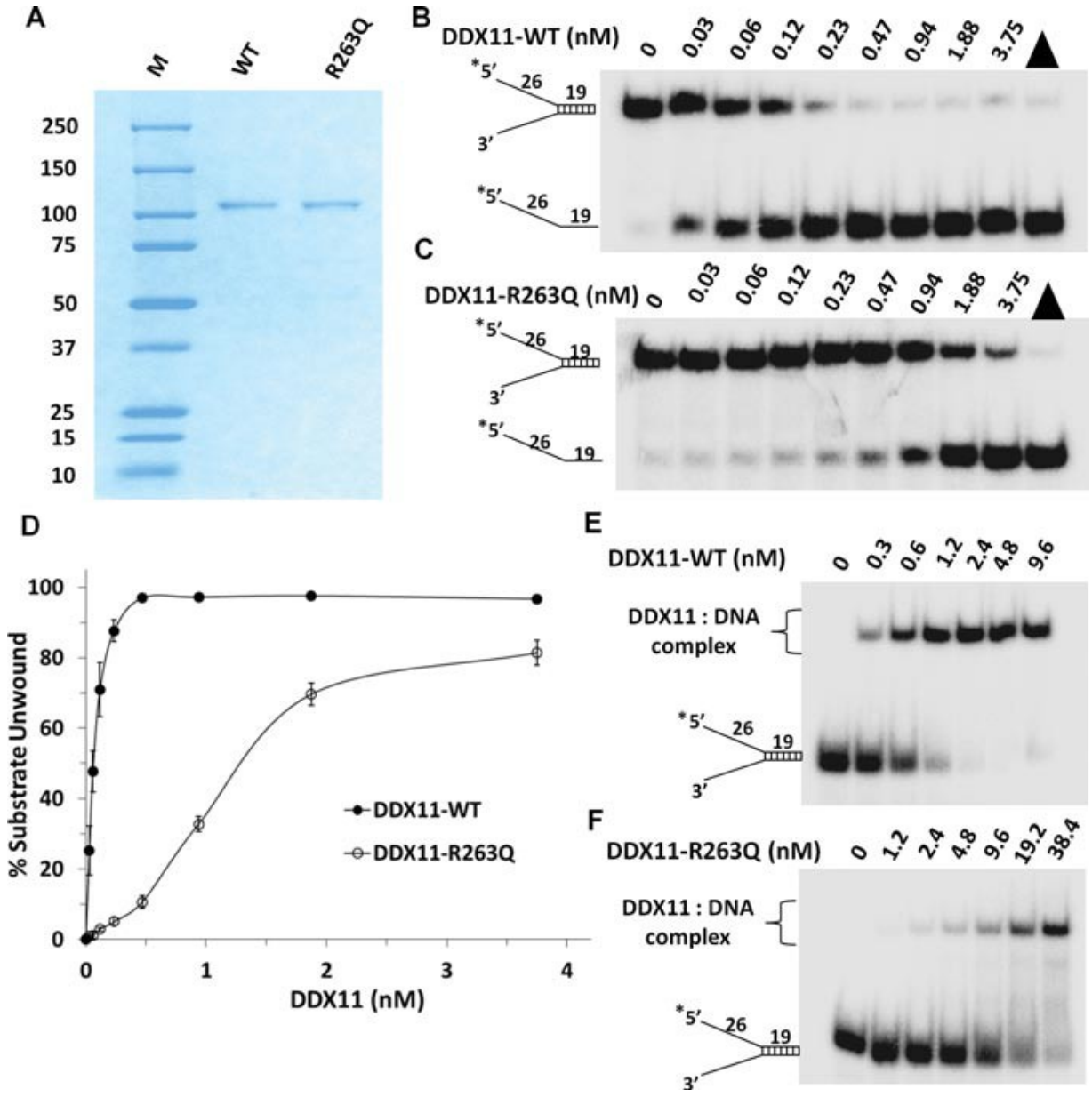

A
M
WT
R263Q
250
150
100
75
50
37
25
15
10
B
DDX11-WT (nM)
0
0.03
0.06
0.12
0.23
0.47
0.94
1.88
3.75
*5′
26
19
3′
C
DDX11-R263Q (nM)
D
% Substrate Unwound
DDX11 (nM)
DDX11-WT
DDX11-R263Q
E
DDX11-WT (nM)
0
0.3
0.6
1.2
2.4
4.8
9.6
DDX11 : DNA complex
F
DDX11-R263Q (nM)
0
1.2
2.4
4.8
9.6
19.2
38.4

Figure II.2

Figure II.2 Purification and biochemical analyses of recombinant DDX11-WT and DDX11-R263Q for helicase activity and DNA binding.

A: The purity of DDX11-WT and DDX11-R263Q was evaluated by Coomassie stained SDS PAGE (predicted MW approximately 110 kD). (M, molecular weight protein standards.)

B, C: Helicase activity of DDX11-WT (B) and DDX11-R263Q (C) on forked duplex DNA substrate. Helicase reactions (20 μl) were performed by incubating the indicated DDX11 protein with 0.5 nM forked duplex DNA substrate at 37◦C for 15 min under standard helicase assay conditions as described in the methods section in the Supporting Information. Products were resolved on native 10% polyacrylamide gels and representative phosphor images are shown. Triangle denotes heat-denatured DNA substrate control, and asterisk denotes 5-^{32}P end label.

D: Quantitative analysis of data from helicase activity of DDX11-R263Q and DDX11-WT on forked duplex DNA substrate is shown. Data represent the mean of at least three independent experiments with standard deviations (SD) indicated by error bars.

E, F: DNA binding by DDX11-WT andDDX11-R263Q as detected by gel mobility shift assays. The indicated concentrations of DDX11-WT (E) and DDX11-R263Q (F) were incubated with 0.5 nM forked duplex DNA substrate on ice for 30 min under standard gel shift assay conditions as described in the Supporting Information. The DNA–protein complexes were resolved on native 5% polyacrylamide gels. Representative phosphor images for DNA-binding assays are shown. Asterisk denotes 5-^{32}P end label

SUPPLEMENTARY SUBJECTS AND METHODS

We recruited a consanguineous Lebanese family comprised of 2 healthy parents and 3 affected siblings showing intellectual disability and various congenital anomalies. Blood samples were obtained from each of the family members with the informed consent of the parents and the approval of the appropriate institutional ethical review board.

Whole-genome genotyping and homozygosity mapping.

Genomic blood DNAs (~2.5□g) from 2 of the affected siblings (IV-1 and IV-3) were used for whole-genome genotyping on the Illumina 660 Quad Chip (Infinium HD). Homozygosity mapping was performed using PLink (v.1.06; http://pngu.mgh.harvard.edu/~purcell/plink/).

Exome capture and massive parallel sequencing.

The exome of one affected sibling (patient V-1) was captured from genomic DNA (3ug), using the Agilent SureSelect Human All Exon Capture 50 Mb Kit, and sequenced (paired-end, 25x 50 bp) using the ABI SOLiD4 (Life Technologies; 6 exomes/slide) at the Child Health Genomics Platform at the Sainte-Justine Hospital Research Center (Montreal, Qc, Canada).

Read mapping and variant analysis.

Sequence reads were mapped (hg18) using the SOLiD Bioscope. After removing PCR duplicates (Picard's MarkDuplicates; http://picard.sourceforge.net/index.shtml), single-nucleotide variants and small insertions/deletions were called by Bioscope's DiBayes and Small Indels tool, respectively (base coverage >= 3x; at least 15% read variant frequency) and annotated using ANNOVAR (Wang et al., 2010).

Drug-induced chromosomal breakage.

Whole-blood cultures (2 dishes/individual) were prepared from 2 affected patients (V-1, V-3) and a healthy control. After incubating for 48 hrs at 37 ºC, the blood cultures (10 mL) were treated with either 30 µl of 1 µg/ mL Mitomycin C (MMC; Sigma) or the vehicle DPBS (Dulbecco's Phosphate-Buffered Saline) and incubated for additional 24 hrs at 37 ºC, followed by a treatment (2 hrs, 37ºC) with 0.1 mL of 10 µg/ mL colchicine (Invitrogen), followed by incubating with a hypotonic shock medium (2 mL fetal bovine serum + 10 mL distilled water) for 7 minutes. The cells were fixed, mounted on a slide, and stained for 12 minutes in a 4% Giemsa (Merck). For each culture, 200 metaphases were analyzed.

DDX11 site-directed mutagenesis.

An DDX11-R263Q mutation was introduced into the 6X His-pCDNA3-3XFLAG plasmid DNA using mutagenic primers (sense: 5'-GGT-CTC-CCT-TGG-CTC-CCA-GCA-GAA-CCT-TTGTG-3' antisense: 5'- CAC-AAA-GGT-TCT-GCT-GGG-AGC-CAA-GGG-AGA-CC-3') and a standard protocol from Quickchange II XL site-directed mutagenesis kit (Stratagene) by Lofstrand labs (Gaithersburg, MD). The open reading frame was sequenced to verify the mutation. The human DDX11 cDNA used herein was provided by the late Dr. Jill M. Lahti (Department of Genetics and Tumor Cell Biology, St. Jude Children's Hospital, Memphis, TN), and requests can be made to Dr. Akira Inoue (St. Jude Children's Hospital).

Recombinant DDX11 protein purification.

DDX11-WT and DDX11-R263Q proteins were purified using a protocol previously described (Wu et al., 2012). Briefly, 6X His-pCDNA3-3XFLAG plasmid containing human DDX11-WT and DDX11-R263Q cDNA was transfected into 293T cells using Lipofectamine 2000 (Invitrogen) as recommended by the manufacturer. Ten to twenty 10-cm plates were grown to ~80% confluence for transfection. Cells were harvested 54 h after transfection by trypsinization and centrifugation. Pellets were washed with cold PBS and cold PBS with proteinase inhibitors (Roche Applied Science) sequentially, resuspended in 5 mL of hypotonic buffer (10 mM Tris HCl pH 7.4, 10 mM KCl, 1.5 mM MgCl2, 1 mM dithiothreitol (DTT), 0.5 mM PMSF, proteinase inhibitors) and incubated on ice for 15 min. Cells were lysed by Dounce homogenization (30 strokes) and centrifuged at 4 °C for 30 min at 2500 x g to separate nuclear pellets from cytosolic fraction. The cytosolic fraction was collected and kept on ice. The nuclear pellets were suspended in 5 mL of nuclear buffer (20 mM Tris HCl, pH 7.4, 0.15 M NaCl, 1.5 mM MgCl2, 10% glycerol, 0.2 mM EDTA, 1 mM DTT, 0.5 mM PMSF and proteinase inhibitors) and incubated in rocking at 4 °C for 30 min. The cytosolic and nuclear fractions were pooled and centrifuged at 43,500 x g for 1 hour at 4 °C. The supernatant (10 mL) was incubated with 0.4 ml of FLAG resin (Sigma) in FLAG buffer (20 mM Tris HCl, pH 7.4, 0.15 M NaCl, 1.5 mM MgCl2, 10 % glycerol,0.05 % Nonidet P 40, 0.2 mM EDTA, 1 mM DTT, 0.5 mM PMSF and proteinase inhibitors) at 4°C for 2 hours. The resin was then washed twice with high salt containing FLAG buffer buffer (20 mM Tris HCl, pH 7.4, 0.5 M NaCl, 1.5 mM MgCl2, 10 % glycerol, 0.05 % Nonidet P 40, 0.2 mM EDTA, 1 mM DTT, 0.5 mM PMSF and proteinase inhibitors) followed by one washing with FLAG buffer. DDX11 protein was eluted with 4 µg/ mL 3X FLAG peptide (Sigma) in elution buffer (25 mM TrisHCl, pH 7.4, 1 mM EDTA, 0.15 M NaCl, 1 mM DTT, 0.01% nonidet P-40, protein inhibitors). The FLAG tagged DDX11 protein was dialysed at 4 °C for 2 hours in elution buffer. Aliquots were frozen in liquid nitrogen and stored at -80 °C. Purified recombinant DDX11 proteins were analyzed for purity by sodium dodecyl sulfate-polyacrylamide gel electrophoresis followed by Coomassie staining. Protein concentration was determined by the Bradford assay using bovine serum albumin (BSA) as a standard. A total of 3 independent purifications were prepared for each wild type and mutant DDX11, each of which was used in the different assays described herein.

Radiometric helicase assays.

The helicase assays were performed as described previously (Wu et al., 2012). Briefly, reaction mixtures (20 µl) contained 25 mM Hepes-NaOH, pH 7.5, 25 mM potassium acetate, 1 mM magnesium acetate, 1 mM DTT, 100 µg/ml BSA, 1 mM ATP, 10 fmol of the fork duplex substrate or 2 fmol of OX-1 G4 DNA substrate and the indicated concentrations of the specific DDX11 protein. Helicase reactions were initiated by addition of DDX11 protein and then incubated at 37 ℃ for 15 minutes. The reactions were stopped by stop buffer (75.5 mM EDTA, 0.3% SDS, 12.5 % glycerol, 0.02% bromophenol blue, 0.02% xylene cyanol blue). For standard duplex substrate, a 10-fold excess of unlabeled oligonucleotide with the same sequence as the labeled strand was included in the stop buffer to prevent reannealing. Reactions were run in 12 % native PAGE for duplex substrate, and 12% native PAGE containing 10 mM KCl for G4 substrate. The reaction product was visualized in PhosphoImager and quantified by

Image quant software. The percentage of helicase substrate unbound was calculated by 100X (P/(S+P) where P is the product and S is the substrate.

Electrophoretic Mobility Shift Assays (EMSA).

DNA binding of DDX11 proteins were performed in a reaction mixture (20 µl) that contained the indicated concentrations of DDX11 and a 0.5 nM concentration of the specified 32P-end-labeled DNA substrate in the same reaction buffer as that used for helicase assays except ATP was omitted. The binding mixtures were incubated on ice for 30 min after the addition of DDX11. After incubation, 3 µl of loading dye (74% glycerol, 0.01% xylene cyanol, 0.01% bromphenol blue) was added to each mixture, and samples were resolved on native 5% (19:1acrylamide:bisacrylamide) polyacrylamide gels at 200 V for 1 h and 30 min at 4 °C. The radiolabeled species were visualized using a PhosphorImager and analyzed with Image Quant software. The specific activity of 32P-gamma ATP used for radiolabeling oligonucleotide to prepare 19 bp forked duplex DNA substrate was 3000 Ci (111TBq)/mmol (Perkin Elmer).

ATP Hydrolysis Assays.

ATP hydrolysis was measured using [γ-32P] ATP (PerkinElmer Life Sciences) and analysis by thin layer chromatography (TLC) on polyethyleneimine-cellulose plates (Mallinckrodt Baker). The standard reaction mixture (20 µl total volume) contained 25 mM Hepes-NaOH, pH 7.5, 25mM potassium acetate, 1 mM magnesium acetate, 1 mM DTT, 100 µg/ml BSA, 250 µM [γ-32P] ATP, and 30 nM DDX11 protein and was incubated at 37 °C. Reactions were quenched with 50mM EDTA (final concentration). The reaction mixture was spotted onto a polyethyleneimine-cellulose TLC plate and resolved using 0.5 M LiCl, 1 M formic acid as the carrier solvent. The TLC plate was exposed to a phosphor imaging cassette for 1 h, visualized using a PhosphorImager, and analyzed with Image Quant software. For experiments to determine Km (ATP) and Vmax, M13mp18 ssDNA was 2.1 nM, the concentration of ATP ranged from 16 to 2000 µM, and the reaction was incubated for 30 min. For determination of kcat, the concentration of ATP was 4.4 mM. Five-microliter aliquots were removed and quenched with 5 µl of 0.1 M EDTA at 0, 7.5, 15, 30 and 45 min, respectively.

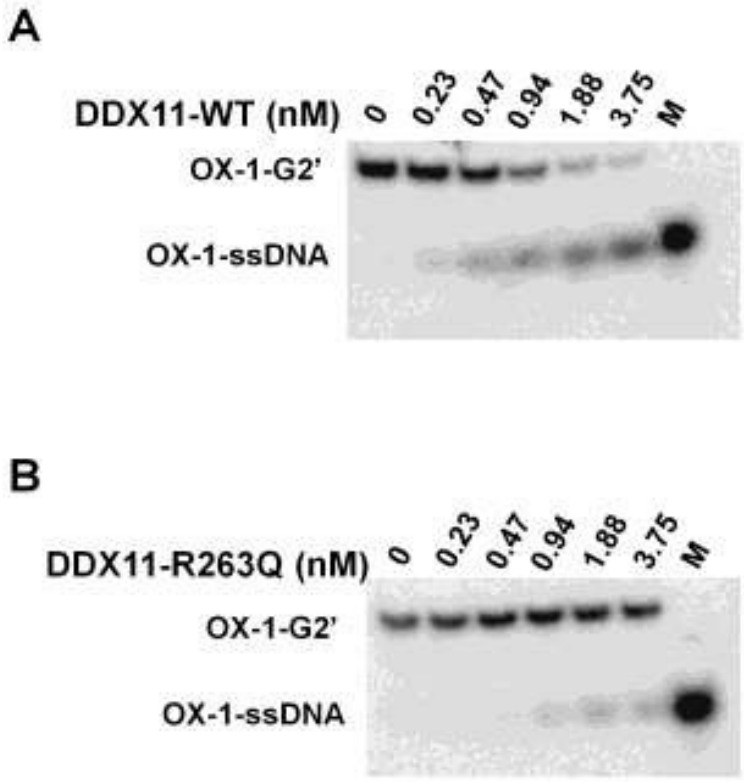


Figure SII.1 Helicase activity of DDX11-WT and DDX11-R263Q on antiparallel G2' G quadruplex DNA substrate.

Helicase reactions (20 µl) were performed by incubating DDX11-WT or DDX11-R263Q with 0.1 nM OX-1-G2'DNA substrate at 37 °C for 15 min under standard helicase assay conditions as described under "Supplementary Subjects and Methods". Products were resolved on native 10% polyacrylamide gels and representative phosphor images of typical gels are shown. *M*, radiolabeled oligonucleotide marker.

Table SII.1 Shared homozygosity regions between patient V-1 and patient V-3.

Chr, chromosome. Markers start and end delimit the regions of shared homozygosity at their corresponding positions (position start and end). Markers positions are relative to the hg18 reference. Homozygosity regions contains > 30 markers and extend over 1 Mb.

Chr	Marker	Marker end	Position	Position end	Size (Mb)
4	rs13114034	cnvi00137	6 575 335	18 621 527	12.0
7	rs4074750	rs1741331	7 006 834	8 767 495	1.8
9	rs715666	rs1060501	114 570 514	116 030	1.5
11	rs590586	rs1790157	71 544 815	74 827 791	3.3
	cnvi000537	rs4937823	130 640 250	133 110	2.5
12	rs12313656	rs1106226	1 305 458	2 517 044	1.2
	rs7134402	rs3819545	12 988 594	46 551 273	33.6
14	rs17127035	rs7151711	53 427 887	78 734 800	25.3
17	rs9910227	rs9652853	51 338 491	52 636 590	1.3
22	rs5996657	rs5998853	22 715 105	32 022 122	9.3

Table SII.2 Prioritization of the variants detected in the exome of patient V-1.

EVS, exome variant server (http://evs.gs.washington.edu/EVS/); (a) only variants that were found in these databases at minor allele frequencies > 0.5% were excluded; (b) based on Sanger sequencing of both DNA strands.

Variant filters applied sequentially	Variant count
Non-synonymous, nonsense, splicing, and coding/splicing	8 087
Homozygous	2 819
In Homozygosity Regions	148
Not in 198 in-house control exomes	5
Not in dbSNP135 or EVS or 1000 Genomes [(a)]	2
Confirmed homozygous in the 3 affected siblings and heterozygous in the parents [(b)]	1 (DDX11 p.R263Q)

Table SII.3 ATPase activity of DDX11-WT and DDX11-R263Q

a. DDX11 proteins were 30 nM final concentration.
b. M13mp18 ssDNA concentration was 2.1 nM. c. ATP concentration was 4.4 mM.

Protein	K_m (mM)[a,b]	V_{max} (nmol min^{-1})[a,b]	k_{cat} (min^{-1})[a,b,c]
DDX11-WT	0.66 ± 0.04	0.56 ± 0.05	220 ± 40
DDX11-R263Q	0.19 ± 0.05	0.066 ± 0.008	12 ± 4

Table SII.4 Clinical features of the patients with DDX11 mutations

a) deafness due to cochlear structural abnormalities; b) tetralogy of Fallot; c) small ventricular septal defects; ID, intellectual disability. * Patient V-2 died at the age of 4 years because of heart failure.

Description	Patient V-1	Patient V-2*	Patient V-3	Lelij et al. (2010)
DDX11 mutation	p.R263Q/ p.R263Q	p.R263Q/ p.R263Q	p.R263Q/ p.R263Q	IVS22+2>T/ c.2689_2691del
Gender	F	M	F	M
Growth retardation	severe	severe	severe	severe
ID	severe	severe	severe	mild
Head circumference, centile	<3rd centile	< 3rd centile	< 3rd centile	< 3rd centile
Facial dysmorphy	+	+	+	+
Cardiac malformations	-	+(b)	-	+(c)
Deafness(a)	+	+	+	+
Abnormal skin pigmentation	-	-	-	+
Clinodactyly /Syndactyly	+/-	-/-	-/-	+/+

CHAPITRE III: Disruption of TBC1D7, a subunit of the TSC1-TSC2 protein complex, in intellectual disability and megalencephaly.

Publié dans: *Journal of Medical Genetics*, 50(11):740-4, 2013

Disruption of TBC1D7, a subunit of the TSC1-TSC2 protein complex, in intellectual disability and megalencephaly.

José-Mario Capo-Chichi,[1] Joseph Tcherkezian,[2] Fadi F Hamdan,[1] Jean Claude Décarie,[3] Sylvia Dobrzeniecka,[4] Lysanne Patry,[1] Marc-Antoine Nadon,[1] Bettina E Mucha,[1] Philippe Major,[1] Michael Shevell,[5] Bouchra Ouled Amar Bencheikh,[4,6] Ridha Joober,[7] Mark E Samuels,[1] Guy A Rouleau,[6] Philippe P Roux,[2,8] Jacques L Michaud[1]

Author affiliations

[1] CHU Sainte-Justine Research Center, Montreal, Quebec, Canada

[2] Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montreal, Quebec, Canada

[3] Department of Medical Imaging, CHU Sainte-Justine, Montreal, Quebec, Canada

[4] CHUM Research Center, Montreal, Quebec, Canada

[5] Department of Neurology and Neurosurgery, McGill University, and Montreal Children's Hospital, Montreal, Quebec, Canada

[6] Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada

[7] Department of Psychiatry, Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada

[8] Department of Pathology and Cell Biology, Université de Montreal, Montreal,Quebec, Canada

Correspondence to :

Jacques L. Michaud,

CHU Sainte-Justine Research Center,

3175 Côte Sainte-Catherine, Montreal, Qc, Canada H3T 1C5.

Phone: (514) 345-4931, ext: 6900; fax: (514) 345-4766

Key words: TBC1D7, mTORC1, intellectual disability, megalencephaly, tuberous sclerosis

Contribution:

J'ai effectué les analyses génétiques ayant permis d'identifier la mutation causale (c.538delT:p.Y180fsX1 dans le gène *TBC1D7)* dans cette famille. J'ai également validé génétiquement l'implication de cette mutation dans la pathologie dans cette famille. J'ai effectué les travaux montrant que la mutation c.538delT entraine la dégradation de TBC1D7 dans les lymphoblastes des patients. Tcherkezian J a effectué les expériences montrant que la mutation p.Y180fsX1 affecte la voie de signalisation mTORC1. Patry L, Nadon MA, Mucha BE, Bencheikh BOA, Joober R, Rouleau GA et Samuels ME ont aidé dans le *design* des travaux. Dobrzeniecka S a genotypé la mutation p.Y180fsX1 dans une cohorte d'individus contrôles. Décarie JC, Major P, Shevell M et Michaud JL ont caractérisé le phénotype des patients dans cette famille. J'ai rédigé la première version du manuscrit, Hamdan FF, Roux PP et Michaud JL ont également contribué à la rédaction et à la révision du manuscrit.

ABSTRACT

Background Mutations in TSC1 or TSC2 cause the tuberous sclerosis complex (TSC), a disorder characterised by the development of hamartomas or benign tumours in various organs as well as the variable presence of epilepsy, intellectual disability (ID) and autism. TSC1, TSC2 and the recently described protein TBC1D7 form a complex that inhibits mTORC1 signalling and limits cell growth. Although it has been proposed that mutations in TBC1D7 might also cause TSC, loss of its function has not yet been documented in humans. **Methods and Results** We used homozygosity mapping and exome sequencing to study a consanguineous family with ID and megalencephaly but without any specific features of TSC. We identified only one rare coding variant, c.538delT: p.Y180fsX1 in TBC1D7, in the regions of homozygosity shared by the affected siblings. We show that this mutation abolishes TBC1D7 expression and is associated with increased mTORC1 signalling in cells of the affected individuals. **Conclusions** Our study suggests that disruption of TBC1D7 causes ID but without the other typical features found in TSC. Although megalencephaly is not commonly observed in TSC, it has been associated with mTORC1 activation. Our observation thus reinforces the relationship between this pathway and the development of megalencephaly.

INTRODUCTION

The tuberous sclerosis complex (TSC) is a multisystemic disorder characterised by the development of hamartomas or benign tumours in various organs, most notably skin, brain and kidneys, as well as a high incidence of epilepsy, intellectual disability (ID) and autism (Crino et al., 2006).TSC is caused by heterozygous loss-of-function mutations in TSC1 (MIM#605284) or TSC2 (MIM#613254). The corresponding proteins form a complex, which functions as a GTPase-activating protein for the Rheb GTPase. Through stimulation of the intrinsic GTPase activity of Rheb, the TSC1–TSC2 complex switches Rheb from its mTORC1-activating, GTP-bound state to its inactive GDP-bound state. In turn, activated mTORC1 phosphorylates many proteins, including translational regulators such as the 40S ribosomal protein S6 kinases (S6Ks) and the 4E-binding proteins (4E-BPs) to promote protein synthesis and cell growth (Garami et al., 2003 ; Tee et al., 2003).

TBC1D7 was recently identified as a third constitutive subunit of the TSC1-TSC2 complex (Nakashima et al., 2007 ; Dibble et al., 2012). Binding of TBC1D7 to TSC1 is required for maintaining the integrity of the TSC1-TSC2 complex, and loss of its function increases mTORC1 signalling, delays induction of autophagy and enhances cellular growth. Sequencing analyses failed to identify mutations in TBC1D7 in a small cohort of individuals with TSC but without detectable mutations in TSC1 or TSC2 (Dibble et al., 2012). Here, we report the identification of a homozygous truncating mutation in TBC1D7 in a family with ID and megalencephaly.

SUBJECTS AND METHODS

Subjects.

The affected individuals (II.1 and II.3) are two siblings born to a consanguineous North African couple in good health. Another sibling is unaffected (II.2). A cohort of 285 North African controls was also studied. Blood samples were obtained from each of these individuals after informed consent and the approval of the appropriate institutional ethic review board.

Homozygosity mapping.

Genomic blood DNA (2.5 mg) from individuals II.1and II.3 was used for whole-genome genotyping on the Illumina 660 Quad Chip (Infinium HD) at the McGill University and Genome Quebec Innovation Center (MUGQIC, Montreal, Canada). Homozygosity mapping was performed using Plink (V.1.06). Shared regions of homozygosity (HR) containing more than 30 consecutive single nucleotide polymorphisms (SNPs) and extending over 1Mb were retained.

Exome capture and sequencing.

The exome of individual II.1 was captured from 3 □g of blood genomic DNA, using the Agilent SureSelect Human All Exon Capture 50 Mb Kit (Mississauga, Ontario, Canada), and sequenced (paired-end, 2×100 bp) using the Illumina HiSeq2000 at the MUGQIC. Sequence processing, alignment and variant calling were done using the Broad Institute Genome Analysis Tool Kit (GATK v2.2) and annotated using Annovar (Wang et al., 2010).

RNA analyses.

Lymphoblastoid cell lines (LCLs) derived from all three siblings and two unrelated healthy individuals were treated either with300 mg/ml of puromycin (Sigma, Toronto, Ontario, Canada) or with vehicle Dulbecco's phosphate-buffered saline (DPBS) for 6 hours at 37°C. Subsequently, total RNA was extracted using the Trizol reagent (Invitrogen, Burlington, Ontario, Canada), reverse-transcribed (1 mg) using the Superscript Reverse Transcriptase-II (Invitrogen), then PCR-amplified using primers targeting exons 5 and 8 of TBC1D7 (NM_001143965.2;ex5F:5'-ggaaatggtggaagatagtgt-3';ex8R:5'-gaccccggtccattcaagctga-3'). β-Actin was also amplified as an internal control.

Protein phosphotransferase assays and western blot analysis.

LCLs derived from all three siblings and two healthy individuals were treated with dimethyl sulfoxide or rapamycin (100 nM) for 1 h prior to being harvested as described previously.(Romeo et al., 2013) Whole-cell lysates were subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and immunoblotted against total and phosphorylated rpS6 (Ser240–244), S6K1 (Cell Signaling Technologies), TBC1D7 (Sigma) andβ-actin (Santa Cruz Biotechnology). For S6K1 kinase assays, immunoprecipitated S6K1 was incubated in a reaction buffer with recombinant glutathione S-transferase–rpS6 fusion proteins and radiolabelled ATP as described elsewhere.7 Resulting samples were subjected to SDS–PAGE, and 32P incorporation was determined using a Fuji PhosphorImager. The data presented are representative of at least three independent experiments.

RESULTS

We identified two siblings with ID, II.1 (9 years) and II.3 (3 years), from a consanguineous North African family (Figure III.1A). The affected siblings showed a similar phenotype with mild ID and increased head circumference (> 97th centile) (see online supplementary information). In subject II.3, the head circumference was already enlarged at birth. The siblings were not dysmorphic but showed a broad forehead, probably due to the associated macrocrania (Figure III.1C, D, G, and H). In addition, the head of subject II.3 was scaphocephalic (Figure III.1H, J). Brain MRI showed normal ventricles and subarachnoid spaces as well as a slightly enlarged corpus callosum in both siblings, indicating that the macrocrania is secondary to an increase in brain volume and not due to increased accumulation of cerebrospinal fluid (Figure III.1E, F, I, J). MRIs did not reveal any cortical tubers or subependymal nodules. There is no history of seizures in the siblings, but EEG recording showed some epileptic activity in the right temporal lobe in subject II.1. Examination of both siblings did not show any skin features of TSC, including facial angiofibroma, shagreen patch or hypomelanotic macule. Serial kidney ultrasound studies performed in subject II.3 showed no signs of angiomyolipoma.

We performed homozygosity mapping in the affected siblings (II.1 and II.3) and identified eight shared regions of homozygosity (see online supplementary table SIII.1). We next performed exome sequencing in one of the affected siblings (II.1). After mapping and removal of duplicate reads, we obtained an average per target base coverage of ~90×, with 95% of the target bases covered at 10× or more. We identified 142 homozygous non-synonymous coding and splicing variants in the shared homozygous intervals. After filtering out variants that were present at a frequency of more than 1.0% in our in-house control exome dataset (n = 200) or in SNP public databases (1000 genomes, NHLBI Go Exome Sequencing Project Exome Variant Server or dbSNP135), only one variant, a truncating mutation in TBC1D7 (NM_001143965.2:exon6: c.538delT: p.Y180fsX1), remained. This mutation was confirmed by Sanger sequencing to be homozygous in the two affected siblings and heterozygous in the parents and in the unaffected sibling (see online supplementary Table S III.2; Figure III.1B). This mutation was absent from our entire in-house exome datasets (>700) and all the above-mentioned

public SNP databases. In addition, genotyping this mutation in 285 control individuals from North African origin failed to identify a single carrier.

We next studied the impact of c.538delT on TBC1D7 protein and mRNA levels using LCLs derived from the two affected subjects, their unaffected sibling and two unrelated healthy individuals. Western blot analysis and reverse transcription PCR showed that cells from the affected siblings did not show any detectable TBC1D7 protein production and mRNA expression, respectively, in contrast to cells from the unaffected sibling or controls, suggesting that c.538delT induces the degradation of TBC1D7 mRNA via the nonsense-mediated decay (NMD) pathway (Figure III.2, see online supplementary figure SIII.1). This possibility is consistent with the fact that direct sequencing of the TBC1D7 cDNA amplified from the heterozygous sibling only showed the presence of the wild-type allele. Moreover, incubating the cells with puromycin, an inhibitor of translation and NMD,(Andreutti-Zaugg et al., 1997) restored the expression of the TBC1D7 mutant allele (see online supplementary figure SIII.1).

Based on the newly described role of TBC1D7 as an upstream regulator of mTORC1, we examined whether cells derived from affected individuals had abnormal levels of mTORC1 activity. For this, we quantified several established mTORC1-dependent signalling events, such as the kinase activity of immunoprecipitated endogenous S6K1, as well as phosphorylation of the S6K1 substrate rpS6 (Ser 240–244). Using this approach, we found that LCLs derived from subjects II.1 and II.3 displayed constitutively high levels of mTORC1 activity compared with cells from the unaffected sibling (II.2) or healthy controls (Figure III.2). In all tested LCLs herein, mTORC1 activity was completely sensitive to rapamycin treatment, consistent with TBC1D7 playing the role of an upstream regulator of mTORC1.

DISCUSSION

We identified a rare homozygous truncating mutation in TBC1D7 in two siblings with ID and megalencephaly. Several observations suggest that this mutation is pathogenic. First of all, exome sequencing showed that it is the only rare coding or splicing variant present in all the regions of homozygosity shared by the affected siblings. Family studies confirmed that the mutation segregates with the phenotype consistent with an autosomal recessive mode of inheritance. Furthermore, this mutation results in the loss of TBC1D7 expression and is associated with increased mTORC1 signalling in cells of the affected subjects, confirming the involvement of TBC1D7 in the regulation of the mTOR pathway. Interestingly, mTORC1 activation has been associated with megalencephaly and ID. For instance, mutations in components of the mTOR pathway, including PTEN, STRADA, AKT3, PIK3R2, PIK3CA and MTOR, have been identified in syndromic and non-syndromic forms of megalencephaly (Liaw et al., 1997 ; Puffenberger et al., 2007 ; Lee et al., 2012 ; Riviere et al., 2012). These different mutations appear to cause megalencephaly by increasing mTORC1 signalling. The documentation of ID and megalencephaly in subjects with mutations in TBC1D7 is thus consistent with what we know about the function of this gene and of the mTOR pathway.

TBC1D7 was recently identified as the third constitutive subunit of the TSC1-TSC2 complex, suggesting that mutations in the corresponding gene might explain some TSC cases. The subjects described here did show ID, which is found in about 50% of patients with TSC (Ehninger, 2013) . Although megalencephaly is not atypical feature of TSC, it has been described in some individuals with that condition (Maloof et al., 1994 ; Griffiths et al., 1998). Moreover, the incidence of macrocrania, which may be indicative of megalencephaly, is greater in individuals with TSC than in the general population (Fidler et al., 2000). However, the affected subjects did not show the specific brain, kidney or skin features associated with TSC. Identification of additional patients with TBC1D7 mutations is needed to better assess the associated phenotypic spectrum and the range of its overlap with TSC.

Acknowledgements

JLM is a National Scientist of the Fonds de Recherche du Québec-Santé, and PPR holds the Canada Research Chair in Signal Transduction and Proteomics. J-MC-C holds a salary award from the Réseau de Médecine Génétique Appliquée du Québec (RMGA). We thank the patients and their family for participating in this study. We also thank the members of the RMGA bioinformatics team (Alexandre Dionne-Laporte, Dan Spiegelman, Edouard Henrion and Ousmane Diallo) for the bioinformatics analysis of the exome-sequencing data.

Contributors

Funding this work was supported by March of Dimes (grant no. 12-FY10–236 to MS and JLM) and the Canadian Institutes for Health Research (grant no. MOP123408 to PPR).

Competing interests None.

Patient consent Obtained.

Ethics approval Ethics committee of Sainte-Justine Hospital Research Center (Montreal, Canada)

Provenance and peer review Not commissioned; externally peer reviewed.

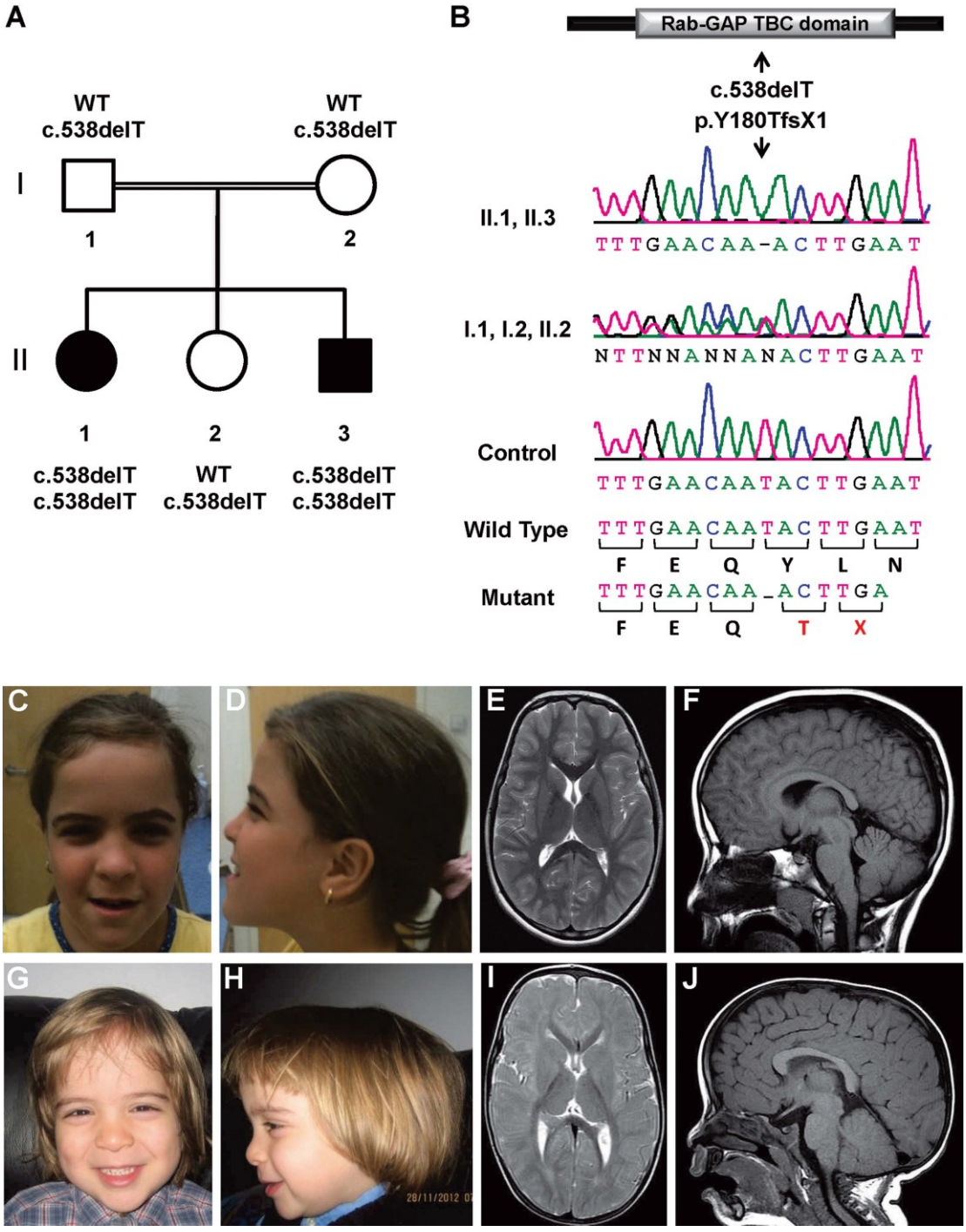

A
WT
c.538delT
WT
c.538delT
I
1
2
II
1
2
3
c.538delT
c.538delT
WT
c.538delT
c.538delT
c.538delT
B
Rab-GAP TBC domain
c.538delT
p.Y180TfsX1
II.1, II.3
TTTGAACAA-ACTTGAAT
I.1, I.2, II.2
NTTNNANNANACTTGAAT
Control
TTTGAACAATACTTGAAT
Wild Type
TTTGAACAATACTTGAAT
F E Q Y L N
Mutant
TTTGAACAA_ACTTGA
F E Q T X
C
D
E
F
G
H
I
J

Figure III.1

Figure III.1 Identification of a homozygous truncating mutation in TBC1D7.

(A) Pedigree of the consanguineous North African family studied here.
(B) Sanger re-sequencing confirming that the c.538delT: p.Y180TfsX1) (NM_001143965.2) mutation in TBC1D7 segregates in this family. Representative chromatograms of the two affected siblings (homozygous for c.538delT), their parents and the unaffected sibling (heterozygous for c.538delT) are shown. (C–J) Pictures and brain MRI images of subjects II.1 (C–F) and II.3 (G–J). Axial T2 images (E and I) showed normal ventricles and subarachnoid spaces. The macrocrania is therefore secondary to an increased brain volume and not due to hydrocephalus or increased subarachnoid spaces. Sagittal T1images (F and J) showed an enlarged corpus callosum. Access the article online to view this figure in colour.

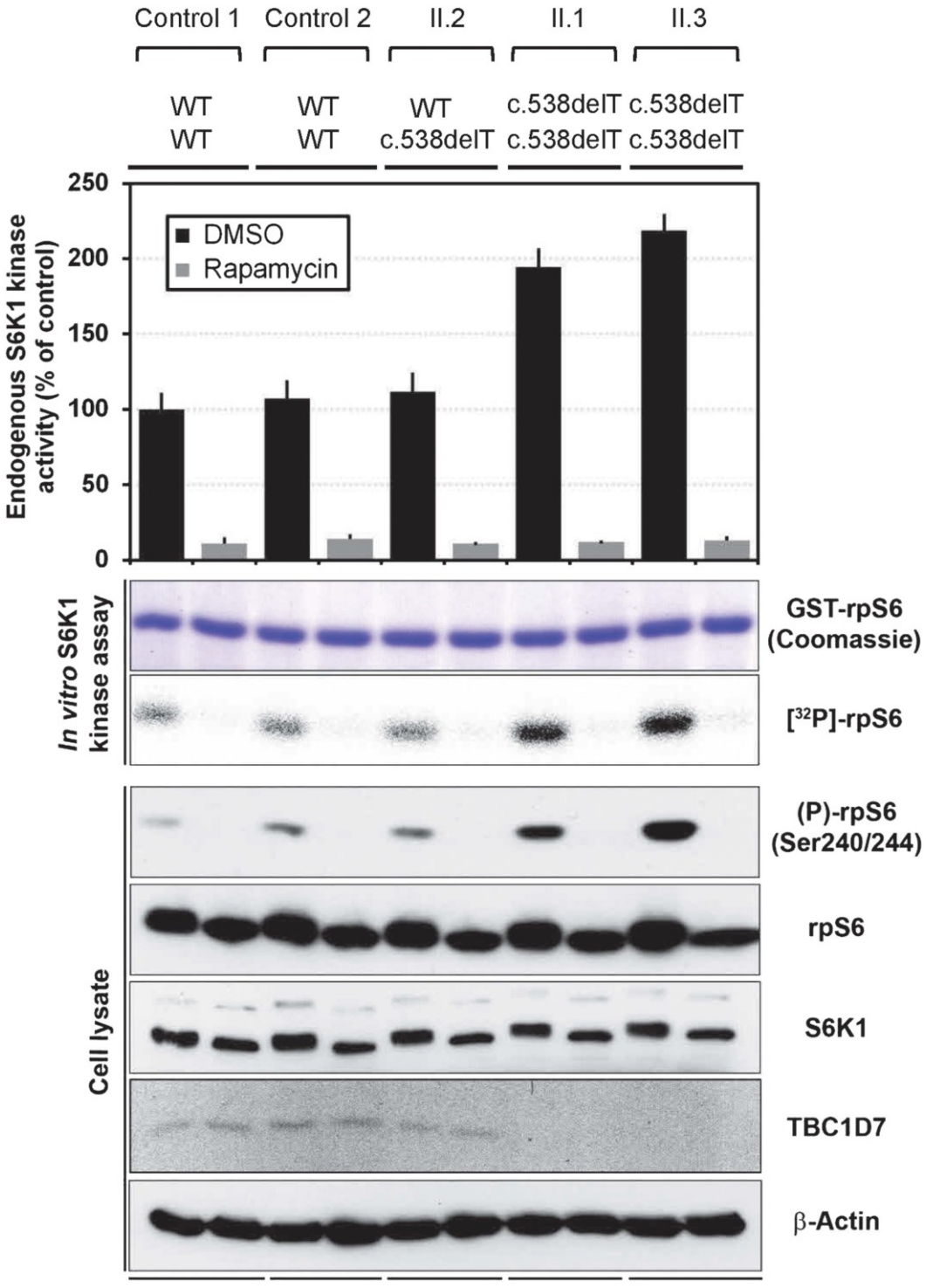

Control 1
Control 2
II.2
II.1
II.3
WT
WT
WT
WT
WT
c.538delT
c.538delT
c.538delT
c.538delT
c.538delT
Endogenous S6K1 kinase activity (% of control)
250
200
150
100
50
0
DMSO
Rapamycin
In vitro S6K1 kinase assay
GST-rpS6 (Coomassie)
[32P]-rpS6
Cell lysate
(P)-rpS6 (Ser240/244)
rpS6
S6K1
TBC1D7
β-Actin

Figure III.2

Figure III.2 Impact of the mutation c.538delT (p.Y180fsX1) on mTORC1 activation.

Assessment of mTORC1 activation in lymphoblastoids cell lines by quantifying the in vitro activity of S6K1 using recombinant glutathione S-transferase (GST)–rpS6 fusion proteins (two upper panels) and blotting total and phosphorylated rpS6, S6K1, TBC1D7 and β-actin in whole-cell lysates. Results are displayed by genotypes of the individual with and without rapamycin treatment. Access the article online to view this figure in colour.

SUPPLEMENTARY MATERIAL

Clinical description

The subjects are siblings born to a consanguineous North African couple in good health. A sister is unaffected. Head circumference of the mother is 55.4 cm (50-75th centile) and of the unaffected sister is 51.5 cm (50th centile) at 6 years of age. We did not have access to the father but he was described as having a "small head".

Subject II.1

Subject II.1 is a 9 year-old girl. Pregnancy was unremarkable except for pre-eclampsia during the third trimester, leading to caesarean section at 34 weeks of gestation. APGAR score was 2 and 7 at 1 and 5 minutes, respectively. At birth, weight was 2.9 kg. The neonatal course was uneventful and the child was discharged from the hospital at two days of life.

Subject II.1 showed global developmental delay. She started to walk at 18 months of age, showed decreased fine motor skills and was slow to talk. Currently, she can produce short sentences of about 4 or 5 words. She can draw circles but not a face. She can recognize some letters and single digits but cannot write them. Formal psychometric assessments were performed at 9 years of age. She performed poorly on the Perceptual Reasoning Index of the Wechsler Intelligence Scale for Children – IV, with a score under the 1st centile. She also obtained low scores (mild-to moderate intellectual disability range) with the Leiter-R Brief IQ. Assessment with the Adaptive Behavior Assessment System-II (ABAS-II) showed a score less than 0.1 centile on the global scale. Overall, the cognitive assessment showed mild intellectual disability. Non-verbal interactions were unremarkable. No seizure was noticed.

At 9 years and 5 months of age, subject II.1 weighed 37.4 kg (75-90th centile) and measured 142.4 cm (90th centile). Her head circumference measured 55.5 cm (97.5th centile). Neurologic examination was unremarkable. Whole-genome comparative hybridization with an array containing 135000 oligonucleotides, and molecular testing for the triple repeat expansion associated with the Fragile X syndrome did not show any abnormality. A comprehensive metabolic work-up, including blood lactate and ammonia, urine creatine and guanidinoacetate measurements, plasma amino-acid and urine organic acid chromatography, as well as evaluation of immunoreactive forms of transferrin after isoelectric focusing in polyacrylamide gels, was unremarkable. EEG recording showed some epileptic activity in the right temporal lobe.

Subject II.3

Subject II.3, a 3 year-old boy, was born at term after an unremarkable pregnancy. APGAR score was 9 and 9 at 1 and 5 minutes, respectively. At birth, weight was 3,7 kg and head circumference was 37.2 cm (95th centile). The neonatal course was uneventful.

Subject II.3 showed global developmental delay. He started to walk at 2 years ½. Currently, he can eat with a spoon and draw spirals but not closed forms. He can say about 10 words but cannot associate words. He understands simple commands and can

designate body parts on demand. Psychometric evaluations performed at the age of 2 years and 7 months with the Wechsler Preschool and Primary Scale of Intelligence (3rd edition), the Mullen Scales of Early Learning, and the ABAS-II showed a profile consistent with mild intellectual disability. Non-verbal interactions were unremarkable. No seizure was noticed.

Kidney ultrasound performed at 2 months of age showed grade IV hydronephrosis with proximal hydroureter on the left side consistent with pyeoloureteral junction obstruction. Open pyeloplasty was performed at 3 months of age.

At 2 years and 11 months of age, subject II.3 weighed 15.6 kg (75th centile) and measured 91.4 cm (10-25th centile). His head circumference measured 54.5 cm (> 97th centile) and his head shape was scaphocephalic. Neurologic examination was unremarkable.

Whole-genome comparative hybridization with an array containing 135000 oligonucleotides, and molecular testing for the triple repeat expansion associated with the Fragile X syndrome did not show any abnormality. A comprehensive metabolic work-up, including blood ammonia, urine creatine and guanidinoacetate measurements, and plasma amino-acid chromatography was unremarkable.

SUPPLEMENTARY FIGURES

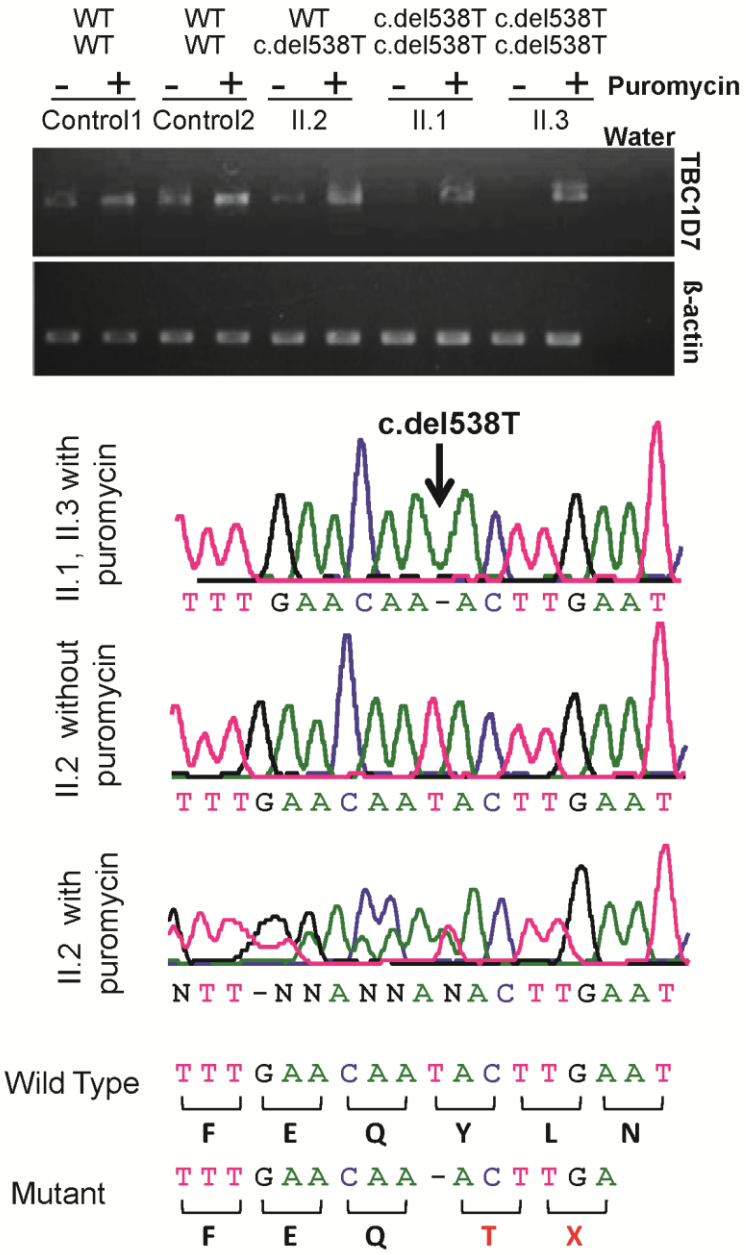


Figure SIII.1

Figure SIII.1 RT-PCR analysis showing nonsense-mediated decay of the *TBC1D7* mutant allele in LCLs from subjects II.1 and II.3.

LCLs from the unaffected sibling (II.2) and two unrelated individuals lacking the c.del538T *TBC1D7* mutation were used as controls. Results are displayed by genotype of the individuals with and without puromycin treatment. TBC1D7 and b-Actin were amplified from LCLs cDNA using gene specific primers. Water was used as a negative control. Lower panel shows the chromatograms of the generated RT-PCR products, with the presence of the mutant *TBC1D7* allele (c.del538T) in the affected siblings (II.1, II.3) only after puromycin treatment.

SUPPLEMENTARY TABLES

Table SIII.1 Shared homozygosity regions between patients II.1 and II.3

Chromosomal coordinates are based on NCBI genome build hg19
Homozygosity regions of minimum 30 SNPs and extending over 1Mb.

Chr	Marker start	Marker end	Position start	Position end	Size (Mb)
1	rs1316342	cnvi0004790	22 680 517	27 533 377	4,9
1	rs697452	rs7549807	191 627 004	204 079 268	12,5
2	rs12466109	rs6436364	214 091 797	223 810 266	9,7
6	rs1179240	rs13201062	89 571	14 390 681	14,3
7	rs763391	rs11760753	68 917 040	69 925 030	1,0
15	rs1668586	rs2957637	42 287 578	43 951 118	1,7
16	rs2271292	rs889792	67 322 857	68 524 308	1,2
20	rs3843781	rs65567	4 352 277	15 041 361	10,7

Table SIII.2 Prioritization of the variants detected in patient II.1's exome

HR, Shared homozygosity regions >30 SNPs and > 1Mb. EVS, exome variant server (http://evs.gs.washington.edu/EVS/) of NHLBI /ESP; (a) only variants that were found in these databases at minor allele frequencies > 1.0% were excluded and the rest were retained. (b) Sanger sequencing in both directions

Variant filters applied sequentially	**Variant**
Non-synonymous, nonsense, splicing, and coding/splicing Indels	13 796
Homozygous	4 981
In Homozygosity Regions	142
Not in 200 in-house control exomes	9
Not in dbSNP135 or EVS or 1000 Genomes[a]	1
Confirmed homozygous in the 2 patients and heterozygous in their parents and the unaffected sibling [b]	1 (*TBC1D7* p.Y180Tfs1X)

CHAPITRE IV: Deficiency of asparagine synthetase causes congenital microcephaly and a progressive form of encephalopathy

Publié dans: *Neuron*, 80(2):429-41, 2013

Deficiency of asparagine synthetase causes congenital microcephaly and a progressive form of encephalopathy.

Elizabeth K. Ruzzo,[1,25] José -Mario Capo-Chichi,[2,25] Bruria Ben-Zeev,[3,4,25] David Chitayat,[5,6] Hanqian Mao,[7] Andrea L. Pappas,[8] Yuki Hitomi,[1] Yi-Fan Lu,[1] Xiaodi Yao,[1] Fadi F. Hamdan,[2] Kimberly Pelak,[1] Haike Reznik-Wolf,[4,9] Ifat Bar-Joseph,[3,4,9] Danit Oz-Levi,[10] Dorit Lev,[4,11,12] Tally Lerman-Sagie,[4,12,13] Esther Leshinsky-Silver,[4,12,14]Yair Anikster,[3,4] Edna Ben-Asher,[10] Tsviya Olender,[10] Laurence Colleaux,[15] Jean-Claude Décarie,[16] Susan Blaser,[17] Brenda Banwell,[18] Rasesh B. Joshi,[8] Xiao-Ping He,[8] Lysanne Patry,[2] Rachel J. Silver,[6] Sylvia Dobrzeniecka,[19] Mohammad S. Islam,[20] Abul Hasnat,[20] Mark E. Samuels,[2] Dipendra K. Aryal,[21] Ramona M. Rodriguiz,[21] Yong-hui Jiang,[22]William C. Wetsel,[21,23] James O. McNamara,[8] Guy A. Rouleau,[19,24] Debra L. Silver,[7] Doron Lancet,[10] Elon Pras,[3,9] Grant A. Mitchell,[2] Jacques L. Michaud,[2,26,*] and David B. Goldstein[1,26,*]

[1] Center for Human Genome Variation, Duke University School of Medicine, Durham, NC 27708, USA. [2] Sainte-Justine Hospital Research Center, Montreal, QC H3T 1C5, Canada. Edmond and Lily Safra Children's Hospital, Sheba Medical Center, 52621 Ramat Gan, Israel. [4] The Sackler School of Medicine, Tel Aviv University, 69978 Tel Aviv, Israel. [5] The Hospital for Sick Children, Division of Clinical and Metabolic Genetics, University of Toronto, Toronto, ON M5G 2L3, Canada. [6] Mount Sinai Hospital, The Prenatal Diagnosis and Medical Genetics Program, University of Toronto, Toronto, ON M5G 1Z5, Canada. [7] Molecular Genetics and Microbiology and The Duke Institute for Brain Sciences. [8] Department of Neurobiology Duke University, Durham, NC 27708, USA. [9] The Danek Gertner Institute of Human Genetics, Sheba Medical Center, 52621 Ramat Gan, Israel. [10] Department of Molecular Genetics, Weizmann Institute of Science, 76100 Rehovot, Israel. [11] Institute of Medical Genetics.[12] Metabolic Neurogenetic Clinic.[13] Pediatric Neurology Unit.[14] Molecular Genetics Laboratory Wolfson Medical Center, 58100 Holon, Israel. [15]Hôpital Necker-Enfants Malades, 75015 Paris, France.[16] Department of Medical Imaging, Sainte-Justine Hospital, Montreal, QC H3T 1C5, Canada.[17] Department of Medical Imaging, The Hospital for Sick Children, University of Toronto, Toronto, ON M5G 2L3, Canada.[18] Division of Neurology, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.[19] Research Center, Centre Hospitalier de l'Université de Montréal, Montreal, QC H2L 2W5, Canada.[20] Department of Clinical Pharmacy & Pharmacology, Faculty of Pharmacy, University of Dhaka, 1000 Dhaka,Bangladesh.[21] Department of Psychiatry and Behavioral Sciences, Mouse Behavioral and Neuroendocrine Analysis Core Facility, Duke University,Durham, NC 27710, USA.[22] Department of Pediatrics and Neurobiology, Duke University School of Medicine, Durham, NC 27710, USA.[23] Department of Cell Biology and Neurobiology, Duke University, Durham, NC 27710, USA.[24] Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada, H3A 2B4. [25] These authors contributed equally to this work.[26] These authors contributed equally to this work

***Corresponding authors:**
Jacques L. Michaud
David B. Goldstein

Contribution:

J'ai effectué les analyses génétiques ayant permis d'identifier les mutations causales (p.A6E et p.R550C dans le gène *ASNS*) dans les familles C et D. J'ai également validé génétiquement l'implication de ces mutations dans la pathologie de ces familles. J'ai effectué le clonage moléculaire des mutations p.A6E /p.R550C ; les constructions générées ont été utilisées dans les études vérifiant l'impact des mutations sur l'expression de ASNS. Ruzzo EK a effectué les analyses génétiques permettant de mettre en évidence la mutation causale (p.F362V dans le gène *ASNS*) dans les familles A et B. Décarie JC, Mitchell GA et Michaud JL ont caractérisé le phénotype des patients de la famille D. Nos collaborateurs en Israël et à Toronto ont caractérisé les phénotypes des patients des familles A, B et C, respectivement. Nos collaborateurs de l'Université de Duke ont validé biologiquement l'impact des mutations p.A6E, p.F362V et p.R550C sur l'expression de ASNS ; ils ont également généré et caractérisé le phénotype d'un modèle murin où Asns est invalidé. Les autres collaborateurs ont aidé dans le *design* de l'étude. Ruzzo EK a rédigé la première version du manuscrit ; j'ai contribué à la rédaction et à la révision du manuscrit avec Hamdan FF, Goldstein DB, Mitchell GA et Michaud JL.

SUMMARY

We analyzed four families that presented with a similar condition characterized by congenital microcephaly, intellectual disability, progressive cerebral atrophy, and intractable seizures. We show that recessive mutations in the *ASNS* gene are responsible for this syndrome. Two of the identified missense mutations dramatically reduce ASNS protein abundance, suggesting that the mutations cause loss of function. Hypomorphic Asns mutant mice have structural brain abnormalities, including enlarged ventricles and reduced cortical thickness, and show deficits in learning and memory mimicking aspects of the patient phenotype. *ASNS* encodes asparagine synthetase, which catalyzes the synthesis of asparagine from glutamine and aspartate. The neurological impairment resulting from ASNS deficiency may be explained by asparagine depletion in the brain or by accumulation of aspartate/glutamate leading to enhanced excitability and neuronal damage. Our study thus indicates that asparagine synthesis is essential for the development and function of the brain but not for that of other organs.

INTRODUCTION

Intellectual disability (ID) affects 2%–3% of the general population and is characterized by a broad range of cognitive deficits. It is usually subdivided into syndromic and non-syndromic forms, depending on whether additional abnormalities are found. Syndromic ID is often accompanied by microcephaly, defined by a head circumference more than two SDs below the age- and sex-adjusted mean. The incidence of microcephaly, as reported in birth defect registries worldwide, varies from 1 to 150 per 100,000 depending upon the range of SD used to define microcephaly and the ethnic population. For example, microcephaly is more prevalent in populations with a high degree of consanguinity (Mahmood et al., 2011) . Causes of congenital microcephaly include metabolic disorders, chromosomal anomalies, and intrauterine infections. However, with the exception of autosomal recessive primary microcephaly (MCPH), the genetic etiology of most congenital microcephaly cases is unknown.

We ascertained four families with a distinct form of severe encephalopathy associated with congenital microcephaly and progressive brain atrophy. Two families were from the same ethnic group, whereas the other two families were independently recognized as presenting with an identical syndrome. Both pairs of families were analyzed independently by exome sequencing. Here we report the clinical features of the affected children and demonstrate that the observed phenotype in all four families can be explained by autosomal recessive deficiency of asparagine synthetase (ASNS).

RESULTS

Identification and Validation of ASNS Mutations

We identified a total of nine children from four families with a severe form of intellectual disability (Table IV.1; Figure IV.1A; Supplemental Experimental Procedures available online). These children were born with a small head circumference and showed progressive microcephaly. Although congenital microcephaly is a consistent feature of this syndrome, the patients do not fit the definition of primary microcephaly (MCPH) (Supplemental Experimental Procedures). Their clinical course was characterized by profound developmental delay and, in a majority of cases, early-onset intractable seizures (Table IV.1; Figure IV.2; Figure SIV.1). Clinical examination revealed axial hypotonia with severe appendicular spasticity in all cases. All affected siblings of family C also showed excessive startle reflex, mimicking hyperekplexia. In addition, several affected individuals from families C and D had episodes of hypothermia. Brain MRI first performed in early infancy showed decreased cerebral volume and size of pons, presumably caused by hypodevelopment and/or atrophy, as well as delayed myelination (Figure IV.2; Figure SIV.1). Some patients also showed gyral simplification. The affected children from two families (C and D) died during the first year of life because of pulmonary aspiration secondary to severe neurological dysfunction, whereas the affected individuals from the other families survived into their third decade.

Families A and B are unrelated but are both of Iranian Jewish ancestry. Targeted exonic regions were captured and sequenced in one affected individual from family A (A.II.1) and two from family B (B.II.2 and B.II.4). We focused on variants that were annotated as having a plausible impact on the function of the resulting gene product (e.g., missense, nonsense, splice site, intron-exon boundary, and coding-disrupting insertion deletions [Indels]). We compared patient exomes to control exomes sequenced in the same facility (n = 261, unrelated samples, not enriched for neurological disorders). Because families A and B belong to the same ethnic community and were the only similar cases identified in Israel to date, we postulated that the causal variant would be a founder mutation in this population shared among all affected individuals in these families. We therefore first focused on homozygous variants that were shared by both siblings in family B (Figure 1A, B.II.2 and B.II.4) and that were uncommon in our control population (Table SIV.1). Since the incidence of this disorder is very low in the

general population, we inspected only variants with a predicted frequency of ≤ 3% in our sequenced control genomes. We found 72 such variants, only three of which were absent in the control population (Table SIV.2). Furthermore, only one of these three variants was also present in homozygous form in the patient from family A (Figure 1A, A.II.1). This variant, located in the asparagine synthetase (*ASNS*) gene, causes a missense change (c.1084T>G) resulting in a phenylalanine to valine substitution at amino acid position 362 (p.F362V; NM_183356). We also performed homozygosity mapping. We found that p.F362V lies in the largest homozygous region and found no additional candidates of interest in the homozygous regions (Supplemental Experimental Procedures; Tables SIV.7–9).

Family C is composed of three affected (C.II.1, C.II.3, and C.II.4) siblings and one healthy (C.II.2) sibling born to consanguineous parents of Bangladeshi origin (Figure 1A). No DNA was available for the first affected child (C.II.1) who had the same clinical manifestations as his affected brothers. Homozygosity mapping showed that the two affected siblings share a total of eight homozygous regions that are >1 Mb in size (Table SIV.3). Exome sequencing performed in one of the affected children (C.II.3) identified 856 rare protein or splice-altering variants (with a frequency ≤ 3% in 169 in-house unrelated exomes, 1000 Genomes Project data set and data from the National Heart, Lung, and Blood Institute [NHLBI] Exome Sequencing Project [ESP]). These included three variants that map to the shared regions of homozygosity; the three variants were Sanger sequenced and all three variants were homozygous in both affected individuals. The parents and the unaffected sibling were heterozygous for two of these variants, whereas the other candidate variant was excluded from further consideration because it was found in a homozygous form in one of the parents. One of the remaining variants, c.1282G>A (p. D428N; NM_017460) in the CYP3A4 gene, is not predicted to affect protein function by SIFT or Polyphen-2 (Ng et Henikoff, 2003 ; Adzhubei et al., 2010) and CYP3A4 encodes a component of cytochrome P450 (subfamily 3A, polypeptide 4), which is predominantly expressed in the liver. Thus, the CYP3A4 variant seemed unlikely to be responsible for this phenotype. The sole remaining variant in this family is c.1648C>T (p. R550C; NM_183356) in ASNS, which is present in the largest region of homozygosity (35 Mb) shared by the two affected children (Table SIV.4).

Family D is a non-consanguineous French Canadian family, consisting of three affected (D.II.1, D.II.2, and D.II.3) and two unaffected (D.II.4 and D.II.5) siblings (Figure IV.1A).

Exome sequencing was performed in two affected (D.II.1 and D.II.2) and two unaffected siblings. In total, 237 rare protein or splice altering variants were present in both affected children (with a frequency ≤ 3% in 169 in-house unrelated exomes, 1000 Genomes Project data set and data from the NHLBI ESP). We excluded from this list X-linked variants that were also present in the unaffected male sibling. We also excluded homozygous or possible compound heterozygous variants that were found in the same form in at least one unaffected sibling. Only two variants (c.1648C>T/p.R550C; c.17C>A/p.A6E; NM_183356), both in ASNS, remained after this filtering process (Table SIV.5).

Critically, in all four families there is complete co-segregation of the identified ASNS mutations/genotypes with disease (Figure IV.1A). Sanger sequencing was used to validate all three mutations (Figure IV.1B).

All three missense mutations are predicted to damage the encoded asparagine synthetase protein by available computer algorithms (SIFT and PolyPhen-2) and all three mutations are absent in dbSNP135, the 1,000 Genomes Project data set, and data from the NHLBI ESP (Table IV.2). To better estimate the frequency of the p.F362V variant in unaffected individuals, we directly genotyped this locus in 1,160 additional controls and failed to detect the mutation. Finally, all three mutations were genotyped in ancestry-matched controls and all remained absent (Table IV.2), with the exception of p.F362V, which has an estimated carrier frequency of 0.0125 in Iranian Jews.

Additionally, we used the sequence data to test for evidence of cryptic relatedness between the patient in family A and the affected siblings from family B and found no indication of elevated identity by descent beyond what is expected for unrelated genomes (data not shown). We also tested whether the p.F362V ASNS variant is found on a common haplotype in all affected individuals of Iranian Jewish origin. Indeed, the ASNS variant was found on the same 1.2 Mb haplotype in both families and this haplotype was very rare (0.8%) in 261 sequenced controls (Supplemental Experimental Procedures; Table SIV.6). This observation is consistent with a single founder origin for p.F362V and subsequent transmission along with the same extended haplotype. We also did not find evidence for homozygote deletions overlapping the ASNS gene in controls (Supplemental Experimental Procedures).

Interestingly, the mutation p.R550C was found in two families of different ethnic backgrounds. This mutation was associated with different haplotypes in each of these families,

suggesting that it arose independently. It should be noted that p.R550C corresponds to a CpG site, which is associated with a higher mutation rate (Nachman et Crowell, 2000) possibly explaining the recurrence of this rare mutation in different populations.

Functional Impact of the Non synonymous Mutations

To test the effect of the identified mutations on ASNS mRNA and protein levels, we generated full-length mutant cDNA constructs (p.A6E, p.F362V, and p.R550C) using PCR-mediated site directed mutagenesis (Figure SIV.2). We then transfected both wild-type and mutant alleles into HEK293 and COS-7 cells and found similarly robust levels of expression of the mRNA corresponding to wild-type and all three mutant alleles (Figure IV.3A). This result indicates that these mutations do not overtly affect mRNA levels, suggesting that they do not influence mRNA stability. For the p.F362V mutation, we also compared wild-type and mutant full-length transcripts, from the patient fibroblasts, to detect any differences in alternative splicing or exon skipping and found no evidence for alternately spliced transcripts (data not shown).

We used two approaches to detect the ASNS protein in transfected cells. First, we used an antibody to human ASNS (Figure SIV.2). This antibody cross-reacted to produce non-specific bands and thus we also used C-terminal FLAG-tagged forms of ASNS to detect the wild-type and all three mutant forms of ASNS (Figure SIV.2) using the anti-FLAG antibody. We found that while high levels of the wild-type protein were easily detected, a dramatic reduction in protein abundance was seen in the HEK293 cells expressing the p.A6E or p.F362V mutant allele. In contrast, cells expressing the p.R550C mutant allele had an increased level of protein abundance compared to wild-type (Figure IV.3B). Consistent with the former observation, a dramatic reduction in ASNS abundance was observed in patient fibroblasts from individual II.1 in family A, harbouring the p.F362V allele (Figure SIV.2). This pattern of protein abundance was also observed in COS-7 cells transfected with empty, wild-type, or mutant vectors (Figure SIV.2). These results suggest that these mutations impair *ASNS* gene function by either reducing protein expression (p.A6E or p.F362V) or reducing functional performance (p.R550C). The mechanism through which the R550C mutation reduces activity remains to be elucidated, but the clinical similarity in presentation of patients suggests that all mutations are loss of function mutations.

We then asked whether these mutations destabilize the protein, targeting it for degradation. We blocked both the ubiquitin-proteasome and the macroautophagy pathways, but neither of these altered ASNS protein abundance (data not shown). We also used Leupeptin to inhibit lysosomal-dependent degradation and this also failed to rescue the p.A6E or p.F362V mutant proteins to wild-type levels (Figure IV.3B), although some experiments did show a trend toward rescue (data not shown).

ASNS encodes the glutamine-dependent asparagine synthetase enzyme (EC 6.3.5.4), which catalyzes ammonia transfer from glutamine to aspartic acid via a □-aspartyl-AMP intermediate. Concordant with this biochemical function, we found that the levels of asparagine were decreased in at least two affected individuals (C.II.3 and D.II.1), whereas glutamine and aspartic acid, both precursors in the ASNS-catalyzed synthesis of asparagine, were mildly elevated in the patients from family B (Table IV.3). These findings are consistent with our in vitro functional studies, emphasizing that the identified mutations have phenotypic consequences.

The mutated amino acid residues in ASNS are located within regions of high sequence conservation among orthologs, from bacterium to man (Figure IV.4A), indicating that these amino acids are likely to be critical for protein function. This is further supported by the inferred positions of the human ASNS mutations in the folded bacterial ortholog (Figure IV.4B; Supplemental Experimental Procedures).

ASNS Expression in the Brain

Cells are capable of both nutritional intake and endogenous synthesis of asparagine, suggesting that ASNS may be dispensable, and raising the question of how loss of ASNS protein or its dysfunction results in a severe, tissue-specific phenotype. ASNS is under complex transcriptional regulation (Greco et al., 1989 ; Chen et al., 2004 ; Li et al., 2006 ; Richards et Kilberg, 2006) and is expressed at low levels in most tissues but is highly expressed in the adult brain, with evidence for a brain-specific splice variant(s) (Figure SIV.3) (Hongo et al., 1996). Consistent with this expression pattern, Asns is expressed in the adult mouse brain (Allen Brain Atlas). In situ hybridization shows that Asns is also expressed in the developing embryonic mouse brain (E14.5), is particularly enriched in the cortical plate where the neurons reside, and is also expressed in the ventricular and subventricular zone layers (VZ

and SVZ) lining the ventricles of the cerebral cortex, where neural progenitors reside (Figure SIV.3) (Visel et al., 2004). RNA-seq of E14.5 cortices also confirms this pattern of expression in the cortical plate and VZ (Ayoub et al., 2011). This expression pattern is similar to that of known microcephaly genes (Bond et al., 2002 ; Jackson et al., 2002), consistent with a role for Asns in cortical development and brain size.

Asns Gene-Trap Mice

We obtained genetically modified mice from EUCOMM in which the Asns genomic locus contains a gene-trap insertion in intron 2(ENSMUST00000115542) containing a splice acceptor site and the LacZ gene (B6NTac; B6N-Asns$^{tm1a(EUCOMM)Wtsi}$/H, termed Asns$^{-/-}$; Figure SIV.4).

Gene traps are frequently hypomorphs, as there can be splicing that skips over the gene-trap cassette, but the degree to which this splicing occurs is construct and gene specific (Adams et Van Der Weyden, 2008) . To determine the extent of Asns expression in the Asns gene-trap mice, we performed a comprehensive Asns mRNA analysis in the brains (cerebral hemispheres) of adult Asns$^{+/+}$, Asns$^{+/-}$, and Asns$^{-/-}$ mice (3 to 4 months of age). First, RT-PCR was used to semi quantitatively assess the existence of Asns mRNA transcripts. RT-PCR spanning from the second exon to the last exon (exon 12) revealed a single band the size of the expected wild-type Asns transcript in all three genotypes (Figure SIV.4). These bands were gel purified and Sanger sequenced, which confirmed that the wild-type transcript was present in all three genotypes and no aberrant splicing events were observed. Additionally, Gapdh RT-PCR was performed as an internal control and the homozygous mice show a decreased abundance of the wild-type Asns transcript compared to the wild-type mice (Figure SIV.4). To quantify this difference, we performed quantitative real-time TaqMan PCR using a probe spanning exons seven and eight. The mRNA levels were significantly different between the three genotypes (one way ANOVA; $p < 0.00001$) (Figure IV.5A). A post-hoc two-tailed t test revealed that both mutant genotypes were significantly different from wild-type mRNA levels ($P_{Asns(+/+)\text{-}Asns(+/-)} = 0.00001$, $*P_{Asns(+/+)\text{-}Asns(-/-)} < 0.00001$) and significantly different from each other ($*P_{Asns(+/-)\text{-}Asns(-/-)} = 0.00083$). We were unable to determine whether there was differential expression of the Asns protein, due to lack of quality and specific mouse anti-Asns antibodies. These data demonstrate that the Asns gene-trap mouse is a hypomorph with ~20% of the

normal level of Asns mRNA being expressed. Given that two of the human mutations lead to decreased protein expression, this mouse provides a reliable model for this phenotype.

We next analyzed the brains of $Asns^{-/-}$ and control ($Asns^{+/+}$ or $Asns^{+/-}$) littermates from embryos and adults. We obtained coronal sections from postnatal day (P) 0 brains and measured cortical area, cortical thickness, and lateral ventricle area for each mouse using rostral-caudal-matched sections (using anatomical landmarks). We found that the cortical thickness and area of the $Asns^{-/-}$ brains were, on average, ~14% thinner and ~5% smaller than their control littermates, respectively. Additionally, the lateral ventricle area in the $Asns^{-/-}$ mice was significantly larger than their control littermates ($p = 0.019$; Figure SIV.4).

Due to the progressive nature of the human disorder, we next evaluated whether adults showed exacerbated brain defects. We generated paraffin-embedded coronal sections from P84 brains of $Asns^{-/-}$ and $Asns^{+/-}$ littermates (three of each genotype, representative sections shown in Figure SIV.4). The use of heterozygous animals was considered suitable because human carriers of ASNS mutations remain unaffected. We analyzed rostral-caudal-matched sections (using anatomical landmarks) from each animal for several parameters. Measurement of the cortical surface area, using methods described by Pulvers and colleagues (Pulvers et al., 2010), showed an ~8% reduction in cortical surface area of $Asns^{-/-}$ mice (Figure IV.5B). A similar reduction (~5%) was observed in the whole-brain surface area of $Asns^{-/-}$ mice (Figure IV.5C). We also observed that the $Asns^{-/-}$ brains had increased lateral ventricles (~95%) relative to control brains (Figure IV.5D). Importantly, the cortical thickness of the $Asns^{-/-}$ mice was significantly reduced compared to the $Asns^{+/-}$ mice ($p = 0.022$; Figure IV.5E).

$Asns^{+/-}$ and $Asns^{-/-}$ mice were assessed with age-matched B6NTac control animals in four behavioural assays. We found no genotype-associated differences in spontaneous locomotor activity, performance on the rotorod, or anxiety-like behaviour in the light-dark emergence test; however, $Asns^{+/-}$ mice were deficient and $Asns^{-/-}$ mice were severely impaired in short- and long-term memory in the novel object recognition task (Supplemental Experimental Procedures; Figure SIV.5). Careful observations of mice during behavioural testing revealed no evidence of seizure activity. To examine the possibility that $Asns^{-/-}$ mice might display epileptiform activity, we conducted prolonged video electroencephalogram (EEG) recordings in chronically implanted $Asns^{-/-}$ mice ($n = 2$) and a wild-type (WT) control ($n = 1$). Neither behavioural nor electrographic seizures were detected in $Asns^{-/-}$ mice or the WT controls.

Taken together, these data indicate that this Asns mouse model recapitulates the human brain phenotype, particularly in the reduced cortical thickness and increased lateral ventricle area.

DISCUSSION

We report that mutations in asparagine synthetase (ASNS) cause a distinct neurodevelopmental disorder characterized by congenital microcephaly, profound intellectual disability, and progressive cerebral atrophy. We found that two of these mutations reduce the abundance of the protein. Finally, we have shown that disrupting this gene in mice creates a model that mimics aspects of the human phenotype, including structural brain abnormalities and learning deficits, albeit with what appears to be a generally milder presentation than observed in humans.

Studies performed on cancer cells showed that asparagine depletion affects cell proliferation and survival (reviewed in (Richards et al., 2006). This is classically illustrated by the effect of asparaginase administration in childhood acute lymphoblastic leukemia. Asparaginase delivery to the bloodstream results in asparagine depletion causing a rapid efflux of cellular asparagine, which is also destroyed. Most cells express sufficient ASNS to counteract this asparagine starvation and survive, but not leukemic cells. Similarly, loss of ASNS activity in thermo sensitive mutant BHK cells leads to cell-cycle arrest as a consequence of a depletion of cellular asparagine (Greco et al., 1989 ; Li et al., 2006).

During development, Asns is expressed in regions where both neural progenitors and postmitotic neurons are present, suggesting that it may function in either or both of these populations. A subset of the brains from our subjects had simplified gyri. Similar features were found in the mutant mice, which showed decreased cortical thickness and enlarged lateral ventricles. These structural abnormalities could be caused in part by aberrations in neural progenitor proliferation during development, resulting from decreased asparagine levels. Asparagine depletion could also cause increased cell death in postmitotic neurons or glial cells, contributing to the progressive atrophy of the brain observed in our subjects.

Strikingly, ASNS deficiency causes severe neurological impairment, without any involvement of peripheral tissues. The concentration of asparagine in the cerebrospinal fluid (CSF) of humans is only ~10% of the concentration found in plasma (Scholl-Burgi et al., 2008) .The poor transport of asparagine across the blood-brain barrier suggests that the brain depends on local de novo synthesis, explaining why the phenotype is essentially neurological.

In addition to ID, a subset of our patients presented with features of hyperexcitability (including epilepsy and hyperekplexia). These features suggest a mechanism that is consistent with the accumulation of aspartate/glutamate in the brain, resulting in enhanced excitability and neuronal damage. While seizures in the patients could reflect enhanced excitability, these could also be secondary to the structural effects of altered proliferation. We cannot exclude the possibility that multiple mechanisms may be contributing to the observed human phenotype. Further analyses of animal and cellular models will help to elucidate the function of ASNS in normal brain development.

Of particular interest is the observation that Asns hypomorphic mice appear to have a milder phenotype than the humans with regards to more modest structural effects on the brain and no evidence of seizures. The ratio for the concentration of asparagine in the CSF to plasma in rats (0.26) (Nishimura et al., 1995) appears to be slightly elevated compared to that of humans [0.081 (Akiyama et al., 2014) to 0.118 (Scholl-Burgi et al., 2008)]. Assuming that the CSF/plasma ratio is similar in mouse and rat, this suggests that the concentration of asparagine is increased in the CSF and interstitial fluid (ISF) of mouse/rat as compared to humans. Thus, asparagine may be more readily available to the $Asns^{-/-}$ mice due to some physiological difference between humans and mice, such as transport at the blood-brain barrier. Alternately, it is possible that low levels of Asns expression in these mice result in a less severe phenotype. It will be of great interest to compare the hypomorph to a complete Asns null animal, which may show an even more dramatic phenotype.

With this report, ASNS deficiency becomes the third example of a recently recognized group of conditions resulting from the inability to synthesize a nonessential amino acid. These conditions all feature severe congenital encephalopathy with microcephaly. The others are glutamine synthetase deficiency (Haberle et al., 2005) and the serine biosynthetic disorders (Van Der Crabben et al., 2013). Although knowledge of ASNS deficiency and of other inborn errors of nonessential amino acid synthesis is incomplete, general considerations regarding diagnosis, disease mechanism, and treatment are in order. In almost every respect, the clinical approach to these diseases is predicted to be the opposite of that recommended for classical aminoacidopathies, which are caused by deficient breakdown of essential amino acids.

Strikingly, every diagnosis of ASNS deficiency was made by molecular genetics, despite extensive previous evaluation of patients that in several cases included amino acid

chromatography of plasma and CSF. Why was ASNS deficiency not suspected on these grounds? The answer may lie in a combination of technical considerations and biology. Compared to most amino acids, the normal levels of asparagine are low, both in plasma (e.g., 50.7 ± 17.7 mmol/l, in children 0–3 years old) and CSF (e.g., 4.0 ± 2.9 mmol/l) (Scholl-Burgi et al., 2008 ; Akiyama et al., 2014). For many reasons, low levels of a metabolite maybe less evident than increases. Abnormally low levels are more easily concealed by variations due to physiological state such as nutrition (which is difficult to standardize in ill newborns or infants) and to machine performance in diagnostic laboratories. In fact, currently used diagnostic technologies cannot discriminate low from normal CSF asparagine levels. In summary, results to date suggest that in patients with unexplained congenital encephalopathy with microcephaly, the absence of a low value does not exclude ASNS deficiency. In the future, an enzyme assay may play an important role in the complete diagnostic evaluation of patients suspected of ASNS deficiency but experience is too limited to conclude. In children with severe congenital encephalopathy and microcephaly, ASNS deficiency should be considered, and molecular diagnosis is the only method with proven reliability.

All three known deficiencies of amino acid biosynthesis present mainly with neurological features. In these conditions, the deficient amino acid becomes essential. Hence, an obvious first consideration for therapy is dietary supplementation, to provide the deficient amino acid to the brain. Plasma levels can usually be substantially increased by dietary supplementation and despite the complex transport systems for amino acids at the brain endothelium; a therapeutic benefit of supplementation has been reported in serine biosynthetic disorders and glutamine synthetase deficiency (Haberle et al., 2012 ; Van Der Crabben et al., 2013).Supplementation with asparagine therefore seems reasonable in ASNS deficiency. However, the prenatal onset of the microcephaly and the early postnatal presentation raise the possibility that such treatment will not be curative unless started prenatally.

The Asns mouse we have analyzed here will provide a model for future comprehensive exploration of the factors influencing phenotypic severity. Comparing this hypomorphic mouse model with a null mouse model will allow us to directly evaluate how residual levels of ASNS activity compare with the absence of ASNS activity, which may inform us about differences in clinical presentation. We can also utilize both animal models when testing the effects of dietary supplementation, which would ensure that a range of ASNS activities were

represented, thus covering the full range of ASNS activities that may also occur in patients. This work therefore sets the stage for evaluation of treatment options in Asns mouse models.

Early diagnosis of ASNS deficiency is now possible. Careful clinical observations and studies of Asns-deficient mice will help define the clinical spectrum and resolve central unanswered issues regarding the pathophysiology of this condition.

EXPERIMENTAL PROCEDURES

Recruitment of Subjects and Collection of Samples.

Families A and B were recruited at Sheba and Wolfson Medical Centers in Israel, family C at The Hospital for Sick Children in Toronto (Canada), and family D at Sainte-Justine Hospital in Montreal (Canada). Blood samples were obtained from most affected individuals, their unaffected siblings, and their parents. The relevant Institutional Review Boards approved the studies and appropriate family members gave written consent.

Sequencing and Variant Identification.

Exome Sequencing in Families A and B.

The Illumina Genome Analyzer IIx platform (Illumina) was used to perform exome sequencing of the three sequenced microcephaly patients (Figure IV.1, family A: II.1 and family B: II.2 and II.4) at the Center for Human Genome Variation (CHGV) at Duke University, Durham, NC. Prior to sequencing, target regions were captured using the SureSelect Human All Exon technology (Agilent Technologies). This technology captures consensus coding sequence exonic regions and flanking intronic regions totalizing ~38 Mb of genomic DNA. The resulting short-sequence reads were aligned to the reference genome [NCBI human genome assembly build 36; Ensembl core database release 50_361 (Hubbard et al., 2009)] using the Burrows-Wheeler Alignment (BWA) tool (Li et Durbin, 2009a). After accounting for PCR duplicates (removed using the Picard software: http://picard.sourceforge.net) and reads that did not align to captured regions of the reference genome, the average coverage for these three samples was ~71x and each sample had > 95% of the bases covered. A base within the 37.8 Mb captured region was defined as covered if ≥ 5 short reads spanned this nucleotide (Table SIV.10).

Genetic differences between each patient genome and the reference genome were identified using the SAM tools variant calling program (Li et al., 2009b), which identifies both single-nucleotide variants (SNVs) and small Indels. We then used the Sequence Variant Analyzer software (SVA) (Ge et al., 2011) to annotate all identified variants. SVA was also used to apply quality control filters to the variants identified by SAM tools. High-quality SNVs were obtained using the following criteria: consensus score ≥ 20, SNP quality score ≥ 20, and reads supporting SNP ≥ 3. High-quality Indels were obtained using the following criteria: consensus score ≥ 20, index quality score ≥ 50, ratio of (reads supporting variant/reads supporting reference): 0.2–5.0, and reads supporting index ≥ 3.

Exome Sequencing in Families C and D.

The exomes of one of the individuals of family C (II.3) and four individuals of family D (II.1, II.2, II.4, and II.5) were captured using the Agilent SureSelect all exon kit V3 (approximately 51.9 Mb of target sequences) and then sequenced in pair ends (2 x 100 bp) on the Illumina HiSeq2000 (v3 chemistry; 3 exomes/lane format) at the McGill University Genome Quebec Innovation Center (Montreal, Canada). The sequences were aligned and the variants were called using GATK (Depristo et al., 2011). After removal of duplicate reads, using Picard, we obtained an average coverage of > 80x per target base, with 95% of the target bases being covered at ≥ 10x. Only variants that meet all the following criteria were considered: base

coverage ≥ 8x, reads supporting the variant ≥ 3, and ratio of reads supporting variant/reads supporting reference ≥ 20%. Variants were then annotated using Annovar (Wang et al., 2010).
Genotyping p. F362V. Genotyping of the p.F362V variant in 1,160 controls was performed in the Center for Human Genome Variation at Duke University (Durham, NC). This was done using a custom TaqMan genotyping assay (Applied Biosystems): forward: 5'-CCTGCGTAAGTTCATCTGATCCTT-3'; reverse: 5'GTATATTCGGAAGAACACAGATAGCGT-3'; probe: 5'-TCCAGAGA [A/C] GATCACC-3'.

Genotyping of the p.F362V variant in 80 Iranian Jewish controls and the non exome-sequenced family members (Figure 1; family A: I.1, I.2, II.2, II.3, and II.4 and family B: I.1, I.2, II.1) was performed at the Gertner Institute of Human Genetics, Sheba Medical Center, Israel. Sanger Sequencing (Figure 1B) or restriction digest with the restriction enzyme Alw26I (data not shown) were used to perform this genotyping. Both methods used the following custom primer sequences: forward: 5'-CTTTCAATTATTTCCAAAAATCAAATC-3, and reverse: 5'-CACTGTCATACTGAAAGATGATAGAAA-3'. These primers resulted in a 286 bp amplicon that targeted the nucleotide of interest.

The p.F362V variant, found in families A and B, was validated in these three samples using all three methods: TaqMan genotyping, Sanger sequencing, and restriction digestion.

Genotyping p.R550C and p.A6E in French Canadian and Bangladeshi Controls.

Sanger sequencing of PCR-amplified products was used to genotype p.R550C and p.A6E variants. The following custom primers were used for p.A6E: forward: 5'-GCCGGTTGAATGTAGAGGTC-3' and reverse: 5'-CCAAAGCAGCAGTTGGTGTA-3'. The following custom primers were used for p.R550C: forward: 5'-GCCATTTTAAGCCATTTTGC-3' and reverse: 5'-TTTCCCTTTTCCTAGCTTACCC-3'. The mutations p.R550C and p.A6E were genotyped in 300 French Canadian healthy controls. In addition, p.R550C was genotyped in 225 Bangladeshi healthy controls.

Clone Preparations .

Full-length cDNA encoding human ASNS was amplified from first-strand cDNA derived from the HEK293 human kidney cell line with an RNeasy plus mini kit (QIAGEN), High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), Phusion HF DNA polymerase (Finnzymes), and a specific primer set (5'-CTCGAGATGTGTGGCATTTGGGCGCT-3 and 5'-CTCGAGCCTAAGCTTTGACAGCTGACT-3').
The cDNA was subcloned into the pCR-Blunt II-TOPO vector (Invitrogen-Life Technologies) and subjected to sequence analysis (pCR-Blunt II-ASNS-WT). Using pCR-Blunt II-ASNS-WT, A6E, F362V, and R550C of ASNS were made by PCR-mediated site-directed mutagenesis using Phusion HF DNA polymerase and a specific primer set (A6E: 5'-GCTGTTTGGCAGTGATGATTG-3' and 5'-TCCCAAATGCCACACATCTC-3'; F362V: 5'-GTCTCTGGAGAAGGATCAGA-3' and 5'-GATCACCACGCTATCTGTGT-3'; R550C: 5'-GCACGCTGACCCACTAC-3' and 5'-AGGCAGAAGGGTCAGTGC-3'), which were phosphorylated by T4 polynucleotide kinase (New England BioLabs). The amplicons were self-ligated using T4 DNA ligase (Promega) and subjected to sequence analysis (pCR-Blunt II-ASNS-A6E, pCR-Blunt II-ASNS-F362V, and pCR-Blunt II-ASNS-R550C). ASNS human cDNA containing each allele was subcloned into the pcDNA3.1$^{(+)}$ vector (Invitrogen-Life

Technologies) using the KpnI and XbaI sites from pCR Blunt II-ASNS-WT, pCR-Blunt II-ASNS-A6E, pCR-Blunt II-ASNS-F362V,or pCR-Blunt II-ASNS-R550C and subjected to sequence analysis (pcDNA3.1$^{(+)}$-ASNS-WT, pcDNA3.1$^{(+)}$-ASNS-A6E, pcDNA3.1$^{(+)}$-ASNSF362V,or pcDNA3.1$^{(+)}$-ASNS-R550C; Figure SIV.2).
Using pcDNA3.1$^{(+)}$-ASNS-WT, pcDNA3.1$^{(+)}$-ASNS-A6E, pcDNA3.1$^{(+)}$-ASNS-F362V, or pcDNA3.1$^{(+)}$-ASNS-R550C, FLAG-tagged-modified ASNS was made by two-step PCR-mediated site-directed mutagenesis using Phusion HF DNA polymerase and specific primer sets (first step: 5'-GACAAGTAGGCTCGAGAAGGG-3' and 5'-GTAGTCAGCTTTGACAGCTGAC-3' ; second step: 5'-GACGATGACAAGTAGGCTCGAGAAGGG-3' and 5'-GTCCTTGTAGTCAGCTTTGACAG-3'), which were phosphorylated by T4 polynucleotide kinase, the amplicons were self-ligated using T4 DNA ligase and subjected to sequence analysis (pcDNA3.1$^{(+)}$-ASNS-FLAG-WT, pcDNA3.1$^{(+)}$-ASNSFLAG-A6E, pcDNA3.1$^{(+)}$-ASNS-FLAG-F362V, or pcDNA3.1$^{(+)}$-ASNS-FLAG-R550C).cDNAs encoding FLAG-tagged human ASNS were subcloned into pcDNA3.1$^{(+)}$ vector again, using the KpnI and XbaI sites and subjected to sequence analysis (Figure SIV.2).

Cell Culture.

RT-PCR.

Empty pcDNA3.1 $^{(+)}$ vector, pcDNA3.1$^{(+)}$-ASNS wild-type, or pcDNA3.1$^{(+)}$-ASNS mutant (p.F362V, p.R550C, or p.A6E) were transfected into the monkeyCOS-7 kidney cell line or human HEK293 kidney cells by lipofection using Lipofectamine 2000 (Invitrogen-Life Technologies). Total RNA was extracted from transfectants using an RNeasy plus mini kit, and first-strand full-length cDNA encoding human ASNS was synthesized using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). RT-PCR to detect ASNS mRNA expression was performed in 25 cycles at 96°C for 30 s, 60°C for 30 s, and 72°C for 30 s using AmpliTaq Gold DNA polymerase (Applied Biosystems) and a specific primer set (5'-TGCACGCCCTCTATGACAAT-3' and 5'-CACCTTTCTAGCAGCCAGTA-3') (Figure IV.3A).

Western Blotting.

Forty-eight hours after transfection, the cells were lysed with RIPA buffer (Sigma-Aldrich) with Protease inhibitor cocktail (Sigma-Aldrich), and the lysates were subjected to SDS-PAGE gel and transferred to a polyvinylidene difluoride membrane (Milipore). The membranes were incubated with anti-FLAG M2 monoclonal antibody (Sigma-Aldrich) or anti-actin antibody (Santa Cruz Biotechnology). Proteins were visualized with the ECL plus western blotting detection system (GE Healthcare). For Leupeptin treatment, 24 hours post transfection, the cells were incubated with 100 mM Leupeptin (Sigma Aldrich). After 8 hours incubation with Leupeptin, the cells were lysed, and FLAG-tagged ASNS were detected as above.

Protein Conservation.

Species and ASNS proteins were from gi P08243 (Human, Homo sapiens), ENSMUSP00000031766 (mouse, Mus musculus), ENSGALP00000015846, (chicken, Gallus gallus), ENSACAP00000012780 (lizard, Anolis carolinensis), ENSXETP00000054608 (frog,

Xenopus tropicalis), ENSTRUP00000013503 (fish, fugu, Takifugu rubripe), FBpp0089009 (fruit fly, Drosophila melanogaster), NP_741864 (worm, nematode, Caenorhabditis elegans), YGR124W (yeast, Saccharomyces cerevisia), and YP_003233213.1 (bacterium, Escherichia coli) (Figure IV.4A). Sequence alignment was performed using ClustalW (Thompson et al., 2002) and alignment editing with the BioEdit software (http://www. mbio.ncsu.edu/bioedit/bioedit.html).The pdb structure was made using Discovery Studio program (http://accelrys.com/products/discovery-studio) (Figure IV.4B).

Mouse Analyses.

Asns mice were obtained from the Eucomm consortium. Mice were maintained by breeding to C57BL/6NTac. Histology at P0 was performed by cryopreservation of tissue, cryosectioning, and hematoxylin and eosin staining. Histology in adult brains was performed by fixation of tissue using formalin perfusion. Tissue was sent to http://www.histoserv.com for paraffin embedding, sectioning, and staining. Analysis of area and thickness was performed by quantifying measurements using ImageJ. The p values for structural measurements were obtained using an unpaired t test and calculations were done using R.

cDNA.

Mouse cerebral hemispheres were carefully dissected. Total RNA was extracted from brain tissue using an RNeasy plus mini kit, and first-strand full-length cDNA encoding human ASNS was synthesized using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems).

Quantitative Real-Time RT-PCR.

Quantitative real-time PCR was done using an Asns gene expression assay, with FAM reporter, spanning exons 7–8 (Mm00803785_m1; Life Technologies) and a Gapdh gene expression assay with VIC reporter (Mm99999915_g1; Life Technologies). Samples were run in triplicate and the standard curve was made using cDNA from a nontest wild-type sample. Twelve mice between 3 and 4 months of age were used for qPCR. Four mice of each genotype were used (Asns+/+, Asns+/- and Asns-/-). One-way ANOVA was used to assess expression differences between the three genotypes ($p < 0.00001$). A post-hoc two-tailed t test was then used to assess genotypic differences in expression (PWT-Asns+/- = 0.00001, *P WT-Asns-/- < 0.00001, *P Asns+/--Asns-/- = 0.00083).

Regular RT-PCR.

RT-PCR to detect Asns mRNA expression was performed in 35 cycles at 96o C for 30 s, 58o C for 30 s, and 72o C for 90 s using AmpliTaq Gold DNA polymerase (Applied Biosystems) and a specific primer set (5'-CAGTGTCTGAGTGCGATGAAGA-3'and 5'-GCGTTCAAAGATCTGACGGTAG-3' (Figure SIV.4). RT-PCR to detect Gapdh mRNA expression was performed in 25 cycles at 96o C for 30 s, 57o C for 30 s, and 72o C for 45 s using AmpliTaq Gold DNA polymerase (Applied Biosystems) and a specific primer set (5'-ACCACAGTCCATGCCATCAC-3' and 5'-CACCACCCTGTTGCTGTAGCC-3') (Figure SIV.4)

Western Blotting.

Two different antibodies were tried for detection of mouse Asns: anti-human-ASNS, which recognizes amino acid residues 506–520 of ASNS (Sigma- Aldrich), and anti-Asns, with

species reactivity in mouse, rat, and human, which recognizes amino acid residues at the C terminus (Abcam). Both were nonspecific (data not shown).

Video EEG Recordings of Mice.

Two adult Asns homozygous mice and one age-matched WT mouse were anesthetized by intraperitoneal injection of Nembutal (60 mg/kg). Under stereotaxic guidance, four monopolar electrodes were implanted into the subdural space over the left and right parietal cortex and occipital cortex for chronic EEG recording. After a 7-day postoperative recovery, EEG activity was recorded with the mouse moving freely in a cage for 80 hours ; animal behaviour was recorded simultaneously with a digital video camera. The EEG and behavioural activities were analyzed by an individual blinded to mouse genotype.

Acknowledgments

We would like to thank all four families for their willingness to participate in this study. J.L.M. is a National Scientist of the Fonds de Recherche du Québec - Santé. E.K.R. is funded by a predoctoral grant from the Epilepsy Foundation and the Jo Rae Wright Fellowship for outstanding women in science (Duke University). J.M.C.-C. holds a salary award from the Réseau de Médecine Génétique Appliquée du Québec (RMGA). We acknowledge the following colleagues for supplying control samples: R. Brown, G. Cavalleri, L. Cirulli, N. Delanty, C. Depondt, V. Dixon, E. Heinzen, J. Hoover-Fong, A. Husain, D. Levy, K. Linney, W. Lowe, J. McEvoy, M. Mikati, J. Milner, A. Need, R. Ottman, R. Radtke, J. Silver, M. Silver, S. Sisodiya, N. Sobriera, D. Valle, and N. Walley. We wish to thank Katherine Whang for helping to section the mouse brains. We wish to thank C. Means and T. Rhodes for helping with the behavioural experiments and J. Zhou and C. Elms for breeding, genotyping, and maintaining the mice. We thank R. Olender and P. Allard for helpful insights. We also thank the members of the RMGA bioinformatics team (Alexandre Dionne-Laporte, Dan Spiegelman, Edouard Henrion, and Ousmane Diallo) for the bioinformatics analysis of the exome sequencing data (families C and D). This research has been funded in part by federal funds from the Center for HIV/AIDS Vaccine Immunology (‘‘CHAVI’’) under a grant from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Grant NumberUO1AIO67854 to D.B.G., and by the March of Dimes (grant no. 12-FY10-236) and Canadian Institutes of Health Research (MOP 106499) to J.L.M. Additional funding provided by: ARRA 1RC2NS070342-01, NIMH Grant RC2MH089915,NINDS Award RC2NS070344, and the Crown Human Genome Center at the Weizmann Institute of Science.

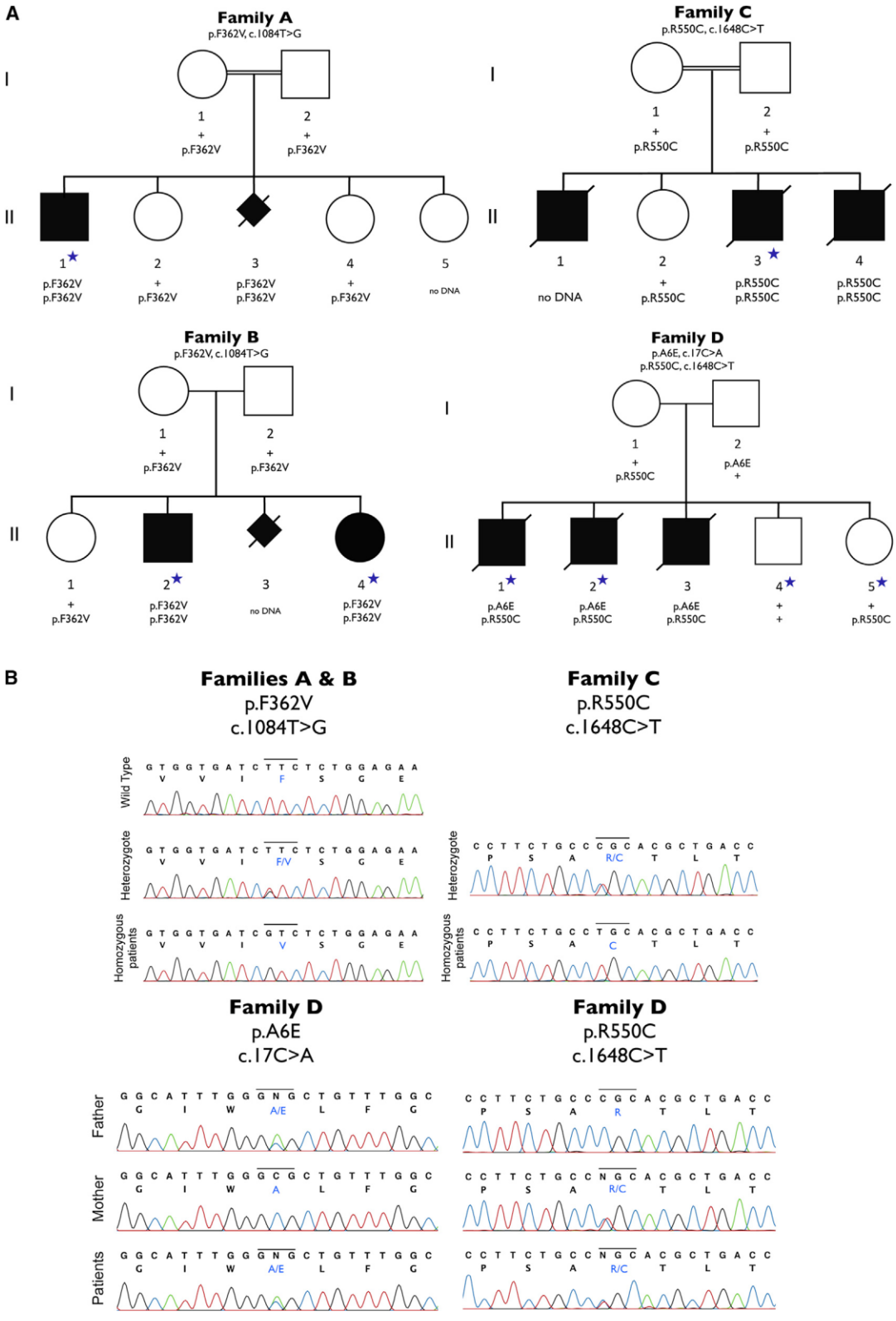

A
Family A
p.F362V, c.1084T>G
I
1
+
p.F362V
2
+
p.F362V
II
1★
p.F362V
p.F362V
2
+
p.F362V
3
p.F362V
p.F362V
4
+
p.F362V
5
no DNA
Family C
p.R550C, c.1648C>T
I
1
+
p.R550C
2
+
p.R550C
II
1
no DNA
2
+
p.R550C
3★
p.R550C
p.R550C
4
p.R550C
p.R550C
Family B
p.F362V, c.1084T>G
I
1
+
p.F362V
2
+
p.F362V
II
1
+
p.F362V
2★
p.F362V
p.F362V
3
no DNA
4★
p.F362V
p.F362V
Family D
p.A6E, c.17C>A
p.R550C, c.1648C>T
I
1
+
p.R550C
2
p.A6E
+
II
1★
p.A6E
p.R550C
2★
p.A6E
p.R550C
3
p.A6E
p.R550C
4★
+
+
5★
+
p.R550C
B
Families A & B
p.F362V
c.1084T>G
Family C
p.R550C
c.1648C>T
Wild Type
Heterozygote
Homozygous patients
Heterozygote
Homozygous patients
Family D
p.A6E
c.17C>A
Family D
p.R550C
c.1648C>T
Father
Mother
Patients
Father
Mother
Patients

Figure IV.1

Figure IV.1 Four families with *ASNS* mutations.

(A) Cosegregation of mutations within each of the four families. Filled symbols represent known or presumed (in the case of aborted fetus) affected individuals. The individuals with a blue star were exome sequenced.
(B) Sanger sequencing confirmation for all three ASNS mutations in the four families. For family D, Sanger sequencing also confirmed inheritance of each mutation from a different parent (compound heterozygote).

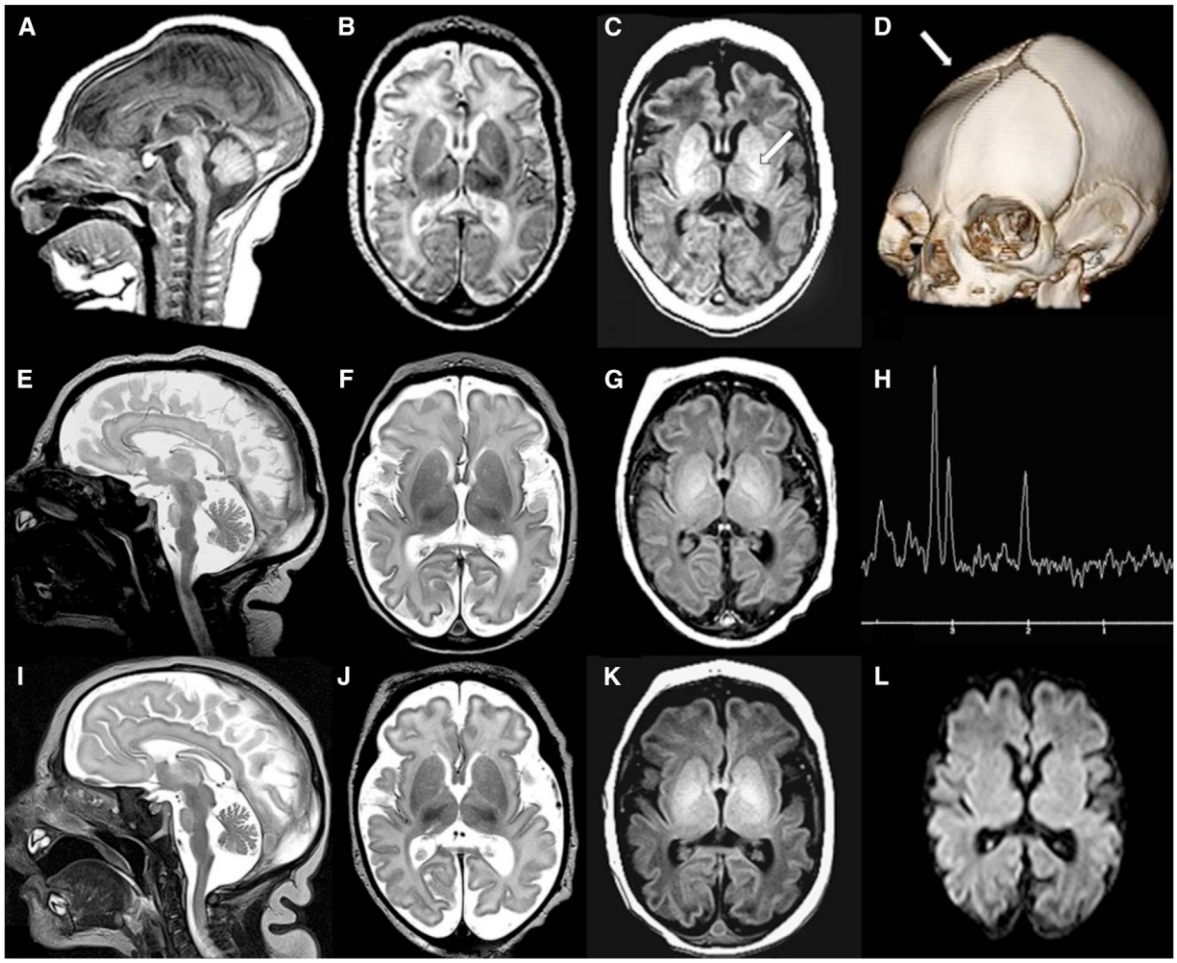


Figure IV.2

Figure IV.2 MRI images from family C.

(A–D) Sibling C.II.1. Sagittal T1W image (A) performed at 13 days of age reveals decreased size of pons. Axial T2W image (B) reveals prominent pericerebral fluid spaces surrounding the brain due to volume loss. White matter is diffusely higher in signal intensity than the cortical ribbon, suggesting delayed myelination. AxialT1W image (C) confirms lack of bright myelin stripe in the posterior limb of internal capsule (arrow). Three-dimensional computed tomography (D) at 2 months of age confirms ridge (arrow) from overlapping sutures, palpable at birth. DNA confirmation was not obtained in this infant. (E–H) Sibling C.II.3. Sagittal T2W image (E) performed at 6 days of age reveals decreased size of pons. Axial T2W image (F) demonstrates prominent pericerebral fluid spaces surrounding the brain due to volume loss. Axial T1W image (G) confirms lack of myelin stripe in posterior limb of the internal capsule (PLIC). Magnetic resonance spectroscopy (MRS) TE144 (H) performed in the basal ganglia is age appropriate (as it was in all three siblings). (I–L) Sibling C.II.4. Sagittal T2W image (I) performed at 1 day of age reveals decreased size of pons. Axial T2W image (J) confirms pericerebral fluid prominence due to cerebral volume loss in a manner similar to siblings. Axial T1W image (K) shows lack of myelin stripe in PLIC. Diffusion-weighted imaging (DWI) (L) is normal (as it was in all three siblings).

See also Figure SIV.1.

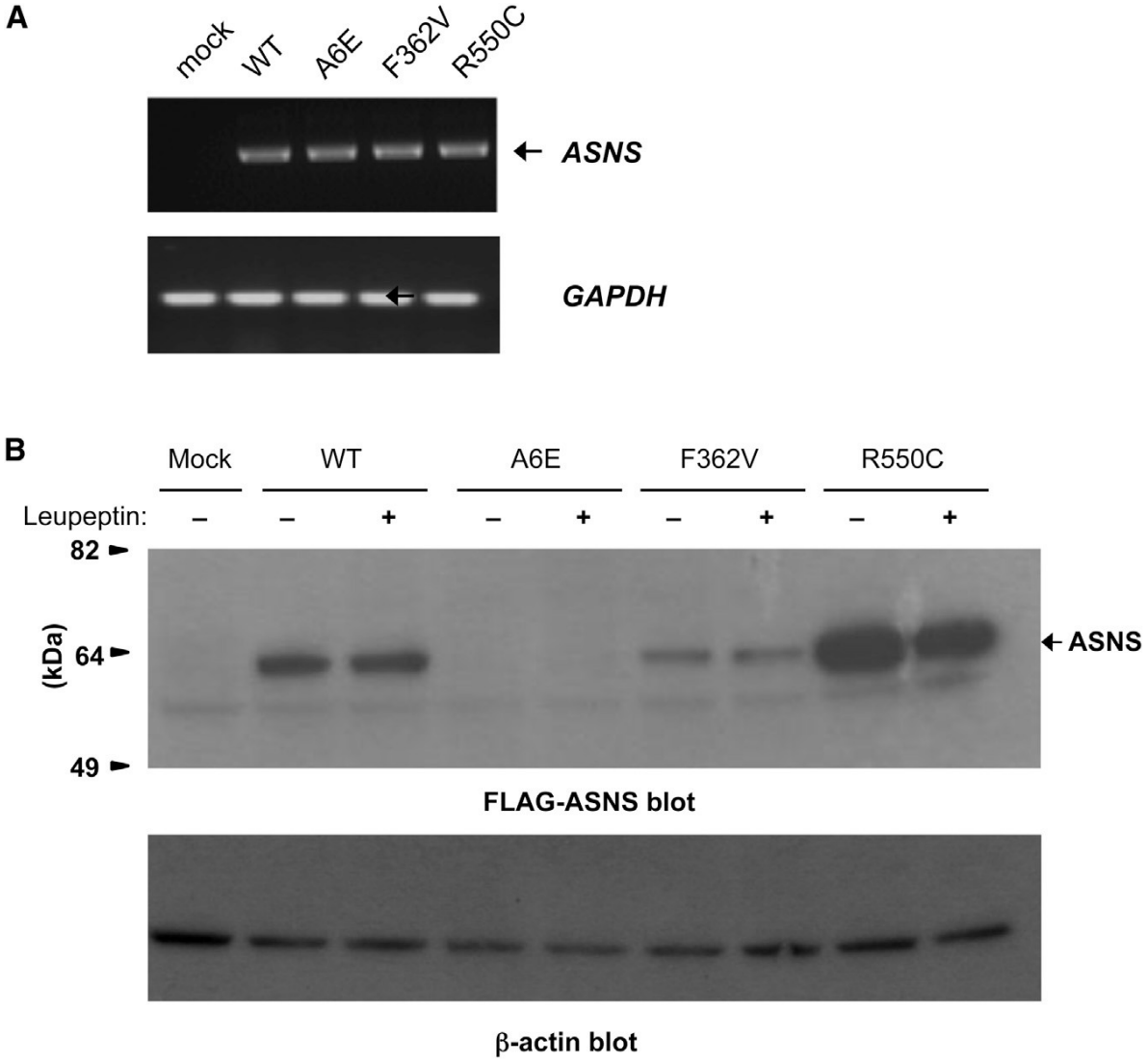
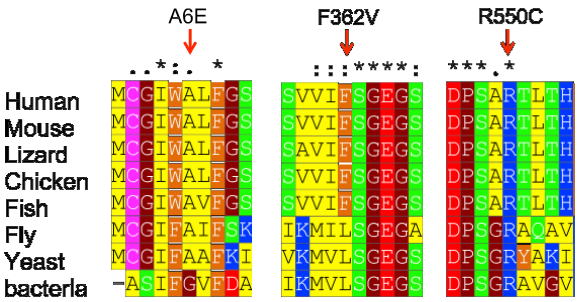

A
mock
WT
A6E
F362V
R550C
ASNS
GAPDH
B
Mock
WT
A6E
F362V
R550C
Leupeptin:
– – + – + – + – +
82
64
49
(kDa)
ASNS
FLAG-ASNS blot
β-actin blot

Figure IV.3

Figure IV.3 Functional impact of *ASNS* mutations.

(A) RT-PCR to detect ASNS mRNA expression in COS-7 cells transfected with empty, wild-type, or mutant vectors. (B) Western blots detecting ASNS-FLAG protein abundance, with or without Leupeptin treatment, in the HEK293 transfectants using an anti-FLAG antibody. β-actin was used as a loading control. See also Figure SIV.2.

A



B

Figure IV.4.

Figure IV.4 Location and conservation of mutated residues in the *ASNS* gene.

(A) A multiple alignment of human ASNS and selected orthologs. Only the regions harboring mutations are shown. On top: ‘‘*’’ represents identical position; ‘‘:’’ represents conserved substitutions; ‘‘.’’ represents semi conserved substitution. Color code is for amino acid type: red, negative; blue, positive; green, polar; yellow, aliphatic; orange, aromatic; brown, helix breaker. (B) The structure of E. coli glutamine-dependent asparagine synthetase B (protein databank ID 1CT9), a bacterial ortholog of human ASNS (A). The N-terminal glutamine aminotransferase domain is colored in pink, and the C-terminal asparagine synthase domain is colored in blue. The residues inferred to be equivalent to the mutated A6E and F362V are highlighted in yellow and the approximate location of the mutated R550C is shown with an asterisk (crystallography could not be determined for the distal end of C-terminal domain). Also, the AMP and glutamine molecules are shown in space-filling style.

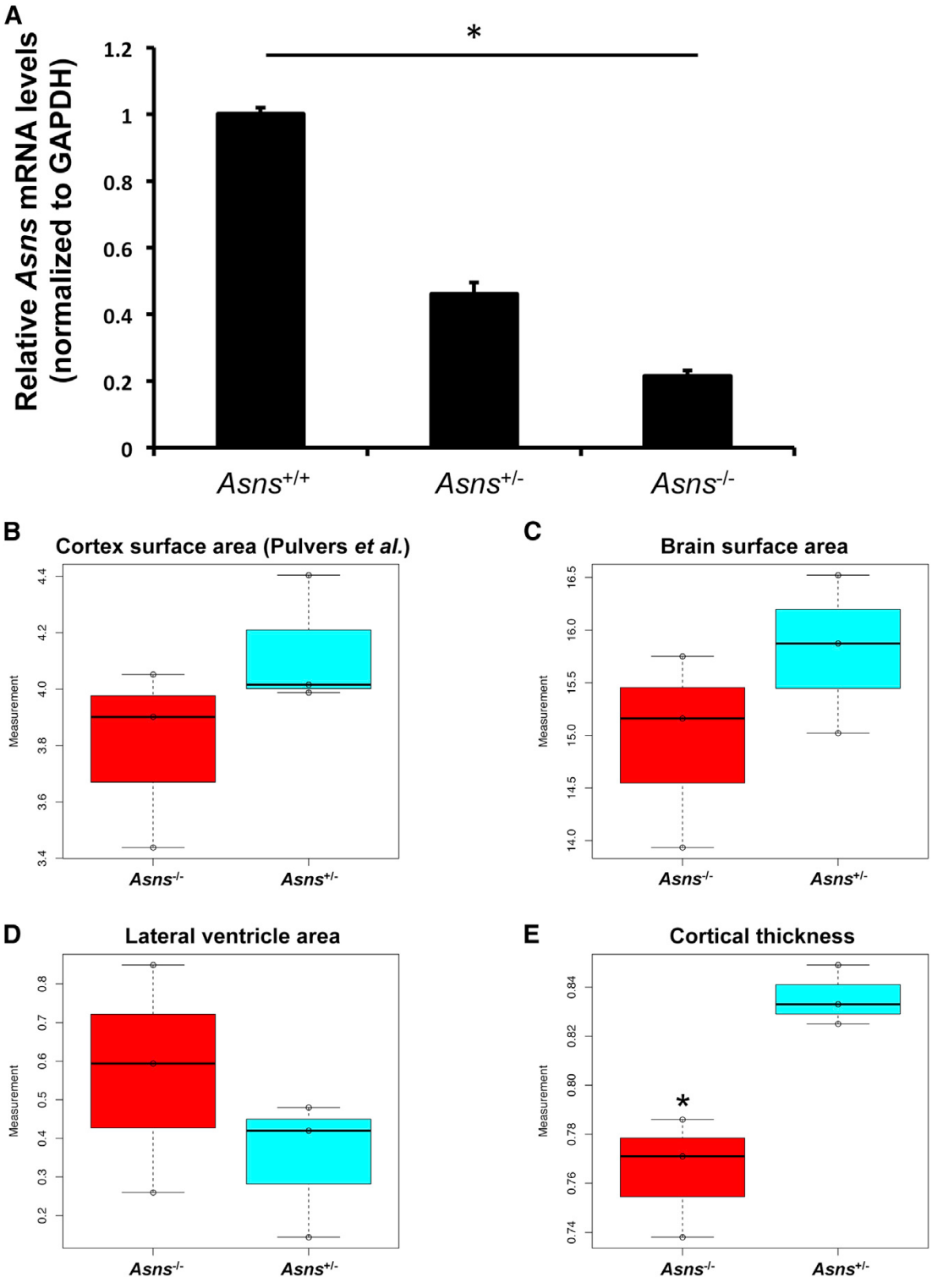

A
Relative Asns mRNA levels (normalized to GAPDH)
1.2
1
0.8
0.6
0.4
0.2
0
*
Asns+/+
Asns+/-
Asns-/-
B
Cortex surface area (Pulvers et al.)
Measurement
Asns-/-
Asns+/-
C
Brain surface area
Measurement
Asns-/-
Asns+/-
D
Lateral ventricle area
Measurement
Asns-/-
Asns+/-
E
Cortical thickness
Measurement
*
Asns-/-
Asns+/-

Figure IV.5.

Figure IV.5 Asns-deficient mice and structural brain abnormalities.

(A) Detection of Asns mRNA (qRT-PCR) in the mouse brain. Four mice of each genotype were used (all 3–4 months of age). The expression differences between the three genotypes are significant (one-way ANOVA test *p < 0.00001). A post-hoc two-tailed t test revealed both mutant genotypes were significantly different from wild type expression ($P_{Asns+/-\_Asns+/-}$= 0.00001,*$P_{Asns+/+\_Asns-/-}$< 0.00001) and significantly different from each other (*$P_{Asns+/-\_Asns-/-}$= =0.00083). Error bars represent SEM. (B–E) Measurements of adult mouse brain (P84) coronal sections, analyzed using ImageJ software (Schneider et al., 2012) , comparing three homozygous mutants to heterozygous littermates. (B) Cortical surface area as measured in (Pulvers et al., 2010). (C) Surface area of the left hemisphere of the brain section, (D) surface area of the lateral ventricle in the left hemisphere of the brain section, and (E) cortical thickness measured from the edge of the hippocampus to the outer cortex. Asterisk (*) indicates a significant difference by an unpaired t test (p = 0.022). Error bars represent the range of the observed data. See also Figures SIV.3, SIV.4, and SIV.5.

Table IV.1 Clinical features of patients with mutations in *ASNS.*

	Family A	Family B		Family C			Family D		
Consanguinity	Yes	No		Yes			No		
Ethnic origin	Iranian Jews	Iranian Jews		Bangladeshi			French Canadian		
Subjects	A.II.1	B.II.2	B.II.4	C.II.1	C.II.3	C.II.4	D.II.1	D.II.2	D.II.3
Genotype	p.F362V/p.F362V	p.F362V/p.F362V	p.F362V/p.F362V	Not determined	p.R550C/ p.R550C	p.R550C/ p.R550C	p.A6E/p.R550C	p.A6E/p.R550C	p.A6E/p.R550C
Gender	Male	Male	Female	Male	Male	Male	Male	Male	Male
Age	14 years	14 years	12 years	4 months†	3 months†	6 months†	9 days†	11 months†	12 months†
HC[1] at birth	31.5 cm (−2.5 SD)	31 cm (−3 SD)	31 cm (−2 SD/−3 SD)	30.5 cm (−3.5 SD)	33 cm (−1 SD/−2 SD)	32 cm (−2 SD)	31.5 cm (−2.5 SD)	31 cm (−3 SD)	28.5 cm (−1 SD/−2 SD)[2]
Developmental delay	Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe
Progressive microcephaly	Yes	Yes	Yes	Yes	Yes	Yes	NA	Yes	Yes
Epilepsy									
Age at onset	1 month	2 weeks	3 weeks	None	None	None	4 days	9 months[5]	8 days
Type of seizures	Spasms, tonic, myoclonic, GTC[3]	Spasms, tonic, myoclonic, GTC[3]	Spasms, tonic, myoclonic, GTC[3]	None	None	None	Tonic, orobuccal	Partial complex	Partial complex
EEG pattern	Hypsarrythmia, MISF[4]	Hypsarrythmia, MISF[4]	Hypsarrythmia, MISF[4]	Disorganized background	Disorganized background	Disorganized background	NA	Suppression bursts, MISF	Suppression bursts, MISF
Clinical examination									
Axial hypotonia	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Appendicular hypertonia	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hyperreflexia	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hyperekplexia	No	No	No	Yes	Yes	Yes	No	No	No
Brain MRI									
Decreased cerebral volume	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decreased size of pons	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
simplified gyri	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

†Deceased; [1]head circumference; [2]born at 33.5 weeks of gestation; [3]generalized tonic-clonic seizures; [4]multiple independent spike foci; [5]tremulous movements and abnormal EEG at 4 days; partial complex seizures with eye deviation and focal epileptic activity at 9 months.

Table IV.2 Mutations identified in *ASNS.*

Nucleotide and amino acid positions are based on the NCBI Reference sequences NM_183356.3 and NP_899199.2, respectively. See also Table SIV.1-10

Family	Genotype	ASNS Modification	Nucleotide Change	Ethnicity	Frequency in Control In-house Exomes/Genomes	Frequency in Ancestry-Matched Controls	PolyPhen-2	SIFT
A	Homozygous	p.F362V	c.1084T>G	Iranian Jewish	0/261	1/80	Damaging (0.95)	Damaging (0.04)
B	Homozygous	p.F362V	c.1084T>G	Iranian Jewish	0/261	1/80	Damaging (0.95)	Damaging (0.04)
C	Homozygous	p.R550C	c.1648C>T	Bangladeshi	0/169	0/225	Damaging (1.00)	Damaging (0.01)
D	Compound heterozyous	p.R550C	c.1648C>T	French Canadian	0/169	0/300	Damaging (1.00)	Damaging (0.01)
		p.A6E	c.17C>A		0/169	0/300	Damaging (0.898)	Deleterious (0.02)

Table IV.3 Measurements of amino acids concentrations in patients blood and urine.

Data were not available for all fields in all patients (NA). The reference concentrations are indicated in parentheses. See also Table SIV.11

	Plasma Amino Acid Levels			Urine Amino Acid Levels		
	Glutamine	Aspartate	Asparagine	Glutamine	Aspartate	Asparagine
B.II.2	1,250 (254–823 μmol/l)	18 (1–24 μmol/l)	57 (23–112 μmol/l)	382 (20–112 μmol/mmol creatinine)	36 (1–10 μmol/mmol creatinine)	21 (0–24 μmol/mmol creatinine)
B.II.4	1,149 (254–823 μmol/l)	2 (1–24 μmol/l)	49 (23–112 μmol/l)	57 (20–112 μmol/mmol creatinine)	40 (1–10 μmol/mmol creatinine)	4 (0–24 μmol/mmol creatinine)
C.II.1	NA	7 (17–21 μmol/l)	NA	NA	NA	NA
C.II.3	NA	NA	12 (16–21 μmol/l)	NA	NA	NA
C.II.4	NA	12 (0–20 μmol/l)	NA	NA	NA	NA
D.II.1	439 (474–736 μmol/l)	7 (4–18 μmol/l)	11 (31–56 μmol/l)	NA	NA	NA
D.II.2	668 (474–736 μmol/l)	9 (4–18 μmol/l)	55 (31–56 μmol/l)	NA	NA	NA

SUPPLEMENTARY EXPERIMENTAL PROCEDURES

Comparison to MCPH genes.

Despite congenital microcephaly being a consistent feature of this syndrome, our patients do not fit the definition of primary microcephaly (MCPH). MCPH disorders are characterized by: a profoundly small head at birth, modest developmental delay (considering the dramatic microcephaly), mild or no seizures, and little or no deterioration. To date, there are seven distinct chromosomal loci for autosomal recessive primary microcephaly: MCPH1-MCPH7 (MIM 251200, 604317, 604804, 604321, 608716, 608393, and 612703, respectively). The causal genes have been identified for all of these loci (Table SIV.8), including the recent discoveries of causal genes in MCPH2 (Bilguvar et al., 2010 ; Nicholas et al., 2010) and MCPH4 (Guernsey et al., 2010). Multiple types of casual mutations (e.g., non synonymous, small deletions) have been identified in each of these genes (Bilguvar et al., 2010 ; Guernsey et al., 2010 ; Kaindl et al., 2010 ; Nicholas et al., 2010) .The observed genetic heterogeneity is not surprising given the broad clinical spectrum of primary microcephaly (Cox et al., 2006).

However, since we still wanted to exclude the possible involvement of previously identified primary microcephaly genes (Table SIV.8), these loci were analyzed with respect to the 1,532 homozygous regions identified in any of the three case genomes (families A and B). There were four homozygous regions that overlapped one of the microcephaly-associated genes, however these regions were never homozygous in more than one patient (Table SIV.9). Parenthetically, exome sequencing has the potential to miss casual variants if the exon harbouring it is not captured or if there is undetectable structural variation. Our analysis, however, suggests that homozygous variants on a single haplotype have not been missed in any of the microcephaly genes.

The new syndrome most resembles a host of neurometabolic disorders (Kelley et al., 2002 ; Hart et al., 2007 ; Siu et al., 2010) that show microcephaly at birth and progressive deterioration, including progressive brain atrophy. Amish microcephaly (MCPHA, OMIM #607196) is one of the most well defined examples of a metabolic disorder, which resembles the new syndrome. MCPHA is characterized by infantile microcephaly and α-ketoglutaric aciduria. However, we noted that Amish microcephaly is more severe than the new syndrome in that Amish microcephaly is typically diagnosed at the second trimester (compared to the third trimester in our patients) and is associated with arrest of brain development in utero (Kelley et al., 2002 ; Siu et al., 2010).

Phenotypic characterization.

Families A & B.

Patients were born at term with a small head circumference (at least two standard deviations below average). From early infancy they showed feeding difficulties, jitteriness, and increased disabling spastic tone, followed by evolution of intractable infantile spasms around the age of three months. The spasms were replaced by frequent myoclonic and tonic seizures, partially responsive to anticonvulsant therapy. By the age of one year, the children failed to reach significant milestones, showed spastic quadriparesis, profound ID, and cortical blindness and had ahead circumference seven standard deviations below average. Physical examination was remarkable for mild dysmorphic features including receding forehead, big fleshy ears and

relatively big hands and feet. Electroencephalogram (EEG) evolved from burst suppression pattern at three months to multiple independent spike foci (MISF) on a slow background. Visual and auditory evoked potentials (VEP, BERA), as well as electroretinogram (ERG), were normal. Brain CT at two months of age showed cortical atrophy without calcifications. Magnetic resonance imaging (MRI) at the age of three months showed diffuse cerebral atrophy, enlarged extra-ventricular spaces with normal size cerebellum and brainstem and at the age of 18 months, progressive cortical and subcortical atrophy with delayed myelination, moderately enlarged lateral ventricles and thin corpus callosum. The brain stem was small, while the cerebellum was only very mildly affected (Figure SIV.1B (ii)). In each of the families one pregnancy was terminated due to arrest of head growth detected by a fetal ultrasound. In the aborted child from family A, a fetal brain MRI was also performed, which identified bilateral cortical dysplasia and cerebellar hypoplasia, both of which were confirmed by autopsy.

Family C.

The patients were three siblings born to a consanguineous Bangladeshi couple in good health. An older sister is unaffected. The parents are second cousins in that the mother's maternal grandmother and the father's maternal grandmother were siblings.

Patient C.II.1.

The couple's first child, a male, was born at term following an uncomplicated pregnancy. Delivery was by C-section due to thick meconium as well as fetal distress. His APGAR score were 5 and 8 at 1 and 5minutes respectively, his birth weight was 3400 g (50th centile) and his head circumference (OFC) was 30.5 cm (-3.5 SD). He was noted to be jittery during the first hour of life with episodes of hyperekplexia. He required oxygen at birth but quickly was weaned. EEG recording at 9 days of age was moderately abnormal due to disorganized and discontinuous background. No epileptiform discharges were seen. Initial MRI showed left transverse sinus thrombosis and cerebral dysgenesis and renal ultrasound showed echogenic kidneys. He was treated with anticoagulants and a CT scan venography at 2 months of age showed that the thrombus cleared. He had no developmental milestones and showed feeding difficulties with episodes of aspirations for which a G-tube was inserted. Repeat brain MRI revealed an immature brain with delayed myelination as well as a left transverse sinus thrombosis. The infant continued having tremor, persistent hyperekplexia and was discharged from the hospital at a month of age only to be re-admitted at 3.5 months of age with respiratory failure. He was intubated and ventilated but multiple attempts at extubation over the course of 3.5 weeks failed due to central respiratory drive failure. Given the infant's futile outlook, the infant life support was withdrawn at 4 months of age. Investigation including chromosome analysis, metabolic work-up and studies for congenital infections showed no abnormalities. The family declined both autopsy and DNA banking.

The couple's second pregnancy resulted in a healthy daughter following an uneventful pregnancy and delivery. At 8years of age she has no health or developmental issues.

Patient C.II.3.

The couple's third pregnancy was followed closely and chorionic villi sampling (CVS) for maternal age revealed a normal male karyotype (46,XY). Serial ultrasounds at 19, 22, 26, and 32 weeks gestation all revealed normal head size at each gestational age. Delivery was by repeat caesarean section at 39 weeks gestation. Birth weight was 3520 g (50th centile) and OFC was 33 cm (-1 SD/-2 SD). Shortly after birth he developed jittery seizure-like

movements and hyperekplexia mirroring those of his late brother. Extensive investigation including plasma amino acids [asparagine value was 12 mmol/L after birth and 17 mmol/L at 2 months (normal 16-21 mmol/L), urinary organic acids, plasma very long chain fatty acids, 7 dehydrocholesterol, transferrin isoelectric focusing, free and total carnitine, serum ceruloplasmin, blood CK, urine mucopolysaccharides and urine oligosaccharides were normal. Chromosome microarray analysis was normal. Initial EEG demonstrated intermittent right and left frontal sharp waves occurring independently of uncertain significance, possibly indicative of underlying cerebral dysfunction. No definitive seizure activity was noted. A repeat EEG done at a month of age was contaminated by a large amount of movement artifact. Within these limitations, the waking background activity did not show any definite abnormality, nor did a brief period of active sleep. Although not definitely abnormal, there were very frequent interictal sharp waves recorded independently over both frontopolar regions with a definite left sided predominance. No ictal epileptiform discharges were recorded. A repeat EEG at 2 months of age was abnormal with discontinuous and asynchronous background and positive Rolandic sharp waves seen over the left more than right hemisphere. No electrographic seizures were seen. Basal auditory evoked potential was normal, visual evoked potential and electroretinogram showed delayed cortical response by 45 msec and somatosensory evoked potential of the median nerve showed absent cortical response bilaterally. EMG was normal. Brain MRIs revealed volume loss of cerebral hemispheres and pons, with normal appearing cerebellum, delayed myelination of posterior limbs of internal capsule and dorsal brainstem, with mild restricted diffusion of both thalami and possibly basal ganglia. There was mildly elevated lactate peak on MRS and no imaging evidence of prenatal asphyxia. The findings were similar to the MRI findings performed on his late brother. At 2 months of age he was noted to have right hemidiaphragm palsy and had increasing respiratory failure. He developed hypothermia and hypercapnia and was removed from respiratory support at 3 months of age.

Patient C.II.4. The couple's fourth pregnancy was again followed closely and CVS for maternal age revealed abnormal male karyotype (46,XY). Fetal ultrasounds at 19, 28, and 30 weeks showed normal head growth. At 32 weeks a slight lag in OFC was noted, with progressive development of microcephaly seen at 32, 34 and 35 weeks gestation by ultrasound. The baby was born at 35.7 weeks gestation via repeat C-section. The APGAR scores were9 and 9 at 1 and 5 minutes respectively. Birth weight was 3230 g (50th centile), birth length was 51.5 cm (50 – 97 centile) and OFC was 32 cm (-2 SD). At 6 hours after birth, he was noted to have abnormal movements with jitteriness and hyperekplexia. He was initially intubated but extubated at 2 days of age and was noted to have stridor. He continued having seizure episodes with back arching and apnea and was treated with clonazepam and phenobarbital. EEG on his first day showed discontinuous background with longer interburst interval than appropriate for age with occasional positive and negative sharp waves over the left central and temporal head regions. There were occasional runs of theta and alpha rhythm of unknown significance. No seizures were captured during this recording. Brain MRI showed microcephaly, delayed myelination and under development of the brain with brainstem hypoplasia. The findings were similar to the MRIs done on the previous siblings. Plasma amino acids and urine organic acids were normal. The anticonvulsive treatment resulted in decrease in his jitteriness and he did not have respiratory problems apart from stridor, which was attributed to laryngomalacia. He was orally fed with frequent vomiting and continued having tremor and jitteriness although this decreased with time. He was transferred to a palliative care facility and died at 6 months of age.

Family D.

The patients were three brothers born to a non-consanguineous French-Canadian couple in good health. An older sister is unaffected.

Patient D.II.1.

The pregnancy and neonatal period were considered normal but on the first routine home visit by a nurse at 5 days of age, he was found to have irregular breathing, rectal temperature of 34.6° to have poor suction at feeds. He was examined in a local center and transferred to CHU Sainte-Justine where he was intubated and became respirator-dependent. He was microcephalic (31.5 cm; -2.5 SD) with central hypotonia and increased tone in the legs and tremulous movements of the extremities. Visceral examination was normal. Glycemia, blood pH, transaminases, and renal function were normal. Fasting plasma amino acids showed asparagine 11 μmol/L (normal range, 31 to 56). Urine organic acids and plasma very long chain fatty acids were interpreted as normal. Transferrin isofocalization was normal. MECP2 gene sequencing and deletion analysis were normal. Cerebral magnetic resonance imaging (MRI) showed microcephaly with reduced gyration. He was extubated and died rapidly thereafter of respiratory failure, aged 9 days. Neuropathological examination revealed microcephaly, although the gyral pattern was preserved grossly on external examination. There was cortical dysgenesis (nests of cells in the subpial region and multiple thin layers of nuclei in the cortex, thinner and more numerous than normal). Also present were absence of the bulbar arcuate nucleus, marked periventricular leukomalacia; and marked gliosis. The visceral examination was normal.

Patient D.II.2.

The patient was born after a normal pregnancy at 39 5/7 weeks by spontaneous vaginal delivery. Fetal MRI at 31.5 weeks gestation and fetal ultrasound at 35 weeks, were normal. On neonatal examination, jerky and tremulous movements and limb hypertonia were noted. Body temperature was repeatedly normal. Head circumference was 31 cm (-3 SD). At 4 days of age an EEG showed deficient electrical activity with bursts of activity with spike activity maximal in the right fronto-centro-temporal regions. Cerebral MRI at 4 days of age revealed a reduction of the normal gyral pattern. The opercularisation of the sylvian fissure corresponded to that of a 34 week old child (at a chronological age of 39 weeks). There was a diffuse hyposignal on T1 imaging, demonstrating a slowness of myelination. The parieto-occipital horns were prominent, especially on the left. The subarachnoid space was prominent. Plasma amino acids were normal including asparagine (55 μmol/L; normal, 31-56), as were urine organic acids, transferrin isofocusing, analysis of subtelomeres and ARX gene sequencing. Ophthalmologic consultation revealed cortical blindness. At age 9 months, partial motor convulsions were reported and focal epileptic activity was seen on EEG; control was achieved with carbamazepine. Examination at 10 months showed a non-dysmorphic child with an absence of motor development and cranial circumference of 40 cm (< 3%). He reacted to noise but showed no ocular pursuit. Divergent strabismus, axial hypotonia and limb hypertonia and hyperreflexia were present. The lower limbs were more severely affected than the upper. He died at age 11months following an upper respiratory infection associated with respiratory acidosis (pCO2, 120; pH, 7.03) and hypothermia (30°C). At autopsy, neuropathology confirmed microcephaly with severe mesial temporal sclerosis and neuronal loss in regions CA3, CA4 and CA1, dysplasia of the olivary nucleus. The spinal cord showed hydromelia with secondary degeneration of motor neurons and reactive gliosis.

Patient D.II.3.

The pregnancy was initially normal. Fetal ultrasound was normal at 25 weeks of gestation. Placental abruption with fetal distress occurred at 33 weeks, and emergency cesarean section was performed. Apgar scores were 4 at one minute, 7 at 5 minutes and 10 at 10 minutes. Birth weight was 2160 g (50-75%), length 46 cm (75%) and head circumference 30 cm (25th centile). Although the patient initially cried vigorously, he showed hypoventilation within hours of birth. Respiratory distress syndrome was diagnosed clinically and on thoracic radiographs. He was intubated. Because of apneas he was treated with caffeine on day two. After resolution of his pulmonary disease, attempts to wean him from continuous positive airway pressure failed because of the development of apnea, cyanosis and bradycardia. Tremulous movements of the extremities were noted. He was transferred to CHU Sainte-Justine on day 6. Convulsions were noted on day 8. He was extubated at 3 weeks of age. At that time he could ingest adequate amounts of formula orally. His head circumference at one month of age was 30.2 cm and at 10 months, 39.5 cm (< 3%). His subsequent clinical course was similar to that of patient 2. His convulsions were controlled by small doses of Phenobarbital. He died at age 12 months of respiratory insufficiency secondary to a respiratory infection, associated with hypothermia.

Haplotype prediction.

To determine if the p.F362V ASNS variant is always found on the same haplotype, all high confidence SNVs (coverage ≥ 10, ≥ 5 reads supporting the variant, SNV quality ≥ 30, SNV consensus ≥ 30) on chromosome 7 were collected from each of the three exome-sequenced patients (Families A and B). Next, the regions on either side of the ASNS variant were examined to find the boundaries of the homozygous stretches containing the ASNS variant in each individual. Finally, the SNVs in this region were compared across samples to obtain the largest possible haplotype that was shared by all three affected patients. This estimated haplotype is approximately 1.2 Mb (build 36, Chr7: 97322654-98488035) and is tagged by 16 SNVs (Table SIV.6). These 16 SNVs were also assessed in the 261 sequenced controls. The genotype for each of these SNVs was input into the fast PHASE program 22 in order to infer the phase of these variants in each control sample. Only 17 controls had all 15 SNVs (the ASNS variant at the outer boundary was not found in any controls), and the majority of these variants were found in heterozygous form. Of the 522 possible haplotypes, fast PHASE determined that 2 haplotypes were identical to the ASNS haplotype (excluding the ASNS variant itself) when the individual haplotype error was minimized, and 4 haplotypes were identical to the ASNS haplotype (again, excluding the ASNS variant itself) when switching error was minimized. All 4 of these control samples are self-identified as Caucasian.

Homozygosity mapping.

To identify all potentially overlapping homozygous stretches between the families A and B, we used the exome sequence data to detect homozygous regions present in these three cases. The largest homozygous region seen in all three case genomes was ~3 Mb and included the ASNS gene. The average size of the remaining shared homozygous regions (n=169), found in these three case genomes, was 131.4 Kb. The sequenced patients from families A and B have an average of 90% of bases covered across the protein-coding portions of the genes in the shared homozygous regions (Table SIV.5). This analysis increases our confidence that the ASNS variant is the causal variant associated with the observed phenotype.

We used all identified high-quality single nucleotide variants to perform homozgyosity mapping. The SNVs in all three of the cases were analyzed using the PLINK "homozygosity mapping" tool (Purcell et al., 2007) .This algorithm considers sliding windows across the genome and screens for homozygous variants in those regions. To account for the non-random spacing of SNVs in the exome data and to detect such regions with high sensitivity, very generous thresholds were applied as outlined below.

The default values provided by PLINK were designed for analysis from dense SNP maps. However, here we were analyzing the SNVs identified by exome-sequencing. Thus before proceeding, all SNVs were filtered for those with very high quality (coverage ≥ 10, ≥ 5 reads supporting the variant, SNV quality ≥ 30, SNV consensus ≥30). We defined a "window" as 10 SNPs within a 1,000 Kb window. One heterozygous SNV was allowed in a window. A homozygous segment was then defined as a region with 10 SNVs in a 1 Kb window.

This analysis identified 1,532 homozygous regions (≥1 Kb) in total, with approximately 500 in each case genome. The two siblings from family B had 341 regions that overlapped by at least 500 base pairs. These regions were then compared to the homozygous regions identified in the third unrelated patient from family A. There were 169 homozygous regions that overlapped by at least 500 bp and they were found in all three case genomes.

The exome data were re-examined by focusing only on regions of homozygosity. This analysis revealed that the case genomes contained only 20 rare and plausibly functional sequence variants (predicted homozygote frequency of ≤ 3% in our sequenced control genomes) in these homozygous regions. The majority of these 20 variants were found in only one of the 3 case genomes. There was only 1 plausibly functional variant (non-synonymous) present in 2 of the three case genomes, however this variant was very common (control MAF of 0.13). Importantly, only the ASNS variant was present in all 3 case genomes from families A and B.

Homozygosity mapping was also done in family C at the McGill University Genome Quebec Innovation Center (Montreal, Canada). Briefly, genomic DNA (~2.5µg) from the blood of the two affected siblings (C:II.3 and II.4) was used to perform whole-genome genotyping using the Illumina 660 Quad Chip (Infinium HD). PLINK (v.1.06) (Purcell et al., 2007) was used to identify shared regions of homozygosity with a size of >1Mb.

Asns in population controls.

We looked in the CHGV dataset of 4,845 high-density genome-wide SNP genotyped controls and found no copy number variations (CNVs), as predicted by PennCNV, which overlapped the ASNS gene (data not shown). Additionally, only one CNV in this region is reported in the Database of Genomic Variants (http://projects.tcag.ca/variation/). The involvement of ASNS in this phenotype would have been refuted, if homozygous deletions encompassing this gene had been observed in this CNV data. Also, ASNS has a halopinsufficiency score of 0.551 (Huang et al., 2010).

Asns conservation and mutation locations in the folded ortholog.

An examination of the E. coli ortholog structure (Larsen et al., 1999) reveals that the inferred p.A6E position is located in the N-terminal domain (which is responsible for glutamine hydrolysis), that faces a pocket identified as the glutamine binding site [(Larsen et al., 1999), Fig. IV.3b]. The inferred p.F362V position is located on the C-terminal domain, near the AMP binding site (Fig.IV.3b). The inferred p.R550C position is located at the distal end of the C terminus which is disordered in the E. coli ortholog and thus the crystallographic structure is not available. However, this region is in the interface between the N-terminal and C-terminal

domains and it might play a role in binding the aspartate substrate (Larsen et al., 1999). Additionally, the predicted deleterious effect of these mutations on the protein functionality (Table IV.2) supports the putative destabilization observed in our heterologous expression experiment, as well as the functionally important disposition of the mutated residues in the homologous 3D structure.

Mouse behavioural testing

Overview.

No genotype differences were discerned for spontaneous locomotor activity in the open field (Figure SIV.5A (i)) or in the overall time spent in movement (Figure SIV.5B (i)). Similarly, no genotype distinctions were seen for anxiety-like behaviour in the open field (Figure SIV.5B (ii)) or in the light-dark emergence test (Figure SIV.5B (iii-iv)). In the accelerating rotorod test, Asns+/- and Asns-/- mice stayed on the rotorod longer than CONT mice over the first three trials of testing (Figure S IV.5A (ii)). However, CONT mice showed motor learning across trials, whereas a ceiling effect probably precluded seeing this change in performance in the mutant mice. Episodic memory was examined in the novel object recognition test where short- (STM) and long-term memory (LTM) were evaluated (Rampon et al., 2000). In the STM test, Asns+/- and Asns-/- mice showed no preference for either the novel or familiar object. In the test for LTM, the Asns-/- mice failed to demonstrate a preference for either object. By comparison, CONT and Asns+/- animals preferred the novel object; however, LTM was still reduced in the Asns-/- mice (Figure SIV.5A (iii)). It should be emphasized that the deficits in STM or LTM by the mutants was not related to the total time spent with the objects or to the total numbers of object contacts (Figure SIV.5B (v-vi)). Collectively, these data indicate that Asns-/- mice are impaired in both STM and LTM. By contrast, there is some sparing of LTM in the Asns+/- animals, suggesting that memory consolidation is delayed in these mutants.

Methods.

Adult male and female B6NTac (Taconic Labs, Hudson, NY) served as controls and were tested with the heterozygous and homozygous mutant mice (B6NTac; B6N Asnstm1a(EUCOMM)Wtsi/H; European Mutant Mouse Archive, Munich, Germany). Adult CONT, Asns+/-, and Asns-/- mice were housed under controlled temperature and humidity conditions under a 14:10 h light:dark cycle (light onset 0800 h) with food and water provided ad libitum. All behavioural assessments were conducted during the light cycle between 1000 and 1700 h. Mice were tested in following order with 3-5 days interposed between tests: light-dark emergence, open field, rotorod, and novel object recognition as described (Carlson et al., 2011). All procedures were approved by the Duke University Institutional Animal Care and Use Committee and were in conducted accordance with NIH guidelines for the care and use of laboratory animals.

All behavioural data are presented as means and SEM and were analyzed with SPSS 20 (IBM North America, New York, New York) using ANOVA (open field, light-dark emergence test), repeated measures ANOVA (RMANOVA; rotarod, novel object recognition test), and the Kruskall-Wallis test for rank-ordered data (percent increases in performance during rotorod testing). Bonferroni corrected pair-wise comparisons were used for aposteriori comparisons where a $p < 0.05$ was considered significant.

The mice were tested in the light-dark emergence test for 5 min and were monitored with Med-PC IV software (Med-Associates, St. Albans, VT); the open field data were collected over 30 min with Versamax software (Accuscan Instruments, Columbus, OH); the accelerating rotorod data were collected over 5 min with Med-Associates rotorod software; and the novel object recognition testing was conducted in Plexiglas arenas with mouse-friendly objects over 10 min and analyzed with TopScan software (Clever Sys Inc., Reston, VA). Preference scores in the novel object recognition test were calculated by subtracting the total time spent with the novel object from the total time spent with the familiar object during each test, and divided by the total time spent with both objects. Positive scores indicated a preference for the novel object, negative scores a preference for the familiar object, and scores approaching zero indicated no preference for either object.

Additional results.

In the open field an ANOVA revealed no significant genotype differences for spontaneous distance traveled (Figure SIV.5A (i)), time spent in overall movement (Figure SIV.5B (i)), or distance moved in the center of the open field (Figure SIV.5B (ii)).

In the rotorod test, a RMANOVA for latency to fall for the within subjects test found the main effect of test-trial to be significant [$F(3,66)=7.832$, $p<0.001$] and the between subjects test revealed a significant genotype effect [$F(2,22)=5.174$, $p<0.014$]. Although the genotype by test-trial interaction was not significant, inspection of the Figure SIV.5A (ii) indicated that the genotypes were differentially responding to the task. When data were expressed as the percent change in latency to fall from the first to last trials; CONT mice showed a 183% increase in motor learning over testing, whereas a small enhancement 25% was observed in the Asns+/- animals and Asns-/- mice had an increase of 51%. A Kruskal-Wallis test for rank-ordered data found a significant difference between the three genotypes [$H(2,22)=6.281$, $p<0.043$]; confirming that performance of the CONT animals was superior to that of the Asns+/- and Asns-/- mice.

In the light-dark emergence test, ANOVA for the latency to emerge from the darkened chamber into the lighted chamber (Figure SIV.5B (iii)) and the time spent in the lighted chamber (Figure SIV.5B (iv)) were not significantly different among genotypes.

In the novel object recognition test, a RMANOVA reported a significant effect of test-session [$F(2,44)=11.209$, $p<0.001$] and a significant test-session by genotype interaction [$F(4,44)=6.178$, $p<0.043$]. Bonferroni corrected pairwise comparisons for the STM test revealed that Asns+/- and Asns-/- mice had significant lower preference scores for the novel object than CONT animals ($ps<0.039$); preference scores between the mutants did not differ (Figure SIV.5A (iii)). In the test for LTM, preference scores were increased for the Asns+/- mice relative to their performance in the STM test. However, their scores were still significantly lower ($p<0.046$) than those for CONT animals. By comparison, Asns-/- mice continued to show no object preference and their scores were significantly lower than those for the Asns+/- and CONT animals ($ps<0.001$). Despite differences in the preference scores, the genotypes of mice did not differ according to time spent exploring the objects (Figure SIV.5B (v)) or total contacts with the objects (Figure SIV.5B (vi). These data indicate that memory function is severely impaired in the Asns-/- mice and is less deficient in the Asns+/- animals.

SUPPLEMENTAL FIGURES

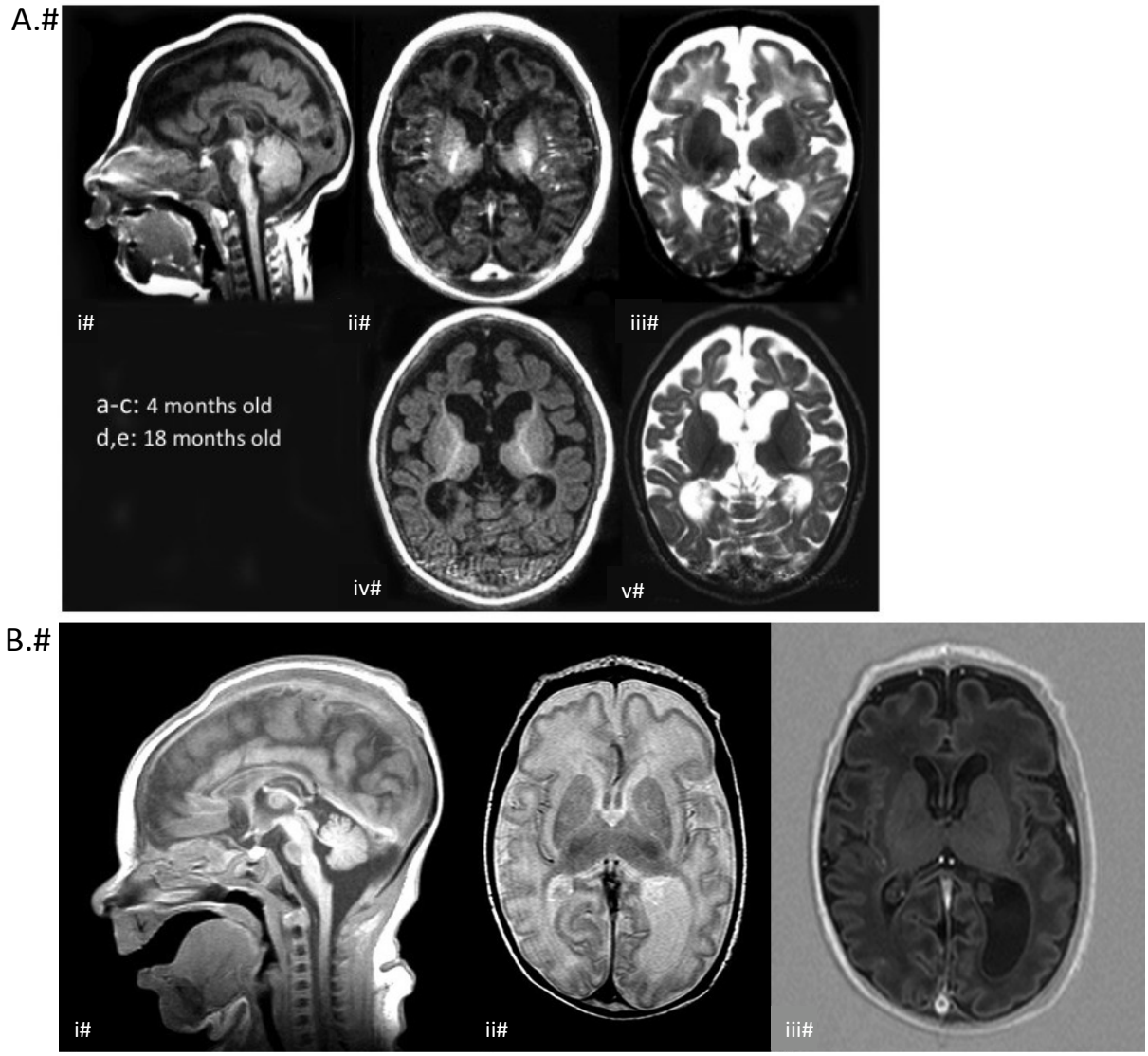


Figure SIV.1

Figure SIV.1 Brain MRIs from additional families.

(Relates to Figure IV.2: MRI images from family C).

A. Brain MRI from patient A.II.1. T1W sagittal image (i) at 4 months of age demonstrates micrognathia, sloping forehead, thin brainstem and relatively normal cerebellum. T1 axial image **(ii)** reveals severe cortical and subcortical atrophy and delay in myelin maturation, with myelin only in the posterior limbs of internal capsule. T2 axial image **(iii)** reveals abnormally increased signal intensity of the white matter for age, prominent ventricles and pericerebral spaces over the frontal convexity. T1W axial image **(iv)** at 18 months of age confirms persistent myelin delay and parenchymal volume loss. Concurrent T2W axial image **(v)** shows abnormal signal and atrophy of the white matter, progressive ventriculomegaly and pericerebral spaces due to volume loss.

B. Brain MRI from patient D.II.2. (i) Sagittal T1 image showing small pons. (ii) Axial T2 image showing decreased brain volume, delayed myelin maturation, and simplified gyral pattern predominantly in the frontal regions. (iii) Axial T1 image confirming the delayed myelination and the other findings seen on the Axial T2 image.

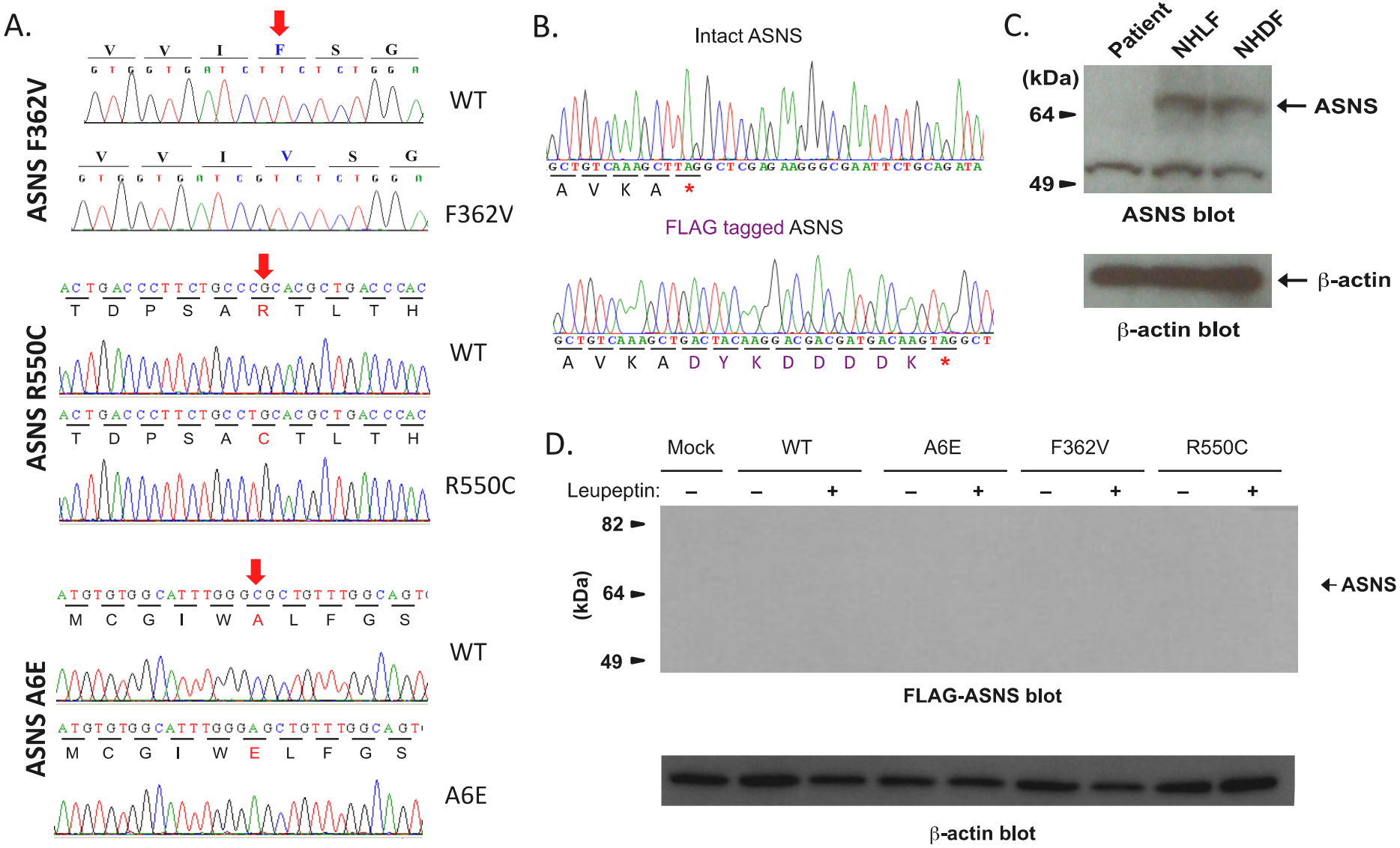

A.
ASNS F362V
WT
F362V
ASNS R550C
WT
R550C
ASNS A6E
WT
A6E
B.
Intact ASNS
FLAG tagged ASNS
C.
Patient
NHLF
NHDF
(kDa)
64
49
ASNS
ASNS blot
β-actin
β-actin blot
D.
Mock
WT
A6E
F362V
R550C
Leupeptin:
82
64
49
(kDa)
ASNS
FLAG-ASNS blot
β-actin blot

Figure SIV.2

Figure SIV.2 In vitro *ASNS* characterization.

(Relates to Figure IV.3: Functional impact of ASNS mutations)

A. Sanger sequencing of *in vitro ASNS* alleles. After PCR-directed mutagenesis, cDNA encoding human *ASNS* containing wild type or mutant alleles was subcloned into pcDNA3.1$^{(+)}$ vector and subjected to sequence analysis.

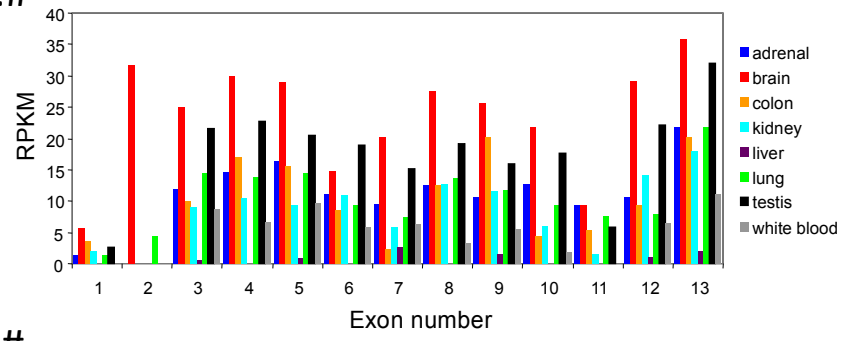
B. Sanger sequencing of FLAG-tagged *ASNS.* A FLAG-tag was introduced into the sequence encoding the C-terminus of ASNS for the wild type and the three mutant forms (A6E, R550C and F362V) and subjected to sequence analysis.

C. ASNS levels in patient fibroblast cells. ASNS in a patient with the F362V mutation (Family A: II.1) was detected using an anti-ASNS antibody. NHDF and NHLF are human fibroblast cell lines. The protein in the patient cells is depleted as was seen with the *in vitro* mutagenesis studies.

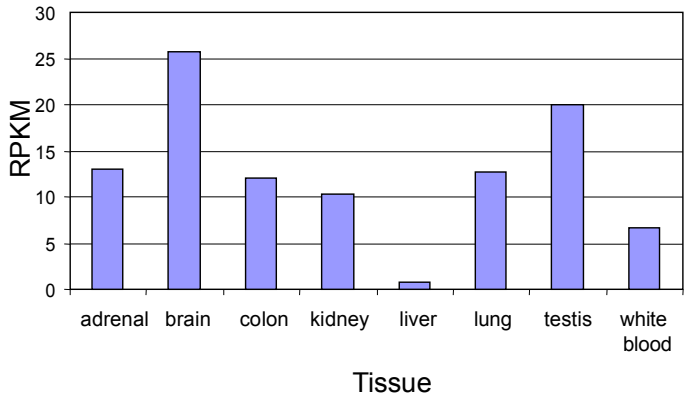
D. ASNS levels in COS-7 cells transfected with empty, wild-type, or mutant vectors. Western Blots detecting ASNS-FLAG protein abundance, with or without Leupeptin treatment, in COS-7 transfectants using an anti-FLAG antibody. β-actin was used as a loading control.

A.#

i.#



ii.#



B.#

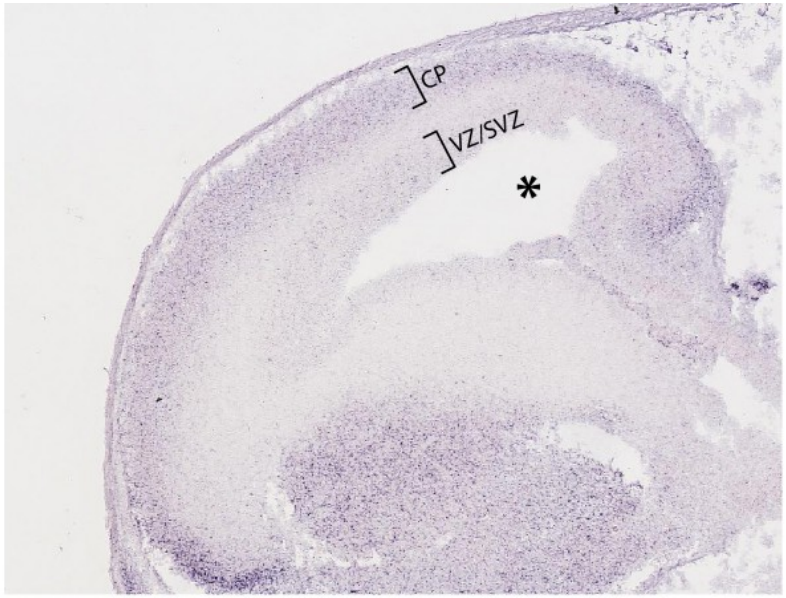


Figure SIV.3

Figure SIV.3 ASNS expression in different tissues and the developing mouse brain.

(Relates to Figure IV.5: Asns deficient mouse and structural brain abnormalities)

A. *ASNS* expression in different tissues. Expression profile of *ASNS* gene (ENST00000394308) using RNA-seq data, extracted from the Human Body Map 2.0 Project of Illumina (downloadable from Integrative Genomics Viewer (IGV) at http://www.broadinstitute.org/igv/). In this database, the RNA from each tissue was obtained from a single individual. Extracted reads were aligned to the human genome (hg19) using Bowtie (Langmead et al., 2009), and BED Tools (Quinlan et Hall, 2010) was applied for counting the number ofreads per exon. Data were normalized to reads per kilobase of exon model per million mapped reads (RPKM) values. (i) Normalized expression per exon, showing an indication of brain-specific expression of exon 2. (ii) Normalized expression for the whole gene. The finding that *ASNS* is more highly expressed in brain and testis is in agreement with Horowitz *et al.* (Horowitz et al., 1968) .

B. *Asns* expression in the developing mouse brain.

Asns mRNA *in situ* hybridization of the developing mouse brain (sagittal section) at E14.5. *Asns expression* is particularly enriched in the cortical plate (CP) and ventricular and sub-ventricular zone layers (VZ/SVZ). The asterisk is labeling the ventricle of the developing mouse cerebral cortex. *In situ* images were taken from (http://www.eurexpress.org/, euxassay_004453).

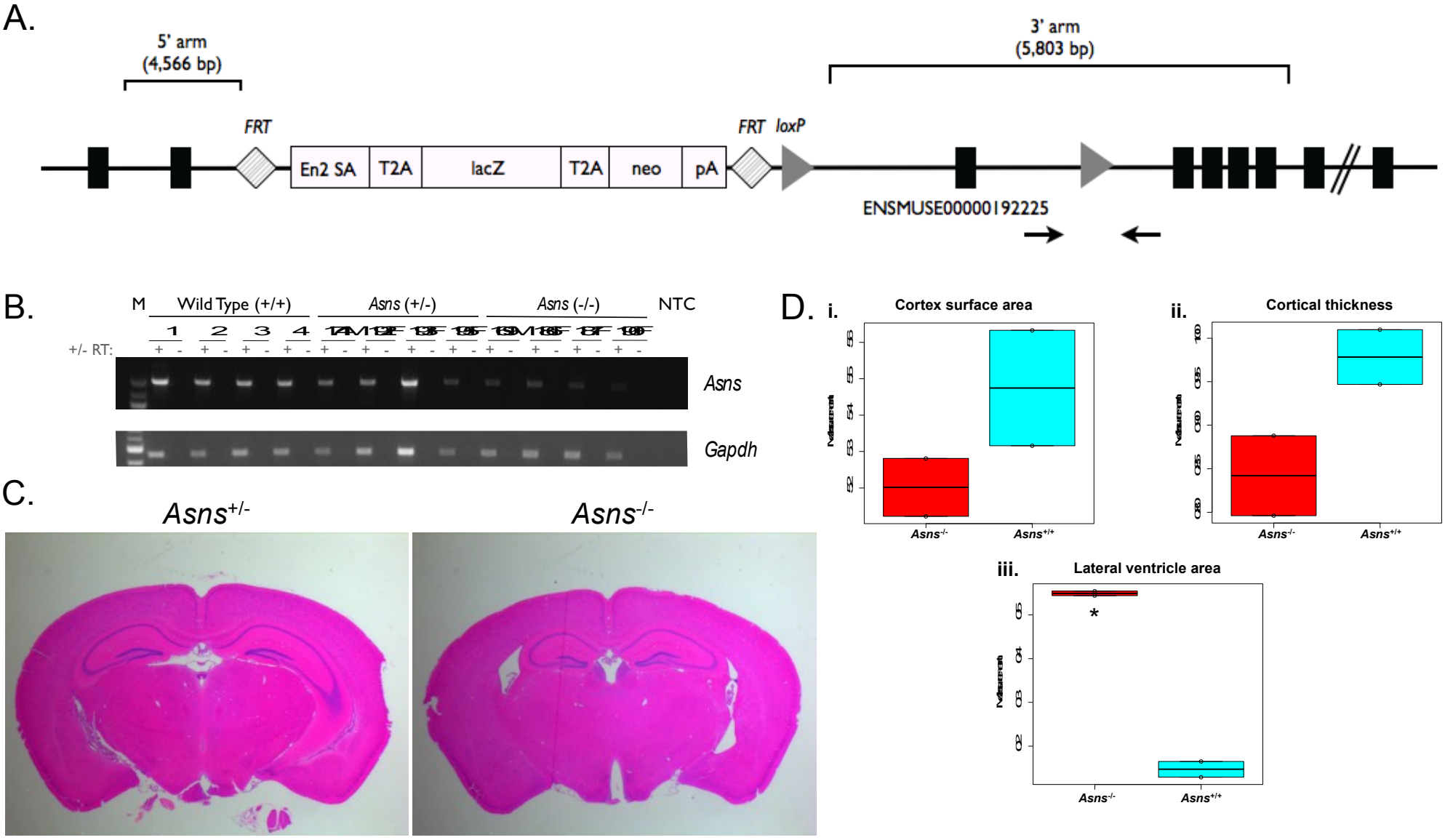

A.
5' arm
(4,566 bp)
3' arm
(5,803 bp)
FRT
FRT
loxP
En2 SA
T2A
lacZ
T2A
neo
pA
ENSMUSE00000192225
B.
M
Wild Type (+/+)
Asns (+/-)
Asns (-/-)
NTC
1
2
3
4
+/- RT:
Asns
Gapdh
C.
Asns+/-
Asns-/-
D.
i.
Cortex surface area
ii.
Cortical thickness
iii.
Lateral ventricle area
*
Asns-/-
Asns+/+

Figure SIV.4

Figure SIV.4 Asns knock out mouse details and phenotyping.

(Relates to Figure IV.5: Asns deficient mouse and structural brain abnormalities)

A. Construct for knocking out *Asns* and genotyping primers. We obtained genetically modified mice from EMMA (http://www.emmanet.org/strains.php). This diagram shows the complete construct including the with a splice acceptor site, after exon 2, for the lacZ gene (B6NTac; B6N-Asnstm1a$^{(EUCOMM)Wtsi}$/H). Mice were genotyped using the following custom primer sequences (black arrows): forward: 5'-TTT TGG TTT GTG TTT CTTCCT G-3' and reverse: 5'-TCA GGA ACG TGA GTG AGT GAG T-3'.

B. *Asns* adult mouse brain RT-PCR exons 2-12. *Asns* RT-PCR with low cDNA input (250 ng) and PCR cycle (35 cycles) was used to assess the existence of *Asns* mRNA transcripts semi-quantitatively. RT-PCR spanned from exon 2 to exon 12 using the specific primer set: 5'- CAG TGT CTG AGT GCG ATG AAG A -3' and 5'- GCG TTC AAA GAT CTG ACG GTA G-3'. *Gapdh* RTPCR with low cDNA input (250 ng) and PCR cycle (25 cycles) was used as a control using the specific primer set: 5'-ACC ACA GTC CAT GCC ATC AC-3' and 5'-CAC CAC CCT GTT GCT GTA GCC-3'.

C. Asns adult mouse brain sections. Coronal sections from one homozygous mutant mouse (-/-) and an age matched (P84) heterozygous littermate (+/-).

D. Postnatal day zero (P0) mouse brain measurements. Measurements were taken using two homozygous mutant mouse and two age matched (P0) wild type littermates.(i) Surface area of the mouse cortex (right hemisphere only) (ii) cortical thickness and (iii) Lateral ventricle area of right hemisphere; * indicates a significant difference by an unpaired t-test ($p=0.019$). Measurement units are arbitrary.

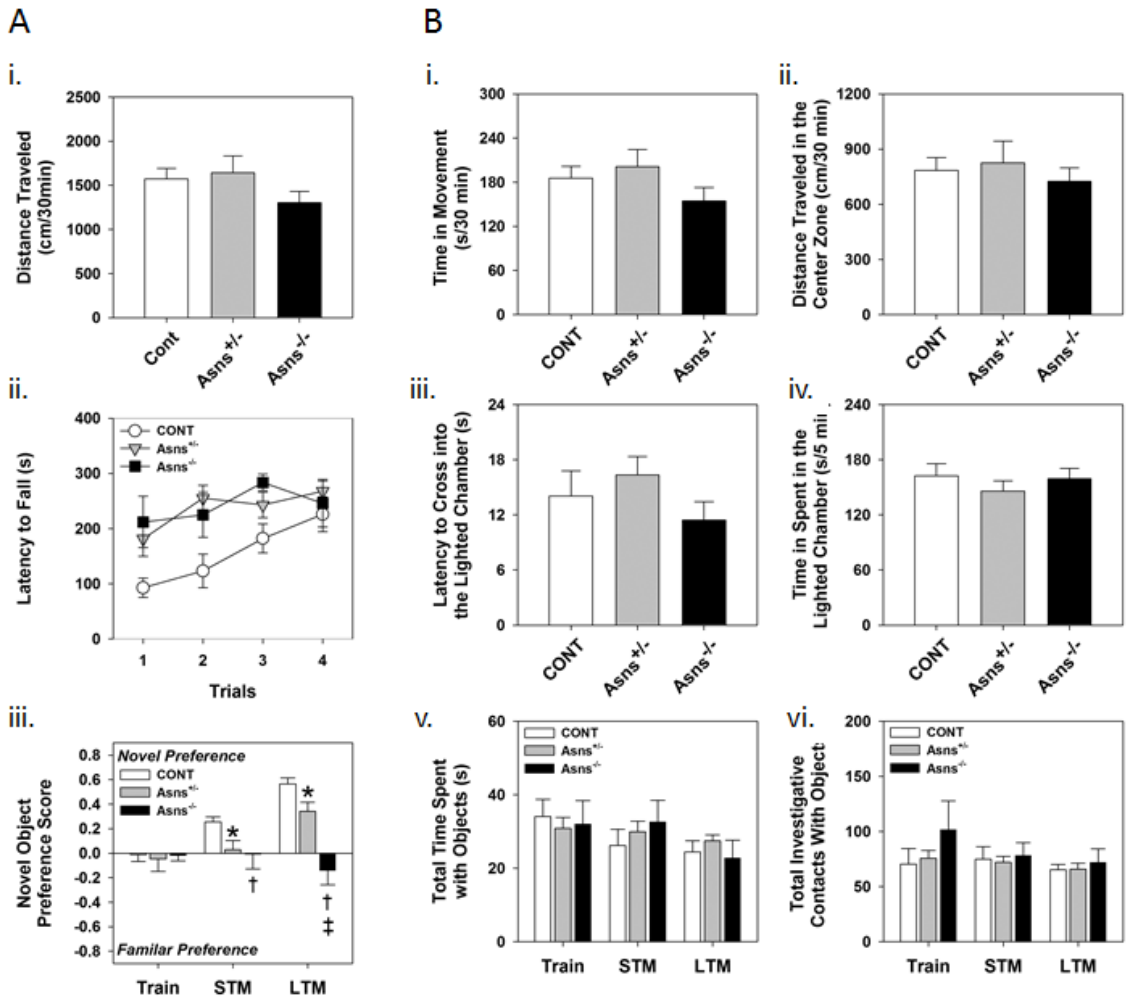

A
i.
Distance Traveled (cm/30min)
2500
2000
1500
1000
500
0
Cont
Asns +/-
Asns -/-
ii.
Latency to Fall (s)
400
300
200
100
0
CONT
Asns+/-
Asns-/-
1
2
3
4
Trials
iii.
Novel Object Preference Score
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
Novel Preference
CONT
Asns+/-
Asns-/-
Familar Preference
Train
STM
LTM
B
i.
Time in Movement (s/30 min)
300
240
180
120
60
0
CONT
Asns+/-
Asns-/-
ii.
Distance Traveled in the Center Zone (cm/30 min)
1200
900
600
300
0
CONT
Asns+/-
Asns-/-
iii.
Latency to Cross into the Lighted Chamber (s)
24
18
12
6
0
CONT
Asns+/-
Asns-/-
iv.
Time in Spent in the Lighted Chamber (s/5 mi
240
180
120
60
0
CONT
Asns+/-
Asns-/-
v.
Total Time Spent with Objects (s)
60
40
20
0
CONT
Asns+/-
Asns-/-
Train
STM
LTM
vi.
Total Investigative Contacts With Object
200
150
100
50
0
CONT
Asns+/-
Asns-/-
Train
STM
LTM

Figure SIV.5

Figure SIV.5 Behavioral analyses of Asns mice.

(Relates to Figure IV.5: Asns deficient mouse and structural brain abnormalities)

SIV.5A. Behavioral analyses of Asns mice. (i) Spontaneous locomotor activity in the open field over 30 min for B6NTac control (CONT), $Asns^{+/-}$, and $Asns^{-/-}$ mice. (ii) Latency to fall from the rotorod across four trials. (iii) Preference scores from the novel object recognition test as assessments during training (Train), short-term (STM), and long-term memory. N=8 CONT, 11 $Asns^{+/-}$, and 6 $Asns^{-/-}$ mice. $^{*}p<0.05$, CONT versus $Asns^{+/-}$ mice, $^{\dagger}p<0.05$, CONT versus $Asns^{-/-}$ mice, $^{\ddagger}p<0.05$, $Asns^{+/-}$ versus $Asns^{-/-}$ mice.

SIV.5B. Behavioral analysis of CONT, $Asns^{+/-}$, and $Asns^{-/-}$ mice.

(i) Time in movement in the open field. (ii) Distance traveled in the center zone of the open field. (iii) Latency to enter the lighted chamber in the light-dark emergence test. (iv) Time spent in the lighted chamber in the light-dark emergence test. (v) Total time spent with objects in the novel object recognition test. (vi) Total investigative contacts with objects. STM = short-term memory, LTM = long-term memory. N=8 CONT, 11 $Asns^{+/-}$, and 6 $Asns^{-/-}$ mice.

SUPPLEMENTAL TABLES

Table SIV.1 All rare variants shared in both patients from family B.

(Relates to Table IV.2: Mutations identified in ASNS)

A summary of all rare variants, with a plausible impact on the function of the resulting gene product, shared in family B (found in both affected individuals Fig. IV.1A: II.2 and II.4). A rare variant was defined as having a predicted homozygote frequency of ≤ 3% in the population of 261 sequenced control genomes.

Functional category	Homozygous	Heterozygous
Essential splice site	1	2
Intron-exon boundary (SNV)	10	32
Intron-exon boundary (indel)	1	9
Non-synonymous coding	58	248
Stop gained	1	5
Coding-disrupted frameshift (indel)	1	8
Coding-disrupted other (indel)	0	2
Total	72	306

Table SIV.2 All rare homozygous functional variants found in both patients from family B.

(Relates to Table IV.2: Mutations identified in ASNS)

The 72 homozygous functional variants found in both II.2 and II.4 of family B. The variants are listed in ascending order based on the variant allele frequency in the sequenced control population. The bolded variant numbers denote the 9 variants that were also observed in the affected individual from family A (Fig. IV.1A, II.1).

Variant	Variant ID	Variant type	Gene symbol	Variant frequency	Variant	Variant ID	Variant type	Gene symbol	Variant frequency
1	7_97322654_C	SNV	ASNS	0.000	37	9_39075808_C	SNV	CNTNAP3	0.117
2	7_107186210_T	SNV	CBLL1	0.000	38	19_38355194_C	SNV	WDR88	0.119
3	7_107475774_C	SNV	LAMB4	0.000	39	19_59416242_A	SNV	LILRB3	0.119
4	7_106296855_T	SNV	PIK3CG	0.002	40	22_44146055_T	SNV	SMC1B	0.121
5	6_41721834_A	SNV	MDFI	0.004	41	X_1497953_T	SNV	ASMTL	0.121
6	12_51248317_A	SNV	KRT74	0.006	42	12_48631243_A	SNV	AQP2	0.123
7	7_106725874_C	SNV	COG5	0.008	43	16_2045401_T	SNV	TSC2	0.123
8	19_38307917_T	SNV	GPATCH1	0.010	**44**	15_19884068_C	SNV	OR4N4	0.123
9	7_104993051_A	SNV	RINT1	0.012	45	6_32740678_A	SNV	HLA-DQB1	0.125
10	7_98997458_G	SNV	ZNF655	0.015	46	22_29862960_T	SNV	PLA2G3	0.127
11	12_51152337_G	SNV	KRT6C	0.025	47	X_1500706_G	SNV	ASMTL	0.127
12	17_35550980_DEL_T	indel	CASC3	0.031	48	17_27649318_A	SNV	RHBDL3	0.129
13	16_2100974_G	SNV	PKD1	0.044	49	X_1506842_G	SNV	ASMTL	0.129
14	17_41416696_G	SNV	MAPT	0.046	50	11_77586662_A	SNV	USP35	0.131
15	7_107004256_A	SNV	DUS4L	0.048	51	22_30919090_T	SNV	RFPL2	0.131
16	6_24511434_T	SNV	MRS2L	0.050	**52**	19_59416243_A	SNV	LILRB3	0.135
17	18_12244956_A	SNV	CIDEA	0.054	53	16_2105631_C	SNV	PKD1	0.137
18	7_102361951_G	SNV	LRRC17	0.054	54	20_10551750_A	SNV	C20orf94	0.138
19	12_51838742_G	SNV	CSAD	0.056	55	3_41852418_C	SNV	ULK4	0.142
20	7_102452856_G	SNV	FBXL13	0.056	56	14_20569961_T	SNV	TPPP2	0.146
21	19_38209355_T	SNV	RHPN2	0.060	57	1_151037145_A	SNV	LCE1D	0.148
22	7_101885015_A	SNV	ALKBH4	0.062	**58**	12_51503256_A	SNV	KRT79	0.148
23	19_13901896_A	SNV	CC2D1A	0.079	59	1_37961428_A	SNV	EPHA10	0.160
24	16_2102362_G	SNV	PKD1	0.083	60	10_18868641_G	SNV	CACNB2	0.160
25	16_2095427_C	SNV	PKD1	0.087	**61**	7_29127129_T	SNV	CPVL	0.162
26	7_99655521_T	SNV	PVRIG	0.102	62	19_37859295_T	SNV	RGS9BP	0.162
27	1_37958776_T	SNV	EPHA10	0.104	63	15_41604696_A	SNV	MAP1A	0.163
28	7_100324690_C	SNV	UFSP1	0.104	64	16_2092388_G	SNV	PKD1	0.165
29	8_17202048_G	SNV	MTMR7	0.104	65	6_31432179_C	SNV	HLA-B	0.169
30	18_59805277_G	SNV	SERPINB8	0.104	**66**	15_19884261_G	SNV	OR4N4	0.169
31	3_62164229_A	SNV	PTPRG	0.108	67	1_75475262_A	SNV	SLC44A5	0.171
32	1_159963696_C	SNV	FCRLB	0.110	68	2_111315429_T	SNV	ACOXL	0.171
33	6_109430212_T	SNV	SESN1	0.110	69	11_77598578_G	SNV	USP35	0.173
34	X_152517671_INS_C	indel	FAM58A	0.112	70	3_47427122_A	SNV	PTPN23	0.175
35	12_51330534_T	SNV	KRT2	0.115	71	2_238675316_C	SNV	ESPNL	0.179
36	6_109587255_C	SNV	C6orf182	0.117	72	17_35133114_G	SNV	ERBB2	0.179

Table SIV.3 Exome sequencing variant filtering in family C.

(Relates to Table IV.2: Mutations identified in ASNS)

Exome sequencing variant filtering in family C. *Rare variants were defined as those present at a frequency ≤ 3% in in-house control exomes (n=169), 1,000 genomes, and NHLBI Exome Sequencing Project.

Filter	Remaining variants
Coding /Splicing	20,439
Remove synonymous (non-splicing)	10,175
Rare*	856
Homozygous	25
In homozygous region	3
Segregating with the disease in the homozygous state	ASNS:NM_183356:exon14:c.C1648T:p.R550C CYP3A4:NM_017460:exon12:c.G1282A:p.D428N
Predicted damaging (SIFT or polyphen-2)	ASNS:NM_183356:exon14:c.C1648T:p.R550C

Table SIV.4 Regions of shared homozygosity in Family C.

(Relates to Table IV.2: Mutations identified in ASNS)

Regions of shared homozygosity (>1Mb) between the affected individuals in Family C (II.3 and II.4).

Chr	Marker start	Marker end	Position start	Position end	Size (Mb)
1	rs3863722	rs272822	35,592,376	36,677,585	1.1
1	rs319950	rs9436447	49,091,687	50,590,732	1.5
3	rs3895736	rs3197999	48,658,467	49,721,532	1.1
3	rs9790150	rs2882429	133,723,028	139,835,865	6.1
4	rs10517277	rs6858830	33,466,803	34,500,856	1.0
5	rs10477652	rs10478752	123,179,118	126,087,378	2.9
7	rs11773446	rs7780168	95,944,485	131,291,677	35.3
12	rs1564121	rs12311684	86,138,548	89,435,177	3.3

Table SIV.5 Exome sequencing variant filtering in family D.

(Relates to Table IV.2: Mutations identified in ASNS)

Exome sequencing variant filtering in family D. Rare variants were defined as those present at a frequency ≤3% in in-house control exomes (n=169), 1000 genomes, and NHLBI Exome Sequencing Project.

	Affected males		Unaffected male	Unaffected female
	II.1	II.2	II.4	II.5
Coding/Splicing	20,336	21,120	19,325	20,252
Remove synonymous (non-splicing)	10,021	10,491	9,793	10,363
Shared by II.1 and II.2	7,830			
Rare*	237			
Homozygous in II.1 & II.2 but *not* homozygous in II.4 or II.5	0			
Variants on chr X in II.1 and II.2 but not in II.4	0			
Genes with ≥2 rare variants in common to II.1 & II.2	14			
Genes with ≥2 rare variants in common to II.1 & II.2 but not in common with II.4 or II.5 (*compound heterozygous*).	ASNS (NM_183356) c.C1648T:p.R550C c.C17A:p.A6E			

Table SIV.6 F362V ASNS shared haplotype in families A and B.

(Relates to Table IV. 2: Mutations identified in ASNS)

The shared F362V *ASNS* haplotype in families A and B. The 16 variants found in the shared p.F362V *ASNS* haplotype. The first SNV is the *ASNS* p.F362V variant

SNV	Coordinate	Alleles
1	chr7:97322654	C/C
2	chr7:97325666	A/A
3	chr7:97326505	T/T
4	chr7:97621997	C/C
5	chr7:97654263	T/T
6	chr7:97660051	A/A
7	chr7:97660146	A/A
8	chr7:97760787	T/T
9	chr7:97760811	A/A
10	chr7:97771537	T/T
11	chr7:97782680	A/A
12	chr7:98298760	A/A
13	chr7:98338892	T/T
14	chr7:98396816	C/C
15	chr7:98487987	C/C
16	chr7:98488035	C/C

Table SIV.7 Summary of exome-sequencing coverage in homozygous regions.

(Relates to Table IV.2: Mutations identified in ASNS)

Summary of exome-sequencing coverage in homozygous regions. There are 672 HGNC recognized protein-coding genes found either partially or fully within the predicted shared homozygous regions (n=170, present in all sequenced patients from Families A and B). All known Ensembl exons (Hubbard et al., 2009), excluding UTRs, were then assessed for coverage, where a base is considered covered if ≥5 short-reads spanned this nucleotide. Each sequenced sample has ≥ 600 of these genes with ≥70% of their bases covered and each sample averages nearly 90% nucleotide coverage across these 672 genes.

Individual	# of genes with ≥ 70% bases covered	# of genes with < 70% of bases covered	Average % of bases covered
Family A (II-1)	600	72	89.03%
Family B (II-2)	601	71	89.01%
Family B (II-4)	602	70	89.01%

Table SIV.8 Seven primary microcephaly loci.

(Relates to Table IV.2: Mutations identified in ASNS)

The previously identified primary microcephaly loci and associated information about casual gene at each loci.

Locus	Location	Gene	Coordinates (NCBI 36)	MIM
MCPH1	8p23	Microcephalin (MCPH1)	chr8:6,251,529-6,493,434	607117
MCPH2	19q13	WD repeat-containing protein 62 (WDR62)	chr19:41,237,623-41,245,393	613583
MCPH3	9q33	CDK5 regulatory subunit-associated protein 2 (CDK5RAP2)	chr9:122,190,968-122,250,225	608201
MCPH4	15q21	Centrosomal protein, 152-KD (CEP152)	chr15:37,900,001-42,700,000	613529
MCPH5	1q31	Abnormal spindle-like, microcephaly-associated (ASPM)	chr1:195,319,997-195,382,287	605481
MCPH6	13q12	Centromeric protein J (CENPJ)	chr13:24,354,412-24,395,085	609279
MCPH7	1p32	SCL/TAL1-interrupting locus (STIL)	chr1:47,488,398-47,552,406	181590

Table SIV.9 Predicted homozygous regions overlapping known primary microcephaly loci.

(Relates to Table IV.2: Mutations identified in ASNS)

Four regions of homozygosity were identified by the PLINK program (Purcell et al., 2007) that overlap one of the known microcephaly loci. None of these regions were found in more than one of the three case genomes.

Individual	Homozygous region start coordinate	Overlap gene	Overlap start coordinate	Overlap stop coordinate	Size of overlap (KB)
Family A (II-1)	chr13:24181657	CENPJ	24354412	24395085	40.674
Family B (II-4)	chr8:6259807	MCPH1	6259807	6364841	105.035
Family A (II-1)	chr8:6289591	MCPH1	6289591	6466364	176.774
Family A (II-1)	chr9:122210554	CDK5RAP2	122210554	122250225	39.672

Table SIV.10 Sequencing coverage for exome-sequenced samples from families A-D.

(Relates to Table IV.2: Mutations identified in ASNS)

A base within the 37.8 Mb (families A and B) or the 52 Mb captured region (families C and D) is defined as covered if ≥5 short-reads spanned this nucleotide. After accounting for PCR duplicates and reads that did not align to captured regions of the reference genome, the average read depth for each sample was greater than 63x and greater than 95% of targeted bases were covered in all eight samples.

Individual	Total covered bases (MB)	% covered bases	Average read depth
Family A (II-1)	36.32	96.1	79.92
Family B (II-2)	36.30	96.0	63.93
Family B (II-4)	36.21	95.8	69.59
Family C (II-3)	51.28	95.0	80
Family D (II-1)	51.64	97.5	144
Family D (II-2)	51.37	95.8	126
Family D (II.4)	51.53	96.3	128
Family D (II.5)	51.67	97.8	89

Table SIV.11 Enzymes of the alanine, aspartate, and glutamine metabolic pathway and associated OMIM disorders.

(Relates to Table IV.3: Measurements of amino acid concentrations in patient blood and urine)

The identification of *ASNS* as the disease-causing gene in our patients also highlights the importance of human enzymes in the alanine, aspartate, and glutamine metabolic pathway in neurodevelopmental disorders. Human enzymes in the alanine, aspartate, and glutamine metabolic pathway (http://www.genome.jp/kegg- bin/show_pathway?hsa00250) and their associated OMIM disorders.

Enzyme	Gene	Associated OMIM disorder	OMIM disorder number	Neurological impairments
1.2.1.24	ALDH5A1	Succinic semialdehyde dehydrogenase deficiency	#271980	yes
1.4.1.3	GLUD1	Hyperinsulinism-hyperammonemia syndrome	#606762	yes
1.4.3.1	DDO	No disease associations in OMIM	-	no
1.4.3.2	IL4I1	No disease associations in OMIM	-	no
1.5.1.12	ALDH4A1	Hyperprolinemia, type II	#239510	yes
2.1.3.2	CAD	No disease associations in OMIM		no
2.4.2.14	PPPAT	No disease associations in OMIM	-	no
2.6.1.1	GOT1	No proven disease associations in OMIM	#614419	no
2.6.1.16	GFPT1	Myasthenia, congenital, with tubular aggregates 1	#610542	no
2.6.1.19	ABAT	GABA-transaminase deficiency	#613163	yes
2.6.1.2	GPT	No disease associations in OMIM	-	no
2.6.1.44	AGXT	Hyperoxaluria, primary, type 1	#259900	no
3.5.1.15	ASPA	Canavan disease	#271900	yes
3.5.1.2	GLS2	No disease associations in OMIM	-	no
3.5.1.3	NIT2	No disease associations in OMIM	-	no
4.1.1.15	GAD1	Cerebral palsy, spastic quadriplegic, 1	#603513	yes
4.3.2.1	ASL	Argininosuccinic aciduria	#207900	yes
4.3.2.2	ADSL	Adenylosuccinase deficiency	#103050	yes
6.3.1.2	GLUL	Glutamine deficiency, congenital	#610015	yes
6.3.4.16	CPS1	Carbamoylphosphate synthetase I deficiency	#237300	yes
6.3.4.4	ADSSL1	No disease associations in OMIM	-	no
6.3.4.5	ASS1	Citrullinemia	#215700	yes
6.3.5.4	ASNS	No disease associations in OMIM		yes
6.3.5.5	CAD	No disease associations in OMIM	-	no

CHAPITRE V: Disruption of CLPB is associated with congenital microcephaly, severe encephalopathy, and 3-methylglutaconic aciduria.

Disruption of *CLPB* is associated with congenital microcephaly, severe encephalopathy, and 3-methylglutaconic aciduria.

José-Mario Capo-chichi[1,*], Sarah Boissel[2,*], Edna Brustein [2], Sarah Pickles[3], Catherine Fallet-Bianco[4], Christina Nassif[1], Lysanne Patry[1], Sylvia Dobrzeniecka[3], Meijiang Liao[2,3], Damian Labuda[1,5], Mark E Samuels[1,5], Fadi F Hamdan[1], Christine Vande Velde[2,3], Guy A Rouleau[6], Pierre Drapeau[2,3], Jacques L Michaud[1,2,5].

[1] CHU Sainte-Justine Research Center, Université de Montréal, Montreal, Canada,
[2] Department of Neurosciences, Université de Montréal, Montreal, Canada,
[3] CHUM Research Center, Université de Montréal, Montreal, Canada
[4] Department of Pathology, CHU Sainte-Justine and Université de Montréal, Montréal, Québec
[5] Department of Pediatrics, Université de Montréal, Montreal, Canada
[6] Montreal Neurological Institute, McGill University, Montreal, Canada,

[*] These authors contributed equally to this work.

Correspondence: Jacques L. Michaud, CHU Sainte-Justine Research Center, 3175 Côte Sainte-Catherine room 5994, Montreal, Qc, Canada H3T 1C5.

Key words: CLPB, encephalopathy, 3-methylglutaconic aciduria

Journal of Medical Genetics, 2015 (*sous-presse*)

Contribution:

J'ai effectué les analyses génétiques ayant permis d'identifier la mutation causale (c.1685delT/ p.I562TfsX23 dans le gène *CLPB)* dans cette famille. J'ai également validé génétiquement l'implication de cette mutation dans la pathologie dans cette famille. J'ai effectué les travaux montrant que la mutation c.1685delT entraine la dégradation de *CLPB* dans les fibroblastes des patientes. J'ai effectué le clonage moléculaire de la mutation p.I562TfsX23 dans des vecteurs d'expression ; ces constructions ont été utilisées dans les études fonctionnelles chez le poisson zèbre. Boissel S et Brustein E ont validé fonctionnellement l'impact de la mutation p.I562TfsX23 chez le poisson zèbre ; elles ont caractérisé le phénotype des mutants *clpb.* Pickles S a étudié l'activité des mitochondries dans les fibroblastes des patientes. Labuda D, Samuels ME, Rouleau GA, Vande Velde C et Meijiang L ont aidé dans le *design* des travaux. Patry L et Dobrzeniecka S ont génotypé les mutations candidates identifiées dans une cohorte d'individus contrôles. Michaud JL et Fallet-Bianco C ont caractérisé le phénotype des patients dans cette famille. J'ai rédigé la première version du manuscrit, Hamdan FF, Boissel S, Drapeau P et Michaud JL ont également contribué à la rédaction et à la révision du manuscrit.

ABSTRACT

Background: The heterogeneous group of 3-methylglutaconic acidurias disorders includes several inborn errors of metabolism that affect mitochondrial function through poorly understood mechanisms. We describe four newborn siblings, from a consanguineous family, who showed microcephaly, small birth weight, severe encephalopathy and 3-methylglutaconic aciduria. Their neurological examination was characterized by severe hypertonia and the induction of prolonged clonic movements of the four limbs upon minimal tactile stimulation. **Methods and results:** Using homozygosity mapping and exome sequencing, we identified a homozygous truncating mutation (p.I562Tfs*23) in *CLPB* segregating with the disease in this family. *CLPB* codes for a member of the family of ATPases associated with various cellular activities (AAA^+ proteins) whose function remains unknown. We found that *CLPB* expression is abolished in fibroblasts from the patients. To investigate the function of this gene, we interfered with the translation of the zebrafish *clpb* ortholog using an antisense morpholino. The *clpb* morphants showed an abnormal touch-evoked response with increased swim velocity and tail beat frequency. This motor phenotype is reminiscent of that observed in the patients and is suggestive of increased excitability in neuronal circuits. Interestingly, knocking down *clpb* reduced the number of inhibitory glycinergic interneurons and increased a population of excitatory glutamatergic neurons in the spinal cord. **Conclusion:** All together, our study suggests that disruption of *CLPB* causes a novel form of neonatal encephalopathy associated with 3-methylglutaconic aciduria.

INTRODUCTION

Severe neurological impairment in the newborn – the so-called neonatal encephalopathies - represents a major diagnostic challenge. Clinical signs tend to be less specific at this early stage of life. Moreover, there are numerous disorders that can lead to such a clinical presentation, reflecting the complexity of brain development and function. Mitochondrial disorders, which are caused either by mutations in the mitochondrial or nuclear genome, represent a vast group of inherited disorders of energy metabolism that are being recognized as a major cause of neonatal encephalopathy (Uziel et al., 2011). Because oxidative phosphorylation is a fundamental pathway of cellular metabolism, these disorders encompass a wide range of symptoms and presentations, severity and outcome.

The heterogeneous group of 3-methylglutaconic aciduria syndromes includes several inborn errors of metabolism biochemically characterized by increased urinary excretion of 3-methylglutaconic acid. One form of 3-methylglutaconic aciduria is an inborn error of leucine catabolism whereas the other forms all affect mitochondrial function through poorly characterized mechanisms (Wortmann et al., 2012). Here, we report a homozygous truncating mutation in the *CLPB* gene in four newborn siblings, from a consanguineous family, who showed microcephaly, severe encephalopathy, hyperlactacidemia and 3-methylglutaconic aciduria. *CLPB* belongs to the Clp/Hsp100 family of ATPases associated with various cellular activities (AAA+ family of proteins). This family has the ability to remodel proteins in an ATP-dependent manner, catalyze protein unfolding, disassembly and disaggregation. *CLPB* function, however, remains unknown. Mitochondrial studies failed to identify any deficits in the fibroblasts of the patients but knock down of native *clpb* using antisense morpholino oligonucleotides (AMO) in the developing zebrafish resulted in morphological and motor phenotypes that are reminiscent of those observed in the patients. Loss of *CLPB* function thus appears to cause a novel form of neonatal encephalopathy with 3-methylglutaconic aciduria.

MATERIALS AND METHODS

Subjects.

Blood samples were obtained from each of the participants after an informed consent and the approval of the appropriate institutional ethic review board.

Homozygosity mapping.

Genomic blood DNA (2.5µg) from individuals II.1 to II.4 was used for whole genome genotyping on the Illumina 660 Quad Chip (Infinium HD) at the McGill University and Genome Quebec Innovation Center (MUGQIC, Montreal, Canada). Homozygosity mapping was performed using PLink (v.1.06). Shared regions of homozygosity (HR) containing >30 consecutive SNPs and extending over 1Mb were retained.

Sequencing of genomic DNA.

The exome of patient II.1 was captured from 3µg of blood genomic DNA, using the Agilent SureSelect Human All Exon Capture 50Mb Kit (Mississauga, ON) and sequenced (paired-end, 2x100bp) using the Illumina Hi2000seq at the McGill University and Genome Québec Innovation Center. Sequence processing, alignment (with a Burrows-Wheeler algorithm; BWA) and variant calling were done using the Broad Institute Genome Analysis Tool Kit (GATK v4) and annotated using Annovar. The mutation c.1685delT:p.I562Tfs*23 in CLPB (NM_030813) was genotyped in the four affected siblings, their parents and 90 East-Asian controls by performing Sanger sequencing with gene specific primers flanking the mutation (Forward: 5'-acagatggacaggaccaggg Reverse: ctccacctttatgtgtgtggg-3').

RNA analyses.

For fibroblast studies, total RNA was extracted using the Trizol reagent (Invitrogen, Burlington, ON) and reverse-transcribed (1µg) using the Superscript Reverse Transciptase-II (Invitrogen). The cDNA was PCR-amplified using primers targeting exons 13 and 16 of CLPB (NM_030813.3; ex13F: 5'-atgagccgtaaccgtattgc-3'; ex16R:5'-tgatgttgtgcctttgcttg-3') flanking the frameshift mutation. □-actin was also amplified as an internal control. For zebrafish studies, total RNA was extracted at 48 hours post fertilization (hpf) using trizol reagent. cDNA was synthetized using random primers with RevertAidTM H Minus First Strand cDNA Synthesis Kit (Fermentas) according to the provided protocol. clpb was amplified by PCR using specific primers flanking the splice junction targeted by the morpholino (Forward 5'-acgagatcgtgtacttcctgc, Reverse taacaactcctgctcgaacg-3'). We used gapdh (Forward 5'-cttggctcctctggctaaagtt, Reverse gtcataccaggagatgagcttgac-3') as an internal control. The PCR products were separated on a 1% agarose electrophoresis gel, purified with the EZ-10 Spin Column PCR Purification Kit (Biobasic inc, BS664) and sequenced on an ABI PRISM 3130.

Protein extraction.

For whole-cell protein extraction, fibroblasts were scraped and lysed into 300 □L of fresh ice-cold RIPA Buffer (25 mM Tris-HCl pH =7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) mixed with protease inhibitors (Roche). Cell lysate was centrifuged at 13000 rpm for 15 min at 4°C and the supernatant was collected as whole-cell extract. Protein concentration was determined by Bradford assay. For zebrafish protein extraction, up to 100 embryos were collected at 48 hpf in 200µL of deyolking buffer (1/2 Ginzburg Fish Ringer without calcium: 55mM NaCl, 1.8mM KCl, 1.25mM NaHCO3). Yolks were disrupted

by pipetting and embryos were shaken at 1100 rpm for 5 min to dissolve completely the yolks. Cells were further pelleted by centrifugation at 300g for 30 sec and the supernatant was carefully discarded. The pellet was lysed into 2µL-per-embryo of a solution of fresh ice-cold RIPA buffer (50mM Tris pH=8, 150mM NaCl, 0.1% SDS, 0.5% Sodium deoxycholate 1% NP40, 1mM PMSF) mixed with a cocktail of protease inhibitors (Roche). Protein concentration was determined by Bradford assay.

Immunoblotting.

10-20 □g of protein were separated by SDS-PAGE (4%, 10%) and transferred onto a polyvinylidene difluoride membrane (PVDF, BioRad). CLPB was detected using a 1:300 dilution of anti-CLPB antibody (SIGMA, HPA039006) raised against aminoacids 589-693 of human CLPB protein. We used a monoclonal mouse antibody (Novus Biological, NB600-501 diluted 1:1000) to detect □-actin as a protein loading control. Primary antibodies were incubated overnight at 4°C. Membranes were washed and incubated 1h at room temperature with secondary antibodies conjugated to horseradish peroxidase goat polyclonal anti-rabbit IgG or rabbit polyclonal anti-mouse IgG (Abcam; diluted 1:10000). Protein detection was done using the Pierce ECL Western Blotting Substrate (Thermo Scientific).

Cloning of CLPB expression vectors.

The full-length open reading frame of human CLPB cDNA (NM_030813.3) was amplified from human brain total RNA using PCR. Amplicons were cloned into the mammalian expression vectors pCS2+ (Addgene, Forward: 5'-aagcggaattcaccatgctggggtccctgg; Reverse: atctctcgagctagatggtgt tgcacacct-3') using gene specific primers. Plasmids were isolated from individual clones and confirmed by restriction digest to contain insert of the right size. Inserts were sequenced to confirm absence of any mutation. The frameshift mutation (NM_030813.3, c.1685delT) was introduced into the CLPB cDNA using QuickChange II site-directed mutagenesis (5'-tgagtttctgggacggacaatgagatcgtctactagtagacgatctattgtccgtcccagaaactca-3') according to the manufacturer's recommendations (Agilent Technologies). Insertion of the c.1685delT frameshift mutation was confirmed by sequencing.

***clpb* knockdown and rescue in zebrafish.**

Experiments were performed on zebrafish (Danio rerio) larvae raised at 28.5°C according to previously established procedures (Kimmel et al., 1995), and in compliance with Canada Council for Animal Care and institutional guidelines. To knockdown the clpb orthologue in zebrafish, a splice-blocking antisense morpholino oligonucleotides (AMO, Gene Tools), was designed to target splice junction 9 of the zebrafish clpb gene (5′-tagacagtattgatcagcacctcgt-3′). Morpholino preparation and injection procedures were performed according to previously described protocols (Kabashi et al., 2010). Briefly, the morpholino was pressure ejected into 1-2 cell stage blastocytes at various concentrations to determine the most penetrant dose (1mM). After establishing the knockdown phenotype using 1mM of AMO, rescue experiments were performed by co-injecting 1mM of AMO with 25 ng/µL of either the WT-HmRNA or mut-HmRNA, which were transcribed from NotI-linearized pCS2+-CLPB constructs using SP6 polymerase with the mMESSAGE Machine Kit (Ambion), followed by a phenol:chloroform extraction and isopropanol precipitation.

Analysis of zebrafish morphology.

Bright field images of 48 hpf larvae were taken with a Flea 2 camera (Point Gray Research) mounted on a dissecting microscope (Olympus xyz). The images were then analyzed offline

using Volocity software (Perkin Elmer) to determine the body length as well as the eyes and head size of each larvae.

Assessment of locomotor behavior.

For all behavior experiments, special care was taken to maintain the E3 medium (5mM NaCl, 0.17mM KCl, 0.33mM CaCl2, and 0.33 mM MgSO4) temperature constant at 28.5 °C. For each batch of injection, raw measurements were normalized to the median of non-injected fish. For the assessment of the touch-evoked response, 52 hpf larvae were placed in the middle of a 150 mm dish filled with E3 medium. Touch response elicited by tactile stimulation delivered to the tail with forceps was recorded from above digitally at 30 Hz (Grasshopper 2 camera, Point Grey Research). Maximum swim velocity was quantified using the manual tracking plug-in of ImageJ (http://rsbweb.nih.gov/ij/). For assessment of tail beat frequency, 52 hpf larvae were immobilized in 1% agarose, dorsal side up, and the tail was liberated allowing a free and full range of motion. Swim behavior was elicited by touching the tail with forceps while video-recording at high speed (120 Hz) using a dissection microscope (Leica MZ16, Wetzlar, Germany). The tail beat frequency was calculated by taking the average of the frequencies of the 10 first free tail beats, which were manually quantified using the Virtual Dub software.

Quantification of glycinergic and glutamatergic interneurons.

Injections of 1mM of clpb AMO were performed in zebrafish pax2:gfp and alx:gfp transgenic lines. Larvae were fixed at 30 hpf in 4% paraformaldehyde (PFA) overnight at 4°C washed in PBS-0.1% tween 20 (PBST) and mounted on a glass slide in PBS. Confocal images were acquired using a Quorum Technologies spinning disk confocal microscope with CSU10B (Yokogawa) spinning head mounted on an Olympus BX61W1 fluorescence microscope and connected to a Hamamatsu ORCA-ER camera. Images were acquired using Volocity software (Improvision). The number of interneurons per somite was calculated offline by taking the average number of gfp-positive cells in 3 consecutive somites (between somite 11 and somite 15).

Statistical analyses on zebrafish experiments.

We used SigmaPlot 11.0 integrated with SigmaStat 3.1 to assess the statistical difference among the various experimental groups. Kruskal-Wallis One Way Analyses of Variance on Ranks (ANOVA) and All Pairwise Multiple Comparison Procedures (Dunn's Method) were performed. Descriptive analyses are presented as following: Mean ± Standard Error to the mean.

RESULTS

We characterized a consanguineous Cambodian family with four female siblings (individuals II.1 - II.4) who were all born with microcephaly (head circumference at 30-30.5 cm) (Figure. V.1A; see detailed clinical description of the siblings in the Supplementary material). Three of them (II-1, II-2 and II-3) also had a small birth weight (under the 3rd centile). All of the siblings showed the same clinical profile. They did not move or breathe spontaneously. Appendicular tonus was greatly increased. In addition, sustained clonic movements of the upper or lower limbs were induced by minimal tactile stimulation and were stopped by gentle physical restraining. Electroencephalography revealed a pattern consistent with burst suppression in all of the siblings. Brain MRI showed gyral simplification in sibling II-2 and was unremarkable in sibling II-3.

Plasma lactate levels were elevated in the first three siblings but were not measured in the last one. Urine organic acid chromatography revealed increased levels of lactate, 3-methyl-glutaconic acid and methyl-glutaric acid in individuals II-1 and II-2. These studies were not performed in the other two siblings. The conjunction of hyperlactacidemia and 3-methylglutaconic aciduria is thus suggestive of a mitochondrial disorder. However, respiratory chain studies showed normal activity levels of pyruvate dehydrogenase, cytochrome c oxidase, and succinate cytochrome c reductase in the fibroblasts of siblings II- 2 and II-3. In the liver of the same individuals, respiratory chain studies showed normal activity levels of cytochrome c oxidase, succinate cytochrome c reductase and citrate synthase but undetectable levels of NADH-cytochrome c reductase, which could be indicative of a severe deficiency of complex I or could represent an artifact because of the tendency of this complex to become rapidly unstable *in vitro*.

Neutropenia was documented in all the siblings (II-1, II-2 and II-3) in whom white blood cells were measured. Coagulation was perturbed with elevated prothrombin time (sibling II-2 and II-4), activated partial thromboplastin time (siblings II-2, II-3, and II-4) and D-dimers levels (II-2, II-3) in some of the siblings. Fibrinogen and anti-thrombin III levels were also low in sibling II-2.

All siblings were ventilator-dependant and died within the first week of life, after the arrest of respiratory support. Neuropathological studies showed the same abnormalities in the four siblings, including: 1) neuronal loss mainly in the hippocampus, the striatum and the pallidum and to a lesser extent in the thalamus, the quadrigeminal tubercles and the pons; 2) diffuse gliosis involving the tegmentum and the ponto-bulbar region; 3) diffuse lesions of the white matter characterized by gliosis and microvacuolisation (with focal cystic degeneration in sibling II-3); and 4) the presence of Alzheimer type II astrocytes in the striatum, pallidum and thalamus. These changes are non-specific and are consistent with a defect in energy metabolism. We did not observe any signs of cortical malformations or neuronal migration defects.

We performed homozygosity mapping in all the affected siblings and identified 11 regions of homozygosity shared among them, the biggest of which was 52 Mb and located on chromosome 11 (41,582,337-93,578,963; NCBI genome build hg19) (Table SV.1). We identified 8 genes associated with neonatal encephalopathy (*ACP2* (MIM#171650), *ALG8* (MIM#608103), *DHCR7* (MIM#602858), *NDUFS3* (MIM#603846), *NDUFS8* (MIM#602141), *NDUFV1* (MIM#161015), *RNASEH2C* (MIM#610330), *SLC35C1* (MIM#605881) that are located in the homozygosity region on chromosome 11. We sequenced all the coding exons and consensus splice sites of these genes in individual II.1 but did not identify any rare homozygous variants.

We next captured and sequenced the exome of individual II.1. After mapping and removing duplicate reads, we obtained an average per target base coverage of ~ 90x, with 95% of the target bases covered at least 10x. We identified 229 functional variants (non-synonymous and insertions/deletions variants in the coding regions; variants affecting the consensus splice sites) in the regions of homozygosity shared by the four affected siblings, 10 of which were rare as they were either not previously reported or present at MAF < 1.0% in our in-house control exome dataset (n=657) or in public SNP databases (1000 genomes, NHLBI Exome Sequencing Project (ESP), or dbSNP) (Tables SV.2 and SV.3). We filtered out four of these rare variants because they were predicted to be benign or tolerated by -four different algorithms (Polyphen-2, SIFT, Mutation Taster and Provean) (Tables SV.2 and SV.3). Two other variants were also excluded from further consideration because recessive

mutations in the corresponding genes (*MYO7A*, *LRP4*) have previously been associated with phenotypes (Usher syndrome, oligo-syndactyly) that are distinct from that found in our patients (Tables SV.2 and SV.4). One of the four remaining variants is a missense (p.D976N) that was found in the *MADD* gene, which encodes the Rab3 GDP-GTP exchange factor (Yoshimura et al., 2010) . We also excluded this variant because mice homozygous for a null allele of *Madd* are hypo-responsive to tactile stimulations while our patients showed the reverse (Tables SV.2 and SV.4) (Tanaka et al., 2001). By Sanger sequencing, we analyzed the three remaining variants in 90 controls of South-Eastern Asiatic origin. One of these variants, located in *SPDYC*, was too frequent (allele frequency: 10%) in this cohort of controls to explain such a rare condition (Table SV.2). The two remaining variants include a missense in *ARHGEF17* (p.T751I) and a truncating mutation in *CLPB* (NM_030813.3, c.1685delT/ p.I562Tfs*23) (Table SV.2). These variants were confirmed by Sanger sequencing to be homozygous in all four affected siblings and heterozygous in their parents and were absent from our entire set of in-house exomes and from all the above-mentioned public SNP databases (Figure V.1A-C). *ARHGEF17* (also known as KIAA0337) encodes a RhoGEF that is predominantly expressed in the heart (Rumenapp et al., 2002). The missense variant in this gene is predicted to be damaging by SIFT but benign by three other prediction tools (Polyphen-2, Mutation Taster, and Provean, Table SV.3). Moreover, the residue affected by this variant is not conserved in several vertebrates including rodents, zebrafish and *Xenopus* (Figure S2). Because the truncating mutation in *CLPB* was more likely to be disruptive, we decided to focus our attention on this candidate gene.

CLPB is a member of the Clp/Hsp100 family of ATPases that are characterized by the presence of core ATP nucleotide binding domains (NBDs) composed of Walker-type A and B consensus regions, arginine finger motifs and sensor-1 and -2 regions (Figure V.1B) (Neuwald et al., 1999 ; Vale, 2000 ; Ogura et Wilkinson, 2001 ; Ogura et al., 2004 ; Hanson et Whiteheart, 2005). CLPB contains only one NBD, which shares remarkable identity with the homologous domain in the prokaryotic Clpb, yeast cytosolic Hsp104 and mitochondrial Hsp78 proteins (Figure SV.3). The structure of human CLPB, however, differs from that of these bacterial and yeast proteins by the presence of four ankyrin repeats at its N-terminal end in place of a second NBD (Figure V.1B). Clpb, Hsp104 and Hsp78 have a unique ability to

regenerate proteins from aggregates by mediating re-solubilization and refolding of damaged proteins (Doyle et Wickner, 2009). These Clp/Hsp100 proteins function as homo-hexameric rings with a central pore through which the substrate proteins are driven due to conformational changes following ATP binding and hydrolysis (Parsell et al., 1994 ; Goloubinoff et al., 1999 ; Mogk et al., 2003 ; Bosl et al., 2006). The biochemical function of CLPB remains unknown.

The c.1685delT/p.I562Tfs*23 mutation is predicted to abolish the C-terminal end of CLPB's NBD (also termed D2-Small) (Figure V.1B). Key residues in this domain have been shown to be important for ATP binding and hydrolysis as well as for oligomerization in most AAA+ proteins (Ogura et al., 2004 ; Hanson et al., 2005 ; Hodson et al., 2012 ; Zeymer et al., 2014). Moreover, truncating the D2-Small domain completely abolishes the oligomerization of the prokaryotic Clpb (Barnett et al., 2000). The NHLBI/ESP exome dataset reports four truncating heterozygous variants in *CLPB* in over 6500 individuals, the most frequent of which was only at 0,8%, thus suggesting that *CLPB* does not accumulate truncating mutations.

We next studied the impact of c.1685delT on *CLPB* mRNA and protein levels in fibroblasts from individuals II.1 and II.2. Reverse transcription PCR (RT-PCR) did not show any detectable expression of *CLPB* mRNA in cells from the affected siblings (Figure V.1D). Western blot analyses using an antibody directed against the portion of CLPB located after the predicted site of truncation did not show any detectable bands (Figure V.1E). These results suggest that c.1685delT abolishes *CLPB* expression probably by inducing nonsense-mediated decay of its mRNA.

We next assessed the biological impact of the c.1685delT mutation by interfering with the expression of the *clpb* zebrafish ortholog, which shares 81% identity with the human CLPB protein. We used splice-blocking antisense morpholino oligonucleotides (AMO) to alter mRNA transcription at splice junction 9 of the zebrafish *clpb,* in proximity to the site of truncation observed in the affected individuals (Figure SV.4A). We tested different concentrations of *clpb* AMO and established that the most penetrant and least toxic dose for zebrafish *clpb* knockdown (KD) experiments was at 1mM. RT-PCR analysis revealed that the expression of wild type (WT) *clpb* mRNA was reduced in larvae injected with *clpb* AMO, compared to the WT non-injected ones. As expected, *clpb* morphants also expressed an abnormally spliced *clpb* transcript (Figure SV.4B). Sequencing of this abnormal transcript

showed that it retained a fragment of intron 9, which resulted in a frameshift and truncation in exon 10 (Figure SV.4A). Western blot results confirmed that the expression of WT clpb protein was reduced in the larvae injected with *clpb* AMO (Figure SV.4C).

clpb morphants were well formed but exhibited some morphological deficits at 48 hours post-fertilization (hpf). They were shorter (3% less in body length) and had smaller eyes (14% less) and head size (3% less) compared to WT non-injected larvae (Figure V.2A-D). Reduction of eye size and body length of *clpb* KD larvae was statistically significant (one way ANOVA, $p<0.05$) (Figure V.2A-C). We assessed the specificity of these deficits by performing rescue experiments whereby *clpb* AMO was co-injected with either WT (WT-HmRNA) or c.1685delT mutant (mut-HmRNA) human *CLPB* mRNA. Expression of WT-HmRNA restored the eye size and body length phenotype of *clpb* KD larvae (R-WT, n=37). In contrast, injecting mut-HmRNA in *clpb* KD larvae (n=50) failed to rescue their morphological deficits (Figure V.2A-D). We observed a similar trend, albeit not significant, for the rescue of the head size phenotype of *clpb* morphants. Overexpression of WT or mutant *CLPB* did not have any effect on these different morphological parameters, excluding a toxic gain of function effect (Figure V.2A-D).

Our patients displayed a striking motor phenotype characterized by severe hypertonia with sustained clonic movements of the upper and lower limbs induced by tactile stimuli. We thus investigated whether *clpb* KD larvae exhibited any abnormal motor behavior. As early as 17 hpf, zebrafish larvae show their first motor activity, consisting of spontaneous coiling of the tail that persists for several hours (Saint-Amant, 2006). After 28 hpf, swimming episodes can be elicited upon tactile stimulation. We did not observe any major difference in the spontaneous motor activity at 19 hpf and 24 hpf between *clpb* KD and WT non-injected larvae. The *clpb* morphants, however, exhibited an exaggerated touch-evoked response at 52 hpf. Indeed, *clpb* mutants (n=106) propelled with a 9% increase in maximum swimming velocity upon tactile stimulation, when compared to WT non-injected larvae (n=114) (Figure V.3A, one way ANOVA, $p < 0.05$). Again, we assessed the specificity of this motor behavior in rescue experiments. While expression of WT-HmRNA restored the motor activity of *clpb* KD larvae (n=47), injecting mut-HmRNA in *clpb* KD larvae failed to rescue the phenotype (n=34) (Figure V.3A) .To better characterize the motor behavior of the *clpb* morphants, we

immobilized the anterior part of their body in agarose and studied the wavering of their tail following tactile stimulation. Consistent with our findings regarding their maximum swim velocity, we observed that *clpb* KD larvae (n=22) responded to tactile stimuli with a 32% increase in tail beat frequency (one way ANOVA, $p < 0.05$), when compared to WT non-injected larvae (n=20) (Figure V.3B-C). *clpb* knockdown in zebrafish thus induces an increase in motor activity upon tactile stimulation that is reminiscent of the phenotype observed in the affected individuals described here.

The motor phenotypes observed in our patients and in the *Clpb* morphants are suggestive of hyperexcitability in motor circuits (Betz, 1991). Glycinergic neurons are the main inhibitory neurons in the zebrafish spinal cord. Hence, we next investigated whether knocking down *clpb* affects the population of glycinergic neurons in zebrafish embryos. We injected *clpb* AMO in larvae expressing a *pax2:gfp* transgene, which mainly labels glycinergic interneurons (Batista et Lewis, 2008). Knocking down *clpb* in *pax2:gfp* larvae significantly reduced the number of *pax2*-expressing neurons (Figure V.4A-E). Rescue experiments showed that *clpb* KD larvae injected with WT-HmRNA recover some of their glycinergic neurons (n=45) (Figure V.4A-E, 42 % decrease; one way ANOVA, $p < 0.05$). Again, expression of mut-HmRNA in *clpb* morphants did not have any effect on their phenotype (n=42) (Figure V.4A-E). We next examined whether *clpb* KD also affects other types of neurons. We studied the impact of a reduction of *clpb* by knocking it down in larvae expressing an *alx:gfp* transgene, which mainly labels ipsilateral descending glutamatergic interneurons (Kimura et al., 2006). Interestingly, knocking down *clpb* increased the number of *alx*-expressing neurons (Figure V.5A-D, 13% increase, one way ANOVA, $p < 0.05$). Overall, our results suggest that decrease of *clpb* causes motor hyperactivity by reducing the ratio of inhibitory to excitatory neurons.

DISCUSSION

We identified a homozygous truncating mutation in *CLPB* in four newborn siblings, who presented with microcephaly, small birth weight and severe encephalopathy. Several observations suggest that this mutation is pathogenic. First, it is the only candidate variant identified by exome sequencing in the regions of homozygosity shared by all four siblings. Second, this mutation is likely causing a loss of *CLPB* function: it is predicted to truncate the protein upstream of a critical functional domain and it abolishes its expression in fibroblasts from the patients. Finally, knocking down *Clpb* expression in the zebrafish induced phenotypes such as reduced growth and increased motor activity following tactile stimulation that are reminiscent of those observed in the patients. Mutant *CLPB* mRNA failed to rescue these phenotypes in the zebrafish, further validating the functional impact of the truncating mutation observed in the siblings.

Our patients showed some signs of mitochondrial dysfunction, as suggested by the increased production of lactate and the presence of 3-methylglutaconic aciduria (Wortmann et al., 2012). In addition, the lesions of the white matter observed in these patients are similar to those described in deficiencies of pyruvate carboxylase, pyruvate dehydrogenase and complexes I and/or IV of the respiratory chain. Interestingly, there are some indications that CLPB might have a mitochondrial function. First, CLPB possesses a mitochondrial signal sequence at its N-terminus end. Second, CLPB is enriched in the mitochondrial proteome of various tissues, particularly in the brain (Pagliarini et al., 2008). Third, *Clpb* expression is up-regulated in mice deficient in *Ant1*, which exchanges the mitochondrially synthesized ATP for the spent cytosolic ADP across the mitochondrial inner membrane (Murdock et al., 1999). Ant1 thus provides the means by which mitochondrial ATP is made available to the cell. It has been suggested that the increase of *Clpb* expression in *Ant1* mutant mice contributes to an effort to produce more mitochondria to compensate for the bioenergetic defects caused by this deficiency (Murdock et al., 1999). Interestingly, the yeast protein Hsp78, which shares great identity with CLPB, is involved in the re-activation of proteins from an aggregated state in the mitochondria (Moczko et al., 1995). One possibility is that CLPB also functions as a mitochondrial chaperone in humans.

We did not notice any deficit in respiratory chain activity in the fibroblasts of the patients (Figures S6 and S7). Moreover, we could not detect any changes in the morphology, membrane potential or superoxide production of the mitochondria in these cells (Figure S5). Although CLPB is produced in fibroblasts, it is possible that its activity is not necessary in these cells or that it is required only under specific stress conditions. We did document undetectable levels of NADH-cytochrome c reductase in the liver of the patients but it is possible that this decrease represents an artifact, as this complex is notoriously unstable.

Knock-down of *clpb* in the zebrafish resulted in increased swim velocity and tail beat frequency following tactile stimulation. *clpb* morphants were also characterized by a reduction of the ratio of inhibitory to excitatory neurons in the spinal cord. Embryonic zebrafish motion consists of coiling of the tail, following alternated contractions and release of trunk muscles, in response to descending excitatory interneurons and subsequent contralateral inhibition by interneurons. (Fetcho, 1992 ; Liao et Fetcho, 2008). We speculate that *clpb* morphants respond to touch with increased speed because their imbalanced inhibitory and excitatory neuronal networks cannot generate the appropriate outputs. The motor phenotype observed in our patients overlaps with that of hyperekplexia (HKPX, MIM#603846), a condition characterized by hypertonia and pronounced startle responses to tactile and acoustic stimuli (Andermann et al., 1980 ; Bakker et al., 2006). HKPX is caused by disruption of glycinergic signaling (Harvey et al., 2008a ; Harvey et al., 2008b). A model of HKPX in the zebrafish is the *bandoneon* (*beo*) mutant, which carries a mutation in one of the glycine receptor gene *glrb2* and shows a complete loss of glycinergic synaptic transmission (Hirata et al., 2005). The *beo* mutants are paralyzed and cannot swim, presumably because their trunk muscles stiffen in the absence of contralateral inhibition from glycinergic interneurons. In contrast, contralateral inhibition of trunk muscles would still occur in *clpb* mutants because of the firing of their residual glycinergic interneurons. Therefore, distinct patterns of disruption of glycinergic circuits could have different impact on motor behavior.

Although genes associated with mitochondrial disorders tend to be ubiquitously expressed, specific gene defects can result in particular clinical phenotypes. We hypothesize that CLPB is specifically required for mitochondrial function in glycinergic cells. CLPB disruption would thus compromise the survival of these cells by affecting their energy

production. In contrast, *clpb* knock-down resulted in an increase of the number of glutamatergic cells in the spinal cord of zebrafish. This increase could represent a secondary effect of the loss of glycinergic interneurons or of other cell types during development. Alternatively, CLPB could be required for both mitochondrial and cytoplasmic processes, affecting multiple pathways including some that are specifically involved in the regulation of cell proliferation or differentiation.

In summary, we describe a novel form of neonatal encephalopathy associated with 3-methylglutaconic aciduria. Loss of function of *CLPB*, a member of the AAA+ family of proteins, is likely the cause of this condition. CLPB appears to be required for mitochondria function, possibly by acting as a protein chaperone. Additional studies will be required to define the precise function of CLPB in the mitochondria and to determine whether CLPB plays other roles during development of the central nervous system.

Acknowledgements

We thank the patients and their family for participating in this study. We also thank the members of the RMGA bioinformatic team (Alexandre Dionne-Laporte, Dan Spiegelman, Edouard Henrion, and Ousmane Diallo) for the primary analysis of the exome sequencing data. We thank G. Laliberté and M. Drits for their help with the zebrafish experiments and Catherine Brunel-Guitton for helpful discussions. This work was supported by March of Dimes (grant no. 12-FY10-236 to M. Samuels and J.L. Michaud). J.L. Michaud is a National Scientist of the Fonds de Recherche du Québec-Santé. J.M. Capo-Chichi holds a scholarship from the Réseau de Médecine Génétique Appliquée du Québec (RMGA). S. Boissel holds a postdoctoral scholarship from the Canadian Institutes of Health Research (CIHR).

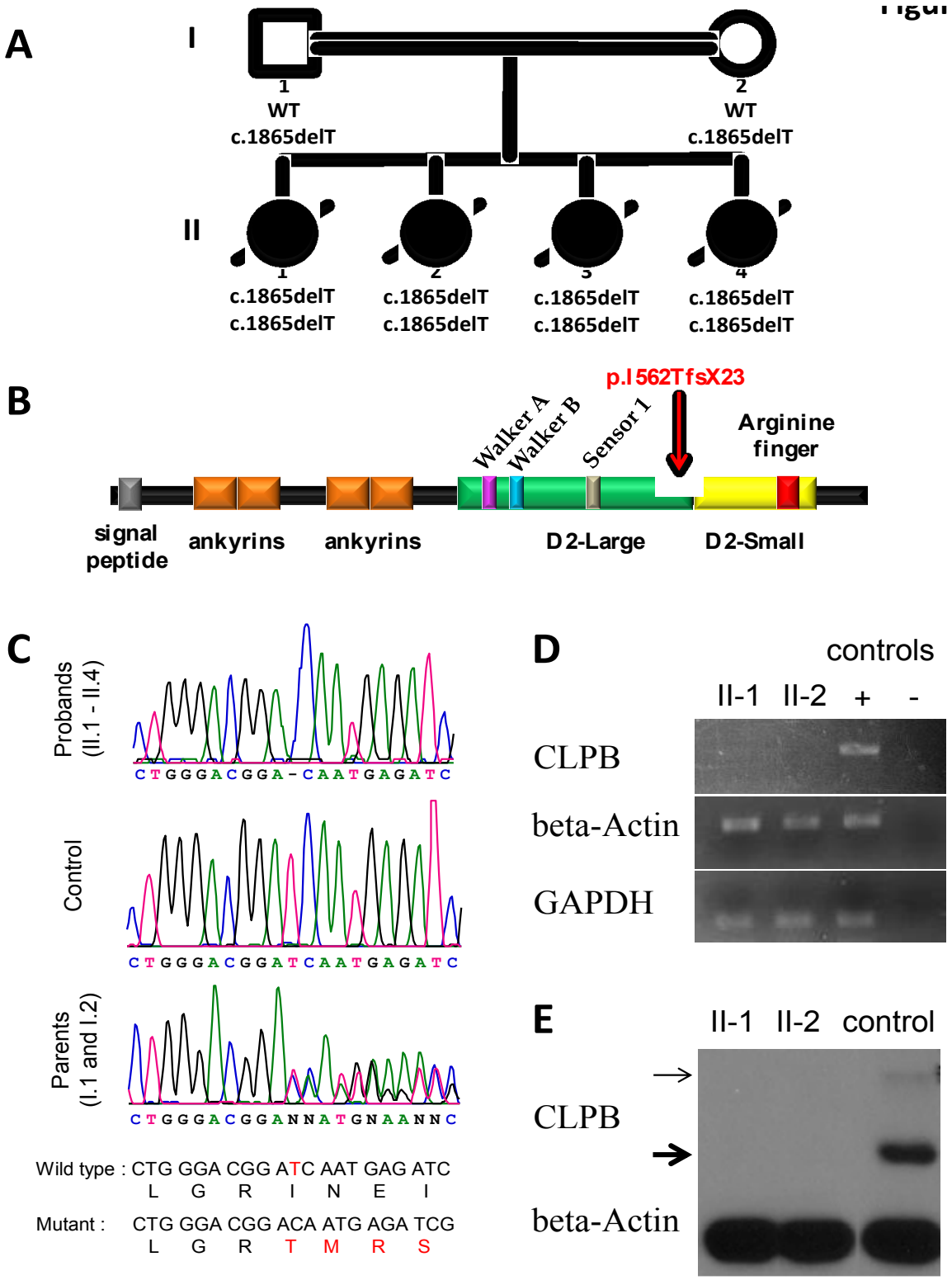

A
I
1
WT
c.1865delT
2
WT
c.1865delT
II
1
c.1865delT
c.1865delT
2
c.1865delT
c.1865delT
3
c.1865delT
c.1865delT
4
c.1865delT
c.1865delT
B
p.I562TfsX23
Walker A
Walker B
Sensor 1
Arginine finger
signal peptide
ankyrins
ankyrins
D2-Large
D2-Small
C
Probands (II.1 - II.4)
C T G G G A C G G A - C A A T G A G A T C
Control
C T G G G A C G G A T C A A T G A G A T C
Parents (I.1 and I.2)
C T G G G A C G G A N N A T G N A A N N C
Wild type : CTG GGA CGG ATC AAT GAG ATC
L G R I N E I
Mutant : CTG GGA CGG ACA ATG AGA TCG
L G R T M R S
D
controls
II-1 II-2 + -
CLPB
beta-Actin
GAPDH
E
II-1 II-2 control
CLPB
beta-Actin

Figure V.1

Figure V.I Identification of a homozygous truncating mutation in *CLPB*.
(A) Pedigree of the family. (B) Schematic representation of CLPB showing the ankyrin repeats as well as the single nucleotide binding domain (NBD-2) composed of D2-Large and D2-Small domains.. p.I562Tfs*23 identified herein truncates the D2-Small domain. Key functional residues in the NBD-2 are also illustrated. (C) Sanger sequencing confirming that the c.1685delT/p.I562Tfs*23 (NM_030813.3) mutation in *CLPB* segregates in this family. Representative chromatograms of the four affected siblings (homozygous for c.1685delT), their parents (heterozygous for c.1685delT) and a healthy unrelated control (Wild-Type; WT) are shown. (D) RT-PCR analysis using gene specific primers to amplify a fragment of *CLPB*. Primary fibroblasts cultures from individuals II.1 and II.2 do not show any detectable expression of *CLPB*. Primary fibroblasts culture from an unrelated healthy individual lacking the c.1685delT mutation in CLPB was used as positive control (+). Water was used as a negative control (-). beta-Actin and GAPDH, amplified using gene specific primers, were used as loading controls. (E) Western blot using an antibody raised against aminoacids 589-693 of human CLPB protein. We did not detect any expression of CLPB in fibroblasts from individuals II.1 and II.2. Arrowheads show two of the main CLPB isoforms recognized by the antibody at 57 and 78 KDa. beta-Actin was used as a loading control.

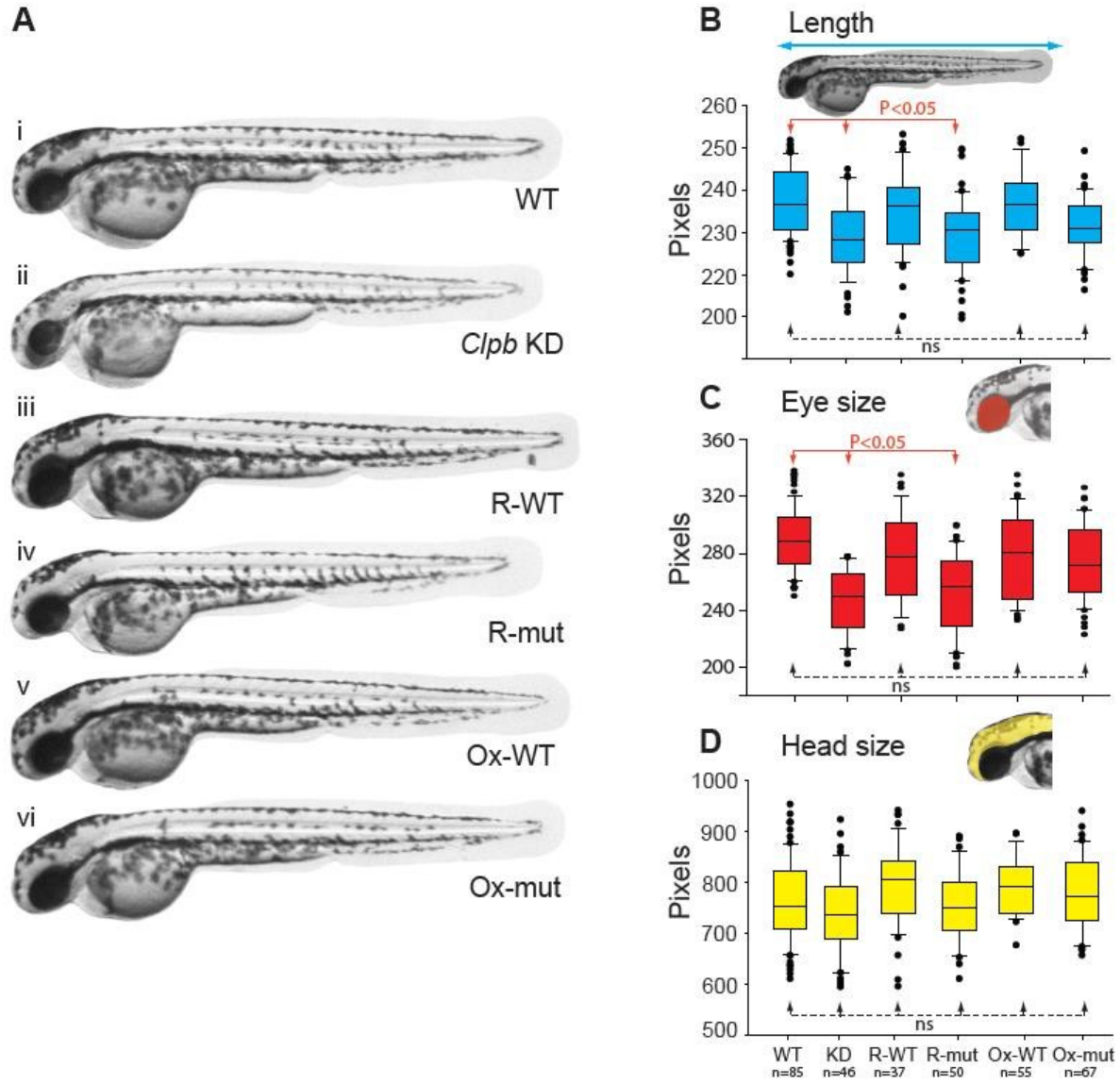

A
i
WT
ii
Clpb KD
iii
R-WT
iv
R-mut
v
Ox-WT
vi
Ox-mut
B
Length
P<0.05
Pixels
260
250
240
230
220
200
ns
C
Eye size
P<0.05
Pixels
360
320
280
240
200
ns
D
Head size
Pixels
1000
900
800
700
600
500
ns
WT n=85
KD n=46
R-WT n=37
R-mut n=50
Ox-WT n=55
Ox-mut n=67

Figure V.2.

Figure V.2 Morphological deficits observed in *clpb* KD larvae.
(A) Representative bright field images of 48 hours-old zebrafish larvae illustrating the phenotype under different experimental conditions: (i) Wild-type (WT) non-injected larvae; (ii) *clpb* knockdown (KD) larvae exhibit significantly smaller eyes and shorter body length but not significantly reduced head size when compared to controls; (iii) Co-injection of AMO with WT human *CLPB* mRNA rescues the *clpb* KD phenotype (R-WT); (iv) Embryos co-injected with AMO and mutant human *CLPB* mRNA exhibit the same morphological deficits as the *clpb* KD larvae (R-mut); (v) Larvae over-expressing solely WT human *CLPB* mRNA are similar to the WT ones (Ox-WT); (vi) Larvae over-expressing solely mutant human *CLPB* mRNA are similar to the WT ones (Ox-mut). (B) Box plot summarizing the distribution of larvae length (pixels) across the different experimental groups, measured as illustrated by the top image inset (blue arrow): Each box plot contains a central line representing the median, the box itself delineating 25-75% of the data, and error bars representing 10-90% of the data range, whereas open circles represent outlying data points. For Statistics see methods. (C) Box plot summarizing the distribution of larvae eye size (pixels) across the different experimental groups, as B. Eye size was measured as illustrated by the top image inset (red). (D) Box plot summarizing the distribution of larvae head size (pixels) across the different experimental groups, as B. Head size was measured as illustrated by the top image inset (yellow).

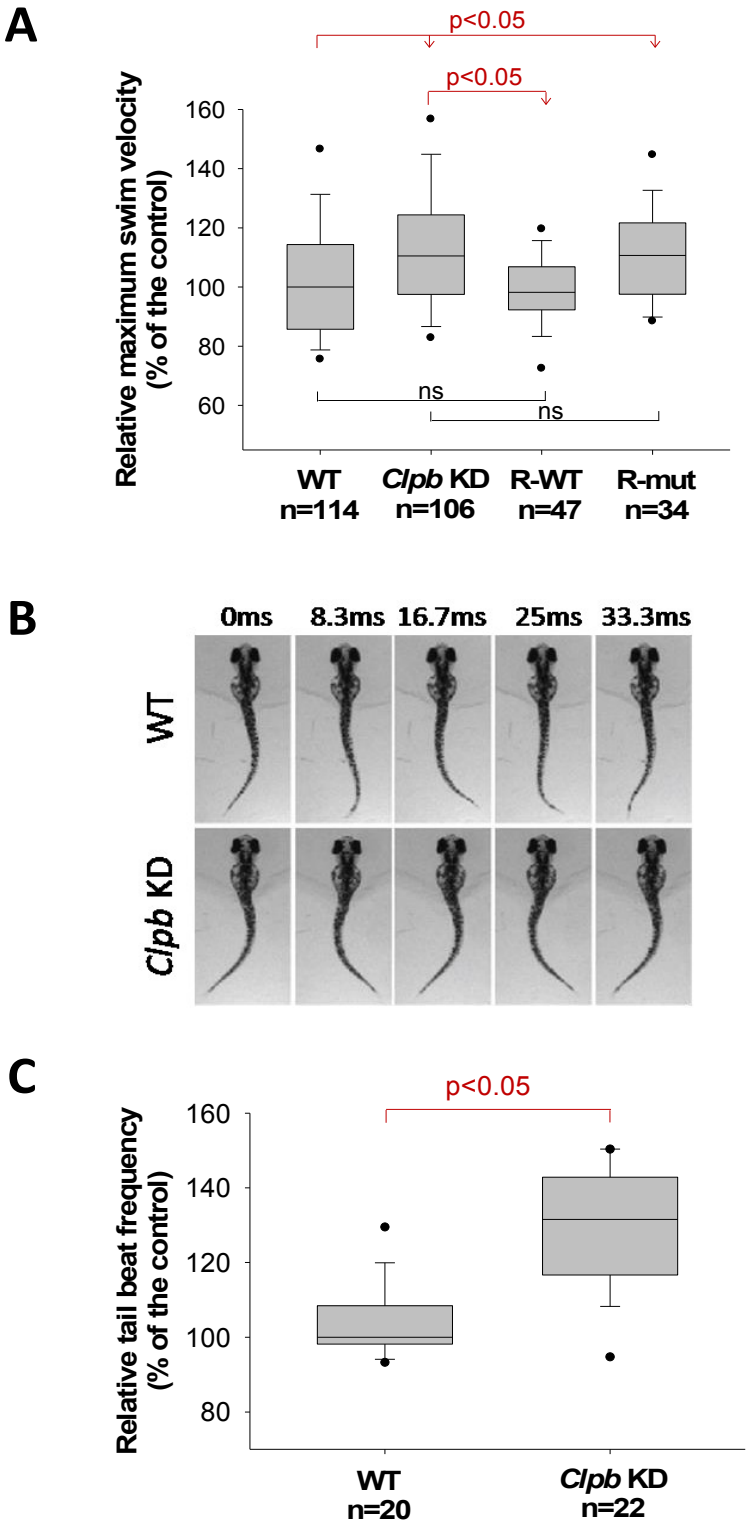

A
p<0.05
p<0.05
Relative maximum swim velocity (% of the control)
160
140
120
100
80
60
ns
ns
WT n=114
Clpb KD n=106
R-WT n=47
R-mut n=34
B
0ms
8.3ms
16.7ms
25ms
33.3ms
WT
Clpb KD
C
p<0.05
Relative tail beat frequency (% of the control)
160
140
120
100
80
WT n=20
Clpb KD n=22

Figure V.3

Figure V.3 c*lpb* KD larvae display an abnormal motor behavior.

(A) Touch-evoked response at 52 hpf. Box plots representing the relative (% of the WT non-injected control) maximum velocity (measured in mm/sec) of zebrafish larvae in the different injection groups. Each box plot contains a central line representing the median, the box itself delineates 25-75% of the data, the error bars represent 5^{th}/95^{th} percentile of the data range, whereas circles represent outlying data points. (B) Video snapshots illustrating tail-beat patterns of WT and *clpb* KD embryos after tactile stimulation at different time frames. (C) Box plot showing the relative (% of the WT non-injected control) tail beat frequency (measured in Hz) of WT and *clpb* KD larvae after mechano-sensory stimulation. WT: Wild-Type. KD: Knockdown. R-WT: Rescue with the WT human *CLPB* mRNA. R-mut: Rescue with the mutated human *CLPB* mRNA. Red arrows: significant difference ($p<0.05$). ns: no significant difference.

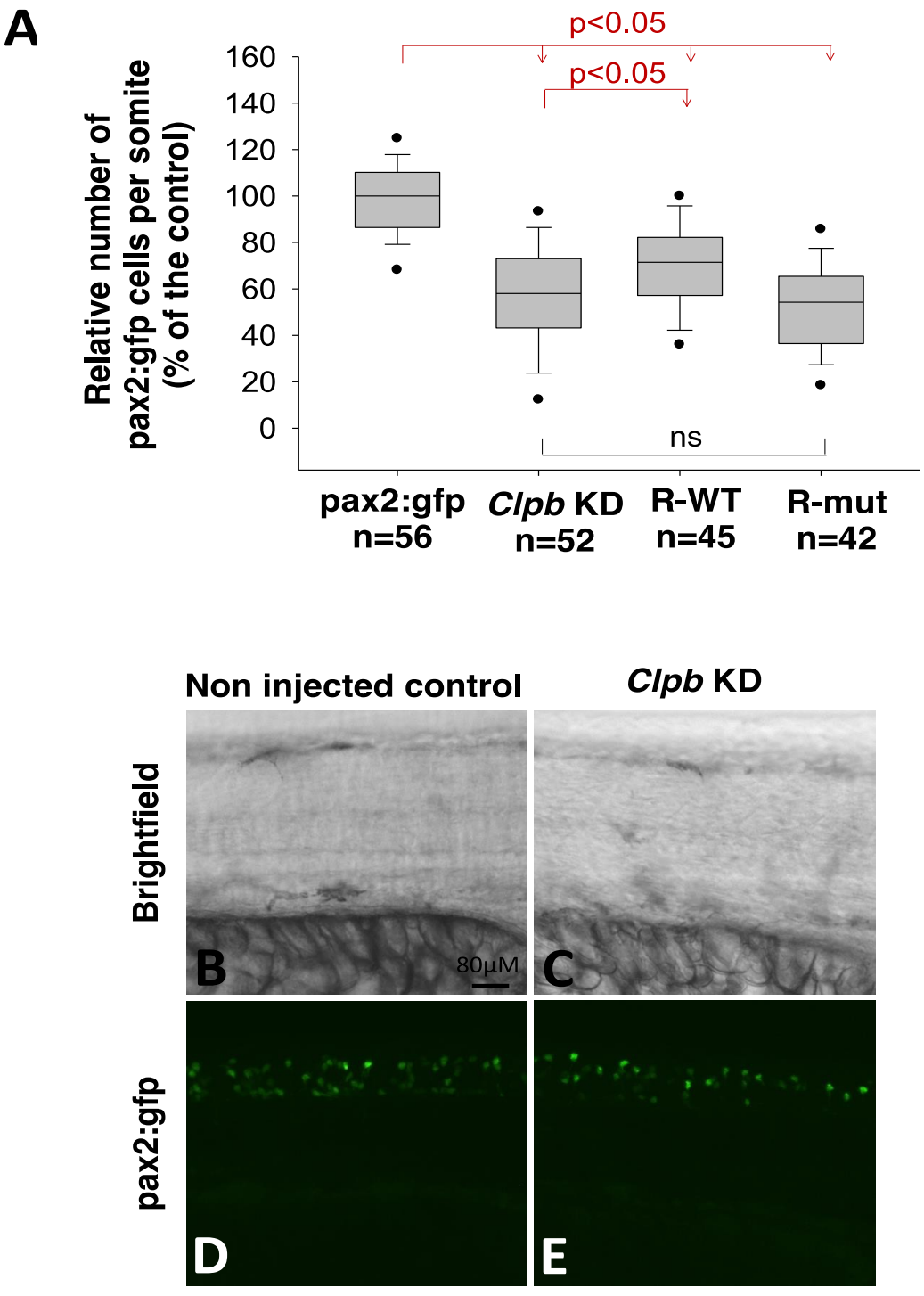

A
p<0.05
p<0.05
Relative number of pax2:gfp cells per somite (% of the control)
160
140
120
100
80
60
40
20
0
ns
pax2:gfp n=56
Clpb KD n=52
R-WT n=45
R-mut n=42
Non injected control
Clpb KD
Brightfield
pax2:gfp
B
C
D
E
80µM

Figure V.4

Figure V.4 Knockdown of *clpb* reduces the population of glycinergic interneurons in pax2:gfp zebrafish transgenic line.
(A) Box plots representing the relative (% of the non-injected pax2: gfp control) number of pax:gfp positive interneurons per somite in the different injection groups. Each box plot contains a central line representing the median, the box itself delineates 25-75% of the data, the error bars represent 5^{th}/95^{th} percentile of the data range, whereas circles represent outlying data points. (B-E) Representative bright-field and GFP confocal images of the spinal cord of non-injected (B and D) and *Clpb KD* (C and E) pax2:gfp transgenic larvae. WT: Wild-Type. KD: Knockdown. R-WT: Rescue with the WT human *CLPB* mRNA. R-mut: Rescue with the mutated human *CLPB* mRNA. Red arrows: significant difference ($p<0.05$). ns: no significant difference.

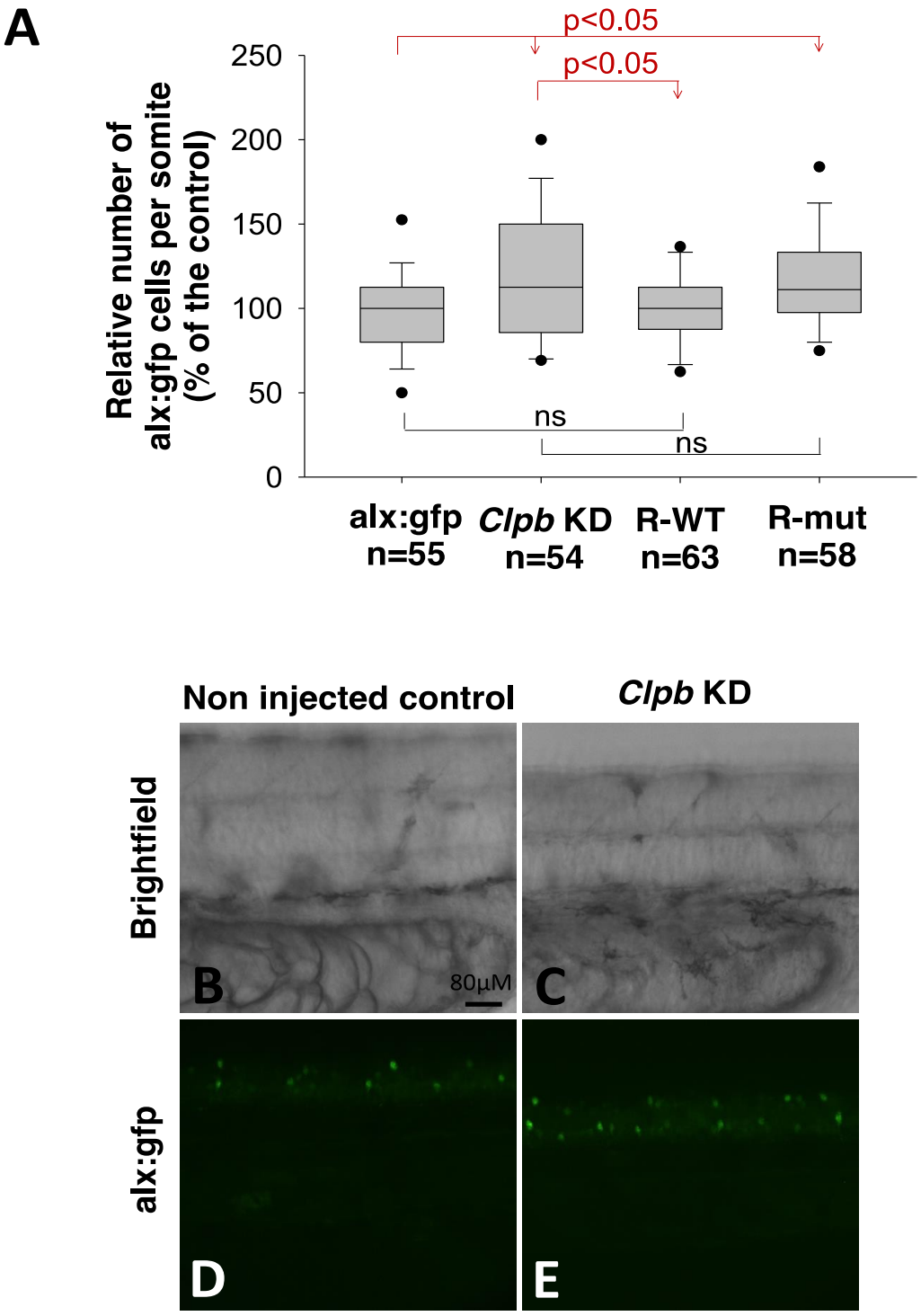

A
p<0.05
p<0.05
Relative number of alx:gfp cells per somite (% of the control)
250
200
150
100
50
0
ns
ns
alx:gfp n=55
Clpb KD n=54
R-WT n=63
R-mut n=58
Non injected control
Clpb KD
Brightfield
alx:gfp
B
C
D
E
80µM

Figure V.5

Figure V.5 Knockdown of *clpb* leads to enlarged population of glutamatergic interneurons in alx:gfp zebrafish transgenic line.
(A) Box plots showing the relative (% of the non-injected alx:gfp control) number of alx:gfp positive interneurons per somite in the different injection groups. Each box plot contains a central line representing the median, the box itself delineates 25-75% of the data, the error bars represent 5^{th}/95^{th} percentile of the data range, whereas circles represent outlying data points. (B-E) Representative bright-field and GFP confocal images of the spinal cord of non injected (B and D) and *Clpb KD* (C and E) alx:gfp transgenic embryos. WT: Wild-Type. KD: Knockdown. R-WT: Rescue with the WT human *CLPB* mRNA. R-mut: Rescue with the mutated human *CLPB* mRNA. Red arrows: significant difference ($p<0.05$). ns: no significant difference.

SUPPLEMENTARY MATERIAL

Clinical description

Sibling II-1.

The pregnancy was characterized by the presence of hydramnios and intra-uterine growth retardation. Sibling II-1 was delivered at 38 weeks of gestation following a vaginal delivery induced because of the presence of signs of fetal distress on monitoring. Apgar score was 1, 4 and 4 at 1, 5 and 10 minutes, respectively. Birth weight was 2090 grams (< 3rd centile) and head circumference was 30 cm (< 3rd centile). No respiratory movements were noted and endotracheal intubation was readily performed. Examination revealed non-reactive pupillae, hypertonia and the remarkable presence of sustained clonic movements of the four limbs induced by tactile stimulation and stopped by physical restraint. EEG showed long period with almost no brain activity interspersed by burst of slow waves and spikes. Echocardiography and abdominal ultrasound were normal.

Blood glucose level was initially 6.9 mmol\L, increased to 29.9 mmol/L and progressively decreased with the administration of insulin. Plasma lactate level was initially 16.35 mmol/L but normalized progressively over the 72 hours. Capillary gas measurements were initially consistent with respiratory acidosis. Urine organic acids showed increased levels of lactate (15345 umol/mmol creatine; ref values: mean: 303 +/- 3 SD), 3-methyl-glutaconic acid (204 umol/mmol creatine; ref values: mean: 56 +/- 3 SD) and methyl-glutaric acid (48 umol/mmol creatine; ref values: mean: 3 +/- 3 SD). Plasma ammonia and aminoacids levels were normal. Blood cell counts showed neutropenia with values oscillating between 2.4 and 3.6 x 10^9 cells /L (normal values: 5.0-18.0 x 10^9 cells/L). Coagulation appeared normal with an International Nationalized Ratio (INR) at 1.17, Prothrombin Time (PT) at 13.1 seconds and fibrinogen levels at 1.65 (normal values: 2.0-4.1 g/L). Cytogenetic studies revealed a normal female karyotype (500 bands).

Sibling II-1 never opened her eyes and did not move or breathe spontaneously. After 5 days of hospitalization, she was extubated and subsequently died.

Sibling II-2.

Pregnancy was characterized by the occurrence of hydramnios during the third trimester. Sibling II-2 was spontaneously delivered at 39 weeks of gestation. Cord blood pH was 7.254 mmHg. Birth weight was 2500 grams (3rd centile) and head circumference was 30.5 cm cm (< 3rd centile). No respiratory movements were noted at birth and the child was readily intubated. Examination showed absence of the photomotor reflex and hypertonia without spontaneous movements but sustained clonic movements of the limbs induced by tactile stimulation and stopped by physical restraint. EEG showed a burst suppression pattern. Brain MRI showed some gyral simplification.

Blood glucose level was initially 10.5 mmol/L and subsequently varied between 4.6 and 15.7 mmol/L. Plasma lactate levels oscillated between 2.39 and 7.51 mmol/L during the hospitalization, with lactate/pyruvate ratios between 18 and 43. Plasma ammonia and aminoacids levels were normal. Urine organic acids showed increased levels of lactate (2749 umol/mmol creatine; ref values: mean: 303 +/- 3 SD), 3-methyl-glutaconic acid (552

umol/mmol creatine; ref values: mean: 56 +/- 3 SD) and methyl-glutaric acid (150 umol/mmol creatine; ref values: mean: 3 +/- 3 SD) and 3-hydroxy-3 methylglutaric acid (396 umol/mmol creatine; ref values: mean: 249 +/- 3 SD).

CSF glucose was 5.7 mmol/L and CSF lactate was 5.26 mmol/L. Concentrations of CSF neurotransmitter metabolites were within reference ranges: 5-hydroxyindoleacetic acid (837 nM; reference values: 208-1159 nM), homovanillic acid (1346 nM; reference values: 337-1299 nM) and 3-O-methyldopa (176 nM; reference values: < 300 nM). In contrast, concentration of neopterin in CSF was greatly elevated (160 nM; reference values: 7-65 nM). Neopterin is a non-specific marker for immune system activation.

Respiratory chain studies performed in the fibroblasts showed normal activity levels (expressed as nmoles/min/mg protein) of pyruvate dehydrogenase (native form: 0.98 +/- 0.27, n=2 versus 1.51 +/- 0.19, n=2, in control; dichloroacetate-activated form: 1.51 +/- 0.19, n=2 versus 1.37 +/- 0.67, n=2, in control), cytochrome c oxydase (7.41 +/- 1.74, n=2 versus 9.19 +/- 0.18, n=2, in control) and succinate cytochrome c reductase (6.76 +/- 0.09, n=2 versus 6.48 +/- 0.48, n=2 in control). Lactate/pyruvate ratios in the fibroblasts of the patient (11.9 +/- 0.7, n=4) and of a control (18.8 +/- 3.6, n=4) were comparable.

Respiratory chain studies in the liver showed normal activity levels (expressed as umoles/g (wet wt) /min) of cytochrome c oxydase (0.7; normal values: 0.6-2.4), succinate cytochrome c reductase (2.46; normal values: 0.6-2.8) and citrate synthase (4.2; normal values: 0.6-5.0) but undetectable levels of NADH-cytochrome c reductase (rotenone sensitive), which could represent an artifact or be indicative of a severe deficiency of complex I.

Blood cell counts showed neutropenia with values oscillating between 0.7 and 4.9 x 109 cells /L (normal values: 5.0-18.0 x 109 cells/L). Coagulation was perturbed with elevated INR (1.94; normal values: 0.93-1.15), PT (21.8 seconds; normal values: 12-14 seconds), Activated Partial Thromboplastin Time (APTT) (54.2 seconds; normal values: 32-42 seconds) and D-dimer levels (1.0 ug/ml; normal values: < 0.25). Fibrinogen (1.59 g/L; normal values: 2.0-4.1 g/L) and anti-thrombin III (0.48 U/ml; normal values: 0.70-1.10 U/L) levels were low.

The patient remained comatose. She was extubated after nine days of life and died over the following hours.

Sibling II-3.

During pregnancy, obstetrical ultrasound showed hyperechogenic bowels and a thickened placenta with some calcifications. Cesarean section was performed at 37 2/7 weeks of gestation because of an abnormal non-stress test. Apgar scores were 3, 5 and 5 at 1, 5 and 10 minutes, respectively. Birth weight was 1840 grams (< 3rd centile) and head circumference was 30 cm (< 3rd centile). Sibling II-3 did not open her eyes upon stimulation, did not move spontaneously and did not react to pain. There was no suction reflex and the photomotor reflex was slow. Appendicular tonus was greatly increased with prolonged clonic movements of the limbs induced by tactile stimulation and stopped by physical restraint. Ophthalmological examination showed that the optic nerves were excavated. EEG showed bursts of theta (4-6 hertz) activity interspersed between periods of low activity. Brain MRI was unremarkable.

Blood glucose levels were initially 9.7 mmol/L and subsequently normalized. Plasma lactate levels oscillated between 2.63 and 6.53 mmol/L with lactate/pyruvate ratios between

14.7 and 35.6. CSF glucose was 3.3 mmol/L and CSF lactate was 3.19 mmol/L. Creatine kinase levels were increased at 1510 U/L (normal values: 15-200).

Respiratory chain studies performed in the fibroblasts showed normal activity levels (expressed as nmoles/min/mg protein) of pyruvate dehydrogenase (native form: 0.92 +/- 0.23, n=4 versus 1.22 +/- 0.32, n=4 in control; dichloroacetate-activated form: 1.68 +/- 0.25, n=4 versus 1.28 +/- 0.28, n=4 in control), pyruvate carboxylase (0.95 versus 0.29 in control), cytochrome c oxydase (4.99 +/- 1.21, n=5 versus 5.80 +/- 0.38, n=5, in control) and succinate cytochrome c reductase (7.00 +/- 0.57, n=5 versus 8.94 +/- 0.62, n=5 in control). Lactate/pyruvate ratios in the fibroblasts of the patient (13.6 +/- 2.8, n=3) and of a control (13.2 +/- 3.7, n=3) were comparable.

Respiratory chain studies in the liver showed normal activity levels (expressed as umoles/g (wet wt) /min) of cytochrome c oxydase (0.99; normal values: 0.6-2.4), succinate cytochrome c reductase (3.71; normal values: 0.6-2.8) and citrate synthase (2.45; normal values: 0.6-5.0) but undetectable levels of NADH-cytochrome c reductase (rotenone sensitive).

Blood cell counts showed neutropenia (0.1 x 109 cells/L; normal values: 5.0-18.0 x 109 cells/L). Coagulation studies showed: normal INR (1.10; normal values: 0.93-1.15), normal PT (13.7 seconds; normal values: 12-14 seconds), prolonged APTT (100.8 seconds; normal values: 32-42 seconds) and elevated D-dimers levels (1.28 ug/ml; normal values: < 0.25). Fibrinogen (2.69 g/L; normal values: 2.0-4.1 g/L) levels were low whereas anti-thrombin III levels were normal (0.76 U/ml; normal values: 0.70-1.10 U/L).

The patient remained comatose. She was extubated after four days of life and died over the following hours.

Sibling II-4.

The pregnancy was unremarkable, except for the presence of a thickened placenta on obstetrical ultrasound. Sibling II-4 was spontaneously delivered at 39 2/7 weeks of gestation. Cord blood pH was 7.20 mmHg. Apgar scores were 1, 2 and 2 at 1, 5 and 10 minutes, respectively. Birth weight was 2770 grams (30th centile) and head circumference was 30 cm (< 3rd centile). Because of a lack of respiratory movements, the newborn was intubated in the delivery room. Sibling II-4 did not open her eyes upon stimulation, did not move spontaneously and did not react to pain. Appendicular tonus was greatly increased with prolonged clonic movements of the limbs induced by tactile stimulation and stopped by physical restraint. EEG showed marked depression of cerebral activity. Blood glucose and lactate levels were 13.4 mmol/L and 3.33 mmol/L, respectively. Coagulation was perturbed with elevated INR (4.70; normal values: 0.93-1.15), PT (44.7 seconds; normal values: 12-14 seconds), and APTT (120.5 seconds; normal values: 32-42 seconds). Sibling II-4 was extubated and died on the day of her birth

SUPPLEMENTARY MATERIAL AND METHODS

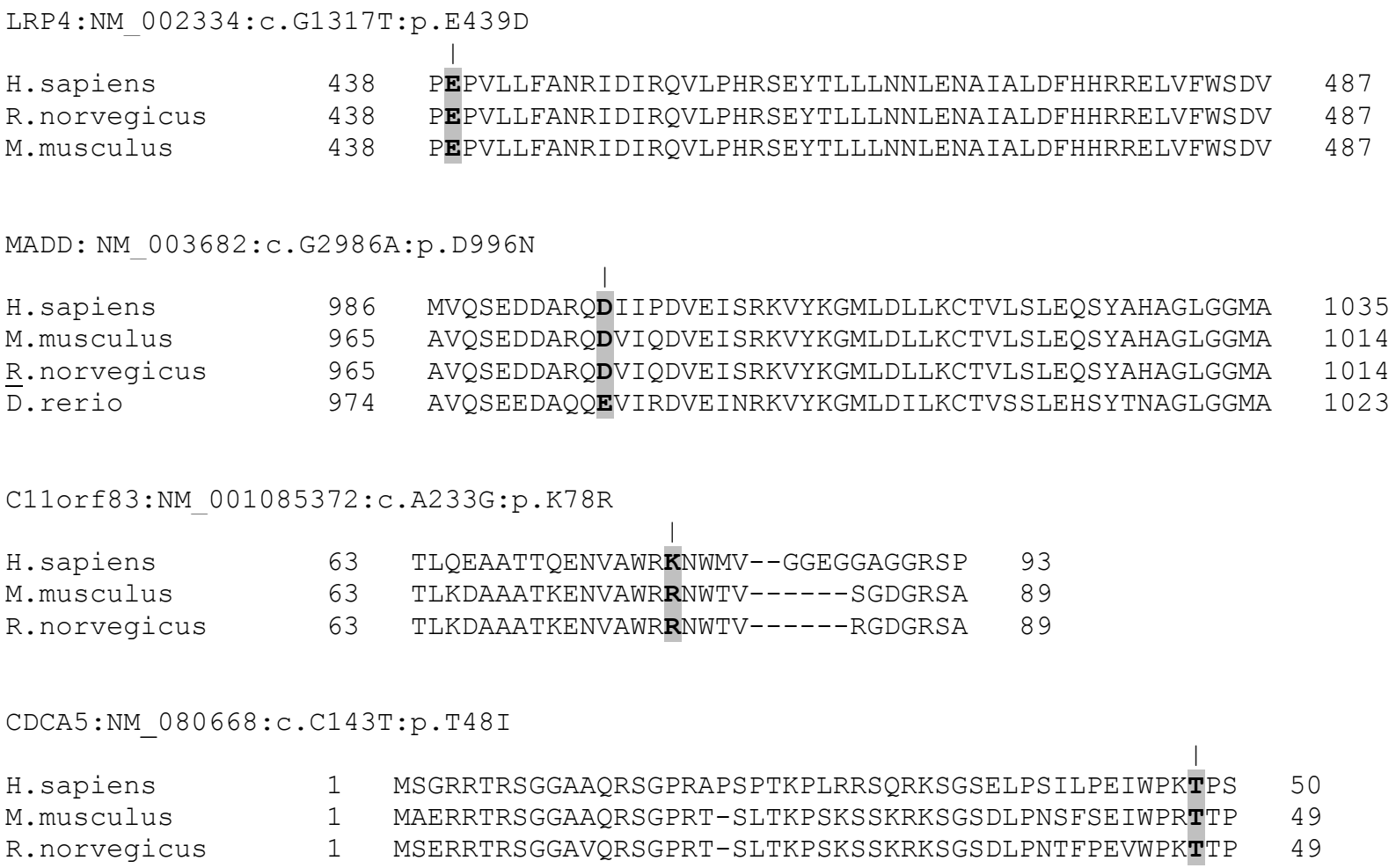
Analysis of mitochondrial function by flow cytometry.

Primary fibroblasts were grown in media supplemented with glucose or galactose respectively. Cells (2 x 10^5) were re-suspended in HBSS (Gibco) and incubated (30 minutes at 37 degrees) with MitoTracker Green (MTG, 100 nM; Invitrogen), to confirm mitochondrial identity, or Tetramethylrhodamine Methyl Ester (TMRM, 100 nM; Invitrogen) to assess mitochondria membrane potential ($\Delta\Psi_m$), or MitoSOX Red (MitoSOX, 5 uM; Invitrogen) to quantify mitochondrial superoxide production. The protonophore carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP, 100 μM; Sigma) was used as a control for $\Delta\Psi_m$ measurements, and the complex III inhibitor, Antimycin A (AA, 100 uM; Sigma) was used as a positive control for mitochondrial superoxide production. 10, 000 events were acquired on a LSR II flow cytometry (BD Biosciences). All data were analyzed with FlowJo (Treestar, Ashland, Oregon). Dyes were selected because of their distinct spectral properties, with minimal to no overlap.

Cell fractioning and protein extraction.

HEK293T and N7 were grown in DMEM medium (Invitrogen) supplemented with fetal bovine serum. Cells were scrapped and lysed into 300 □L of fresh ice-cold RIPA Buffer (25 mM Tris-HCl pH =7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) mixed with protease inhibitors (Roche). Cell lysate was centrifuged at 13000 rpm for 15 min at 4°C and the supernatant was collected as whole-cell extract. Mitochondrial and cytoplasmic proteins were extracted from fresh cell culture according to the manufacturer's recommendations (ZmTech Scientifique). Protein concentration was determined by Bradford assay.

SUPPLEMENTARY FIGURES



EHBP1L1:NM_001099409:c.G1667A:p.G556D

```
                                |
H.sapiens          551   GAISRGLGGWEAEAGGSGDLETETEVVGLEVLGTQEKEVEGSGFPE--TR   598
M.musculus         540   GVKSRDQRAQEAEVGESRVLETEAEWVPWEVIGTSKTDAGIPESLDTEAG   589
R.norvegicus       540   GVVSRDQKAQEAEVGEPRVLETEAEWVPWEVIGTPETNAGSLESPDIETG   589
```

TPCN2:NM_139075:c.G889A:p.V297M

```
                                          |
H.sapiens          290   -------IFFIV------------FTVIGSLFLMNLLTAIIYSQFRGYLM   320
M.musculus         274   -------LFFIV------------FTLIGSLFLMNLLTAIIYNQFRGYLM   304
R.norvegicus       259   ----------------------------GSLFLMNLLTAIIYNQFRGYLM   280
D.rerio            303   -------IFFIL------------FSVFGTYLLMNLMTAIIYNQFRGYLL   333
```

CLPB:NM_030813:c.1685delT:p.I562fsx23

```
                                     |
H.sapiens          544   VIRPILKAHFRRDEFLGRINEIVYFLPFCHSELIQLVNKELNFWAKRAKQ   593
M.musculus         514   VIRPILKAHFRRDEFLGRINEIVYFLPFCHSELIQLVNKELNFWAKRAKQ   563
R.norvegicus       514   MIRPILKAHFRRDEFLGRINEIVYFLPFCHSELIQLVNKELNFWAKRAKQ   563
D.rerio            255   VIRPILKAHFRRDEFLGRINEIVYFLPFCHSELLQLVSRELHYWAKKAKQ   304
```

ARHGEF17:NM_014786:c.C2252T:p.T751I

```
                                  |
H.sapiens          742   HRAATSEEPTGFSVDSNLLGSLSPKTGLPATSAMDEGLTSGHSDWSVGSE   791
M.musculus         735   HRAATSEEPSGFSVDSNLLGSLNSKTGLPVTPAMDEGLTSGHSDWSVVSE   784
R.norvegicus       736   HRVASSEEPSGFSVDSNLLGSLNSKTGLPVTPTMDEGLTSGHSDWSVVSE   785
```

```
MYO7A:NM_001127180:exon32:c.G4309A:p.A1437T
                                        |
H.sapiens          1427   TLEKWAQLAIAAHKKGIYAQRRTDAQKVKEDVVSYARFKWPLLFSRFYEA   1476
M.musculus         1427   NLEKWAQLAIAAHKKGIYAQRRTDSQKVKEDVVNYARFKWPLLFSRFYEA   1476
R.norvegicus       1427   NLEKWAQLAIAAHKKGIYAQRRTDAQKVKQDVVNYARFKWPLLFSRFYEA   1476
D.rerio            1426   TVERWAQFIMAAHKKGVYTQKRVDPQKVKEEVVDFARYKWPLLFSRFYEA   1475
```

Figure SV.1 HomoloGene alignment of residues affected by the rare mutations found in individual-II.1's exome.

Alignments of protein segments around the mutations were generated when reference sequences of their homologs in mouse (M.musculus), rat (R.norvegicus) and zebrafish (D.rerio) were available. In bold and grey highlight are residues affected by the mutations. HomoloGene: http://www.ncbi.nlm.nih.gov/homologene /

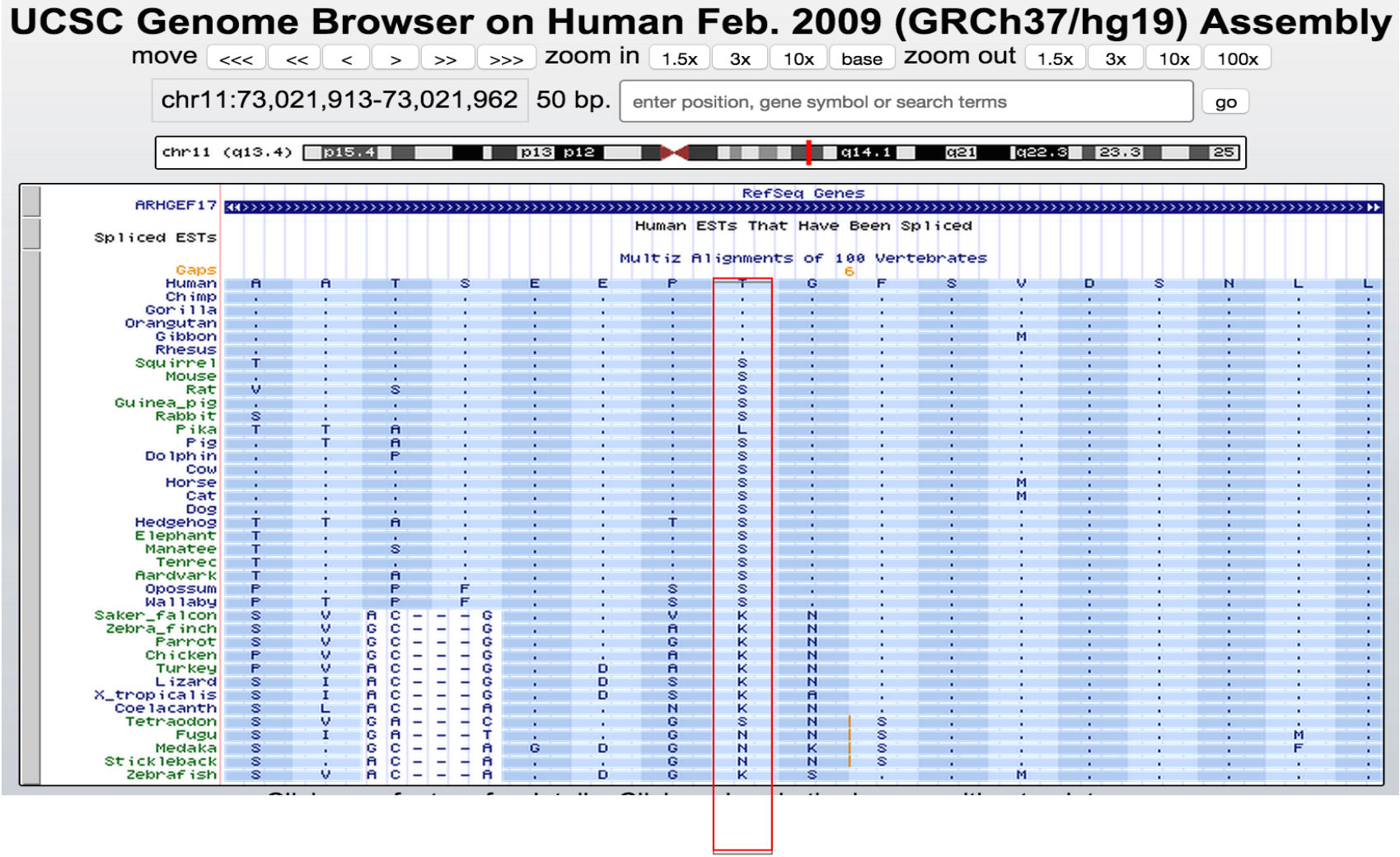
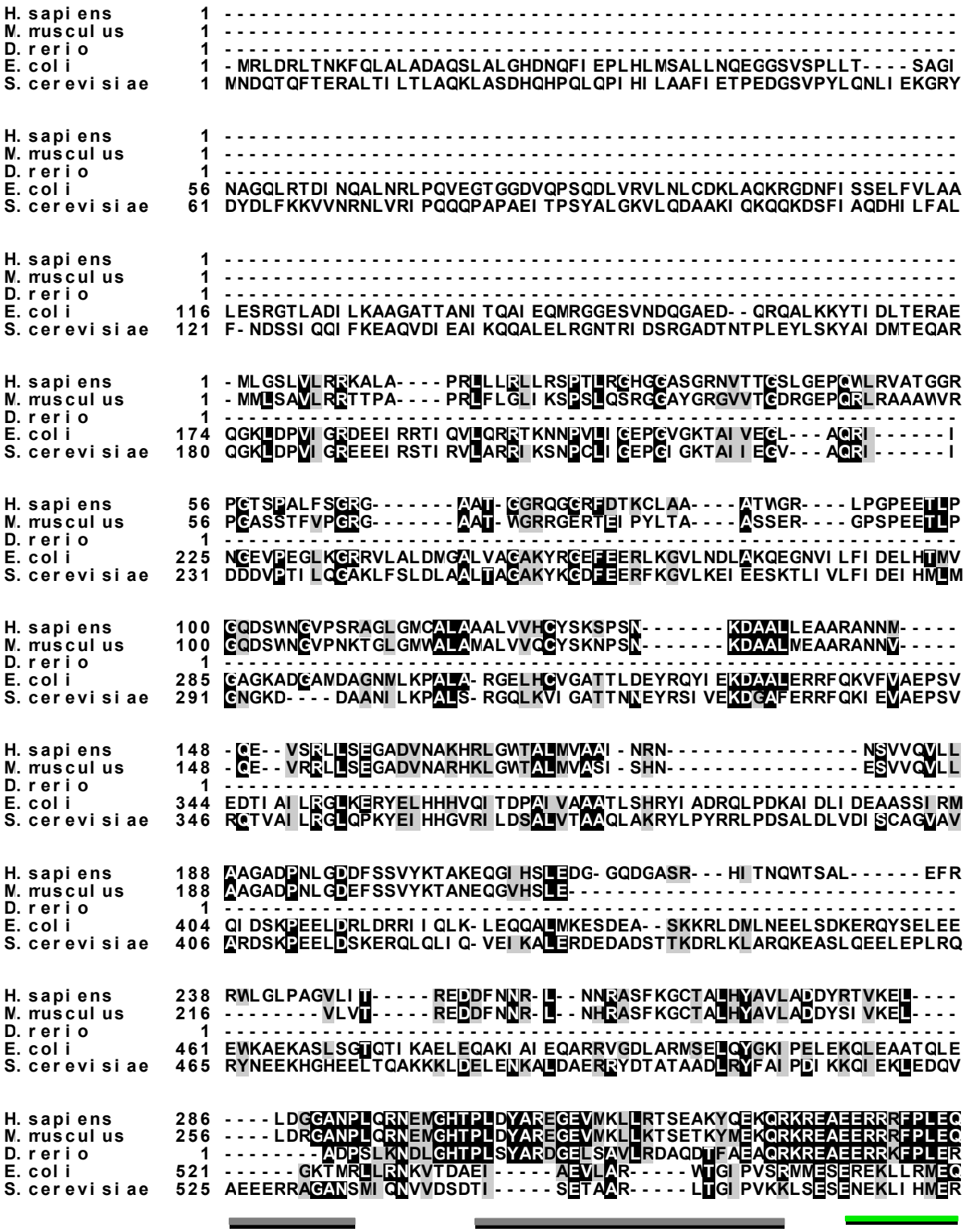


Figure SV.2 Multiple alignment T751 amino acid residue in *ARGHEF17*.
Snapshot image of UCSC Genome Browser showing multiple alignment of a segment of ARHGEF17 in vertebrates. Conservation of T751 residue is indicated in red box. Dots indicate that the residue is conserved in the corresponding vertebrate. UCSC Genome Browser: http://genome.ucsc.edu/


H. sapiens 1
M. musculus 1
D. rerio 1
E. coli 1 - MRLDRLTNKFQLALADAQSLALGHDNQFIEPLHLMSALLNQEGGSVSPLLT - - - - SAGI
S. cerevisiae 1 MNDQTQFTERALTILTLAQKLASDHQHPQLQPIHILAAFIETPEDGSVPYLQNLIEKGRY
H. sapiens 1
M. musculus 1
D. rerio 1
E. coli 56 NAGQLRTDINQALNRLPQVEGTGGDVQPSQDLVRVLNLCDKLAQKRGDNFISSELFVLAA
S. cerevisiae 61 DYDLFKKVVNRNLVRIPQQQPAPAEITPSYALGKVLQDAAKIQKQQKDSFIAQDHILFAL
H. sapiens 1
M. musculus 1
D. rerio 1
E. coli 116 LESRGTLADILKAAGATTANITQAIEQMRGGESVNDQGAED - - QRQALKKYTIDLTERAE
S. cerevisiae 121 F - NDSSIQQIFKEAQVDIEAIKQQALELRGNTRIDSRGADTNTPLEYLSKYAIDMTEQAR
H. sapiens 1 - MLGSLVLRRKALA - - - - PRLLLRLLRSPTLRCHGGASGRNVTTGSLGEPQWLRVATGGR
M. musculus 1 - MMLSAVLRRTTPA - - - - PRLFLGLIKSPSLQSRGCAYGRGVVTGDRGEPQRLRAAAWVR
D. rerio 1
E. coli 174 QGKLDPVIGRDEEIRRTIQVLQRRTKNNPVLIGEPGVGKTAIVEGL - - - AQRI - - - - - - I
S. cerevisiae 180 QGKLDPVIGREEEIRSTIRVLARRIKSNPCLIGEPGIGKTAIIEGV - - - AQRI - - - - - - I
H. sapiens 56 PGTSPALFSGRG - - - - - - - AAT - GGRQGGRFDTKCLAA - - - - ATWGR - - - - LPGPEETLP
M. musculus 56 PGASSTFVPGRG - - - - - - - AAT - WGRRGERTEIPYLTA - - - - ASSER - - - - GPSPEETLP
D. rerio 1
E. coli 225 NGEVPEGLKGRRVLALDMGALVAGAKYRGEFEERLKGVLNDLAKQEGNVILFIDELHTMV
S. cerevisiae 231 DDDVPTILQGAKLFSLDLAALTAGAKYKGDFEERFKGVLKEIEESKTLIVLFIDEIHMLM
H. sapiens 100 GQDSWNGVPSRAGLGMGALAAALVVHGYSKSPSN - - - - - - - KDAALLEAARANNM - - - - -
M. musculus 100 GQDSWNGVPNKTGLGMWALAMALVVQGYSKNPSN - - - - - - - KDAALMEAARANNV - - - - -
D. rerio 1
E. coli 285 GAGKADGAMDAGNMLKPALA - RGELHCVGATTLDEYRQYIEKDAALERRFQKVFVAEPSV
S. cerevisiae 291 GNGKD - - - - DAANILKPALS - RGQLKVIGATTNNEYRSIVEKDGAFERRFQKIEVAEPSV
H. sapiens 148 - QE - - VSRLLSEGADVNAKHRLGWTALMVAAI - NRN - - - - - - - - - - - - - - - NSVVQVLL
M. musculus 148 - QE - - VRRLLSEGADVNARHKLGWTALMVASI - SHN - - - - - - - - - - - - - - - ESVVQVLL
D. rerio 1
E. coli 344 EDTIAILRGLKERYELHHHVQITDPAIVAAATLSHRYIADRQLPDKAIDLIDEAASSIRM
S. cerevisiae 346 RQTVAILRGLQPKYEIHHGVRILDSALVTAAQLAKRYLPYRRLPDSALDLVDISCAGVAV
H. sapiens 188 AAGADPNLGDDFSSVYKTAKEQGIHSLEDG - GQDGASR - - - HITNQWTSAL - - - - - - EFR
M. musculus 188 AAGADPNLGDEFSSVYKTANEQGVHSLE -
D. rerio 1
E. coli 404 QIDSKPEELDRLDRRIIQLK - LEQQALMKESDEA - - SKKRLDMLNEELSDKERQYSELEE
S. cerevisiae 406 ARDSKPEELDSKERQLQLIQ - VEIKALERDEDADSTTKDRLKLARQKEASLQEELEPLRQ
H. sapiens 238 RWLGLPAGVLIT - - - - - REDDFNNR - L - - NNRASFKGCTALHYAVLADDYRTVKELL - - - -
M. musculus 216 - - - - - - - - VLVT - - - - - REDDFNNR - L - - NHRASFKGCTALHYAVLADDYSIVKELL - - - -
D. rerio 1
E. coli 461 EWKAEKASLSGTQTIKAELEQAKIAIEQARRVGDLARMSELQYGKIPELEKQLEAATQLE
S. cerevisiae 465 RYNEEKHGHEELTQAKKKLDELENKALDAERRYDTATAADLRYFAIPDIKKQIEKLEDQV
H. sapiens 286 - - - - LDGGANPLQRNEMGHTPLDYAREGEVMKLLRTSEAKYQEKQRKREAEERRRFPLEQ
M. musculus 256 - - - - LDRGANPLQRNEMGHTPLDYAREGEVMKLLKTSETKYMEKQRKREAEERRRFPLEQ
D. rerio 1 - - - - - - - - ADPSLKNDLGHTPLSYARDGELSAVLRDAQDTFAEAQRKREAEERRKFPLER
E. coli 521 - - - - - - GKTMRLLRNKVTDAEI - - - - - AEVLAR - - - - - WTGIPVSRMMESEREKLLRMEQ
S. cerevisiae 525 AEEERRAGANSMIQNVVDSDTI - - - - - SETAAR - - - - - LTGIPVKKLSESENEKLIHMER

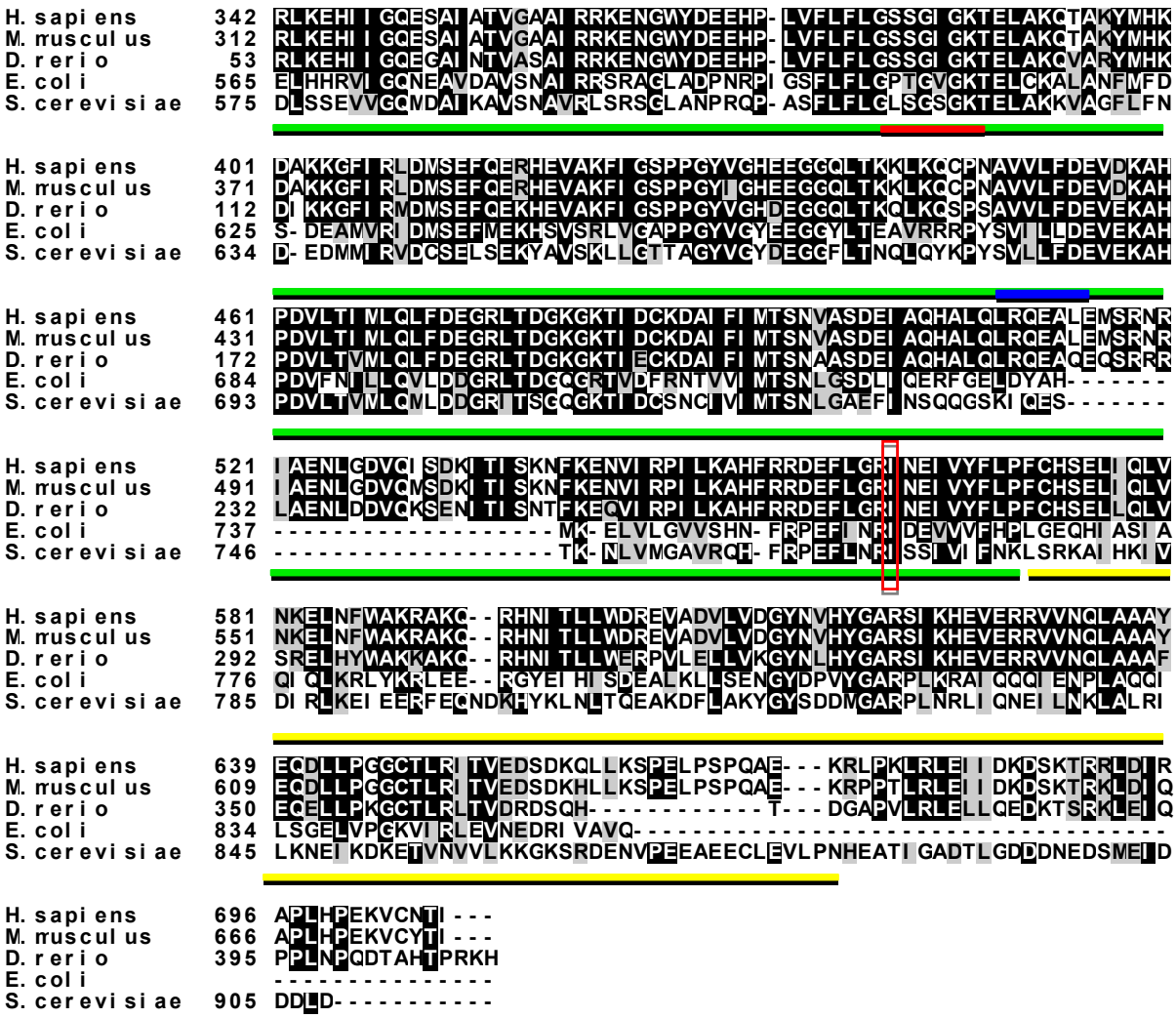


Figure SV.3 Multiple alignment of CLPB.
Functional domains of CLPB are identified based on multiple alignments with Escherichia Coli clpB as annotated from Uniprot (P63284). Note that CLPB possess a single Nucleotide binding domain composed of D2-Large (green bar) and D2-Small domains (yellow bars). Walker type-A and B motifs are represented by red and black bars respectively. p.562Tfs*23 truncation is boxed in red. Unlike Hsp104 and Clpb , CLPB possess ankyrins repeats in its N-terminal end (grey bars).
Uniprot : http://www.uniprot.org.

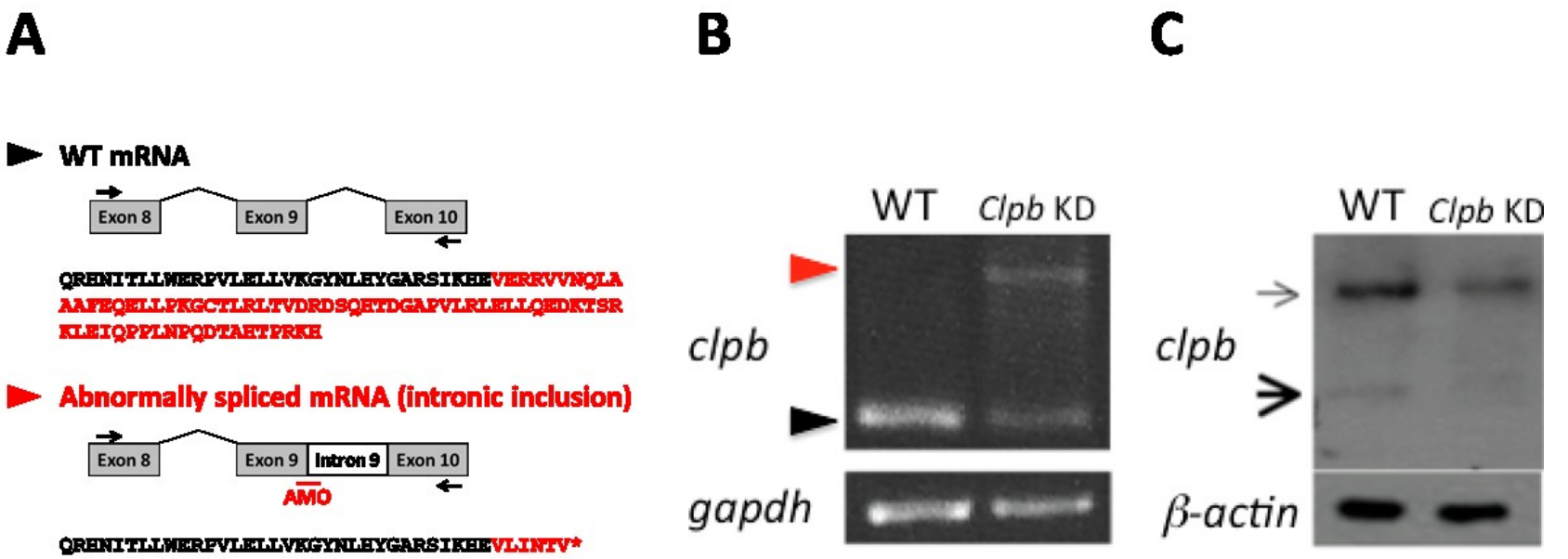


Figure SV.4 *clpb* knockdown in zebrafish using splice-site blocking antisense morpholino oligonucleotides.
(A) Schematic representation of exon 8 to exon 10 of WT and abnormally spliced *clpb* mRNA. Arrows indicate the position of the gene specific primers used in reverse transcriptase PCR (RT-PCR). Red bar indicates the location of the AMO. Amino acids sequences displayed below the cartoons indicate the premature truncation in abnormally spliced mRNA due to intron at splice junction 9. (B) RT-PCR analysis in WT and KD *clpb* zebrafish larvae at 48 hpf. *clpb* KD larvae were injected with 1mM AMO. The black arrowhead indicates normal *clpb* PCR product at 242 bp; the red arrowhead shows the abnormally spliced product. *gapdh* was used as a loading control. (C) Expression of clpb protein in WT and KD zebrafish larvae. Arrowheads indicate main clpb isoforms recognised by the epitope of the antibody used at 57 KDa (strong arrowhead) and 75 KDa (thin arrowhead). AMO: antisense morpholino oligonucleotides, KD: Knockdown, WT: Wild-Type.

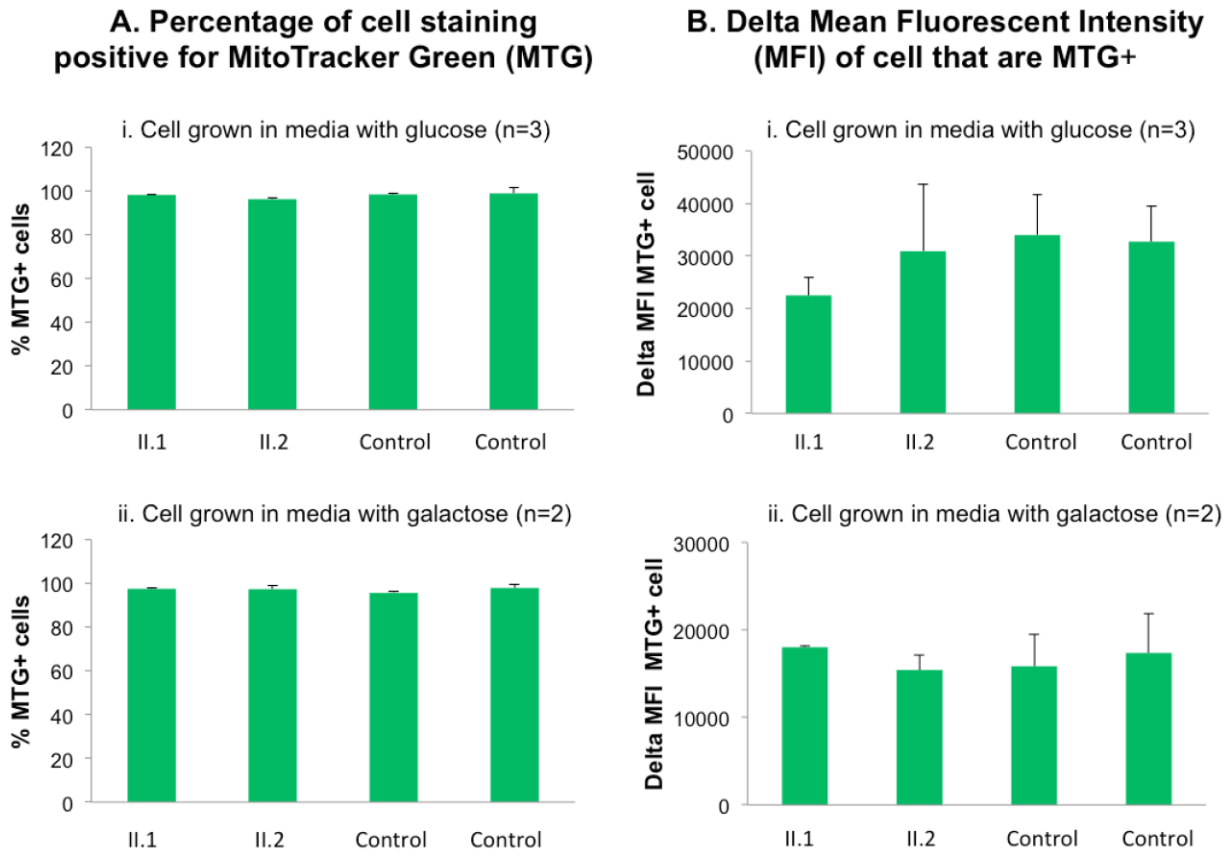


Figure SV.5 Measure of mitochondrial identity and volume in primary fibroblast cells.

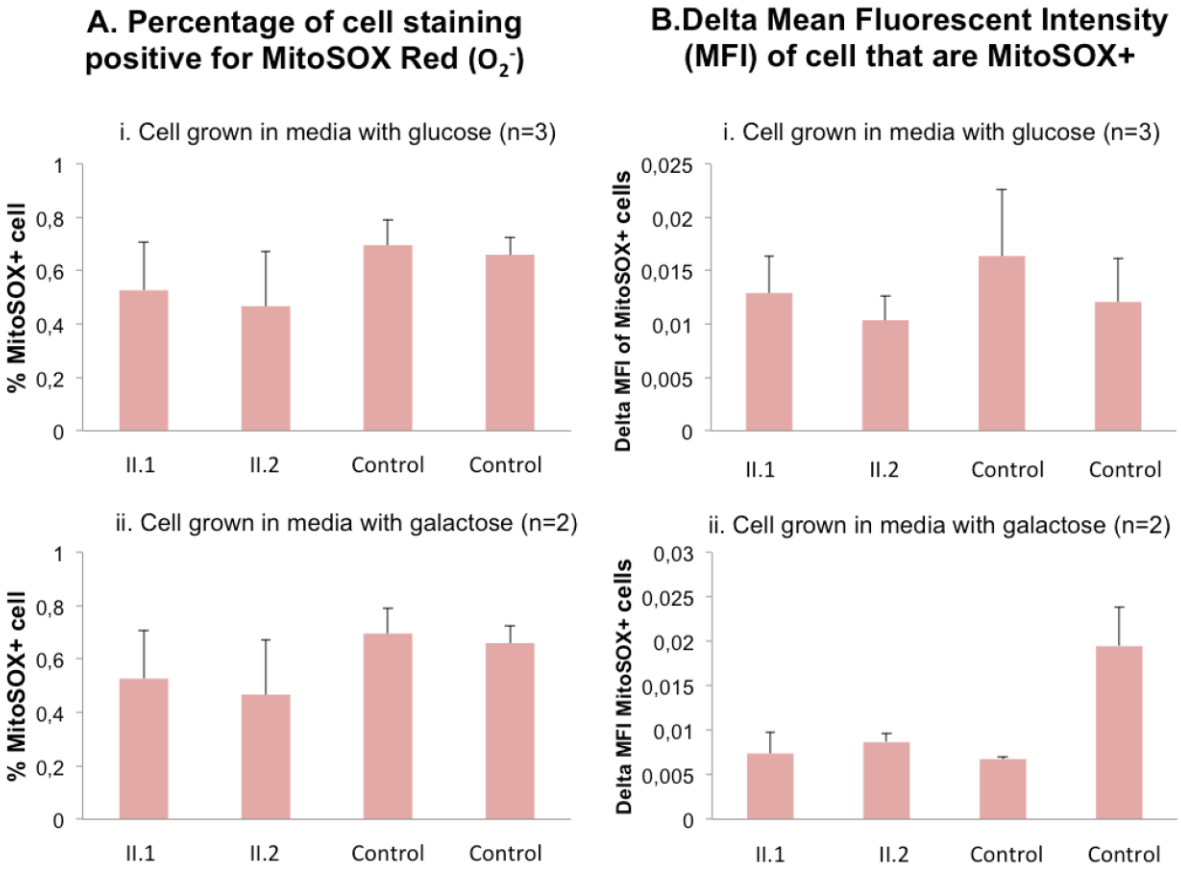


Figure SV.6 Measure of mitochondrial transmembrane potential in primary fibroblast cells.

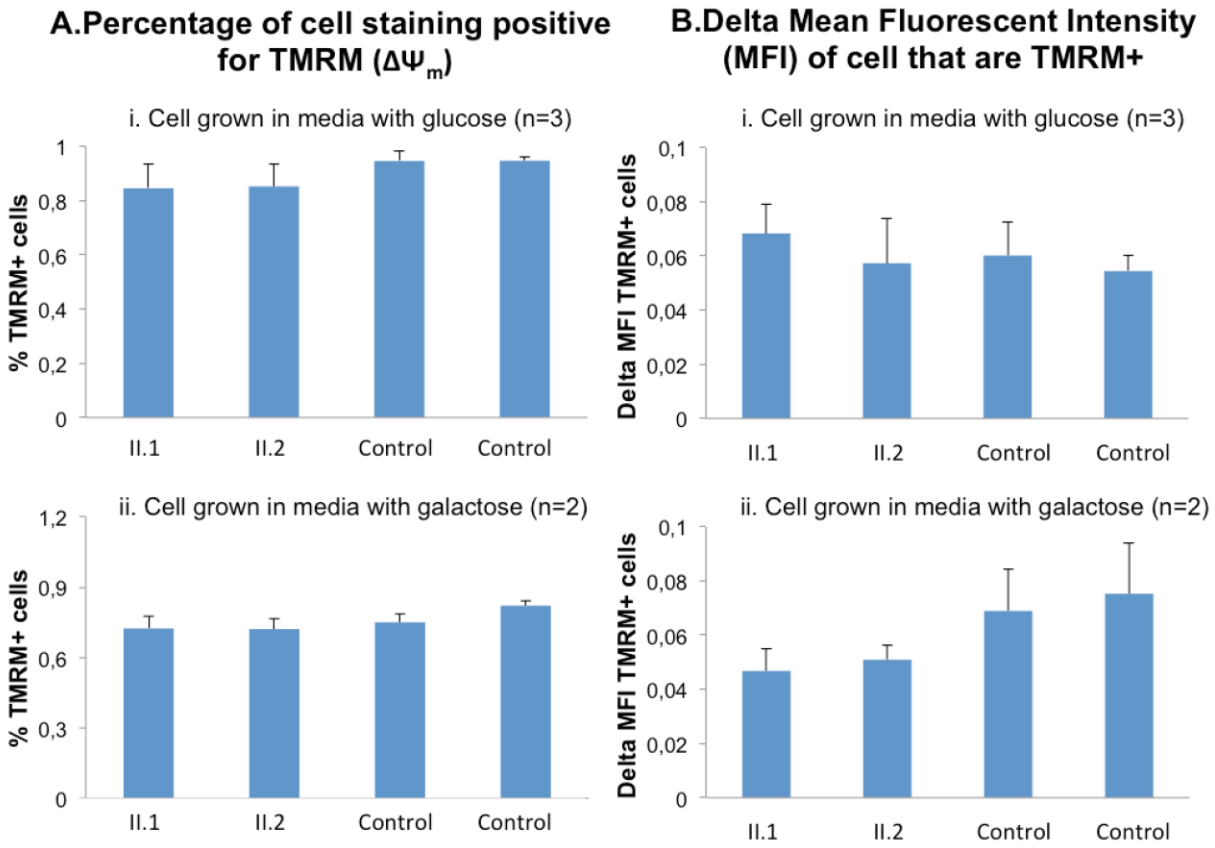


Figure SV.7 Measure of mitochondrial superoxide production in primary fibroblast cells.

SUPPLEMENTARY TABLES

Table SV.1 Regions of homozygosity shared between affected individuals.
Regions of homozygosity > 1Mb and 30 consecutive SNPs. chr, chromosome. Markers start and end delimit the regions of shared homozygosity at their corresponding positions (position start and end). Markers positions are relative to the hg19 reference.

Chr	Marker Start	Marker End	Position Start (bp)	Position End (bp)	Size (Mb)
1	rs10399889	rs7529595	52 010 203	53 076 663	1,1
2	rs4848940	rs1563337	125 342 613	126 715 414	1,4
3	rs2228561	rs1138536	48 628 014	50 153 356	1,5
7	rs6466686	rs10953910	118 636 817	120 017 847	1,4
10	rs2136295	cnvi0004872	132 574 570	133 972 265	1,4
11	rs12788207	rs592928	41 582 337	93 578 963	52,0
13	rs9520170	rs7338116	88 960 710	90 036 166	1,1
15	rs17184382	rs7173216	63 792 486	65 023 156	1,2
17	cnvi0008543	rs792766	49 897 984	51 111 510	1,2
18	rs10513946	rs280986	1 468 409	4 449 084	3,0
22	rs470113	rs28572577	40 729 614	42 522 312	1,8
	rs9616080	rs6009870	47 018 166	50 154 682	3,1

Table SV.2 Prioritization of the variants detected in individual-II.1's exome.

(a) Only variants that were found in these databases at minor allele frequencies < 1.0% were retained.
EVS : Exome Variant Server, http://evs.gs.washington.edu/EVS/ ; 1000 genomes : http://www.1000genomes.org
(b) Four variants predicted to be benign/tolerated or neutral by all four algorithms tested (Polyphen-2, Mutation Taster and Provean) were filtered out (See Table SV.3). Three other variants were also excluded from further consideration, since the corresponding genes are associated with non-relevant clinical manifestations or phenotypes in Humans and mouse knockout models (see Table SV.4).
(c) One variant (SPDYC: NM_001008778:exon7:c.C860A:p.P287H) was found at MAF 10% in 90 controls of South-Eastern Asiatic origin genotyped.
(d) Sanger sequencing in both directions.

	Variant filters applied sequentially	**Variant count**
	Non-synonymous, nonsense, splicing, and coding/splicing Indels	11 471
1-	Homozygous	4 863
2-	In all the homozygosity regions	229
3-	Not in 657 in-house control exomes	18
4-	MAF<0,01 in EVS, 1000 Genomes [(a)]	10
5-	Mutations in candidate genes [(b)]	3
6-	MAF<0.01 in 90 ethnically matched controls [(c)]	2
7-	Confirmed homozygous in the 4 affected siblings and heterozygous in the parents [(d)]	2 (*CLPB* p.562fs*23 *ARHGEF17* p.T75I)

Table SV.3 Rare homozygous mutations found in individual II-1's exome.

Variants listed here are found in all the homozygosity regions shared by the affected individuals of this family (Step 4 of Table SV.2) . When possible, predictions of the pathogenicity of these variants are given according to four differents algorithms.
Polyphen-2 : http://genetics.bwh.harvard.edu/pph2/;
SIFT/Provean : http://provean.jcvi.org/index.php ;
Mutation Taster : http://www.mutationtaster.org/

	Gene	Variant	Polyphen-2	SIFT	Provean	Mutation Taster
1	*LRP4*	p.E439D	BENIGN	TOLERATED	NEUTRAL	DISEASE CAUSING
2	*MADD*	p.D976N	BENIGN	DAMAGING	DELETERIOUS	DISEASE CAUSING
3	*C11orf83*	p.K78R	BENIGN	TOLERATED	NEUTRAL	POLYMORPHISM
4	*CDCA5*	p.T48I	BENIGN	TOLERATED	NEUTRAL	POLYMORPHISM
5	*SPDYC*	p.P287H	DAMAGING	DAMAGING	NEUTRAL	POLYMORPHISM
6	*EHBP1L1*	p.G556D	BENIGN	TOLERATED	NEUTRAL	POLYMORPHISM
7	*TPCN2*	p.V297M	BENIGN	TOLERATED	NEUTRAL	POLYMORPHISM
8	*CLPB*	p.I562fs	--	--	--	DISEASE CAUSING
9	*ARHGEF17*	p.T751I	BENIGN	DAMAGING	NEUTRAL	POLYMORPHISM
10	*MYO7A*	p.A1437T	BENIGN	TOLERATED	DELETERIOUS	DISEASE CAUSING

Table SV.4 Genes identified with non-relevant clinical manifestations or phenotypes in mouse knockout models.

The genes listed here were among the genes remaining at step 4 of the prioritization of the variants detected in individual-II.1's exome (Table SV.2). Information on gene functions were obtained from OMIM and the MGI mouse knockout database. OMIM : http://www.ncbi.nlm.nih.gov/omim/ ; MGI mouse knockout database : http://www.informatics.jax.org/

Gene	*LRP4*	*MADD*	*MYO7A*
OMIM	Cenani-Lenz syndactyly syndrome (CLSS) [MIM:212780] Sclerosteosis 2 (SOST2) [MIM:614305]	no entry	Usher syndrome 1B (USH1B) [MIM:276900], Deafness, autosomal recessive, 2 (DFNB2) [MIM:600060], Deafness, autosomal dominant, 11 (DFNA11) [MIM:601317], also associated to Leber congenital amaurosis (LCA,(Wang et al., 2011))
Mouse knockout database	Malformed digits, brachydactyly,syndactyly	die shortly after birth due to respiratory failure, hypo-responsive to tactile stimulation, depletion of synaptic vesicles at the neuromuscular junction	Head-shaking, circling and deafness

CHAPITRE VI: SÉQUENÇAGE D'EXOME DANS LA DÉFICIENCE INTELLECTUELLE

RESUMÉ

Nous avons mis de l'avant l'utilisation du WES dans l'étude des bases génétiques de la DI dans 15 familles additionnelles ; une transmission de la DI selon un mode autosomique récessif est suspectée dans ces familles (Figure VI.1). Le tableau VI.1 présente une description sommaire des 15 familles étudiées dans le chapitre VI; ces familles ont été analysées en suivant un protocole analogue à celui décrit dans la section *Supplementary Subject and Methods* du chapitre II. Nous avons adopté différentes approches expérimentales (Figure I.7), afin d'identifier les déterminants génétiques de la DI dans ces familles (Tableau VI.2). Dans certaines des familles analysées, nous avons réalisé une cartographie de l'autozygome (Figure I.4) ; les régions d'homozygotie partagées entre les individus affectés de ces familles sont rapportées dans le tableau VI.3.Le rendement diagnostic du WES dans ces dans les 20 familles étudiées est présenté dans le tableau VI.4.

RÉSULTATS

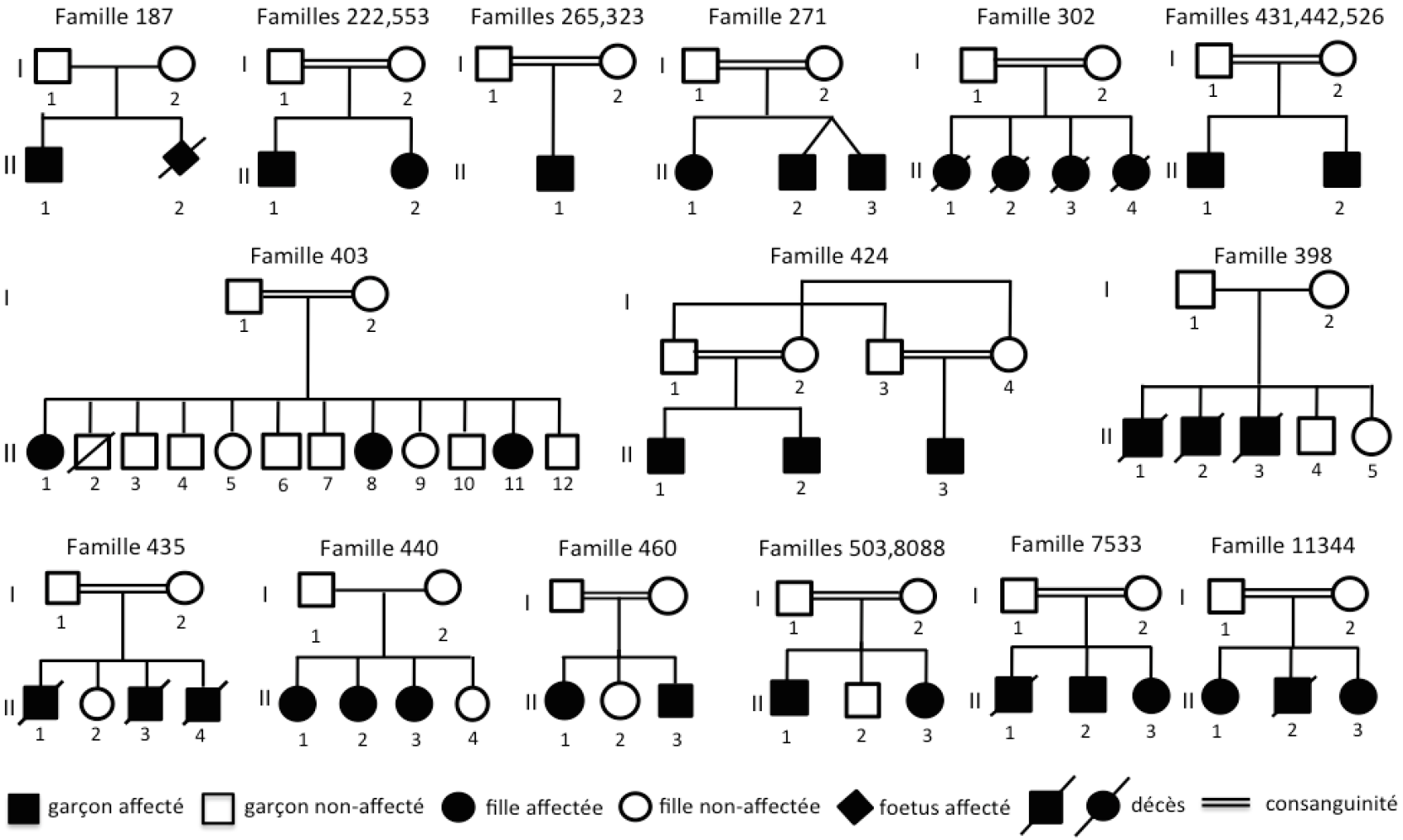


Figure VI.1 Pédigrée des 20 familles analysées dans ce projet de recherche.

Tableau VI.1 Description des 15 familles étudiées dans le chapitre VI.
DI : Déficience intellectuelle, DI-NS: Déficience intellectuelle non-syndromique.

Famille	Origine ethnique	Individus affectés	Consanguinité	Phénotype
187	Canadienne-Française	2	non	Syndrome de Malpuech
222	Nord-africaine	2	oui	DI-NS
265	Pakistanaise	1	oui	DI+microcéphalie
271	Libanaise	3	oui	DI-NS, épilepsie, myoclonies
323	Arménienne	1	oui	DI-NS + tétralogie de Fallot
403	Pakistanaise	3	oui	DI-NS
424	Canadienne-Française	3	oui	Ataxie cérébelleuse + retard dans le développement
431	Turque	2	oui	DI-NS
440	Haïtienne	3	non	DI-NS
442	Libanaise	2	oui	DI-NS + Diabète insipide néphrogénique
503	Libanaise	2	oui	DI-NS
526	Canadienne-Française	2	non	DI-NS
553	Libanaise	2	oui	DI-NS
8088	Libanaise	2	oui	Retard dans le développement + dysplasie squelettique

Tableau VI.2 Approches expérimentales utilisées dans l'exploration des causes génétiques la déficience intellectuelle.

WES: Séquençage d'exome. Les 20 familles étudiées dans le cadre de ce projet de recherche sont inclues dans ce tableau. Les approches expérimentales utilisées (a) à (f) sont illustrées à la Figure I.7; une description détaillée de ces approches est également fournie. Nous avons priorisé l'analyse des mutations récessives homozygotes et hétérozygotes composites, parce qu'une transmission récessive de la maladie est suspectée dans ces familles.

Famille	Individus affectés	Individus séquencés	Approches expérimentales priorisées
187	2	1	approche f
222	2	1	approches b et c
265	2	1	approches b et c
271	3	2	approches b et c
302	4	1	approches b et c
323	1	1	approche f
398	3	3	approche d
403	3	2	approches b et c
424	3	2	approches b et c
431	2	2	approches b et c
435	3	1	approches b et c
440	3	3	approche d
442	2	1	approches b et c
460	2	1	approches b et c
503	2	2	approche d
526	2	2	approche d
553	2	2	approche d
7533	2	1	approches b et c
8088	2	1	approche f
11344	3	2	approches b et c

Tableau VI.3 Tableau récapitulatif des régions d'homozygotie cartographiées dans les familles analysées dans le chapitre VI.

Famille	Chr	Marqueurs		Position (pb)		Taille (Mb)
222	4	rs2636670	rs7696619	143 234 593	144 388 923	1,2
222	7	rs10232307	rs805785	116 369 187	123 219 972	6,9
222	11	rs1399588	rs7936255	37 857 857	38 998 451	1,1
222	13	rs1981058	rs28676337	39 833 403	41 290 779	1,5
222	15	rs11632611	rs11632639	72 095 711	73 349 633	1,3
222	22	rs5757373	rs13056665	39 320 828	49 552 044	10,2
265	1	rs4584419	rs1188002	57 109 660,00	58 484 116,00	1,4
265	1	rs4638118	rs16837628	237 525 382,00	239 156 403,00	1,6
265	1	rs16837628	rs11581448	239 156 403,00	242 009 816,00	2,8
265	2	rs2680837	rs10106770	222 471 516,00	235 832 763,00	13,4
265	2	rs10106770	rs11895915	235 832 763,00	237 032 607,00	1,2
265	4	rs6855169	rs12496398	87 190 830,00	115 156 497,00	28,0
265	4	rs12496398	rs10957824	115 156 497,00	122 211 103,00	7,0
265	4	rs10957824	rs1533236	122 211 103,00	123 564 458,00	1,4
265	4	rs10020796	rs6858578	153 519 781,00	154 682 058,00	1,2
265	5	rs17792757	rs1486849	5 723 396,00	19 493 684,00	13,8
265	5	rs840459	rs73604424	121 424 413,00	128 371 441,00	7,0
265	5	rs73604424	rs318401	128 371 441,00	143 356 921,00	15,0
265	8	rs7003742	rs2569201	97 061 202,00	109 657 233,00	12,6
265	8	rs2569201	rs12804886	109 657 233,00	113 890 126,00	4,2
265	8	rs12804886	rs13251655	113 890 126,00	119 760 559,00	5,8
265	9	rs12684332	rs1902482	132 157 537,00	141 077 352,00	9,0
265	12	rs1157480	rs793167	30 373 004,00	31 493 505,00	1,1
265	14	rs3898026	rs8015187	20 420 338,00	22 053 340,00	1,6
265	16	rs12920545	rs11074377	10 994 803,00	19 280 924,00	8,3
265	20	rs6140378	rs761319	7 654 939,00	11 970 685,00	4,3
271	1	kgp8069155	rs6666410	173 369 724	175 072 789	1,7
271	2	kgp7085901	kgp14773887	89 121 178	97 135 703	8,0
271	3	kgp1503508	kgp1482803	98 879 786	100 180 646	1,3
271	5	kgp11244383	kgp11239904	23 720 508	24 807 491	1,1
271	5	kgp22434617	kgp22700386	129 711 240	131 337 520	1,6
271	9	kgp3375557	kgp10657859	44 881 297	69 881 429	25,0
271	10	rs12358699	rs12766584	45 553 729	47 543 535	2,0
271	11	kgp11023211	kgp8846702	88 762 668	90 073 685	1,3
271	12	kgp10505054	rs11048586	20 646 575	26 726 859	6,1
271	15	kgp9177783	rs8040193	20 005 287	22 755 185	2,7
271	15	kgp19841333	kgp19879389	43 452 023	44 723 819	1,3
271	15	rs7402085	kgp5879026	72 084 834	73 172 584	1,1
271	16	kgp16523446	kgp16479121	31 948 520	34 274 265	2,3
271	17	kgp11813284	rs223151	1 091 895	29 926 222	28,8
271	20	kgp19235734	kgp5480833	42 474 858	50 460 246	8,0

Tableau VI.3 (suite 1)

Famille	Chr	Marqueurs		Position (Hg19)		Taille (Mb)
403	1	rs2753319	rs10753244	25 452 671	31 683 500	6,2
403	1	rs2482850	rs10874016	62 368 811	79 580 281	17,2
403	3	rs344386	rs317531	3 021 663	4 101 356	1,1
403	3	rs11917072	rs501217	165 267 908	166 343 800	1,1
403	4	rs6848548	rs4834232	127 825 712	129 024 273	1,2
403	5	rs10045093	rs17340123	59 808 255	60 854 916	1,0
403	6	rs9355850	rs16890289	158 070 192	159 846 743	1,8
403	10	rs10828165	rs11013558	18 349 535	23 823 654	5,5
403	11	rs956267	rs504847	92 742 186	94 404 315	1,7
403	22	rs17742907	rs8137706	18 890 615	27 773 685	8,9
424	6	rs6459154	rs4288209	56 423 255	57 465 895	1,0
424	11	rs10768552	rs7113549	39 877 303	71 830 217	32,0
431	1	rs7544630	kgp15866056	147 382 550	149 044 448	1,7
431	3	rs4078466	rs2526395	47 126 543	50 188 886	3,1
431	10	kgp12534008	kgp21706383	39 112 575	42 431 158	3,3
431	11	kgp8907413	kgp12902338	30 774 737	32 060 658	1,3
431	14	kgp12190198	kgp775301	102 108 558	103 225 591	1,1
431	14	kgp19546290	kgp19472791	106 002 833	107 073 718	1,1
431	16	kgp7564490	kgp10118995	31 946 713	33 598 707	1,7
442	2	rs28672994	kgp8274669	95 395 757	96 421 021	1,0
442	3	rs11712666	rs13086798	11 619 958	41 626 139	30,0
442	3	kgp6852929	kgp18061951	46 593 892	47 914 989	1,3
442	5	rs17630080	rs315805	132 819 346	169 746 734	36,9
442	6	kgp701001	kgp17146843	115 375 485	116 436 706	1,1
442	7	kgp10275144	kgp2914124	57 883 558	63 306 567	5,4
442	8	rs351572	rs2517168	16 021 468	17 457 420	1,4
442	9	rs10441809	rs7852479	43 581 253	71 034 203	27,5
442	9	kgp18479020	kgp18380372	122 999 873	124 081 045	1,1
442	10	rs3829203	rs7922051	3 868 226	8 555 318	4,7
442	11	rs669976	rs12223637	64 573 589	76 104 519	11,5
442	12	kgp19051204	kgp18855742	83 863 228	85 163 334	1,3
442	15	kgp5643865	kgp19950951	84 194 103	85 384 813	1,2
442	16	kgp7564490	kgp22822345	31 946 713	33 579 417	1,6
442	17	rs2586841	rs8076446	71 815 050	77 723 584	5,9

Tableau VI.3 (suite 2)

Famille	Chr	Marqueurs		Position (Hg19)		Taille (Mb)
11344	1	SNP_A-2189523	SNP_A-4223597	92 130 959	94 351 641	2,2
11344	1	SNP_A-1831013	kgp4540826	99 889 438	101 199 147	1,3
11344	1	kgp22777993	kgp823688	142 603 938	144 936 353	2,3
11344	1	kgp22730900	kgp15866056	147 405 193	149 044 448	1,6
11344	2	SNP_A-4207914	rs1643189	14 012 242	16 926 263	2,9
11344	2	SNP_A-1923396	rs3884626	82 079 119	83 557 502	1,5
11344	3	kgp2400525	kgp17546267	50 586 704	51 851 780	1,3
11344	4	rs9760810	kgp9135534	8 731 330	9 916 662	1,2
11344	4	kgp20832578	kgp2293907	106 926 894	107 957 885	1,0
11344	6	SNP_A-2225560	SNP_A-1837523	63 092 140	65 372 450	2,3
11344	6	kgp2495736	rs7745274	126 386 344	127 509 157	1,1
11344	7	kgp13613108	kgp8018476	61 055 273	62 531 835	1,5
11344	7	kgp5536782	rs4717530	68 907 942	69 935 641	1,0
11344	8	rs10955176	kgp845757	99 467 856	101 346 492	1,9
11344	9	rs7855071	kgp18553579	65 629 772	70 987 152	5,4
11344	10	rs2277205	SNP_A-4233074	114 206 683	115 373 864	1,2
11344	15	rs12595720	rs4775002	56 324 451	58 190 887	1,9
11344	15	kgp19857610	kgp19900845	72 043 372	73 649 171	1,6
11344	16	SNP_A-2299264	rs811330	76 466 593	84 813 708	8,3
11344	16	rs8048908	kgp16430222	31 918 418	34 258 165	2,3
11344	20	SNP_A-1967706	SNP_A-1885489	9 933 041	12 578 804	2,6
11344	21	SNP_A-2017497	SNP_A-2017728	28 730 368	30 117 231	1,4

Les régions d'homozygotie (RH) cartographiées contiennent un minimum de 25 marqueurs génétiques consécutifs ; elles s'étendent sur un minimum d'1 Mb. Sauf indication contraire, les marqueurs indiqués délimitent les régions d'homozygotie partagées entre au moins deux individus affectés (voir Figure VI.1) ; leurs positions chromosomiques correspondantes sont indiquées selon la version hg19 génome humain. Chr : chromosomes ; pb : paires de bases ; Mb : méga paires de bases. Dans les familles 271, 424, les RH présentées ont été obtenues à partir des individus II.1 et II.3.

SYNTHÈSE

Nous avons utilisé le WES, afin d'explorer les bases génétiques de la DI dans 20 familles consanguines ou non-consanguines (Figure VI.1, Tableau VI.1), où une transmission récessive de la maladie est suspectée. Dans l'analyse de ces familles, nous étions confronté à certaines difficultés ; une description détaillée de l'analyse des différentes familles est fournie dans les lignes suivantes.

Famille 187 : Il s'git d'une famille Canadienne-Française non-consanguine (Figure VI.1) avec un garçon présentant un phénotype s'apparentant à celui du Syndrome de Malpuech (3MC ; Base de données Orphanet ; ORPHA2453). La deuxième grossesse de la mère a été interrompue, parce que le fœtus présentait également une fente labio-palatine, caractéristique de 3MC. Nous ne disposons pas d'ADN pour cet individu. 3MC décrit une maladie génétique rare caractérisée par un retard de croissance prononcé, une dysmorphie faciale, des malformations urogénitales et une forme de DI. Des mutations récessives dans les gènes *MASP1* [MIM:257920] et *COLEC11* [MIM:265050] ont été identifiées chez des patients avec 3MC. Les dysfonctions dans ces gènes mettent en évidence des défaillances dans la voie de signalisation de la lectine (Rooryck et al., 2011) . Nous avons capturé et séquencé l'exome de l'individu II.1 ; dans notre analyse, nous nous sommes intéressés aux variants rares identifiés dans dès gènes impliqués dans la voie de biosynthèse de la lectine. Nous n'avons identifié aucune mutation rare dans les gènes *MASP1* et *COLEC11*. Nous avons identifié en tout 24 variants rares (homozygotes et hétérozygotes composites) chez cet individu, en considérant uniquement une transmission de la maladie selon un mode autosomique récessif. Aucun de ces variants n'affecte un gène dont les fonctions biologiques sont associées à la voie de signalisation de la lectine. Il n'est pas exclu que d'autres mécanismes cellulaires soient impliqués dans la pathophysiologie de cette famille. Par ailleurs ; la transmission de la maladie pourrait se faire selon un autre modèle ; il nous, par exemple, était impossible de mettre en évidence les mutations dominantes *de novo*, n'ayant séquencé que l'exome du patient.

Famille 222 : Il s'agit d'une famille consanguine Nord-Africaine (Figure VI.1) avec un garçon et une fille présentant une forme de DI-NS. Nous avons cartographié les RH partagées entre les individus affectés de cette famille (Tableau VI.3) ; parallèlement nous avons

séquencé l'exome de l'individu II.2. Nous avons identifié une seule mutation rare dans la totalité des RH partagées entre les individus affectés de cette famille ; la mutation c.C502T/p.R618W identifiée dans le gène *NUP50* ségrégue avec une transmission récessive dans cette famille. Cette mutation n'a pas été retrouvée dans notre cohorte d'exomes d'individus contrôles (n>1000) ; elle est également absente des bases de données publiques 1000 Genomes (n= 1000), Exome Variant Server (n>6500) et dbSNP138, ce qui suggère qu'elle pourrait avoir un impact pathogénique sur la fonction du gène. En effet, le résidu R168 est hautement conservé chez les vertébrés ; la mutation p.R168W est prédite comme dommageable par les logiciels Polyphen-2, SIFT et Provean.

NUP50 (ou *NPAP60*) code pour une composante du complexe du pore nucléaire; elle intervient dans le transport des protéines NLS-CARGO du cytoplasme vers le noyau de la cellule, en interagissant avec l'importine-alpha (Matsuura et Stewart, 2005 ; Stewart, 2006). NUP50 est exprimée dans différents tissus, particulièrement dans le *cerebellum* et l'hippocampe. Les souris dans lesquelles le gène *Nup50* a été invalidé meurent avant terme ; elles présentent des anomalies du tube neural, ainsi qu'un retard prononcé dans le développement (Smitherman et al., 2000) . NUP50 jouerait donc un rôle dans le développement de l'embryon, plus précisément au niveau du cerveau ; plusieurs syndromes associés à la DI impliquent des anomalies survenant tôt au niveau du développement de l'enfant (chapitre I section 4.1).

On distingue deux isoformes fonctionnels de NUP50 chez l'Homme; NUP50-S stabilise la liaison de l'importine-alpha à la protéine NLS-CARGO, tandis que NUP5O-L promeut la relâche de l'importine-alpha du NLS-CARGO (Ogawa et al., 2010) . Les rôles antagonistes de NUP50-S et NUP50-L suggèrent que des variations dans l'expression de NUP50 peuvent affecter l'efficacité du transport des protéines NLS-CARGO vers le cytoplasme. Plusieurs de ces protéines nucléaires, en l'occurrence les facteurs de transcription, interviennent dans le développement du cerveau ; il est donc possible d'envisager que des dysfonctions dans NUP50 pourraient expliquer la pathophysiologie de la DI dans cette famille. Une validation fonctionnelle de l'impact de la mutation p.R168W sur NUP50 sera nécessaire, afin d'établir les bases moléculaires de ce gène dans la DI.

La mutation p.R168W affecte les deux isoformes de NUP50. Nous avons cloné les isoformes sauvages et mutants (p.R168W) de NUP50 dans des vecteurs d'expression. Nous avons surexprimé ces constructions dans des cellules HEK293T ; les résultats préliminaires, par immunobuvardage de type Western, ne montrent aucune différence dans l'expression de NUP50-S, NUP50-L sauvages et mutants. Il se pourrait que la mutation p.R618W affecte plutôt la fonction de NUP50. Un bon test de validation serait d'évaluer l'impact fonctionnel de cette mutation sur le transport nucléaire ou cytoplasmique de protéines NLS-CARGO. NUP50 régule également le recyclage de l'importine-alpha vers le cytoplasme (Matsuura et al., 2005 ; Stewart, 2006) ; nous envisageons d'étudier l'effet de p.R618W sur la localisation de l'importine-alpha. Par ailleurs, le modèle du poisson zèbre a été mis de l'avant dans l'étude des anomalies tubes neural et autres maladies neurodévelopmentales; il se prêterait bien à l'étude du rôle de NUP50 dans le développement du poisson zèbre.

Famille 265 : Il s'agit d'une famille consanguine Pakistanaise (Figure VI.1) avec un garçon présentant une forme de DI associée à une microcéphalie. Nous avons cartographié l'autozygome de ce garçon (Tableau VI.3) ; parallèlement nous avons séquencé son exome. En tout, nous avons identifié 5 RHs de plus de 12 Mb et 5 autres RHs de 5 à 9 Mb chez le garçon affecté de la famille 265 (Tableau VI.3) ; nous avons identifié une trentaine de mutations récessives (homozygote ou composées génétiques) rares dans la totalité de ces RHs. Aucune de ces mutations n'affecte un gène connu de microcéphalie, suggérant que la cause de la maladie dans cette famille est due à un gène qui n'a pas encore été identifié. Dans cette famille nous avons mis en évidence un nombre considérable d'allèles récessifs ; il nous était impossible de valider fonctionnellement chacune de ces mutations individuellement. Avec d'autres individus (affectés et/ou non-affectés) dans la famille, il aurait été possible de réduire le nombre de RH candidates et/ou la liste de variants à analyser.

Famille 271 : Il s'agit d'une famille consanguine (Figure VI.1) Libanaise avec 3 enfants manifestant des crises d'épilepsie. Nous avons cartographié les RH partagées entre les individus affectés de cette famille (Tableau VI.3, RH partagées entre individus II.1 et II.3 uniquement), parallèlement nous avons séquencé l'exome des individus II.1 et II.3 . Les jumeaux II.2 et II.3 de la famille 271 ne présentent pas les mêmes manifestations cliniques ; l'individu II.2 est également hypotonique, contrairement aux autres patients de la famille 271.

Nous n'avons identifié aucune mutation homozygote rare dans les RH partagées entre les 3 individus affectés de cette famille. Nous avons donc analysé les mutations partagées entre les individus II.1 et II.3, lesquels présentent les mêmes traits cliniques ; ces individus ont 6 mutations homozygotes rares en commun. Nous avons identifié deux bons gènes candidats pour expliquer l'épilepsie chez les individus II.1 et II.3 ; les mutations identifiées dans les gènes *KCNAB3* (p.K52R) et *SMG6* (p.G86R) ne sont pas retrouvées chez l'individu II.2. *KCNAB3* code pour une sous-unité des canaux potassique voltage dépendant ; des dysfonctions dans ces types de canaux ont été mis en évidence dans plusieurs affections neurologiques incluant l'épilepsie (Lerche et al., 2005 ; Lerche et al., 2013) . SMG6 intervient dans les processus de NMD, un mécanisme de contrôle de la qualité de l'ARNm chez les eucaryotes. Des défaillances dans les processus de NMD se traduisent par une accumulation de mutations tronquant ; des dysfonctions dans ces mécanismes de contrôle cellulaire contribuent dans l'épileptogénèse (Kang et al., 2009) . Des études plus poussées seront nécessaires, afin d'élucider les bases moléculaires de l'épilepsie dans la famille 271.

Famille 302 : (voir chapitre V)

Famille 323 : Il s'agit d'une famille consanguine Arménienne (Figure VI.1) avec un patient présentant une forme de DI associée à des malformations cardiaques (la tétralogie de Fallot). Nous avons séquencé l'exome de cet individu. En tout, nous avons identifié une quarantaine de mutations autosomiques récessives (homozygotes et hétérozygotes composites) dans l'exome de cet individu. Dans notre analyse, nous avons porté une attention particulière aux gènes qui pourraient avoir une fonction biologique dans le développement du cerveau et du cœur; il était impossible de valider fonctionnellement toutes les mutations rares identifiées, étant donné le grand nombre d'allèles impliqués. Nous n'avons pu identifier la cause de la maladie chez cet individu.

Familles 398 et 435 : (voir chapitre IV; Ruzzo et al., 2013)

Famille 403 : Il s'agit d'une famille consanguine Pakistanaise (Figure VI.1, pédigrée complet non représenté) composée de 8 fratries avec un total de 14 individus affectés. La fratrie représentée à la figure VI.1 est composée de 3 filles présentant une forme de DI-NS; nous avons cartographié les RH partagées entre les 3 filles (Tableau VI.3). Nous avons observé

certaines variabilités dans les traits cliniques, manifesté par les patients de cette famille. La famille 403 est caractérisée par de multiples (au moins trois) boucles de consanguinité ; de fait, il n'est pas exclu qu'il puisse y avoir plus d'un allèle muté impliqué dans la pathophysiologie de cette famille. Nous avons identifié 8 mutations homozygotes rares, dans la totalité des RH partagées entre les individus affectés de la famille 403. Nous n'avons pas identifié de candidats potentiels, parmi les gènes affectés par des mutations homozygotes. Par ailleurs, nous avons généré une liste de 25 gènes potentiellement affectés par des mutations hétérozygotes composites. À cause du grand nombre d'allèles mutés identifiés, il nous est actuellement impossible d'élucider la cause de la DI dans la famille 403.

Famille 424 : Il s'agit d'une famille consanguine Canadienne-Française (Figure VI.1) avec trois garçons ataxiques, présentant également un retard prononcé dans le développement. L'individu II.2 présente un phénotype plus sévère que les autres patients; nous avons mis en évidence une anomalie cytogénétique *de novo* pouvant expliquer la cause de la maladie chez ce garçon. Nous avons cartographié les RH, partagées entre les individus II.1 et II.3, puis nous avons séquencé l'exome de l'individu (II.3). En tout, nous avons identifié 4 mutations homozygotes rares dans la totalité des RH ; une de ces mutations (c.6034 +1G>A) identifiée dans le gène *SPTBN2* [MIM : 615386], peut être mise en cause dans la pathologie de cette famille. Une délétion de 5 pb dans *SPTBN2* a été récemment identifiée dans une famille avec ataxie cérébelleuse, DI et retard de développement (Elsayed et al., 2014) . Nous avons montré que la mutation c.6034 +1G>A affecte l'épissage de *SPTBN2* et résulte en une troncation prématurée de la protéine. La mutation c.6034+1G>A ségrégue avec une transmission récessive de la maladie dans la famille 424 ; elle n'est toutefois pas retrouvée à l'état homozygote chez l'individu II.2 , lequel est porteur d'une anomalie cytogénétique *de novo*.

Familles 431 et 553 :

Famille 431 : Il s'agit d'une famille consanguine Turque (Figure VI.1) avec deux garçons présentant une forme de DI-NS. Nous avons cartographié les RH partagés entre ces individus, puis nous avons séquencé l'exome du patient II.1. Nous n'avons identifié aucune mutation homozygote rare dans la totalité des RH partagées entre les patients de cette famille. Nous

avons identifié deux mutations hétérozygotes composites rares (p.T156M et p.T2492M dans le gène *RNF213*) ségrégant avec une transmission récessive dans la famille 431.

Famille 553 : Il s'agit d'une famille consanguine Libanaise (Figure VI.1) avec deux cas de DI-NS. Dans cette famille, nous avons procédé au séquençage de l'exome des deux individus affectés; puis nous nous sommes intéressés aux mutations partagées entre eux. Nous avons porté une attention particulière aux mutations récessives ; en tout, nous avons identifié 5 variants récessifs rares (1 homozygote et 4 composées génétiques) ségrégant dans la famille 553. Deux de ses mutations p.K471E et p.T4638I affectent également le gène *RNF213*.

Au total, nous avons mis en évidence 4 mutations dans le gène *RNF213,* dans deux familles avec DI-NS. Des études pan-génomique, suggèrent qu'un allèle de RNF213 (p.R4859K) prédisposerait à la maladie de Moyamoya (MMD, Orphanet : ORPHA2573), une condition génétique caractérisée par une sténose des artères cérébrales du cerveau (Kamada et al., 2011 ; Liu et al., 2011) . MMD est généralement associée à des symptômes diabétiques ; les souris chez lesquelles *Rnf213* a été invalidé présentent également un phénotype associé à la progression du diabète (Scott et Smith, 2009 ; Bober et al., 2010 ; Kobayashi et al., 2013) . Les 4 allèles identifiés dans *RNF213* n'ont pas été retrouvés dans notre cohorte d'exomes d'individus contrôles ou dans les autres bases de données publiques (1000 Genomes, Exome Variant Server et dbSNP138). RNF213 est une protéine de plus de 5200 acides aminés, il est possible que les 4 allèles identifiés dans ce gène ne soient que des allèles rares, sans potentiel pathogénique. La fonction biologique de RNF213 est inconnue ; des analyses bioinformatiques suggèrent que RNF213 intervient dans le processus d'ubiquitination et qu'elle jouerait un rôle dans l'angiogenèse des vaisseaux sanguins. Générer des constructions de ce gène pour effectuer des études de validation fonctionnelle serait particulièrement laborieux, étant donné la taille des transcrits (>15 Kb). Une alternative serait de cloner les domaines fonctionnels (d'après les prédictions) de ce gène, particulièrement ceux affectés par les mutations identifiées. Les défauts dans l'angiogenèse associés à des dysfonctions dans RNF213 pourraient être également modélisés dans des études chez le poisson zèbre.

Famille 440 : Il s'agit d'une famille Haïtienne non-consanguine (Figure VI.1) avec trois cas de DI-NS. Nous avons capturé et séquencé l'exome des 4 filles de cette famille (les 3 individus affectés et leur sœur non-atteinte). Nous avons identifié 5 mutations homozygotes

rares partagées entre les patientes et non retrouvées dans l'exome de leur sœur non-atteinte. Cette famille n'étant pas caractérisée par une union consanguine, nous avons priorisé l'analyse des variants hétérozygotes composites. En tout, nous avons identifié deux mutations rares dans le gène *C10orf90* (p.I315V/c.2061+5T>C) ségrégant avec une transmission récessive de la maladie dans la famille 440. La fonction biologique de C100rf90 n'est pas connue ; une étude montre qu'il serait impliqué dans la formation des cils (Knorz et al., 2010) . C10orf90 partage des homologies avec ALMS1 [MIM:203800]; des dysfonctions dans ALMS1 causent le syndrome d'Alström, une ciliopathie parfois associée la DI (Collin et al., 2002 ; Hearn et al., 2005). C10orf90 est un bon candidat pour expliquer l'origine de la DI dans cette famille.

Famille 442 : Il s'agit d'une famille consanguine Libanaise (Figure VI.1) avec deux garçons avec DI-NS, œdèmes cérébraux et diabète insipide. Nous avons cartographié les RH partagées entre les individus affectés, puis nous avons séquencé l'exome du patient II.1. Cette famille est caractérisée par une forte homogénéité génétique, avec 3 RH de plus de 27 Mb et 4 autres RH entre 4 et 10 Mb (Tableau VI.3) Nous avons identifié 11 mutations homozygotes rares dans la totalité des RH ségréguant avec une transmission récessive dans cette famille. L'une de ces mutations (p.E311V) affecte le gène *CAND2*, un bon candidat pour expliquer la pathophysiologie de la DI dans la famille 442.

Les membres de la famille CAND (c*ullin-associated and neddylation-dissociated protein*), préviennent l'ubiquitination des protéines et leur dégradation subséquente dans le protéasome (Min et al., 2003) .Typiquement, cette classe de protéines inhibent la formation du complexe de dégradation SCF, en se liant aux cullines. Par exemple, CAND2/TIP120B régule le processus de myogenèse, en se liant à la culline-1; ce faisant, elle inhibe la dégradation de la myogénine par le complexe SCF (Shiraishi et al., 2007)

Le diabète insipide est causé par une déficience dans la sécrétion de l'hormone antidiurétique ou Vasopressine (AVP). Les patients avec cette pathologie ne parviennent pas à concentrer leurs urines, ils souffrent de polyurie et n'arrivent pas à étancher leur soif. Des mutations dans le gène codant pour l'AVP [MIM:125700] sont communément mises en cause chez les individus avec diabète insipide (Blackett et al., 1983 ; Bahnsen et al., 1992) . Le gène *CUL5* ou culline-5 code pour un récepteur d'AVP. Des études montrent que *Cul5* est surexprimé dans le cortex cérébral, l'hypothalamus et les reins du rat, lorsque l'animal est

privé d'eau ; *Cul5* jouerait donc un rôle dans le maintien de l'homéostasie en situation de stress osmotique (Byrd et al., 1997 ; Ceremuga et al., 2001 ; Ceremuga et al., 2003) . CAND2 est exprimée dans plusieurs types cellulaires, particulièrement au niveau du cortex et de l'hypothalamus de la souris (*Allen Brain Atlas*) ; mais son rôle n'a été étudié que dans les muscles. CAND1, l'autre membre de la famille CAND, régule l'expression de certaines protéines cibles en médiant leur ubiquitination ; l'interaction de CAND1 a été confirmée avec tous les membres des cullines (CUL1 à 5). Il est possible que CAND2 contrôle l'expression d'AVP en agissant sur son récepteur CUL5. La fonction biologique de CAND2 doit être encore investiguée, afin de confirmer son rôle dans la pathophysiologie de la famille 442.

Famille 460 : (voir chapitre III, Capo-Chichi et al., 2013b).

Famille 503 : Il s'agit d'une famille consanguine Libanaise (Figure VI.1) avec deux cas de DI-NS. Dans cette famille, nous avons procédé au séquençage de l'exome des deux individus affectés; puis nous nous sommes intéressés aux mutations partagées entre eux. Nous avons identifié 10 variants homozygotes rares ségrégant avec la DI dans cette famille; deux de ces variants ont été identifiés dans des gènes de connus de DI : p.R167W dans le gène *CCBE1* [MIM:235510] et p.R95Q dans le gène *PIGN* [MIM:614080]) (Alders et al., 2009 ; Maydan et al., 2011) . Les traits cliniques présentés par les individus affectés de la famille 503, s'apparentent à ceux manifestés par les patients avec des mutations dans *PIGN*. Des dysfonctions dans PIGN sont à l'origine de la pathogenèse de la famille 503, nous œuvrons à valider l'impact biologique de la mutation p.R167W sur la fonction de PIGN.

Famille 526 : Il s'agit d'une famille Canadienne-Française non-consanguine (Figure VI.1) avec deux cas de DI-NS. Nous avons utilisé un procédé analogue à celui de la famille 503, afin d'étudier la cause de la DI dans cette famille. En tout, nous avons mis en évidence 12 allèles récessifs rares (6 homozygotes et 6 hétérozygotes composites) partagés entre les individus II.1 et II.2. Nous n'avons identifié aucun gène candidat pouvant expliquer la pathophysiologie de la DI, dans cette famille.

Famille 7533 : (voir chapitre II, Capo-Chichi et al., 2013b).

Famille 8088 : (voir Annexe I, Mehawej et al., 2014)

Famille 11344 : Il s'agit d'une famille consanguine Libanaise (Figure VI.1) avec trois enfants présentant un retard prononcé dans le développement et une défaillance hépatique. Une description détaillée des traits phénotypiques des patients de cette famille a été rapportée par Mégarbané et al., 2008 . L'un des garçons de cette famille est décédé ; nous n'avions pas assez d'ADN pour compléter toutes les analyses chez cet individu (II.1). Nous avons cartographié les RH partagées entre les individus II.2 et II.3 (voir tableau VI.3) ; nous avons également séquencé l'exome de ces patients. En tout, nous avons identifié une dizaine de variants rares pouvant être transmsis selon un mode autosomique récessif, dans cette famille. Des études de validation génétique et fonctionnelle plus poussées seront nécessaires, afin de mettre en évdence les déterminants génétiques impliquant dans la pathologie de cette famille.

Tableau VI.4 Exploration des causes génétiques de la déficience intellectuelle par séquençage d'exome.

Dans l'ordre, les numéros de transcrits d'ARNm des gènes candidats identifiés sont : NUP50 (NM_007712); CLPB (NM_030813.3) ; RNF213 (NM_001256071); c10ORF90 (NM_001004298).

Famille	Phénotype	Cause génétique identifiée	Mutations dans gènes candidats	Cause génétique non identifiée
187	Syndrome de Malpuech	-	-	x
222	DI-NS	-	*NUP50* p.R618W	-
265	Microcéphalie primaire	-	-	x
271	DI, épilepsie, myoclonies	-	-	x
302	Encéphalopathie néonatale + acidurie 3-méthylglutaconique	*CLPB* p.I562TfsX23	-	-
323	DI+ tétralogie de Fallot	-	-	x
398	Encéphalopathie néonatale + hyperekplexie	ASNS p.6AE/ p.R550C (Ruzzo et al., 2013)	-	-
403	DI-NS	-	-	x
424	Ataxie cérébelleuse + retard dans le développement	*SPTBN2* c.6034+1G>A (Elsayed et al., 2014)	-	-
431	DI-NS	-	*RNF213* p.T156M/p.T2492M	-

Tableau VI.4 (Suite)

Famille	Phénotype	Cause génétique identifiée	Mutations dans gènes candidats	Cause génétique non identifiée
435	Encéphalopathie néonatale + hyperekplexie	ASNS p.R550C (Ruzzo et al., 2013)	-	-
440	DI-NS	-	*C10orf90* p.I315V/c.2061+5T>C	-
442	DI-NS + Diabète insipide néphrogénique	-	-	x
460	DI + mégalencéphalie	TBC1D7 p.Y180fsX1 (Capo-Chichi et al., 2013b)	-	-
503	DI + épilepsie	*PIGN* p.R95Q (Maydan et al., 2011)	-	-
526	DI-NS	-	-	x
553	DI-NS	-	*RNF213* p.K471E/p.T4638I	-
7533	Syndrome de cassure chromosomique de Varsovie	DDX11 p.R263Q (Capo-Chichi et al., 2013a)	-	-
8088	Retard dans le développement + dysplasie squelettique	MAGMAS p.N76D (Mehawej et al., 2014)	-	-
11344	Retard dans le développement + défaillance hépatique	-	-	x

CHAPITRE VII: DISCUSSION GÉNÉRALE

Nous avons étudié l'origine de la DI dans vingt familles où une transmission de la maladie selon un mode autosomique récessif était suspectée. En tout, nous avons été en mesure d'élucider la cause de la DI dans huit des familles analysées (Tableau VI.4). D'abord, dans trois de ses familles (424,503 et 7533), nous avons mis en évidence des mutations dans des gènes déjà associés à la DI. Dans la famille 7533 (Chapitre II, Capo-chichi et al.,2013a), nous avons rapporté le second cas de patients avec syndrome de cassure de Varsovie (Van Der Lelij et al., 2010a) ; notre étude a confirmé l'implication du gène *DDX11* dans ce syndrome, en plus de contribuer à en élargir le spectre phénotypique. Dans la famille 424 (Chapitre VI), nous avons identifié une mutation tronquant SPTBN2; *SPTBN2* [MIM : 615386] a été récemment mis en cause dans la pathologie d'une famille avec ataxie cérébelleuse, DI et retard de développement (Elsayed et al., 2014) . Nous avons mis en évidence une mutation dans *PIGN* à l'origine de la pathogenèse dans la famille 503. En effet, les traits cliniques de nos patients s'apparentent à ceux d'individus avec syndrome d'anomalies congénitales multiples-hypotonie-épilepsie-1 (MCAHS1) ; ces patients ont également des mutations dans le gène *PIGN* [MIM :614080] (Maydan et al., 2011) .

Dans cinq des familles étudiées (302, 398, 435,460 et 8088), nous avons identifié de nouveaux gènes de DI. Dans la famille 460 (Chapitre III), nous avons associé des dysfonctions dans le gène *TBC1D7* à une condition caractérisée par la mégalencéphalie et la DI (Capo-Chichi et al., 2013b). Des cas additionnels avec mutation dans *TBC1D7* ont été récemment rapportés (Alfaiz et al., 2014) ; confirmant l'implication de TBC1D7 dans la pathophysiologie de la DI. Dans les familles 398 et 435 (Chapitre IV), nous avons mis en évidence des mutations dans le gène *ASNS* chez six patients présentant une forme sévère d'encéphalopathie néonatale. Nos collaborateurs de l'université de Duke ont mis en évidence cinq cas additionnels avec dysfonctions dans le gène *ASNS*. Notre étude a montré que des défauts dans la voie de biosynthèse de l'Asparagine sont mis en cause dans la pathophysiologie de ces familles avec encéphalopathie sévère (Ruzzo et al., 2013). Une autre groupe a récemment identifié des patients, avec phénotypes similaires aux nôtres, et ayant également des mutations dans *ASNS* (Ben-Salem et al., 2014) . Dans la famille 8088, nous avons pu associer des dysfonctions dans le MAGMAS à une condition caractérisée par un retard prononcé dans le développement et une dysplasie osseuse sévère ; en tout nous avons identifié deux familles

avec des patients ayant des phénotypes similaires et porteurs de la même mutation dans *MAGMAS* (Annexe 1, Mehawej C et al., 2014). Finalement, dans la famille 302 (Chapitre V), nous avons mis en évidence une mutation tronquant dans CLPB chez quatre sœurs avec encéphalopathie néonatale sévère, hyperekplexie et acidurie 3-méthylglutaconique.

Nous avons identifié de bons candidats pouvant expliquer l'origine de la DI dans quatre des familles analysées ; des analyses plus poussées seront nécessaires, afin d'élucider la cause génétique de la DI dans les huit autres familles où nous n'avons pu mettre en évidence des gènes candidats (Tableau VI.4). Nous avons généré une liste de tous les variants rares identifiés dans ces douze familles. Nous confrontons cette liste à celles de nos collaborateurs à travers le monde, afin d'identifier des cas additionnels présentant des phénotypes similaires à ceux de nos patients. Il s'agira d'un argument majeur, afin de confirmer l'implication des gènes identifiés dans la pathophysiologie de la DI. Parallèlement, nous avons entrepris de valider l'impact fonctionnel de certaines mutations identifiées dans de bons gènes candidats.

Dans une étude similaire à la nôtre, Najmabadi et ses collaborateurs ont exploré l'origine de la DI dans 136 familles consanguines multiplexes (composées d'au moins deux individus affectés). Ils ont effectué une cartographie des RH, parallèlement au WES, et ont concentré leurs analyses sur les mutations homozygotes identifiées dans l'autozygome des individus affectés (Najmabadi et al., 2011) . L'étude rapporte 50 nouveaux gènes candidats de DI ; dans 26 des familles analysées, l'origine de la maladie était en lien avec des gènes connus de DI. Le rendement diagnostic de ce projet de recherche est de 40% (8/20) ; les résultats de notre étude ont été répliqués par d'autres groupes, confirmant l'implication des 8 gènes identifiés dans la pathologie de la DI. Les nouveaux gènes identifiés dans l'étude de Najmabadi sont de bons candidats pour expliquer la DI. En revanche, à ce jour, la majorité de ces gènes n'ont été rapportés que dans une seule famille ; une validation génétique chez des patients avec des phénotypes similaires doit être faite, afin de confirmer l'implication de ces gènes dans la pathophysiologie de la DI. L'étude de Najmabadi se base presqu'exclusivement sur des modèles de prédiction ; une validation de l'impact fonctionnel des différentes mutations sera nécessaire, afin de confirmer leur impact pathogénique sur la protéine mutée.

Tout comme l'équipe de Najmabadi, nous avons mis de l'avant le WES dans l'exploration des bases génétiques de la DI ; les NGS ont été également utilisées dans

l'exploration des bases moléculaires d'autres maladies génétiques. Typiquement, ces technologies ont fait leurs preuves dans l'étude de maladies monogéniques, particulièrement celles impliquant des allèles rares hautement pénétrants. En revanche, des traits complexes, polygéniques ou multifactoriels traduisant des mutations faiblement pénétrantes seraient plus difficilement approchables par les techniques de NGS, comme le WES.

Nos travaux mettent en évidence certains défis liés à l'étude de la DI, notamment la grande hétérogénéité génétique de la DI et la non-homogénéité clinique de la DI-NS. Les NGS peuvent être utilisées dans l'évaluation clinique de patients avec DI ; certaines considérations doivent être prises en compte, afin d'augmenter leur potentiel diagnostique. Par exemple, une caractérisation phénotypique approfondie des patients est indispensable au succès du WES ; incidemment, dans six des douze cas non résolus, nous n'avons pu détecter chez nos patients d'autres traits distinctifs en dehors de la DI. Ces cas de DI-NS sont particulièrement difficiles à étudier, comme en témoigne l'étude de Najmabadi. Dans certains cas, une première vague d'investigations par des approches cytogénétiques seraient à considérer, avant l'application du WES ; étant donné le potentiel de détection limité des NGS pour la mise en évidence des lésions macroscopiques ou submicroscopiques. Les patients avec des défauts dans des voies mécanistiques bien caractérisées pourraient être priorisés pour le WES. Il serait alors, plus aisé de cibler, dans les analyses, les gènes intervenants dans la voie de signalisation défectueuse. Une autre alternative serait de prioriser la capture et le séquençage de panels de gènes susceptibles d'être mutés dans la maladie à l'étude. Le séquençage d'une telle fraction de gènes ne pourrait être d'emblée appliqué dans l'étude de conditions fortement hétérogènes comme la DI. Il pourrait, par contre, être mis à contribution dans l'étude de certaines IEMs ou encore de syndromes dont les bases moléculaires ont déjà été investiguées. À noter que les librairies de ces gènes-cibles devront être constamment mises à jour dans les panels, au fur et à mesure que de nouveaux gènes causaux seront identifiés ; sans quoi, le diagnostic clinique pourrait être faussé. Les panels de gènes permettent de prévenir l'occurrence de découvertes fortuites, plus courantes dans d'autres applications des NGS, comme le WES.

Identifier une mutation rare, même dans un bon gène candidat pour expliquer la pathophysiologie d'une maladie donnée ne suffit pas pour poser un diagnostic clinique. Il est nécessaire de valider biologiquement l'impact de la mutation identifiée sur la fonction du gène

en question, pour pouvoir confirmer sa pathogénicité. Les variants identifiés ne sont pas validés, d'emblée ; de sorte que certains d'entre eux peuvent être faussement associés à la pathogenèse d'une maladie. Par exemple, une étude a mis en évidence 406 mutations rapportées comme causales ; desquelles 27% sont en réalité des polymorphismes ou des variants dont le potentiel pathogénique n'avait pas été investigué (Bell et al., 2011) . Des tests fonctionnels ne sont pas forcément disponibles pour effectuer la validation fonctionnelle des mutations ; particulièrement en clinique, où les ressources sont également limitées. Des efforts sont fournis, afin de regrouper des mutations causales de différentes maladies dans des bases de données ; ces mutations sont validées biologiquement et/ou génétiquement dans des patients présentant les mêmes traits phénotypiques . À l'ère du séquençage à haut-débit, il n'est pas rare d'identifier des mutations qui ne sont rapportées dans aucune base de donnée. Les laboratoires de génétique moléculaire sont soumis, à juste titre, à des règles plus stringentes dans l'identification et l'interprétation des mutations. Typiquement, les variants rapportés, en clinique, sont répartis en différentes catégories : pathogénique, potentiellement pathogénique, bénin, etc. Une classe de variants particulièrement difficiles à interpréter sont les variants dits de signification inconnue ou VUS (variant *of unknown significance*). Des considérations éthiques et légales sont associées à l'utilisation des NGS, en génétique. Des lignes directrices permettent de régir, dans la mesure du possible, les informations qui doivent être communiquées au patient. Des situations conflictuelles peuvent survenir ; le consentement du patient ne s'accordent pas nécessairement avec celui du clinicien , notamment au sujet de la prise en charge du patient ou encore des services génétiques qui pourraient être promulgués aux autres membres de sa famille.

L'application des NGS dans les laboratoires de diagnostic moléculaire, est encore mise à rude épreuve. Certaines améliorations doivent être apportées, afin d'implémenter ces nouvelles technologies en clinique. Parmi les difficultés les plus couramment rencontrées dans l'usage des NGS, nous pouvons citer :

a.Les difficultés techniques liées au processus de capture d'exome

Connaissances sur le génome humain.

Moins de 10% des séquences du génome humain ont été caractérisées (Rizzo et Buck, 2012) ; c'est le cas notamment des régions codantes, lesquelles doivent être continuellement annotées.

De fait, les librairies utilisées pour la capture sont incomplètes, elles doivent être constamment mises à jour. La composition des librairies utilisées pour la capture varie également selon le manufacturier. Par exemple, les libraires des trois plateformes commerciales les plus populaires , à savoir Nimblegen, Agilent et Illumina ciblent respectivement 64.1 Mb, 51.1 Mb et 62.08 Mb de matériel génétique dans l'exome; elles n'ont toutefois que 26.2 Mb de séquences en commun (Chilamakuri et al., 2014) .

Composition du génome humain.

L'efficacité de capture varie selon les propriétés biochimiques des séquences ciblées. La composition des séquences en Guanine et Cytosine (GC) a un impact considérable sur l'efficacité de capture, particulièrement lorsque le taux de GC est inférieur à 30% ou supérieur à 70% (Dohm et al., 2008 ; Gnirke et al., 2009 ; Chilamakuri et al., 2014). Typiquement , 5 à 10% des séquences de l'exome sont mal couvertes ou ne peuvent carrément être capturées (Bamshad et al., 2011) .

Autres biais dans la capture.

Une proportion significative du matériel capturé est composée de séquences non-ciblées (séquence introniques, intergéniques, ARN non codants ou ADN mitochondrial); de tels contaminants peuvent affecter les analyses subséquentes. Par ailleurs, l'efficacité de capture est plus faible pour les séquences situées à la jonction exons-introns ; ce qui peut avoir un impact sur la détection des mutations, particulièrement celles qui affectent les sites consensus d'épissage. Un autre problème récurrent dans le processus de capture est la génération d'artéfacts de PCR (ou *PCR duplicates*). On estime que 10 à 15 % du matériel capturé est composé d'amplicons identiques d'un même fragment d'ADN ; de telles séquences n'apportent aucune information supplémentaire , elles ne peuvent donc être retenues dans le traitement des données (Clark et al., 2011) .

b.Les difficultés techniques dans le traitement des données de séquençage

Accessibilité des ressources.

En dépit des efforts de vulgarisation, les NGS demeurent encore largement inaccessibles, parce qu'elles nécessitent un support logistique considérable et des connaissances bioinformatiques non-négligeables.

Usage des bases de données.

Le traitement des résultats de séquençage dépend grandement des informations provenant de différentes bases de données. Il revient à l'utilisateur de s'entourer des outils, lui permettant de faire un bon usage des informations recueillies. Ces informations sont parfois incomplètes ou carrément erronées; on estime qu'environ 15% des données répertoriées dans dbSNP sont des faux-positifs (Day, 2010). Il faut savoir choisir la base de données qui est la mieux adaptée au devis de notre étude. Certaines bases de données comme le NHLBI-ESP (*US National Heart, Lung, and Blood Institute Exome Sequencing Project*) et ExAC (*Exome Aggregation Consortium*) Browser prennent en compte des données sur individus affectés par des maladies génétiques; ces derniers ne peuvent être strictement considérés comme des contrôles, selon le phénotype des patients à l'étude.

Qualité des données.

Environ 97% des séquences peuvent-être alignées au génome. Selon la qualité des données, après le tri des séquences de mauvaises qualités, les *PCR duplicate* et autres artéfacts de capture, l'efficacité d'alignement peut varier de 71 à 40%, selon la technologie utilisée (Chilamakuri et al., 2014) . Des erreurs récurrentes inhérentes à la structure du génome sont rencontrées dans l'alignement des séquences ; par exemple, les séquences répétées en tandem, dupliquées ou homologues ne peuvent être efficacement alignées au génome. Le séquençage des homopolymères constitue une autre source de faux positifs couramment rapportée (Shendure et al., 2008) .

Détection des variants.

Séquençage d'exome versus séquençage du génome.

Jusqu'ici le WES a été principalement préféré au séquençage du génome (WGS), parce que l'exome constitue la fraction du génome qui a été la plus étudiée. Il n'est toutefois pas exclu que la cause d'une maladie implique des lésions génétiques dans des séquences, en dehors de l'exome. En fait, d'après les analyses pan génomiques , plus de 80% des polymorphismes associés à des maladies génétiques sont identifiés hors de l'exome ; ce qui suggère que les régions non codantes renferment une bonne proportion des déterminants génétiques de ces maladies (Alexander et al., 2010 ; Manolio, 2010) . À cet effet, dans une étude menée sur 100 patients avec DI sévère, l'équipe de De Ligt a obtenu un rendement diagnostic du WES de 16

%. En procédant par WGS, Gilissen et ses collaborateurs ont été en mesure d'élucider la cause génétique de la DI dans 21 des 50 cas de DI sévères non résolus dans l'étude de De Ligt (De Ligt et al., 2012 ; Gilissen et al., 2014) . L'application du WGS contribue à palier aux biais de capture, notamment en ce qui concerne les jonctions introns-exons. Les approches de WES ont été majoritairement utilisées pour mettre en évidence des mutations ponctuelles ; ces technologies doivent être encore optimisées pour la détection d'autres types de lésions génétiques comme les CNVs, les grandes insertions-délétions (*indels*), les évènements de fusion de gènes et autres réarrangements structuraux (translocations, inversion). Le WGS s'avère également être un meilleur outil pour caractériser ces autres types de mutation.

c.Difficulté liées au processus d'identification de la mutation causale

Variants candidats à analyser.

Le potentiel de découverte des NGS est illimité, de sorte qu'il n'est pas rare qu'on identifie plusieurs gènes candidats; et ce même lorsque des approches de cartographie sont utilisées en parallèle. Par exemple dans les familles 265 et 403 (Figure VI.1), nous avons mis en évidence plusieurs dizaines de variants candidats, et ce quand bien même une cartographie de l'autozygome avait été préalablement réalisée. Dans ces familles, il était impossible d'identifier la cause de la DI, étant donné le grand nombre d'allèles mutés mis en jeu. De plus, dans certaines des familles étudiées, il n'y avait pas d'individus non-affectés ; de sorte qu'il n'était pas possible de réduire le nombre de variants candidats, dans des études de ségrégation.

Choix de l'approche d'analyse.

L'analyse des résultats de NGS, doit être basée sur une hypothèse (transmission autosomique récessive, dominante, liée au chromosome X) ; il ne faut toutefois pas se restreindre à un seul modèle, parce que la cause génétique du trait à l'étude peut impliquer un mode de transmission non suspecté. Par exemple, Najmabadi et ses collaborateurs ont étudié l'origine de la DI dans 136 familles consanguines ; ils n'ont considéré que les mutations récessives homozygotes (Najmabadi et al., 2011) . De fait, ils n'ont pu identifier la cause de la DI dans 40% des familles analysées ; d'autres modes de transmission de la maladie peuvent être mis en jeu dans les cas de DI non-élucidés dans cette étude. Dans nos analyses, nous n'avons pas tenu compte des mutations dominantes ou des somatiques ; il est également possible que des anomalies cytogénétiques non détectables par WES ou array-CGH, puissent être mises en

cause dans certains des cas non résolus de ce projet de recherche. Même dans les familles consanguines, il n'est pas exclu que la maladie soit transmise selon d'autres modèles ; par exemple, dans la famille 424, nous avons mis en évidence une anomalie cytogénétique *de novo* à l'origine de la DI chez le patient II.2 ; cette lésion génétique n'a pas été détectée chez les autres individus affectés de la famille.

Priorisation de gènes candidats.

La principale difficulté liée à l'analyse des données de séquençage se situe au niveau de la priorisation des variants candidats (filtre IV, Figure I.6). Il s'agit d'appliquer les bons paramètres, afin d'éliminer habilement les faux positifs, pour ne retenir que la (les) mutation(s) causale(s). Des critères de filtre trop stringents peuvent biaiser les analyses. Par exemple, on ne pourrait d'emblée filtrer tous les variants rapportés dans une cohorte d'individus contrôles. En effet, certaines maladies rares, particulièrement celles où une transmission autosomique récessive est suspectée, peuvent mettre en jeu des mutations causales avec une faible fréquence allélique.

Les difficultés de priorisation des gènes candidats, mettent en évidence les limites de nos connaissances sur le génome humain. Il arrive souvent que le gène candidat identifié n'ait pas été caractérisé ou que ces fonctions biologiques ne soient pas connues. Par ailleurs, il est également coutume d'utiliser des logiciels, afin de prédire le potentiel délétère des mutations non-synonymes. Les algorithmes utilisés diffèrent d'un logiciel à un autre, de sorte qu'il n'est pas rare d'obtenir des résultats contradictoires pour une même mutation donnée (Tableau SV.3). Par ailleurs , bien qu'elles soient considérées comme ayant un faible potentiel pathogénique, les mutations synonymes peuvent affecter la stabilité et l'expression de l'ARNm ; inversement plusieurs mutations non-sens, généralement classées parmi les mutations ayant le plus haut potentiel délétère, n'ont aucun impact pathogénique (Tarpey et al., 2009 ; Zheng et al., 2014) .

d.Difficulté dans la validation fonctionnelle de la mutation causale

Les outils bioinformatiques contribuent grandement dans l'analyse des données de NGS, avec un pouvoir de prédiction de 70-8% (Castellana et Mazza, 2013) , mais le meilleur moyen d'identifier la cause génétique d'une maladie reste encore de valider fonctionnellement

l'impact biologique de toutes les mutations candidates. Différents paradigmes se prêtent à ce type d'études. Le modèle du poisson zèbre a été mis de l'avant dans la validation fonctionnelle d'allèles associés à des maladies génétiques (Davis et al., 2014) . De façon générale, l'utilisation de modèles animaux se prête mieux dans l'étude de traits bien caractérisés et dont les manifestations phénotypiques peuvent être également aisément analysées dans le modèle choisi. Par exemple, si des tests comportementaux permettent de modéliser les maladies neurologiques chez la souris et le rat ; la DI peut être difficilement étudiée chez le poison zèbre. Les familles 265 avec microcéphalie primaire, 302 avec hyperekplexie et microcéphalie congénitale , 323 avec tétralogie de Fallot et 442 avec diabète insipide nephrogénique présentent des traits caractéristiques qui peuvent être plus facilement modélisés chez le poisson zèbre . D'autres facteurs doivent être pris en considération, lorsqu'on fait usage de modèles animaux, notamment le pourcentage d'homologie du gène à l'étude et le nombre de copies du gène dans le génome du modèle choisi.

Les études fonctionnelles peuvent être également menées dans des systèmes cellulaires. L'étude du patron d'expression de gènes ou l'analyse de leurs profils transcriptomiques peut permettre d'évaluer l'impact biologique d'une mutation. Le choix du type de cellules est déterminant ; dans certains cas on doit se résoudre à utiliser des modèles cellulaires qui ne reflètent pas forcément la pathologie à l'étude. Par exemple, si les tissus sanguins permettent de mieux étudier certains types de cancers comme la leucémie ; il est beaucoup plus laborieux d'avoir accès à des échantillons de cerveau pour analyser les maladies neurologiques.

Les études classiques d'expression (RT-PCR ou immunobuvardage de type Western) donnent des indices sur le potentiel pathogénique d'un variant ; pour vraiment valider l'impact d'une mutation, il faut effectuer des analyses plus approfondies sur la fonction du gène. Le test de validation doit être adapté selon le rôle biologique du gène; le potentiel de découverte des NGS étant illimité, il n'est pas rare que les connaissances requises soient en dehors du domaine d'expertise du laboratoire dans lequel l'étude est menée. Les analyses sont d'autant plus compliquées, lorsque le gène identifié n'a pas encore été caractérisé, comme c'est le cas dans les familles 302, 431, 440, 553 et 11344.

e.Évaluation phénotypique des patients

L'application du WES comme outil de diagnostic clinique requiert une caractérisation phénotypique approfondie des individus affectés. Poser un diagnostic n'est pas une tâche aisée, particulièrement lorsque le trait à l'étude inclus plusieurs manifestations cliniques. Il peut être difficile de discriminer le phénotype prédominant, parmi tous les symptômes manifestés ; dans certains cas les traits phénotypiques peuvent varier selon la progression ou la sévérité de la maladie. Dans les formes syndromiques de DI, il n'est pas évident de déterminer si la maladie est causée par un problème neurologique, une défaillance métabolique ou des dysfonctions dans l'un des multiples mécanismes pouvant être impliqués. L'évaluation phénotypique du patient peut varier selon la spécialité ou l'intérêt du clinicien ; une caractérisation complète d'une maladie complexe requiert généralement l'intervention de plus d'un spécialiste.

L'évaluation du patient est déterminante dans la priorisation des gènes candidats ; elle permet de mieux cerner les voies de signalisation les plus susceptibles d'être défectueuses chez le patient. Identifier des cas additionnels présentant les mêmes traits cliniques que le patient à l'étude, constitue le meilleur argument pour établir un lien de causalité entre une mutation et une maladie génétique donnée. Dans le contexte de maladies complexes comme la DI, il est plus ardu d'identifier d'autres individus affectés ; particulièrement dans la DI-NS. Ces maladies impliquent bon nombre de cas isolés, de sorte qu'il n'est pas aisé d'identifier des cas additionnels; de plus, en l'absence d'une caractérisation phénotypique approfondie, ces cas sont d'autant plus difficiles à identifier. Certaines bases de données (*Human Gene Mutation Database* - HGMD, *Online Mendelian Inheritance in Man* – OMIM) peuvent aider à discriminer, avec certaines réserves, des gènes déjà associés à des maladies génétiques (Mckusick, 1998 ; Stenson et al., 2003) . Ces bases de données ne sont pas toujours à jour ; elles peuvent être parfois erronées. Il faut également tenir compte du fait que des lésions génétiques dans un même gène puissent se traduire par plus d'un phénotype ; en plus de cette hétérogénéité clinique, il faut composer avec la grande hétérogénéité génétique de la DI. Par exemple , nous ne pouvons exclure d'emblée que *RNF213* soit mis en cause dans la pathophysiologie des familles 431 et 553 , même si ce gène a déjà été associé à MMD .

Poser un diagnostic clinique est d'autant plus compliqué, lorsque les traits phénotypiques varient d'un individu à un autre ; et ce au sein d'une même famille (familles 271 et 403). Les mariages entre individus apparentés peuvent mettre en jeu plusieurs boucles de consanguinité (familles 403 et 424). Dans ce type d'unions, il peut y avoir plus d'un allèle causal transmis. On peut également imaginer qu'il y ait un haplotype conférant le phénotype majeur manifesté, tandis que les autres marqueurs génétiques transmis serviraient davantage de modificateurs. De telles variabilités dans les symptômes présentés par les patients d'une même famille peuvent venir fausser les analyses. Dans ces cas de figures, il est difficile de déterminer si les individus affectés ont un même génotype, ou si certains d'entre eux doivent être analysés séparément.

CONCLUSION

Les NGS sont entrain en train de révolutionner l'univers du diagnostic moléculaire, en permettant la recherche directe de mutations à l'échelle de l'exome ou du génome humain. Avec la baisse des coûts de séquençage, le WGS est en voie de remplacer WES, promettant également de combler certaines des lacunes de sa technique-soeur. Ces technologies sont graduellement intégrées, dans l'évaluation et la prise en charge des patients dans les cliniques de génétique. Des applications des NGS sont également utilisées dans d'autres sphères; elles permettent notamment de réaliser des tests non-invasifs par la détection de l'ADN fœtal circulant ou encore de l'ADN tumoral libre. Des considérations légales et éthiques sont associées à l'usage des NGS en clinique, notamment en ce qui concerne les VUS, les découvertes fortuites et leurs implications économiques et sociales. Le pouvoir des NGS est indéniable ; certaines améliorations doivent être néanmoins apportées, afin de faire progresser ces technologies et de les utiliser à leur plein potentiel.

ANNEXE I: The Impairment of MAGMAS Function in Human Is Responsible for a Severe Skeletal Dysplasia

Publié dans *Plos genetics* 2014 May 1;10(5)

Cybel Mehawej[1,2], Agnès Delahodde[3], Laurence Legeai-Mallet[2], Valérie Delague[4,5], Nabil Kaci[2], Jean-Pierre Desvignes[4,5], Zoha Kibar[6,7], José-Mario Capo-Chichi[6,7], Eliane Chouery[1], Arnold Munnich[2], Valérie Cormier-Daire[2], André Mégarbané[1]*

1.Unité de Génétique Médicale et Laboratoire International associé INSERM à l'Unité UMR_S 910, Faculteé de Médecine, Université Saint-Joseph, Beirut, Lebanon
2.Département de Génétique, Unité INSERM U781, Université Paris Descartes-Sorbonne Paris Cité, Fondation Imagine, Hôpital Necker Enfants Malades, Paris, France
3.University of Paris-Sud, CNRS, UMR 8621, Institute of Genetics and Microbiology, Orsay, France
4.Inserm, UMR_S 910, Marseille, France
5.Aix Marseille Université, GMGF, Marseille, France
6.Center of Excellence in Neuroscience of Université de Montréal, Centre de Recherche du CHU Sainte-Justine, Montréal, Canada
7.Department of Obstetrics and Gynecology, Université de Montréal, Montréal, Canada

Corresponding author:

André Mégarbané

Competing Interests: The authors have declared that no competing interests exist.

Contribution:

J'ai analysé les résultats de séquençage d'exome de la famille 8088 avec Kibar Z. Mégarbané A, Mehawej C, Delague V et Desvignes JP ont analysé les données de séquençage de l'autre famille étudiée dans cet article. Mégarbané A, Cormier-Daire V, Delahodde A, Legeai-Mallet L et Munnich A ont partcipé au design des expériences. Kaci N, Delahodde A, Delague V et Chouery E ont fourni le matériel et les réactifs utilisés dans cette étude. Mégarbané A, Mehawej C et Cormier-Daire V ont rédigé le manuscrit.

ABSTRACT

Impairment of the tightly regulated ossification process leads to a wide range of skeletal dysplasias and deciphering their molecular bases has contributed to the understanding of this complex process. Here, we report a homozygous mutation in the mitochondria-associated granulocyte macrophage colony stimulating factor-signaling gene (MAGMAS) in a novel and severe spondylodysplastic dysplasia. MAGMAS, also referred to as PAM16 (presequence translocase-associated motor 16), is a mitochondria-associated protein involved in preprotein translocation into the matrix. We show that MAGMAS is specifically expressed in trabecular bone and cartilage at early developmental stages and that the mutation leads to an instability of the protein. We further demonstrate that the mutation described here confers to yeast strains a temperature-sensitive phenotype, impairs the import of mitochondrial matrix pre-proteins and induces cell death. The finding of deleterious MAGMAS mutations in an early lethal skeletal dysplasia supports a key role for this mitochondrial protein in the ossification process.

INTRODUCTION

During embryonic development and postnatal growth period, most longitudinal bone growth occurs via endochondral ossification, a process in which chondrocytes produce a cartilage template that will ultimately be replaced by bone (Karsenty et al., 2009). Disturbance of this tightly regulated process results in skeletal dysplasias (SD), which are an extremely diverse and complex group of rare genetic diseases (Baitner et al., 2000). Causative mutations have now been identified in over 230 different genes in more than 400 unique skeletal phenotypes (Ikegawa, 2006 ; Warman et al., 2011). However, the genetic basis of over 100 different entities remains to be determined. Among SD, spondylodysplastic dysplasia is a generic descriptive term for a heterogeneous group of SD characterized by severe vertebral abnormalities and distinctive skeletal or extra- skeletal features (Ikegawa, 2006 ; Warman et al., 2011). The outcome is variable from early lethality to long-term survival. The last international classification recognizes at least five distinct entities within this group of rare disorders, i) achondrogenesis type 1A [MIM 200600] due to mutations in TRIP11 [MIM 604505], which encodes the thyroid hormone receptor interactor 11, ii) Schneckenbecken dysplasia [MIM 296250] due to mutations in SLC35D1 [MIM 610804], encoding the solute carrier family 35 member D1, also known as UDP- Galactose transporter-related 7 iii) spondylometaphyseal dysplasia (SMD) Sedaghatian type [MIM 250220], a heterogeneous condition that is due in some cases, to mutations in SBDS [MIM 607444] involved in the processing of ribosomal RNA, iv) fibrochondrogenesis [MIM 228520], due in some cases to COL11A1 mutations and v) opsismodysplasia [MIM 258480], due to mutations in INPPL1 [MIM 600829] that encodes the inositol polyphosphate phosphatase like-1 (Huber et al., 2013). The wide range of proteins involved in this group of SD points out the complexity of the ossification process.

Although different entities within the group of spondylodysplas- tic dysplasia are well defined, a few cases remain unclassified. Recently, Mégarbané et al. have reported on two unrelated consanguineous Lebanese families with four affected cases presenting a novel type of early lethal spondylometaphyseal dysplasia (Megarbane et al., 2008a ; Megarbane et al., 2014) . To elucidate the molecular basis of this entity, we performed exome sequencing in these two

distinct families and identified a homozygous mutation in MAGMAS (Mitochondria-associated granulocyte macrophage colony stimulating factor- signaling molecule) in patients from both families.

RESULTS

Patients

Two unrelated consanguineous Lebanese families with four affected cases presenting a novel type of early lethal spondylodys- plastic dysplasia, recently reported by Mégarbané et al. (Megarbane et al., 2008a ; Megarbane et al., 2014) , were included in this study. The main clinical features were pre- and postnatal growth retardation, developmental delay, dysmorphic features, prominent abdomen, narrow thorax with short ribs and short limbs. Skeletal features included severe platyspondyly at birth that improved somewhat over time, square iliac bones, horizontal acetabulae, short long bones with abnormal modeling, widening of the distal femoral metaphyses and delayed epiphyseal ossification (Figure A.1). In the first reported family (Family F1, Figure A.2 ; Megarbane et al., 2008a) both affected sibs died at an early age from respiratory insufficiency (F1-IV.1 at age 9 months and F1-IV.3 at age 2 years) while heart failure was the cause of the early death, at age 2 years, in patients (F2-IV.2 and F2-IV.3) of the second family (Family F2: individuals IV.2 and IV.3; Megarbane et al., 2014) (Figure A.2).

Molecular findings

Exome sequencing has been performed in Family 2, in three individuals: the proband (F2-IV.3) and her two unaffected parents (F2-III.1 and F2-III.2) (Figure A.2). Considering the recessive inheritance of the disease in the pedigree and consanguinity (F = 1/16), we assumed identity by Descent (Lander et al., 1987) . Using an in-house bioinformatics tool, we realized identity by descent filtering of exome data and identified 1037 homozygous variants in the proband, which were also heterozygous in both parents. Of these 1037 variants, 916 were exonic and/or splice site mutations, of which 434 were either non-synonymous and nonsense single nucleotide variants, or small insertions/deletions leading to a shift in the reading frame.

In order to further refine the list of candidate nucleotide variations, additional filtering was performed on what we called the ‘‘dbSNP13700’’ database (i.e. dbSNP137, from which we have removed mutations that are reported as pathogenic alleles, as well as all referenced

variants with frequencies lower than 1%, or without any frequency information). Additional filtering was also performed using our in-house exome database that includes all homozygous variants detected by exome sequencing of 14 Lebanese unaffected individuals. We ended up with a list of 16 candidate homozygous by descent variants in the proband F2- IV.3.

In family 1, exome sequencing has been performed only in the proband (F1-IV.3, Figure A.2). Using the same analysis pipeline and after all filtering steps, we were able to identify 59 homozygous variants in the genes coding regions of this patient.

In total, patients F1-IV.3 and F2-IV.3 shared 13872 homozygous variants, of which only 3 exonic variations were both homozygous by descent in the proband from family 2 (F2-IV.3), and homozygous in the proband from family 1 (F1-IV.3), after applying the different filters (Table SA.1). Of those three variants, only one missense mutation (NM_016069: c.226A.G; p.Asn76Asp) in the PAM16 gene (presequence translocase-associated motor 16), that is more commonly referred to as MAGMAS (Mitochondria-associated granulocyte macrophage CSF-signaling molecule), was predicted to have pathogenic effects using different bioinformatics prediction softwares, such as: Mutation Taster (disease causing variation with a probability value of 0.99999957), SIFT (damaging effect, score 0.006) (Kumar et al., 2009b) , PROVEAN (deleterious effect, score 24.021) (Choi et al., 2012) and PolyPhen-2 (possibly damaging effect, score 0.728 HumDiv). By Sanger sequencing, we confirmed the segregation of the c.226A.G transition in MAGMAS with the disease in both families (Figure A.3A): the mutation was homozygous in the patients, heterozygous in the parents and in the unaffected sibs in both families. We also demonstrated that the variation was absent in a set of 550 Lebanese control chromosomes. Protein sequence alignments showed that the Asparagine at codon 76 in MAGMAS is well conserved among mammals and vertebrates (Figure A.3B).

The two patients described here are not related. However, they display the same homozygous missense mutation, thus suggesting the existence of a founder mutation. In order to verify this hypothesis, we used both exome data and genotyping of STR markers. By analyzing exome data, we were able to identify 11 regions of shared homozygosity on chromosome 16. However, only three regions displayed a size over 100 kb, of whom two were containing or contiguous to the MAGMAS locus (Chr16: 4015729–102036 (~3 Mb) and Chr16: 8728764–

8840025 (~1.1 Mb)). Also, by using PLINK (Purcell et al., 2007) , we were able to reconstruct haplotypes from exome data of both patients, and here again, we identified a 3.6 Mb shared homozygous region located between SNP markers rs3730119 and rs4616299 on chromosome 16p13.3 (Chr16: 4,057,603–7,657,432). Genotyping of microsatel- lite markers allowed to refine this region and define a minimal ancestral common homozygous haplotype spanning 1.9 Mb between markers D16S758 and D16S243 containing MAGMAS (Figure SA.1). Altogether, our results demonstrate that the c.226A. G transition in MAGMAS (p.Asn76Asp) is likely to be pathogenic in both families. MAGMAS is located on chromosome 16p13.3 and is composed of five exons. It encodes MAGMAS, an essential component of the pre-sequence translocase-associated protein import motor (PAM) that regulates pre-protein translocation into the mitochondrial matrix (Frazier et al., 2004) .

Gene expression analysis

A naturally occurring read-through transcription between the neighboring CORO7 (coronin 7) and MAGMAS genes on chromosome 16 is stated in the Genome UCSC browser (NM_001201479). To determine if the fusion protein is involved in the pathogenesis of the severe skeletal manifestations observed in the affected cases, we checked the expression of CORO7, MAGMAS and CORO7-MAGMAS fusion transcript in human cultured control cells (fetal chondrocytes, osteoblasts and skin fibroblasts) by Reverse Transcriptase PCR. CORO7, MAGMAS and CORO7-MAGMAS were all expressed in skin fibroblasts. MAGMAS expression in bone and cartilage was confirmed. On the other hand, neither CORO7 nor CORO7-MAGMAS was expressed in significant amounts in chondrocytes and osteoblasts (Figure A.4A). This result was confirmed by quantitative real-time RT-PCR (Figure A.4B) : The ratio of MAGMAS in chondrocytes versus in fibroblasts is 0.82 whereas the ratios of CORO7 and CORO7- MAGMAS in chondrocytes versus in fibroblasts are 0.24 and 0.12, respectively.

Immunolocalization of MAGMAS protein in wild type growth plates

To further confirm the specific involvement of MAGMAS in skeletogenesis, we performed immunohistochemical staining of fixed femurs isolated from wild type (WT) mice at different developmental stages from E16.5 to 2 weeks of age. Type X collagen staining was also

performed in order to mark the terminally differentiated hypertrophic chondrocytes. MAGMAS protein was expressed in bone and cartilage. Expression of the protein was mostly observed in the hypertrophic zone of the cartilage mainly at postnatal ages and in the primary spongiosa of trabecular bone. Its expression in proliferative chondrocytes and in the resting zone of the cartilage was not detected (Figure A.5).

Functional complementation

Human and murine forms of MAGMAS present a very high degree of homology. Of the 125 amino acids comprising the mature protein, 120 are identical (Jubinsky et al., 2001) . The evolutionary conservation of MAGMAS from yeast to mammals (Peng et al., 2005) and the lethal phenotype of its deletion (pam16D) in Saccharomyces cerevisae (Winzeler et al., 1999) and Drosophila (Becker et al., 2001) imply a fundamental role of this protein in cell growth and development. Previous studies have shown that human MAGMAS exhibits a complete growth complementation of deleted pam16 yeast cells at all temperatures (Sinha et al., 2010) .

In order to obtain in vivo evidence of the deleterious nature of the identified mutation, we performed a yeast complementation assay using genetic approaches as described previously by Sinha et al. (Schatz et Dobberstein, 1996) . We first found that at 28°C (301.15°K), both WT and mutant MAGMAS genes were able to complement the growth defect of the pam16D cells since spores expressing $MAGMAS_{WT}$ and $MAGMAS_{Asn76Asp}$ were viable when spotted onto G418 selective medium (Figure A.6A-B). To figure out if the Asn76Asp mutant confers a deleterious growth phenotype at non-permissive temperatures, tenfold serial dilutions of pam16D spores expressing WT or mutant MAGMAS were subsequently spotted onto fermentable (YPD) and non-fermentable (YPG) growth media and incubated at different temperatures for 3 days. As shown in Figure A.6C, the mutant strain pam16D-$MAG_{Asn76Asp}$ that presented a normal growth phenotype on both YPD and YPG media at 28°C (301.15°K), grew more slowly than pam16D-MAG_{WT} on YPD at 34°C (307.15°K) and 36°C (309.15°K). A more severe growth defect of pam16D-$MAG_{Asn76Asp}$ on YPG was evidenced by a very slow-growth at 34°C (307.15°K), and inviability at 36°C (309.15°K). To confirm this temperature-sensitive growth pheno- type, we secondly tested the ability of $MAGMAS_{Asn76Asp}$ to rescue the inviability of the pam16D strain by plasmid shuffling

using the haploid YPH499pam16D strain transformed with a plasmid encoding the yeast WT PAM16 (yPAM16). Unlike WT MAGMAS expressing strains, the mutant pam16D-MAG$_{Asn76Asp}$ exhibited an obvious slow-growth at 34°C and was unable to grow at 36°C (309.15°K) on YPD (Figure A.6D). On the other hand, pam16D- PAM16-MAG$_{Asn76Asp}$ cells carrying a WT copy of yPAM16 in addition to the mutant MAGMAS expressing plasmid, showed a normal growth phenotype at all temperatures.

Protein import and MAGMAS expression in yeast extracts

MAGMAS is an essential component of PAM, a peripheral mitochondrial inner membrane complex, that is crucial for preprotein translocation into the mitochondrial matrix (Schatz et al., 1996 ; Frazier et al., 2004) . Preproteins targeted to the matrix cross the outer mitochondrial membrane through the TOM complex then pass through the inner membrane via the TIM23^{MOTOR} complex which is composed of the TIM23 translocase associated with the PAM complex (Wagner et al., 2009) (Figure SA.2). To assess the transport function of the mutant MAGMAS$_{Asn76Asp}$, we evaluated the protein translocation activity by following the mitochondrial import of the Hsp60 protein. Non-processed mitochondrial Hsp60 was accumulated in significant amounts in the pam16D-MAG$_{Asn76Asp}$ strain when the mutant phenotype was induced (37°C, 310.15°K), but was undetectable in WT pam16D-PAM16 and pam16D-MAGMAS expressing cells (Figure A.7A). We then evaluated the expression levels of MAGMAS protein in the different strains, by performing a Western blot using Human anti-MAGMAS antibody. Immu- noblotting detected a lower level of mutant MAGMAS when the strain was shifted at the non-permissive temperature, compared to the amount of MAGMAS$_{WT}$ in the same condition (Figure A.7B).

Mitochondrial morphology and peroxisomal biogenesis defects in the MAGMAS$_{Asn76Asp}$ expressing strain

Previous studies in mammalian cells, yeast and drosophila have shown the crucial role of MAGMAS and its respective homologs in mitochondria biogenesis (Frazier et al., 2004 ; Sinha et al., 2010 ; Roy et al., 2012) . Altered mitochondria mor- phology observed in pam16 yeast mutants and in Blp-knowdown drosophila cells led us to study the aspect of these organelles in our mutant strains (Roy et al., 2012 ; Short et al., 2012) . Microscopic

examination of mitochondria revealed unhealthy features in the mutant strain consisting in punctuate and highly fragmented mitochondria that differed from the normal reticulated mitochondrial network observed in WT MAGMAS expressing cells. This unusual aspect was more obviously seen in the mutant strains shifted at the non-permissive temperature (Figure A.8A).

Furthermore, the recent finding of impaired peroxisomes biogenesis in pam16 yeast mutants (Short et al., 2012) prompted us to check the morphology of these organelles in our mutant strains. A remarkable aspect (vacuole labeling), indicative of pexophagic degradation, was only observed in pam16D-MAG$_{Asn76Asp}$ cells grown at the non-permissive temperature compared to pam16D- MAG$_{WT}$ cells (Figure A.8B).

DISCUSSION

Here, we report the identification of a homozygous missense MAGMAS mutation, (p.Asn76Asp), in patients from two unrelated Lebanese families, affected with a rare lethal spondylometaphyseal dysplasia recently described by Mégarbané et al. (Megarbane et al., 2008a ; Megarbane et al., 2014) . Byeconstructing haplotypes from exome and STR genotyping data from patients F1-IV.3 and F2-IV.3, we were able to identify a minimal common ancestral homozygous haplotype around MAGMAS, spanning 1.9 Mb on chromosome 16p13.3, suggesting the existence of a founder mutation in gene. MAGMAS, also referred to as PAM16, was first identified in 2001 by Jubinsky et al. (Jubinsky et al., 2001) , as a gene encoding a novel mitochondria-associated protein involved in granulocyte macrophage colony stimulating factor (GM-CSF) signal transduction and expressed by nearly all mammalian cells. The observation of a MAGMAS mutation in a severe chondrodysplasia supports a specific role of MAGMAS protein in the endochondral ossification process. Its expression in the notochord and the somites has been reported at day 12.5 embryo and in the cartilage of the developing ribs at day 14.5 embryo (Jubinsky et al., 2003) . However, its involvement in skeletogenesis has never been described to date. Our findings demonstrate that MAGMAS is expressed in trabecular bone and cartilage and more specifically by differentiated chondrocytes localized in the hypertophic zone and by osteoblasts at early developmental stages underlining its key role in skeletogenesis. On the other hand, the implication of the fusion protein CORO7-MAGMAS in the pathogenesis of the skeletal dysplasia has been ruled out, as the related transcript was not well expressed in human fetal chondrocytes and human osteoblasts.

Besides the skeletal features, respiratory insufficiency (F1-IV.1 and F1-IV.3) or heart failure (F2-IV.2 and F2-IV.3) were observed, leading to early deaths of the patients (Megarbane et al., 2008a ; Megarbane et al., 2014) . Of note, an elevated MAGMAS level was detected in murine heart and lung in previous studies (Jubinsky et al., 2003) . These features may also support a key role of MAGMAS in heart and lung.

MAGMAS consists of two main functional domains: a N-terminal hydrophobic region

including a membrane association domain (M) and a predicted mitochondria targeting region (T), and a C-terminal J-like domain (J) characterized by three a-helical segments (Frazier et al., 2004 ; D'silva et al., 2005) . Importantly, the p.Asn76Asp mutation lies in the helix II of MAGMAS J-like domain. The latter domain forms, with the J-domain of DNAJC19 (ortholog of yeast Pam18 which is another essential component of PAM), a stable heterodimeric subcomplex (Chacinska et al., 2009 ; Elsner et al., 2009 ; Wagner et al., 2009) (Figure SA.2). Mutations in DNAJC19 preventing MAGMAS/DNAJC19 dimerization have been shown to lead to dilated cardiomyopathy with ataxia syndrome (DCMA) (Davey et al., 2006 ; Sparkes et al., 2007) . Considering the localization of the p.Asn76Asp mutation in the J-like domain of MAGMAS, a similar phenotype was likely to be observed in our reported patients. Cardiomyopathy was observed exclusively in patients of the second family and ataxia was not present but all 4 cases died early before 2 years of age. On the other hand, the absence of DCMA in affected cases may also suggest a redundancy in the function of the subcomplex MAGMAS/DNAJC19 in some tissues.

We further demonstrated that the p.Asn76Asp mutation confers to yeast strains a temperature-sensitive phenotype characterized by a very slow-growth at 34°C (307.15°K) and inviability at 36°C (309.15°K), on both fermentable and non-fermentable growth media. Remarkably, the deleterious effect of the mutation was only observed at a homozygous state, which is in correlation with the recessive mode of inheritance of the spondylodysplastic dysplasia reported in this study.

Our findings of impaired import of mitochondrial-matrix proteins in the mutant strains shifted at non-permissive temper- ature confirmed the deleterious nature of the identified mutation. Likewise, it has been demonstrated that aa alterations in MAGMAS J-like domain resulted in temperature sensitivity and in vivo protein import defects in yeast cells (D'silva et al., 2008 ; Sinha et al., 2010) . The low expression level of MAGMAS detected in the mutant strain at the non-permissive temperature indicates that the p.Asn76Asp mutation confers protein instability. Additionally, we observed unusual punctuated and highly fragmented mitochondria together with enhanced pexophagy in the mutant MAGMAS strain at the non-permissive temperature, supporting a cell death induction in the yeast mutant strain. These data are

consistent with the finding of Roy et al., in Drosophila (Roy et al., 2012) and confirm the crucial role of MAGMAS for cell integrity and viability.

The association between a mitochondrial protein and a skeletal dysplasia is striking as skeletal manifestations are rarely seen in mitochondrial disorders. To date, the only severe skeletal dysplasia identified as having a possible mitochondrial basis is the anauxetic dysplasia [MIM 607095], which is allelic to two milder forms of SD, namely metaphyseal dysplasia Mc Kusick type [MIM 250250] and metaphyseal dysplasia without hypotrichosis [MIM 250460]). The gene responsible for these SD is RMRP [MIM 157660], which encodes the RNA component of the mitochondrial RNA processing endoribonuclease (MRP) complex (Ridanpaa et al., 2001 ; Bonafe et al., 2002 ; Thiel et al., 2005) . Of note, RNase MRP is not exclusively found in the mitochondria but also in the nucleus. Previous studies have shown that mitochondrial and nucleolar RNase MRP have identical RNA components but distinct enzymatic activities and protein components (Lu et al., 2010) . The recent identification of mutations in POP1 (Processing of precursor 1) [MIM 602486], which encodes a protein component of the nuclear ribonucleoprotein complex in two siblings with anauxetic dysplasia (Glazov et al., 2011) may support the involvement of the nuclear ribonucleoprotein complex rather than the mitochondrial complex in the etiology of these severe skeletal dysplasias. Similarly, besides its role in preprotein import into mitochondria, Magmas may also be involved in additional important mechanisms as suggested by Short et al. (Short et al., 2012) . The detection of different human transcripts of MAGMAS, with some lacking the mitochondrial transit peptide (mTP) leader sequence (Peng et al., 2005) may support a novel extra mitochon- drial activity of this protein.

In conclusion, this is the first report linking a mutation in MAGMAS with a human developmental disorder and supporting a key role for this mitochondrial protein in the ossification process.

MATERIALS AND METHODS

Patients

Informed consents for participation, sample collection and photographs publication were obtained from both families in compliance with the ethics guidelines of ‘‘Conseil de recherche de l’Université Saint-Joseph’’ (number PTS5). Peripheral blood samples were collected from all available family members.

DNA extraction and exome sequencing

Genomic DNA was extracted from peripheral blood samples by standard salt-precipitation methods (Miller et al., 1988) . Exome sequencing was carried-out on 3 individuals from family 2 (F2-III.1, F2-III.2 and F2-IV.3) and the proband from family 1 (individual F1-IV.3) (Figure A.2). Targeted exome sequencing, library preparation, capture and sequencing were performed by the French National Genotyping Center (CNG). Exomes were captured and enriched using the in solution Agilent SureSelect Human All Exon kit v3.0 (Sulonen et al., 2011) and then sequenced on an Illumina HiSeq2000, using a paired-end 100-bp read sequencing protocol. Image analysis and base calling were performed using the Illumina Data Analysis Pipeline Software 1.5 with default parameters. Raw data were mapped to the current built of the human genome (hg19) by using BWA 0.75 (Li et al., 2009a) . Variant calling was subsequently performed using GATK 2.5.2 (Mckenna et al., 2010) and annotation was done with ANNOVAR (Wang et al., 2010) . In order to predict the deleterious effect of the identified sequence variations, different bioinformatics tools were applied; such as MutationTaster (http://www.mutationtaster.org/), SIFT (http://sift.bii.a-star.edu.sg/) (Kumar et al., 2009b) , PROVEAN (http://provean. jcvi.org/index.php/) (Choi et al., 2012) and PolyPhen-2 (http://genetics.bwh. harvard.edu/pph2/).

Linkage analysis

Searching for shared homozygous regions using exome data and reconstructing haplotypes were realized using the free software package PLINK (Purcell et al., 2007) .

Genotyping of STR markers

The following STR markers flanking the PAM16 gene and spanning a 8.7 Mb region on chromosome 16p13.3, were amplified using fluorescently labelled primers: D16S521 (AFMA1 39WG1), D16S3024 (AFMA134XC1), D16S3070 (AFMA33Y H1), D16S3027

(AFMA154WC9), D16S758 (CHLC.ATA24C09), D16S3084 (AFMB080YH5), D16S3072 (AFMB015WA9), D16S 3134 (AFMA059WF9), D16S510 (AFM312VD5), D16S423 (AFM 249YC5), D16S423 (AFM249YC5), AFM339XG1 (RH4035) and D16S3135 (AFMA059XC1). Marker heterozygosity rate ranged from 0.71 (D16S521 and D16S510) to 0.87 (D16S3082 and D16S3027). PCR amplifications were performed in a 30-ml reaction mix, containing 50 ng of genomic DNA, 0.5 units of Taq polymerase (Invitrogen), 1.5 mM $MgCl_2$, 5 pmol of each primer, and 0.2 mM of each dNTP. After an initial dena- turation step of 5 min at 96uC, 35 cycles of amplification (94°C for 30 sec; 55°C for 30 sec; 72°C for 30 sec) were performed, followed by a final elongation step of 5 min at 72°C. After amplification, PCR products were subjected to denaturing electrophoresis on an ABI PRISM 3100 DNA Sequencer (Applied Biosystems).

Capillary sequencing

Genomic and cDNA sequences of MAGMAS (also known as PAM16) were obtained from UCSC Genomic Browser on Human February 2009 (NM_016069). Primers used for PCR amplification were designed using Primer3 software (http://frodo.wi.mit.edu) to amplify the region surrounding the mutation detected by exome sequencing in exon 4. PCR products were purified by exonuclease I/Shrimp Alakaline Phosphatase treatment (ExoSAP-IT; Fisher Scientific SAS, Illkirch, France) according to the manufacturer's instructions and both strands were sequenced using the Big Dye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems). Sequence reactions were purified on Sephadex G50 (Amersham Pharmacia Biotech, Foster City, CA) and capillary electrophoresis was performed on Genetic Analyser 3100 (Applied Biosystems). Electropherograms were analyzed on the Sequence Analysis Software version 5.2 (Applied Biosystems) and aligned with the WT MAGMAS gene sequence using ChromasPro version 1.22 (Technelysium, Queensland, Australia).

Reverse Transcriptase PCR

Total RNAs were extracted from human primary cultured control cells (fetal chondrocytes, osteoblasts and skin fibroblasts) via the RNeasy Mini Kit (Qiagen). Complementary DNA (cDNA) was synthesized using the SuperScript II Reverse Transcriptase (Invitrogen Life Technologies, Carlsbad, CA, USA) according to the manufacturer's protocol. Sequences of MAGMAS (NM_016069), CORO7 (NM_001201472), CORO7-MAGMAS (NM_001201479)

and GAPDH (NM_002046) were obtained from UCSC Genomic Browser on Human February 2009 and specific primers for cDNA amplification were designed through the Primer3 software as follows: MAGMAS-F: 5'-TACCTGGCCCAGATCATTGT-3', MAGMAS-R: 5'-GCATCTGCCCTTTTTCTCTG-3'; CORO7-F: 5'-GTCCTGGTGTACTGGGCATT-3', CORO7-R: 5'-AGAGTCAGGGTCCAGCAGAG-3'; CORO7-MAGMAS-F: 5'-CTCAGCGCAGTACCTGGAAG-3', CORO7-MAGMAS-R: 5'-GCCACCCACGGATTTATCAT-3' and GAPDH-F: 5'-ATGTTCGTCATGGGTGTGAA-3', GAPDH-R: 5'-TTCAGCTCAGGGATGACCTT-3'.

Quantitative Real–Time RT-PCR

Reactions were conducted in a 96-well plate with the ABI 7500 Sequence Detection System (Applied Biosystems). Primers against MAGMAS, CORO7 and CORO7-MAGMAS were designed using Primer Express software (Applied Biosystems) (Figure SA.3). PCR was performed in a 20-ml reaction volume containing 10 ml Power SYBR-GreenPCR Master Mix of SYBR Green PCR Master Mix buffer (Applied Biosystems, Foster City, CA, USA), 10 pmol forward and reverse primers and 5 ng of RNAse H-treated (Invitrogen Life Technologies) cDNAs obtained from reverse transcription of DNAseI-treated total RNA extracted from control fibroblasts and chondrocytes cDNA. The reaction cycling conditions were 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and at 60°C for 1 min. Sequence Detection Software was used for exporting the threshold cycle data and further analyzing differences in threshold cycle values (DCt) between the test locus and the control locus. The expression values are normalized to that of GAPDH. For each gene, the expression ratio for a sample was calculated as the ratio between the average mRNA signal in the sample and the signal from fibroblasts

Immunohistochemical staining

Femurs isolated from WT mice (at different developmental stages including E16.5-day, birth and 2 weeks of age) were fixed with 4% paraformaldehyde and embedded in paraffin. Femur sections were stained with Hematoxylin and eosin (H&E) using standard protocol for histological analysis or were subjected to immunohistochemical staining. For immunohistochemistry, sections were stained with Ab specific to MAGMAS (Abcam plc., UK) at 1/200 dilution or with Ab specific to Col X (Quartett) at dilution 1/20 using DAKO

EnVision kit. Images were captured with an Olympus PD70- IX2- UCB microscope. All experimental procedures were approved by the French Animal Care and Use Committee.

Plasmids, yeast strains and growth conditions

Plasmids and yeast strains used in this study are described in Table A.1. The yeast expression vector pRS415-WT-MAGMAS, encoding full-length human WT (WT) MAGMAS under the TEF promoter was provided by D'Silva et al. (Sinha et al., 2010) . The pRS415- sn76Asp-MAGMAS plasmid carrying the point mutation p.Asn76Asp was generated from pRS415-WT-MAGMAS through a site directed mutagenesis using the QuickChange protocol (Stratagene) according to the manufacturer's recommendations. Direct sequencing was performed to confirm the successful introduction of the mutation. For construction of yeast strains expressing the human WT/mutated MAGMAS, we used Saccharomyces cerevisiae Dpam16 strains (provided by Delahodde A). For the yeast complementation assay, the diploid strain BY4743 carrying the deleted pam16 allele (pam16D::kanMX6) in a heterozygous state was first transformed with pRS415-WT-MAGMAS or pRS415- Asn76Asp-MAGMAS plasmids, accordingly and then subjected to sporulation and tetrad dissection using standard techniques (Schiestl et Gietz, 1989) . Spores from a single ascus were first grown at 28°C (301.15°K) in YPD media then replica-plated to G418 plates for the selection of spores carrying the deleted pam16 allele and to Leucine drop-out plate for the selection of pRS415MAGMAS positive spores. The haploids BY-pam16D-MAG$_{WT}$ and BY-pam16D-MAG$_{Asn76Asp}$ were thus generated and spotted onto YPD and YPG to analyze their phenotype. For the construction of YPH499-pam16D- pPAM16-MAG$_{WT}$ and YPH499-pam16D-pPAM16-MAG$_{Asn76Asp}$, the haploid YPH499pam16D::ADE2 strain containing a plasmid encoding the yeast PAM16 gene (kindly provided by Guiard B), was transformed either with pRS415-WT-MAGMAS or pRS415- Asn76Asp-MAGMAS plasmids, respectively then patched onto the selection plates (lacking leucine and uracile). Transformants were subsequently replicated to 5-Fluoroorotic (5-FOA) containing plates, recovered and grown at 28°C (301.15°K), thus generating YPH499-pam16D-MAG$_{WT}$ and YPH499-pam16D-MAG$_{Asn76Asp}$ strains that have lost the plasmid containing the yeast PAM16 gene and the URA3 marker gene (Sikorski et Boeke, 1991) . Ten fold serial dilutions of the different generated strains were subsequently spotted onto YPD plates then incubated at different

temperatures for 3 days. Unless otherwise noted, yeast cultures were grown at 28°C (301.15°K). YPD (1% bactopeptone, 1% yeast extract, and 2% glucose) and YPG (1% bactopeptone, 1% yeast extract, and 2% glycerol) were used as rich fermentable and non-fermentable growth media respectively. W0 (0.67% yeast nitrogen base without aa and 2% glucose) supplemented with appropriate nutritional requirements according to the strains was used as a minimal medium. For solid media, 2% bacto agar (Difco) was added.

Protein import and western blotting

For the import activity analysis, YPH499pam16D-PAM16, YPH- pam16D-MAG_{WT} and YPH-pam16D-$MAG_{Asn76Asp}$ cells were first grown to $OD_{600\ nm}$ of 1.0 at 28°C (301.15°K) in rich medium then shifted to 37°C (310.15°K), for 8 h (28800 s) to induce non-permissive conditions. Cell pellets were suspended in ice-cold TCA buffer (TrisHCl pH 8 20 mM, Ammonium Acetate 50 mM, EDTA 2 mM, protease inhibitor solution, and TCA10%) then lysed by adding glass beads and shaking vigorously for 300s (5 fold 60s in ice between). After centrifugation of 600 s (13,000 rpm) at 4°C (277.15°K), the proteins were solubilized in TCA-Laemmli loading buffer (SDS/glycerol stock solution 48%, Tris EDTA 40%, b-mercaptoethanol 0.5%, PMSF 0.2%, protease inhibitor solution 0.2%) and separated by SDS-PAGE. The proteins were subsequently transferred to a nitrocellulose membrane that was first blocked, then incubated 1 h at room temperature with a blocking solution containing the monoclonal HSP60 Ab (provided by Dujardin G.) at 1/15000 dilution. After several washes with TBS1X (Tris Buffered Saline 1X), the membrane was immersed for 1 h at room temperature with HRP – conjugated secondary anti-mouse Ab (at 2/15000 dilution in a blocking solution) then washed again with TBS1X. Secondary antibodies bound to the membrane were detected by chemiluminescence using the Thermo Scientific PierceECL 2 according to the manufacturer's instructions. The same protocol was used for the immunodetection of MAGMAS protein in YPH499, YPH-pam16D-MAG_{WT} and YPH-pam16D-$MAG_{Asn76Asp}$strains. For this purpose, we incubated the membrane with a blocking solution containing the polyclonal anti-MAGMAS Ab (Abcam plc., UK) at a 1/1000 final concentration. Therefore, the secondary Ab used in this case was HRP–conjugated secondary anti-rabbit Ab (at 2/15000 dilution in a blocking solution). Phosphoglycerate kinase (PGK) and Porin were detected using anti-PGK (Invitrogen) and anti-Porin (Invitrogen),

accordingly.

Fluorescence microscopy

In order to label mitochondria, YPH-pam16D-MAG$_{WT}$ and YPH-pam16D-MAG$_{Asn76Asp}$ strains were transformed with the pYX232-mtGFP plasmid that encodes the mitochondrial matrix- targeted GFP (Westermann et Neupert, 2000) and grown for 2 days at 28°C (301.15u°K). Cells were secondly incubated overnight at the non-permissive temperature to induce the phenotype. For the visualization of peroxisomes, BY-pam16D-MAG$_{WT}$ and BY-pam16D-MAG$_{Asn76Asp}$ were transformed with pUG34DsRed.SKL encoding the red fluorescent protein DsRed containing the C-terminal peroxisomal targeting signal type 1 (PTS1, DsRed-SKL). Transformants were grown for 2 days at 28°C (301.15°K) in minimal medium. Harvested cells were washed in Phosphate-buffered saline 1X (PBS1X) then transferred to microscopic slides for imaging. For observations at high temperature, cells grown 2 days at 28°C (301.15°K) in minimal medium, were shifted to 37°C (310.15°K), for 6 hours then harvested, washed and transferred to slides for visualization. The slides were examined with a DMIRE2 microscope (Leica, Deerfield, IL). For GFP fluorescence, filters used were 450/490 nm for excitation and 500/550 nm for emission whereas for Ds.Red fluorescence, filters were 542/582 nm for excitation and 604/644 nm for emission. Images were captured using a CCD camera (Roper Scientific, Tucson, AZ) and Metamorph software (Universal Imaging, WestChester, PA) was used to deconvolute Z- series and treat the images.

Acknowledgments

We express our deepest gratitude and sympathy to the families for their full cooperation throughout the study. We are grateful to Mark Lathrop and to the Centre National de Génotypage (CNG) for exome sequencing. We are grateful to D'Silva, P., Guiard B., Winzeler E.A., Dujardin G. and Westermann B. for their generous gifts of plasmids, antibodies and yeast strains. We thank Le Goff C. for her guidance and assistance in the directed mutagenesis experiment.

CM was supported by the "Agence Universitaire de la Francophonie". We also thank the "Association Médicale Franco-Libanaise" and the "Conseil de Recherche de l'Université Saint-Joseph" for their funding. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Figure A.1 Radiological features of the patients F2-IV.3 and F1-IV.3.

Radiographs of patient F2-IV.3 at 9 months **(A, C, E)** and F1-IV.3 at birth **(B, F)** and at 3 months **(D)** show platyspondyly, square iliac bones, and delayed epiphyseal ossification.

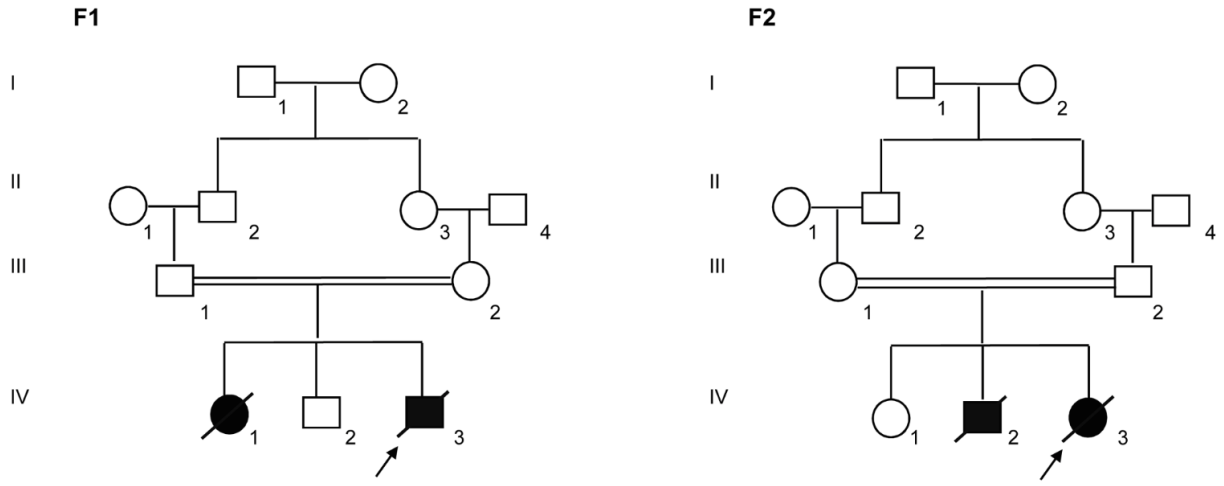
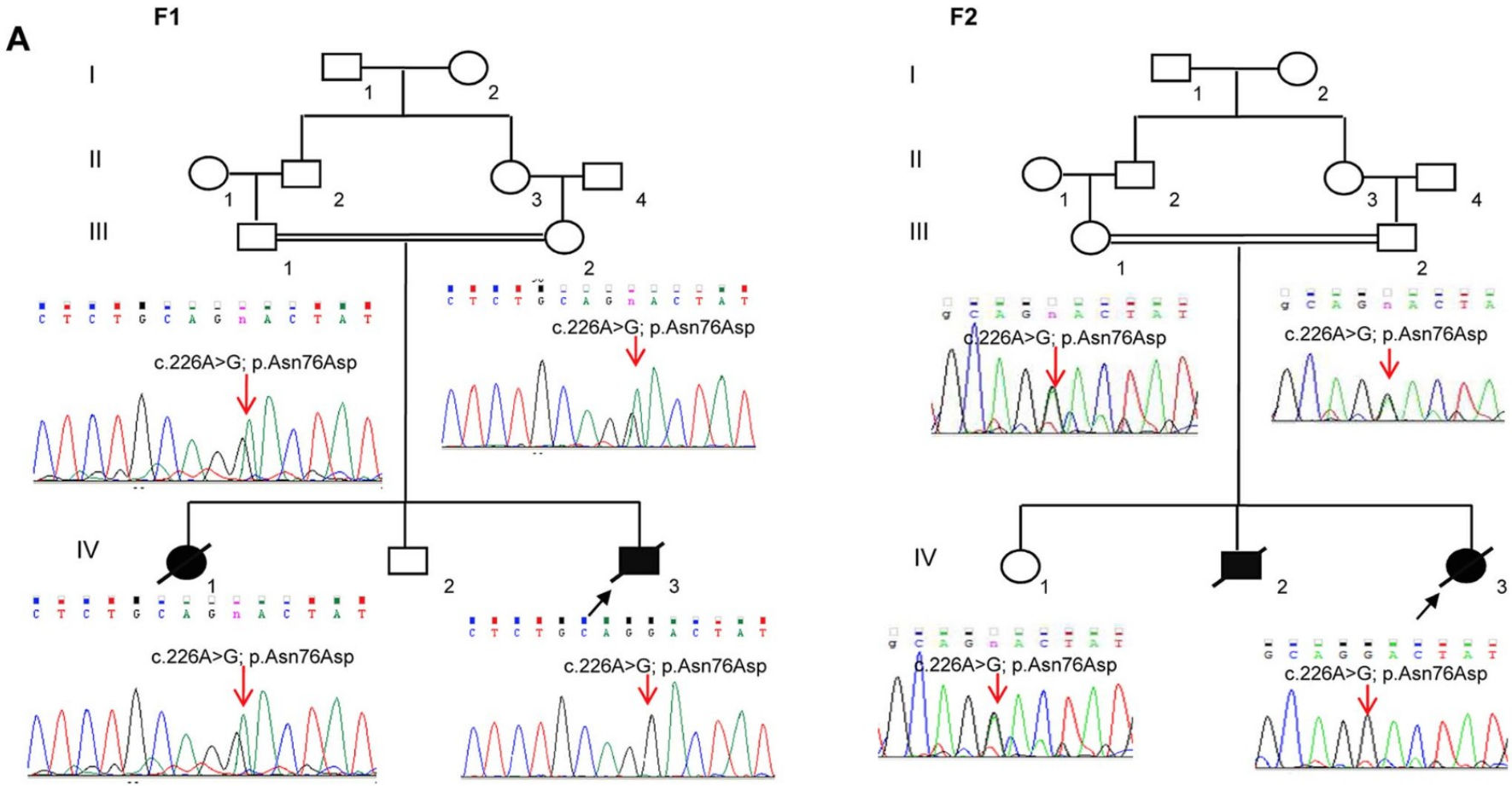


Figure A.2 Pedigrees of the families F1 and F2 included in this study.

Blackened symbols indicate affected individuals.


A
F1
I
II
III
IV
1
2
3
4
c.226A>G; p.Asn76Asp
c.226A>G; p.Asn76Asp
c.226A>G; p.Asn76Asp
c.226A>G; p.Asn76Asp
F2
I
II
III
IV
1
2
3
4
c.226A>G; p.Asn76Asp
c.226A>G; p.Asn76Asp
c.226A>G; p.Asn76Asp
c.226A>G; p.Asn76Asp

B

Figure A.3 The missense MAGMAS mutation c.226A.G (p.Asn76Asp).

(A) Sequencing chromatograms showing the segregation of the c.226A.G transition in MAGMAS with the disease in both families F1 and F2. n: heterozygous peak. **(B)** Multiple alignments between human MAGMAS protein and several orthologs showing that the Asparagine at codon 76 in MAGMAS is well conserved among mammals and vertebrates.

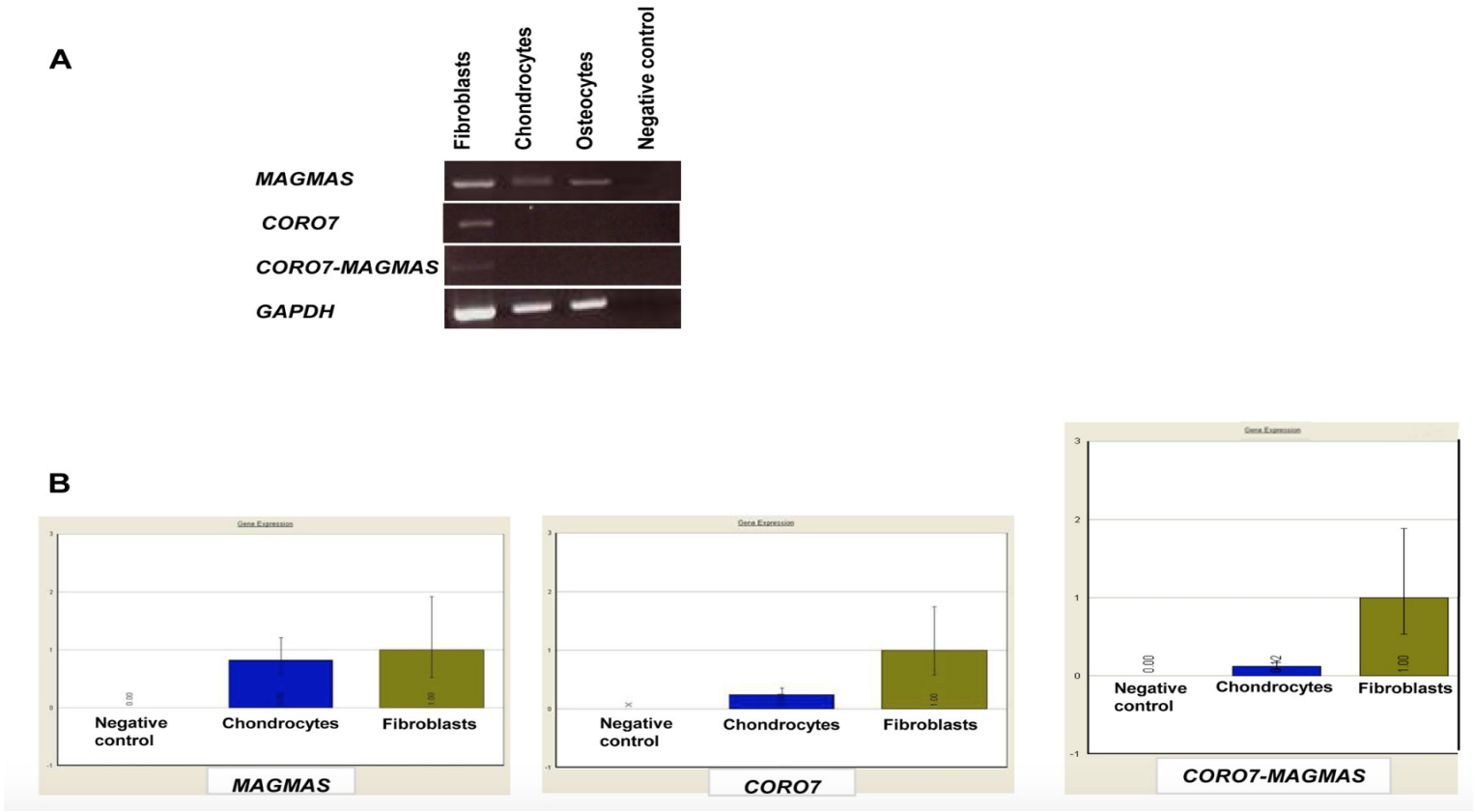


Figure A.4 Transcript expression analysis of MAGMAS, CORO7 and CORO7-MAGMAS. (A) Expression analysis of MAGMAS, CORO7 and CORO7- MAGMAS by RT-PCR in control fibroblasts, chondrocytes and osteoblasts, showing that, contrary to MAGMAS, neither CORO7 nor CORO7-MAGMAS was significantly expressed in chondrocytes and osteoblasts. (B) Quantitative expression analysis of MAGMAS, CORO7 and CORO7-MAGMAS showing a very low level of expression of CORO7 and CORO7-MAGMAS in chondrocytes compared to that of MAGMAS in the same sample.

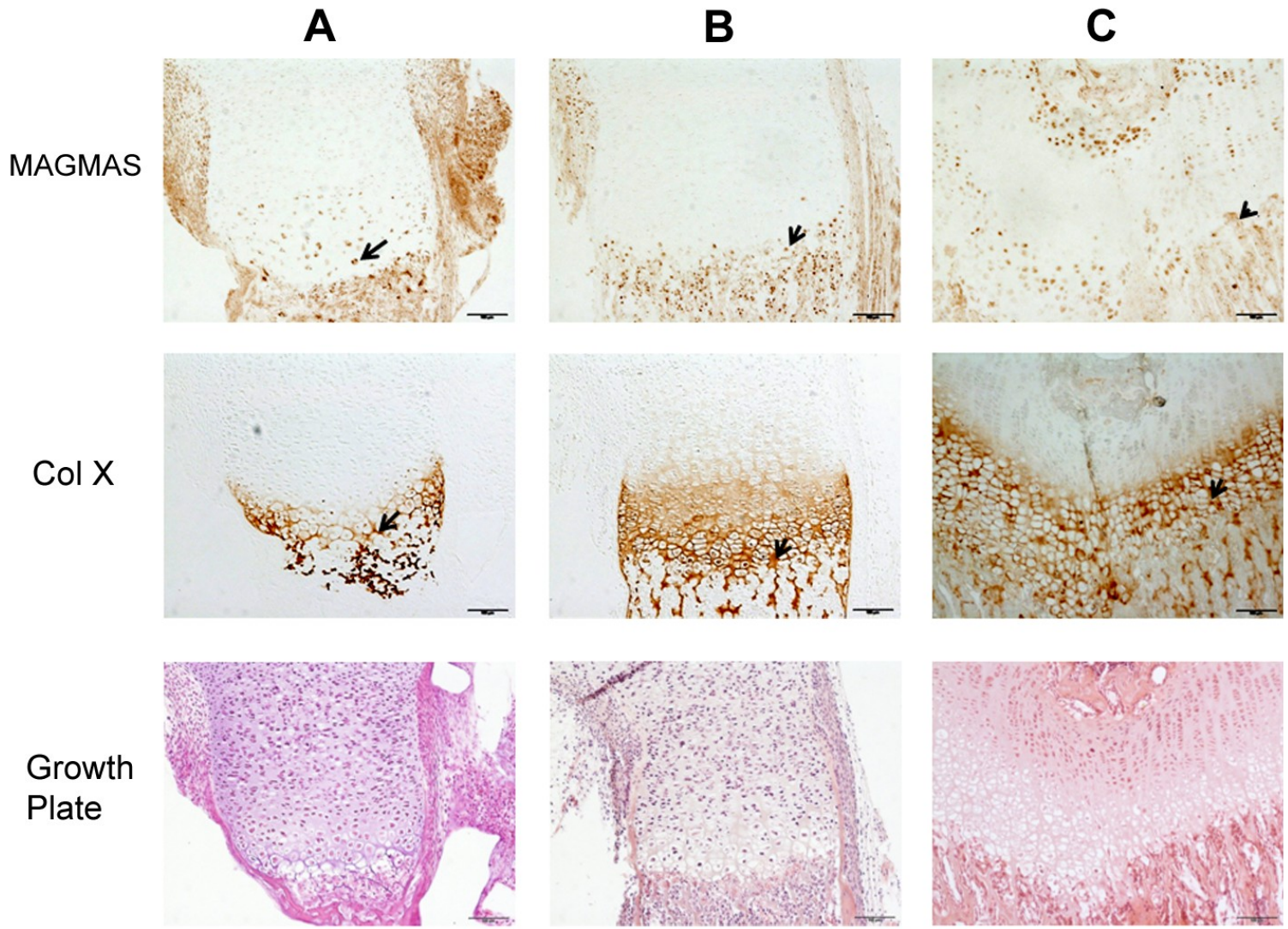


Figure A.5

Figure A.5 MAGMAS localization in control growth plates of mice.

Distal femurs of wild-type mice at different developmental stages; (A) 16.5- day embryo; (B) Newborn and (C) 2 week-old; were fixed and stained with anti-MAGMAS antibodies. Sections were also stained with anti-CoX (marker of differentiated hypertrophic chondrocytes) and with HES. MAGMAS expression is reflected by the brown precipitate resulting from the peroxidase reaction. MAGMAS was detected mainly in hypertrophic chondrocytes (black arrows) and osteocytes of control growth plates at the different developmental stages. Its expression was also detected in some type X collagen –negative chondrocytes that are initiating the process of differentiation.

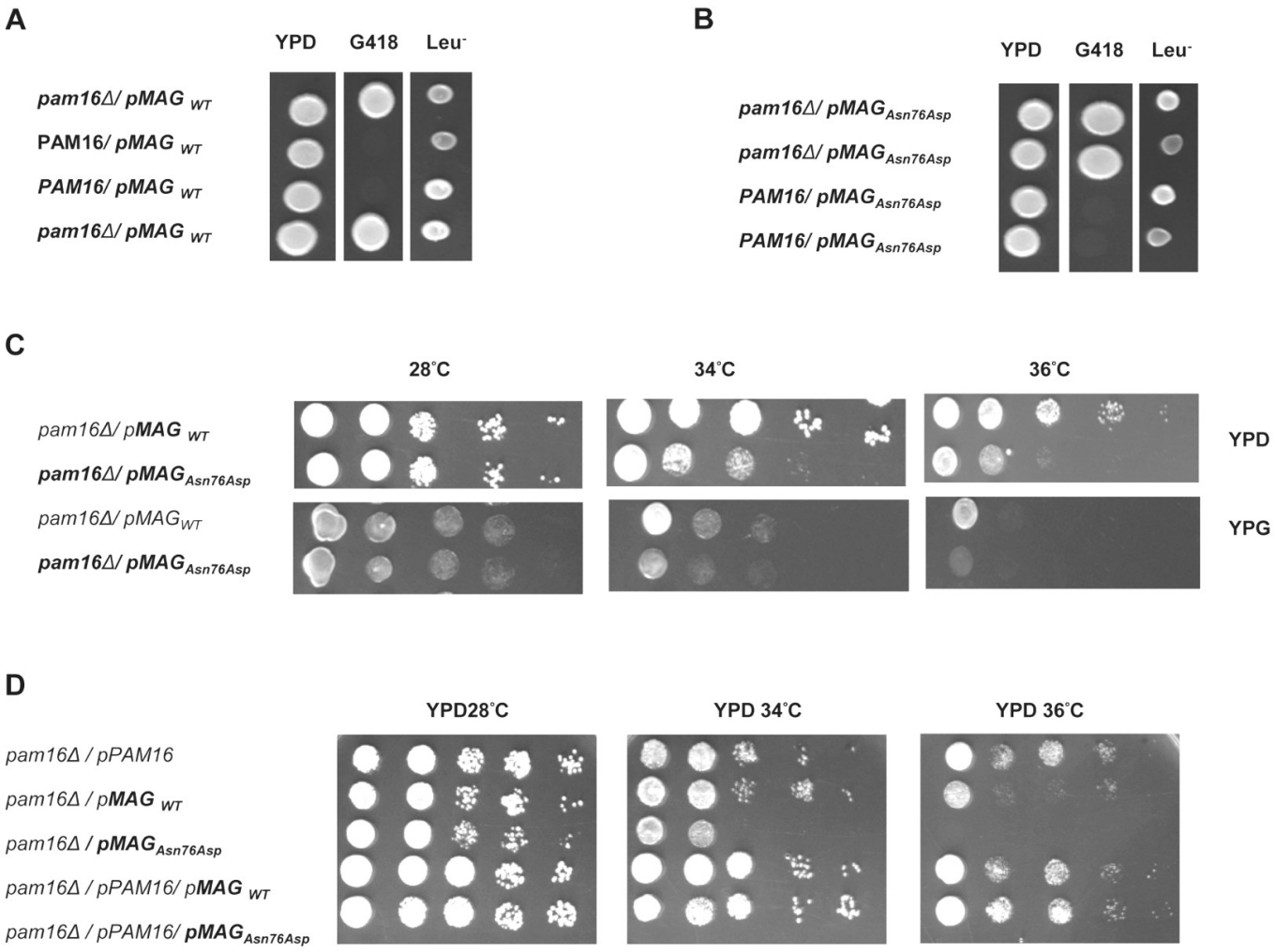

A
YPD
G418
Leu-
pam16Δ/ pMAG WT
PAM16/ pMAG WT
PAM16/ pMAG WT
pam16Δ/ pMAG WT
B
YPD
G418
Leu-
pam16Δ/ pMAGAsn76Asp
pam16Δ/ pMAGAsn76Asp
PAM16/ pMAGAsn76Asp
PAM16/ pMAGAsn76Asp
C
28˚C
34˚C
36˚C
pam16Δ/ pMAG WT
pam16Δ/ pMAGAsn76Asp
pam16Δ/ pMAGWT
pam16Δ/ pMAGAsn76Asp
YPD
YPG
D
YPD28˚C
YPD 34˚C
YPD 36˚C
pam16Δ / pPAM16
pam16Δ / pMAG WT
pam16Δ / pMAGAsn76Asp
pam16Δ / pPAM16/ pMAG WT
pam16Δ / pPAM16/ pMAGAsn76Asp

Figure A.6

Figure A.6 In vivo growth analysis.

(A–C) Diploid BY4743Dlip3/+ cells were transformed with **(A)** pRS415-WT-MAGMAS and **(B)** pRS415-Asn76Asp- MAGMAS, and then subjected to sporulation followed by tetrad dissection analysis. Spores from a single ascus grown on YPD were spotted onto: G418 for the selection of spores carrying the deleted pam16 gene (pam16D::kanMX6) and onto Leu drop-out media for the selection of MAGMAS positive spores. **(C)** Growth complementation analyses showing the temperature-sensitive growth phenotype of the mutant pam16D-MAG$_{Asn76Asp}$ in the BY4743 genetic background at 34uC (307.15uK) and 36uC (309.15uK) on both fermentative (YPD) and non-fermentative (YPG) media. **(D)** Haploid YPH499 pam16D/pPAM16 cells transformed with pRS415-WT-MAGMAS and pRS415-Asn76Asp-MAGMAS were tested for their ability to rescue the inviability of the pam16D strain by plasmid shuffling at different temperatures.

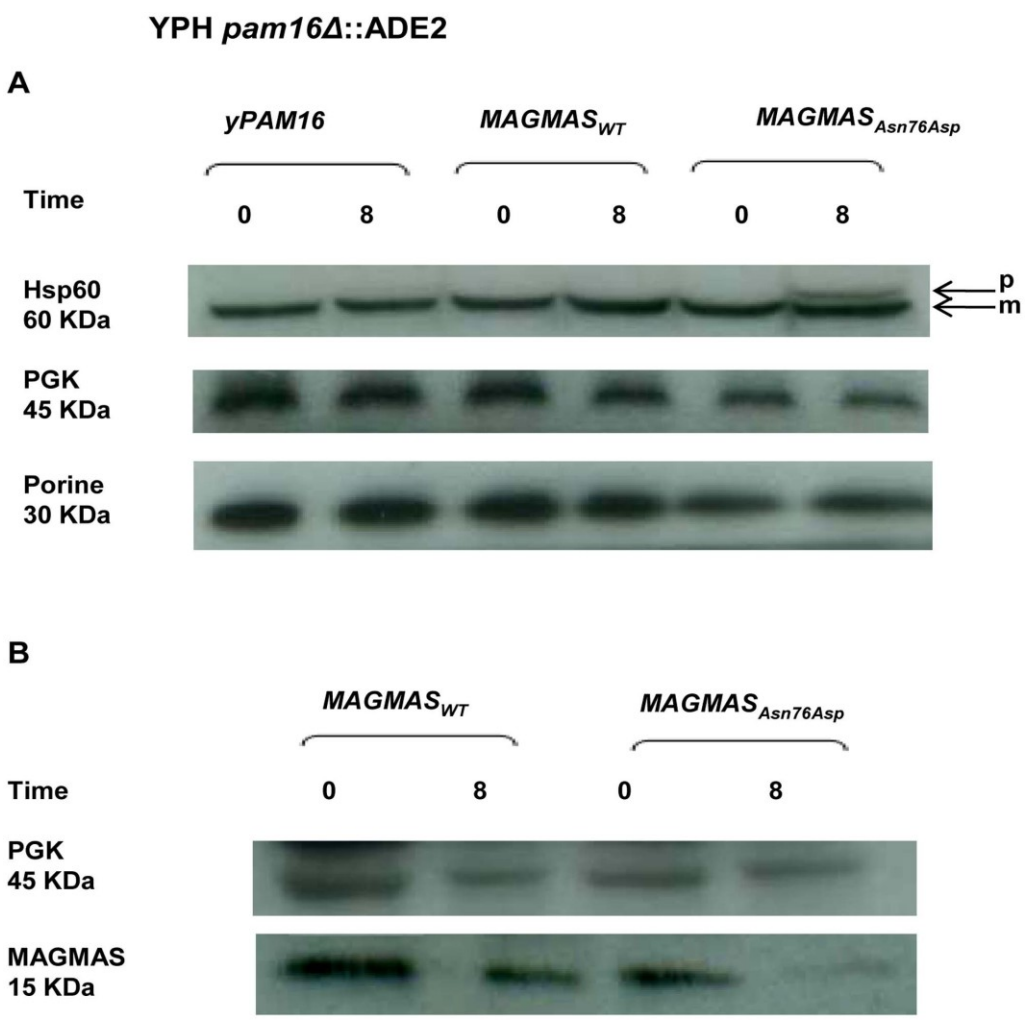

YPH pam16Δ::ADE2
A
yPAM16
MAGMAS WT
MAGMAS Asn76Asp
Time
0
8
0
8
0
8
Hsp60
60 KDa
p
m
PGK
45 KDa
Porine
30 KDa
B
MAGMAS WT
MAGMAS Asn76Asp
Time
0
8
0
8
PGK
45 KDa
MAGMAS
15 KDa

Figure A.7

Figure A.7 Preprotein translocation and MAGMAS expression analyses in yeast cells.

pam16D-MAG$_{WT}$ and pam16D-MAG$_{Asn76Asp}$ strains in the YPH499 genetic context, expressing wild-type and mutant MAGMAS, accordingly were constructed by plasmid shuffling (see Materials and Methods). Strains were grown at 28uC (301.15uK) then shifted for 8 h (28800 s) at 37uC (310.15uK) to induce the mutant phenotype. (A) Analysis of protein translocation in yeast cells: Non-processed precursor form and mature Hsp60 were detected by Western blotting using anti-Hsp60 antibodies. The non-processed Hsp60 precursor accumulated in the mutant strain after induction of the phenotype. Only the mature form of Hsp60 was detected in wild-type yPAM16 and MAGMAS expressing yeast cells. PGK was used as a loading control and Porin as a mitochondrial loading control. p: precursor form of Hsp60, m: mature form of Hsp60. (B) Immunoblotting using Human anti-MAGMAS antibodies detected a lower level of MAGMAS protein in the mutant strain when shifted at non-permissive temperature, compared to the amount in MAGMAS$_{WT}$ expressing strains.

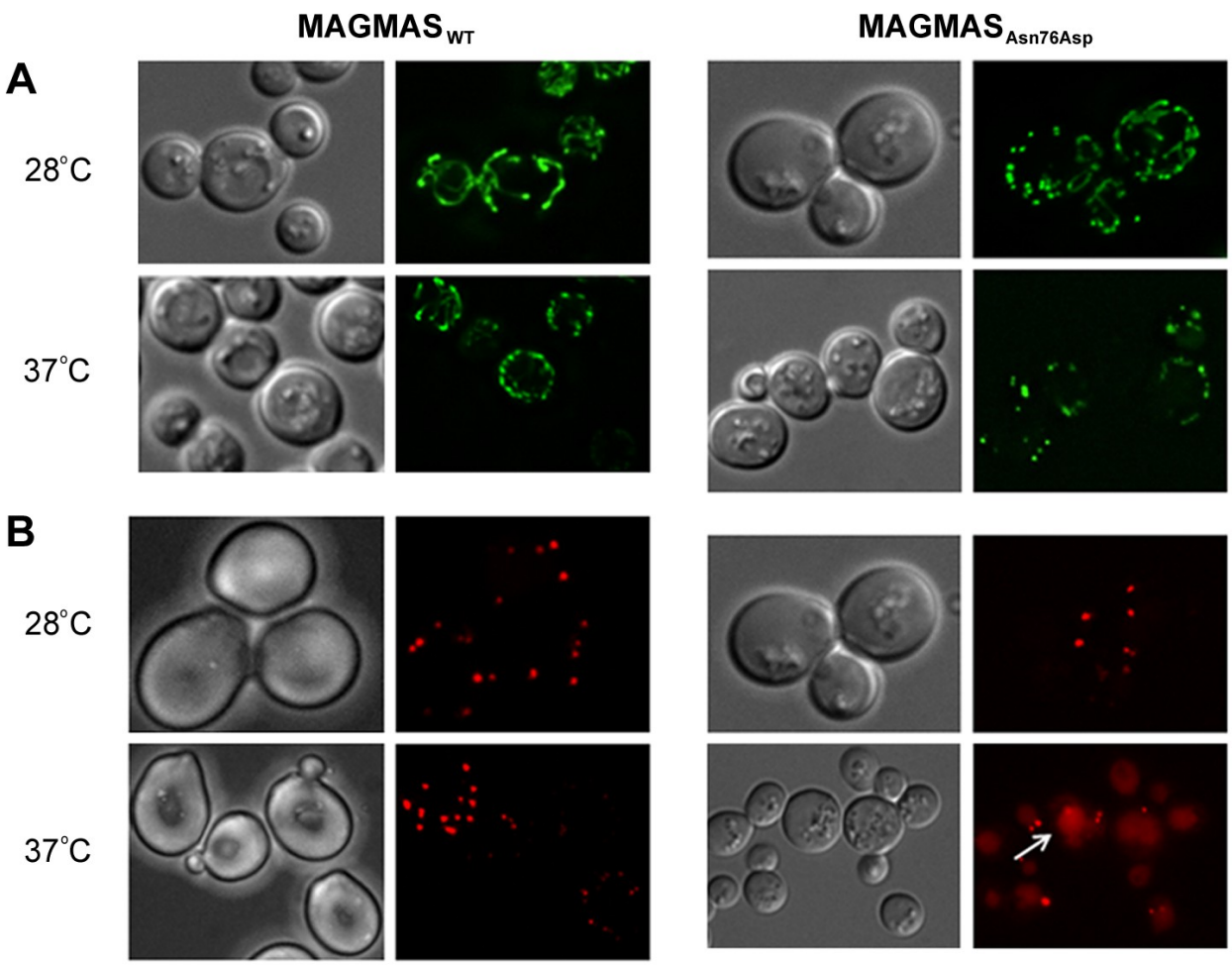


Figure A.8 Morphological defects of the yeast MAGMAS mutant strain.

(A) For mitochondria visualization, YPH499-pam16D-MAG$_{WT}$ and YPH499-pam16D-MAG$_{Asn76Asp}$ strains were transformed with pYX232-mtGFP plasmid; microscopy filters used were 450/490 nm for excitation and 500/550 nm for emission. Unlike the normal reticulated mitochondrial network observed in wild-type MAGMAS expressing cells, highly fragmented mitochondria were detected in the mutant strains grown at both permissive and non-permissive temperatures. (B) For peroxisomes visualization, BY- pam16D-MAG$_{WT}$ and BY-pam16D-MAG$_{Asn76Asp}$ were transformed with the pUG34DsRed.SKL reporter plasmid; filters were 542/582 nm for excitation and 604/644 nm for emission. A red halo (white arrow) indicative of pexophagic degradation was observed by mutant cells when grown at the non permissive temperature.

Table A.1 Yeast strains and plasmids used in this study.

Yeast strain or Plasmid	Genotype	Source
BY4743*pam16Δ::kanMX6/PAM16*	*leu2Δ/leu2Δ his3Δ/his3Δ ura3Δ/ura3Δ lys2Δ/LYS2 met15Δ/MET15 pam16Δ::kanMX6/PAM16*	Winzeler, E.A. [15]
YPH499*pam16Δ::ADE2-pPAM16*	*leu2Δ ade2Δ ura3Δ his3Δ trp1Δ lys2Δ pam16Δ::ADE2 [pPAM16 URA3]*	Guiard, B. (kind gift)
BY-*pam16Δ*-MAG$_{WT}$	*leu2Δ/leu2Δ his3Δ/his3Δ ura3Δ/ura3Δ lys2Δ/LYS2 met15Δ/MET15 pam16Δ::kanMX6[pRS415-WT-MAGMAS]*	This study
BY-*pam16Δ*-MAG$_{Asn76Asp}$	*leu2Δ/leu2Δ his3Δ/his3Δ ura3Δ/ura3Δ lys2Δ/LYS2 met15Δ/MET15 pam16Δ::kanMX6 [pRS415-Mut-MAGMAS]*	This study
YPH499-*pam16Δ*-MAG$_{WT}$	*leu2Δ ade2Δ ura3Δ his3Δ trp1Δ lys2Δ pam16Δ::ADE2 [pRS415-WT-MAGMAS]*	This study
YPH499-*pam16Δ*-MAG$_{Asn76Asp}$	*leu2Δ ade2Δ ura3Δ his3Δ trp1Δ lys2Δ pam16Δ::ADE2 [pRS415-Mut-MAGMAS]*	This study
YPH499-*pam16Δ*-pPAM16-MAG$_{WT}$	*leu2Δ ade2Δ ura3Δ his3Δ trp1Δ lys2Δ pam16Δ::ADE2 [pPAM16 URA3] [pRS415-WT-MAGMAS]*	This study
YPH499-*pam16Δ*-pPAM16-MAG$_{Asn76Asp}$	*leu2Δ ade2Δ ura3Δ his3Δ trp1Δ lys2Δ pam16Δ::ADE2 [pPAM16 URA3] [pRS415-Mut-MAGMAS]*	This study
pRS415-WT-MAGMAS	Amp^R*CEN LEU2 TEF MAGMAS$_{WT}$*	D'Silva P. *et al.* [17].
pRS415-Asn76Asp-MAGMAS	Amp^R*CEN LEU2 TEF MAGMAS$_{Asn76Asp}$*	This study
pYX232-mtGFP	*Amp^{R}2 micronTP1 preSu9 GFP TRP1*	Westermann B. and Neupert W. [41]
pUG34 DsRed.SKL	Amp^R*MET25DsRed.SKL HIS3*	Kuravi K.*et al.* [42]

SUPPLEMENTARY FIGURES

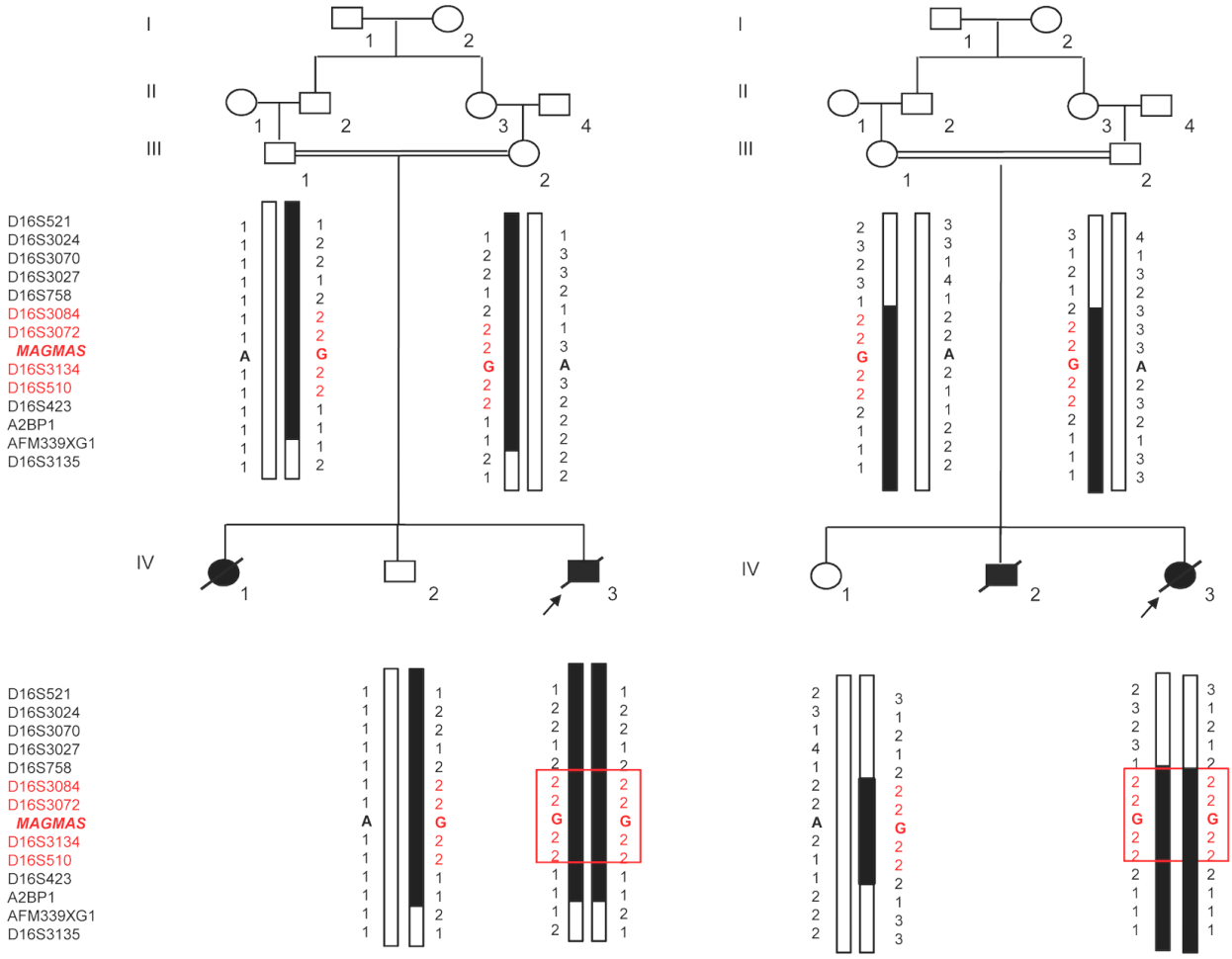


Figure SA.1 Pedigrees and haplotypes of families F1 and F2.

Markers are reported from telomere (top) to centromere (bottom) on chromosome 16p13.3. Blackened symbols represent affected individ- uals. Disease-bearing chromosome is represented in black. The red box represents the minimal ancestral homozygous haplotype spanning 1.9 Mb between markers D16S758 and D16S243 and shared between F1-IV.3 and F2-IV.3. The normal (A) or mutated (G) allele of MAGMAS is indictated within the haplotype.

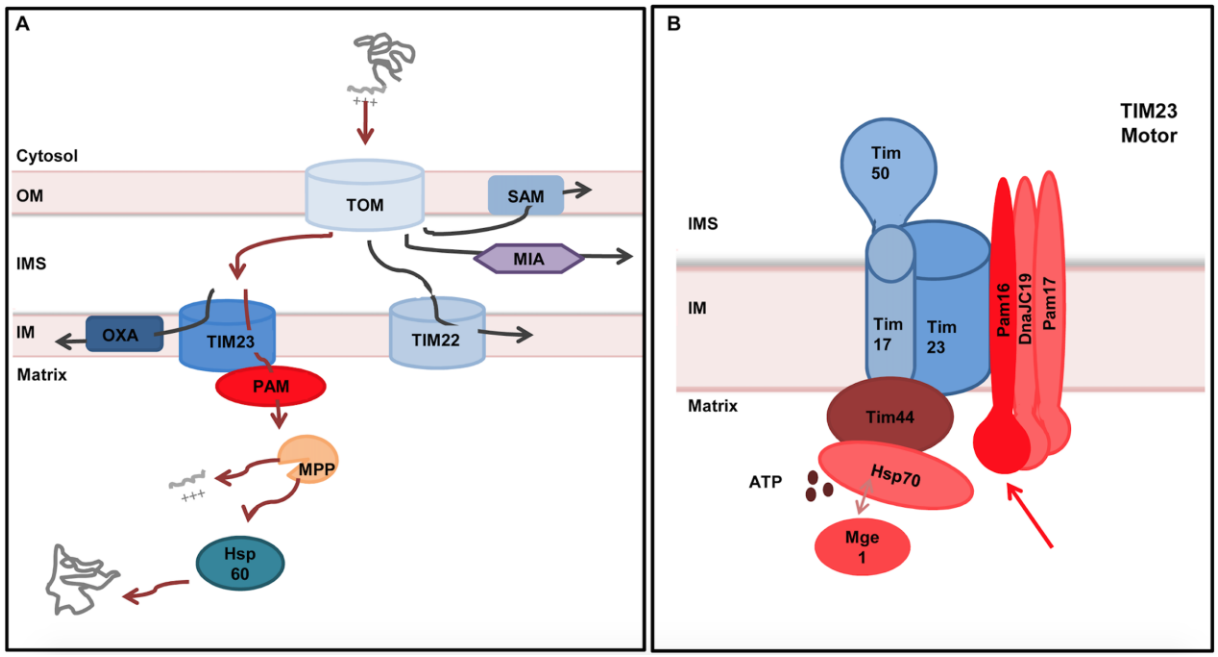


Figure SA.2 Protein import pathways into mitochondria.
(A) Several sophisticated transport machineries mediate recognition, import, sorting and assembly of preproteins into a specific subcompartment of the mitochondria. Preproteins targeted to the matrix cross the outer mitochondrial membrane through the TOM complex then pass through the inner membrane via the $TIM23^{MOTOR}$. Once processed into the matrix, preproteins are folded to their active forms after the cleavage of their presequences by the mitochondrial processing peptidase (MPP). HSP60 assists in the correct folding and assembly of the imported proteins.
(B) $TIM23^{MOTOR}$: the translo- case involved in preproteins import into the mitochondrial matrix. It is composed of TIM23 translocase associated with the PAM complex (In red). The ATP-dependant core of PAM, mtHSP70, drives the translocation and the unfolding of preproteins. The regulation of the position and activity of mtHSP70 is mediated by five other subunits of the PAM complex, consisting of a soluble nucleotide exchange factor MGE1 and four membrane bound co-chaperones TIM44, DNAJC19, PAM16 and PAM17. Via its C-terminal J-like domain, PAM16 (red arrow) interacts with DNAJC19 thus enabling its tethering to the translocon. IM, inner membrane; IMS, intermem- brane space; OM: outer membrane.

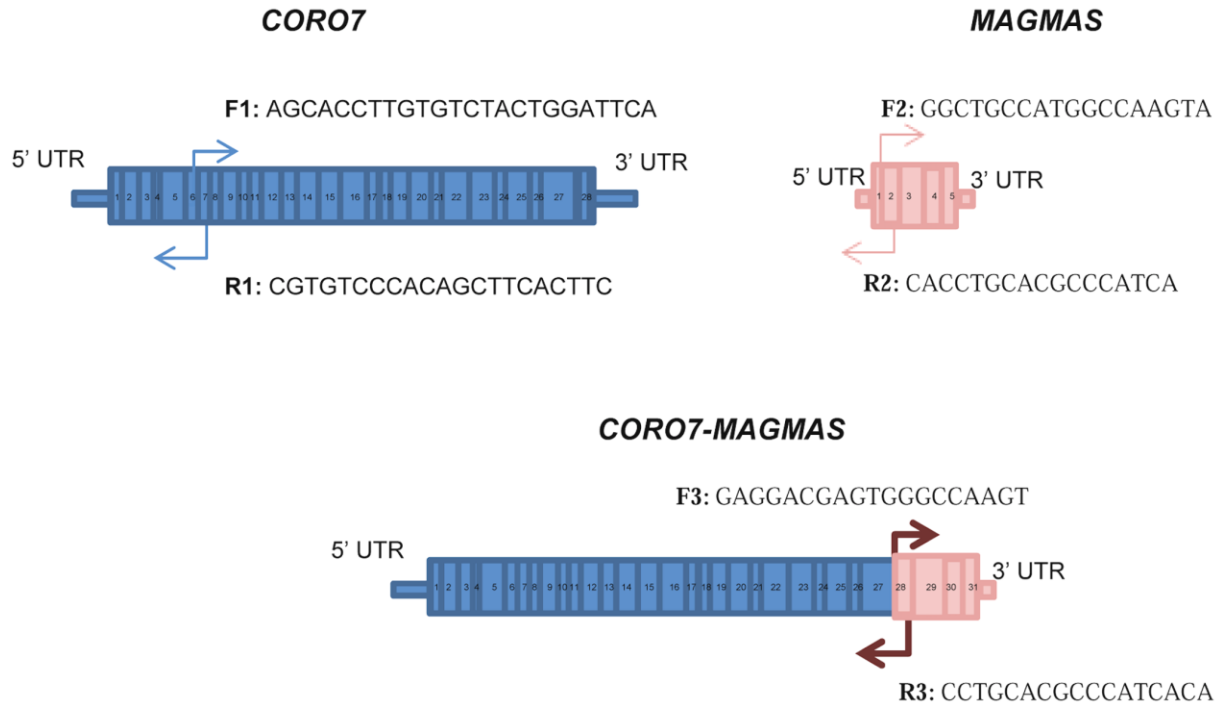


Figure SA.3 Primers used for expression analysis by quantitative Real-Time RT-PCR.
F: Forward primer. R: Reverse primer. F1, R1 are specific to the CORO7 transcript; F2, R2 to the MAGMAS transcript and F3, R3 to the CORO7-MAGMAS transcript

SUPPLEMENTARY TABLE

Table SA.1 The list of common exonic variations shared at a homozygous state between both probands F1-IV.3 and F2-IV.3. SNV: Single Nucleotide Variation.

Chr	Start	End	Gene	Function	Exonic Function	Amino Acid change
Chr9	88937852	88937854	ZCCHC6	Exonic	nonframeshift deletion	ZCCHC6:NM_001185074:exon9: c.2442_2444del:p.814_815del, ZCCHC6:NM_001185059:exon13 c.2811_2813del:p.937_938del,ZCCHC6: NM_024617:exon13:c.2811_2813del:p. 937_938del
Chr16	4390986	4390986	PAM16, CORO7-PAM16	Exonic	nonsynonymous SNV	PAM16:NM_016069:exon4:c.A226G:p. N76D, CORO7-PAM16:NM_001201479: exon30 :c.A2995G:p.N999D
Chr21	31913982	31913982	KRTAP19-6	Exonic	frameshift deletion	KRTAP19-6: NM_181612:exon1:c.171delC:p.F57fs

BIBLIOGRAPHIE

1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME and Mcvean GA (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

Abu-Safieh L, Alrashed M, Anazi S, Alkuraya H, Khan AO, Al-Owain M, Al-Zahrani J, Al-Abdi L, Hashem M, Al-Tarimi S, Sebai MA, Shamia A, Ray-Zack MD, Nassan M, Al-Hassnan ZN, Rahbeeni Z, Waheeb S, Alkharashi A, Abboud E, Al-Hazzaa SA and Alkuraya FS (2013). "Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes." Genome Res **23**(2): 236-247.

Adams DJ and Van Der Weyden L (2008). "Contemporary approaches for modifying the mouse genome." Physiol Genomics **34**(3): 225-238.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS and Sunyaev SR (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.

Akiyama T, Kobayashi K, Higashikage A, Sato J and Yoshinaga H (2014). "CSF/plasma ratios of amino acids: reference data and transports in children." Brain Dev **36**(1): 3-9.

Alders M, Hogan BM, Gjini E, Salehi F, Al-Gazali L, Hennekam EA, Holmberg EE, Mannens MM, Mulder MF, Offerhaus GJ, Prescott TE, Schroor EJ, Verheij JB, Witte M, Zwijnenburg PJ, Vikkula M, Schulte-Merker S and Hennekam RC (2009). "Mutations in CCBE1 cause generalized lymph vessel dysplasia in humans." Nat Genet **41**(12): 1272-1274.

Alesi V, Bertoli M, Barrano G, Torres B, Pusceddu S, Pastorino M, Perria C, Nardone AM, 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**(2): 384-387.

Alexander RP, Fang G, Rozowsky J, Snyder M and Gerstein MB (2010). "Annotating non-coding regions of the genome." Nat Rev Genet **11**(8): 559-571.

Alfaiz AA, Micale L, Mandriani B, Augello B, Pellico MT, Chrast J, Xenarios I, Zelante L, Merla G and Reymond A (2014). "TBC1D7 mutations are associated with intellectual disability, macrocrania, patellar dislocation, and celiac disease." Hum Mutat **35**(4): 447-451.

Alkuraya FS (2010). "Autozygome decoded." Genet Med **12**(12): 765-771.

Alkuraya FS (2013). "The application of next-generation sequencing in the autozygosity mapping of human recessive diseases." Hum Genet **132**(11): 1197-1211.

Alkuraya FS, Cai X, Emery C, Mochida GH, Al-Dosari MS, Felie JM, Hill RS, Barry BJ, Partlow JN, Gascon GG, Kentab A, Jan M, Shaheen R, Feng Y and Walsh CA (2011). "Human mutations in NDE1 cause extreme microcephaly with lissencephaly [corrected]." Am J Hum Genet **88**(5): 536-547.

Allen KM, Gleeson JG, Bagrodia S, Partington MW, Macmillan JC, Cerione RA, Mulley JC and Walsh CA (1998). "PAK3 mutation in nonsyndromic X-linked mental retardation." Nat Genet **20**(1): 25-30.

American Psychiatric Association. (1994). Diagnostic criteria from DSM-IV. Washington, D.C., The Association.

American Psychiatric Association. (2000). Electronic DSM-IV-TR plus. Washington, D.C., American Psychiatric Association,**:** 1 CD-ROM.

Andermann F, Keene DL, Andermann E and Quesney LF (1980). "Startle disease or hyperekplexia: further delineation of the syndrome." Brain **103**(4): 985-997.

Anderson AD and Weir BS (2007). "A maximum-likelihood method for the estimation of pairwise relatedness in structured populations." Genetics **176**(1): 421-440.

Andreutti-Zaugg C, Scott RJ and Iggo R (1997). "Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques." Cancer Res **57**(15): 3288-3293.

Ayoub AE, Oh S, Xie Y, Leng J, Cotney J, Dominguez MH, Noonan JP and Rakic P (2011). "Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing." Proc Natl Acad Sci U S A **108**(36): 14950-14955.

Bahnsen U, Oosting P, Swaab DF, Nahke P, Richter D and Schmale H (1992). "A missense mutation in the vasopressin-neurophysin precursor gene cosegregates with human autosomal dominant neurohypophyseal diabetes insipidus." EMBO J **11**(1): 19-23.

Baitner AC, Maurer SG, Gruen MB and Di Cesare PE (2000). "The genetic basis of the osteochondrodysplasias." J Pediatr Orthop **20**(5): 594-605.

Bakker MJ, Van Dijk JG, Van Den Maagdenberg AM and Tijssen MA (2006). "Startle syndromes." Lancet Neurol **5**(6): 513-524.

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA and Shendure J (2011). "Exome sequencing as a tool for Mendelian disease gene discovery." Nat Rev Genet **12**(11): 745-755.

Barkovich AJ, Hevner R and Guerrini R (1999). "Syndromes of bilateral symmetrical polymicrogyria." AJNR Am J Neuroradiol **20**(10): 1814-1821.

Barnett ME, Zolkiewska A and Zolkiewski M (2000). "Structure and activity of ClpB from Escherichia coli. Role of the amino-and -carboxyl-terminal domains." J Biol Chem **275**(48): 37565-37571.

Basel-Vanagaite L (2007). "Genetics of autosomal recessive non-syndromic mental retardation: recent advances." Clin Genet **72**(3): 167-174.

Basel-Vanagaite L, Attia R, Yahav M, Ferland RJ, Anteki L, Walsh CA, Olender T, Straussberg R, Magal N, Taub E, Drasinover V, Alkelai A, Bercovich D, Rechavi G, Simon AJ and Shohat M (2006). "The CC2D1A, a member of a new gene family with C2 domains, is involved in autosomal recessive non-syndromic mental retardation." J Med Genet **43**(3): 203-210.

Basel-Vanagaite L, Taub E, Halpern GJ, Drasinover V, Magal N, Davidov B, Zlotogora J and Shohat M (2007). "Genetic screening for autosomal recessive nonsyndromic mental retardation in an isolated population in Israel." Eur J Hum Genet **15**(2): 250-253.

Bashi J (1977). "Effects of inbreeding on cognitive performance." Nature **266**(5601): 440-442.

Batista MF and Lewis KE (2008). "Pax2/8 act redundantly to specify glycinergic and GABAergic fates of multiple spinal interneurons." Dev Biol **323**(1): 88-97.

Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, Brudno M, Knoppers B, Marcadier J, Dyment D, Adam S, Bulman DE, Jones SJ, Avard D, Nguyen MT, Rousseau F, Marshall C, Wintle RF, Shen Y, Scherer SW, Consortium FC, Friedman JM, Michaud JL and Boycott KM (2014). "FORGE Canada Consortium: Outcomes of a 2-Year National Rare-Disease Gene-Discovery Project." Am J Hum Genet **94**(6): 809-817.

Becker S, Gehrsitz A, Bork P, Buchner S and Buchner E (2001). "The black-pearl gene of Drosophila defines a novel conserved protein family and is required for larval growth and survival." Gene **262**(1-2): 15-22.

Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, Sheth V, Woodward JE, Peckham HE, Schroth GP, Kim RW and Kingsmore SF (2011). "Carrier testing for severe childhood recessive diseases by next-generation sequencing." Sci Transl Med **3**(65): 65ra64.

Ben-Salem S, Gleeson JG, Al-Shamsi AM, Islam B, Hertecant J, Ali BR and Al-Gazali L (2014). "Asparagine synthetase deficiency detected by whole exome sequencing causes congenital microcephaly, epileptic encephalopathy and psychomotor delay." Metab Brain Dis.

Bentley DR (2006). "Whole-genome re-sequencing." Curr Opin Genet Dev **16**(6): 545-552.

Betz H (1991). "Glycine receptors: heterogeneous and widespread in the mammalian brain." Trends Neurosci **14**(10): 458-461.

Bilguvar K, Ozturk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalnizoglu D, Tuysuz B, Caglayan AO, Gokben S, Kaymakcalan H, Barak T, Bakircioglu M, Yasuno K, Ho W, Sanders S, Zhu Y, Yilmaz S, Dincer A, Johnson MH, Bronen RA, Kocer N, Per H, Mane S, Pamir MN, Yalcinkaya C, Kumandas S, Topcu M, Ozmen M, Sestan N, Lifton RP, State MW and Gunel M (2010). "Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations." Nature **467**(7312): 207-210.

Billuart P, Bienvenu T, Ronce N, Des Portes V, Vinet MC, Zemni R, Roest Crollius H, Carrie A, Fauchereau F, Cherry M, Briault S, Hamel B, Fryns JP, Beldjord C, Kahn A, Moraine C and Chelly J (1998). "Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation." Nature **392**(6679): 923-926.

Bittles A (2001). "Consanguinity and its relevance to clinical genetics." Clin Genet **60**(2): 89-98.

Bittles AH and Black ML (2010). "Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases." Proc Natl Acad Sci U S A **107 Suppl 1**: 1779-1786.

Blackett PR, Seif SM, Altmiller DH and Robinson AG (1983). "Familial central diabetes insipidus: vasopressin and nicotine stimulated neurophysin deficiency with subnormal oxytocin and estrogen stimulated neurophysin." Am J Med Sci **286**(3): 42-46.

Bober MB, Khan N, Kaplan J, Lewis K, Feinstein JA, Scott CI, Jr. and Steinberg GK (2010). "Majewski osteodysplastic primordial dwarfism type II (MOPD II): expanding the vascular phenotype." Am J Med Genet A **152A**(4): 960-965.

Boda B, Alberi S, Nikonenko I, Node-Langlois R, Jourdain P, Moosmayer M, Parisi-Jourdain L and Muller D (2004). "The mental retardation protein PAK3 contributes to synapse formation and plasticity in hippocampus." J Neurosci **24**(48): 10816-10825.

Bonafe L, Schmitt K, Eich G, Giedion A and Superti-Furga A (2002). "RMRP gene sequence analysis confirms a cartilage-hair hypoplasia variant with only skeletal manifestations and reveals a high density of single-nucleotide polymorphisms." Clin Genet **61**(2): 146-151.

Bond J, Roberts E, Mochida GH, Hampshire DJ, Scott S, Askham JM, Springell K, Mahadevan M, Crow YJ, Markham AF, Walsh CA and Woods CG (2002). "ASPM is a major determinant of cerebral cortical size." Nat Genet **32**(2): 316-320.

Book JA (1957). "Genetical investigations in a North Swedish population; the offspring of first-cousin marriages." Ann Hum Genet **21**(3): 191-221.

Bosl B, Grimminger V and Walter S (2006). "The molecular chaperone Hsp104--a molecular machine for protein disaggregation." J Struct Biol **156**(1): 139-148.

Botstein D and Risch N (2003). "Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease." Nat Genet **33 Suppl**: 228-237.

Botta E, Nardo T, Lehmann AR, Egly JM, Pedrini AM and Stefanini M (2002). "Reduced level of the repair/transcription factor TFIIH in trichothiodystrophy." Hum Mol Genet **11**(23): 2919-2928.

Braverman N, Steel G, Obie C, Moser A, Moser H, Gould SJ and Valle D (1997). "Human PEX7 encodes the peroxisomal PTS2 receptor and is responsible for rhizomelic chondrodysplasia punctata." Nat Genet **15**(4): 369-376.

Brunelli S, Faiella A, Capra V, Nigro V, Simeone A, Cama A and Boncinelli E (1996). "Germline mutations in the homeobox gene EMX2 in patients with severe schizencephaly." Nat Genet **12**(1): 94-96.

Bundey S and Alam H (1993). "A five-year prospective study of the health of children in different ethnic groups, with particular reference to the effect of inbreeding." Eur J Hum Genet **1**(3): 206-219.

Byrd PJ, Stankovic T, Mcconville CM, Smith AD, Cooper PR and Taylor AM (1997). "Identification and analysis of expression of human VACM-1, a cullin gene family member located on chromosome 11q22-23." Genome Res **7**(1): 71-75.

Capo-Chichi JM, Bharti SK, Sommers JA, Yammine T, Chouery E, Patry L, Rouleau GA, Samuels ME, Hamdan FF, Michaud JL, Brosh RM, Jr., Megarbane A and Kibar Z (2013a). "Identification and biochemical characterization of a novel mutation in DDX11 causing Warsaw breakage syndrome." Hum Mutat **34**(1): 103-107.

Capo-Chichi JM, Tcherkezian J, Hamdan FF, Decarie JC, Dobrzeniecka S, Patry L, Nadon MA, Mucha BE, Major P, Shevell M, Bencheikh BO, Joober R, Samuels ME, Rouleau GA, Roux PP and Michaud JL (2013b). "Disruption of TBC1D7, a subunit of the TSC1-TSC2 protein complex, in intellectual disability and megalencephaly." J Med Genet **50**(11): 740-744.

Carlson BR, Lloyd KE, Kruszewski A, Kim IH, Rodriguiz RM, Heindel C, Faytell M, Dudek SM, Wetsel WC and Soderling SH (2011). "WRP/srGAP3 facilitates the initiation of spine development by an inverse F-BAR domain, and its loss impairs long-term memory." J Neurosci **31**(7): 2447-2460.

Castellana S and Mazza T (2013). "Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools." Brief Bioinform **14**(4): 448-459.

Ceremuga TE, Yao XL and Mccabe JT (2001). "Vasopressin-activated calcium-mobilizing (VACM-1) receptor mRNA is present in peripheral organs and the central nervous system of the laboratory rat." Endocr Res **27**(4): 433-445.

Ceremuga TE, Yao XL, Xia Y, Mukherjee D and Mccabe JT (2003). "Osmotic stress increases cullin-5 (cul-5) mRNA in the rat cerebral cortex, hypothalamus and kidney." Neurosci Res **45**(3): 305-311.

Chacinska A, Koehler CM, Milenkovic D, Lithgow T and Pfanner N (2009). "Importing mitochondrial proteins: machineries and mechanisms." Cell **138**(4): 628-644.

Chakrabarti L and Davies KE (1997). "Fragile X syndrome." Curr Opin Neurol **10**(2): 142-147.

Chen H, Pan YX, Dudenhausen EE and Kilberg MS (2004). "Amino acid deprivation induces the transcription rate of the human asparagine synthetase gene through a timed program of expression and promoter binding of nutrient-responsive basic region/leucine zipper transcription factors as well as localized histone acetylation." J Biol Chem **279**(49): 50829-50839.

Chilamakuri CS, Lorenz S, Madoui MA, Vodak D, Sun J, Hovig E, Myklebost O and Meza-Zepeda LA (2014). "Performance comparison of four exome capture systems for deep sequencing." BMC Genomics **15**: 449.

Chiurazzi P, Schwartz CE, Gecz J and Neri G (2008). "XLMR genes: update 2007." Eur J Hum Genet **16**(4): 422-434.

Choi Y, Sims GE, Murphy S, Miller JR and Chan AP (2012). "Predicting the functional effect of amino acid substitutions and indels." PLoS One **7**(10): e46688.

Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, Butte AJ and Snyder M (2011). "Performance comparison of exome DNA sequencing technologies." Nat Biotechnol **29**(10): 908-914.

Collin GB, Marshall JD, Ikeda A, So WV, Russell-Eggitt I, Maffei P, Beck S, Boerkoel CF, Sicolo N, Martin M, Nishina PM and Naggert JK (2002). "Mutations in ALMS1 cause obesity, type 2 diabetes and neurosensory degeneration in Alstrom syndrome." Nat Genet **31**(1): 74-78.

Corbett MA, Bahlo M, Jolly L, Afawi Z, Gardner AE, Oliver KL, Tan S, Coffey A, Mulley JC, Dibbens LM, Simri W, Shalata A, Kivity S, Jackson GD, Berkovic SF and Gecz J (2010). "A focal epilepsy and intellectual disability syndrome is due to a mutation in TBC1D24." Am J Hum Genet **87**(3): 371-375.

Cox J, Jackson AP, Bond J and Woods CG (2006). "What primary microcephaly can tell us about brain growth." Trends Mol Med **12**(8): 358-366.

Crane DI (2014). "Revisiting the neuropathogenesis of Zellweger syndrome." Neurochem Int **69**: 1-8.

Crino PB, Nathanson KL and Henske EP (2006). "The tuberous sclerosis complex." N Engl J Med **355**(13): 1345-1356.

Curry CJ, Stevenson RE, Aughton D, Byrne J, Carey JC, Cassidy S, Cunniff C, Graham JM, Jr., Jones MC, Kaback MM, Moeschler J, Schaefer GB, Schwartz S, Tarleton J and Opitz J (1997). "Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics." Am J Med Genet **72**(4): 468-477.

D'adamo P, Menegon A, Lo Nigro C, Grasso M, Gulisano M, Tamanini F, Bienvenu T, Gedeon AK, Oostra B, Wu SK, Tandon A, Valtorta F, Balch WE, Chelly J and Toniolo D (1998). "Mutations in GDI1 are responsible for X-linked non-specific mental retardation." Nat Genet **19**(2): 134-139.

D'silva PR, Schilke B, Hayashi M and Craig EA (2008). "Interaction of the J-protein heterodimer Pam18/Pam16 of the mitochondrial import motor with the translocon of the inner membrane." Mol Biol Cell **19**(1): 424-432.

D'silva PR, Schilke B, Walter W and Craig EA (2005). "Role of Pam16's degenerate J domain in protein import across the mitochondrial inner membrane." Proc Natl Acad Sci U S A **102**(35): 12419-12424.

Dan B and Boyd SG (2003). "Angelman syndrome reviewed from a neurophysiological perspective. The UBE3A-GABRB3 hypothesis." Neuropediatrics **34**(4): 169-176.

Daoud H, Suhail H, Szuto A, Camu W, Salachas F, Meininger V, Bouchard JP, Dupre N, Dion PA and Rouleau GA (2012). "UBQLN2 mutations are rare in French and French-Canadian amyotrophic lateral sclerosis." Neurobiol Aging **33**(9): 2230 e2231-2230 e2235.

Davey KM, Parboosingh JS, Mcleod DR, Chan A, Casey R, Ferreira P, Snyder FF, Bridge PJ and Bernier FP (2006). "Mutation of DNAJC19, a human homologue of yeast inner mitochondrial membrane co-chaperones, causes DCMA syndrome, a novel autosomal recessive Barth syndrome-like condition." J Med Genet **43**(5): 385-393.

Davis EE, Frangakis S and Katsanis N (2014). "Interpreting human genetic variation with in vivo zebrafish assays." Biochim Biophys Acta.

Day IN (2010). "dbSNP in the detail and copy number complexities." Hum Mutat **31**(1): 2-4.

De Ligt J, Willemsen MH, Van Bon BW, Kleefstra T, Yntema HG, Kroes T, Vulto-Van Silfhout AT, Koolen DA, De Vries P, Gilissen C, Del Rosario M, Hoischen A, Scheffer H, De Vries BB, Brunner HG, Veltman JA and Vissers LE (2012). "Diagnostic exome sequencing in persons with severe intellectual disability." N Engl J Med **367**(20): 1921-1929.

Depristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, Mckenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D and Daly MJ (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." Nat Genet **43**(5): 491-498.

Dibble CC, Elis W, Menon S, Qin W, Klekota J, Asara JM, Finan PM, Kwiatkowski DJ, Murphy LO and Manning BD (2012). "TBC1D7 is a third subunit of the TSC1-TSC2 complex upstream of mTORC1." Mol Cell **47**(4): 535-546.

Ding H, Schertzer M, Wu X, Gertsenstein M, Selig S, Kammori M, Pourvali R, Poon S, Vulto I, Chavez E, Tam PP, Nagy A and Lansdorp PM (2004). "Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein." Cell **117**(7): 873-886.

Dohm JC, Lottaz C, Borodina T and Himmelbauer H (2008). "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing." Nucleic Acids Res **36**(16): e105.

Doyle SM and Wickner S (2009). "Hsp104 and ClpB: protein disaggregating machines." Trends Biochem Sci **34**(1): 40-48.

Drews CD, Yeargin-Allsopp M, Decoufle P and Murphy CC (1995). "Variation in the influence of selected sociodemographic risk factors for mental retardation." Am J Public Health **85**(3): 329-334.

Dubaele S, Proietti De Santis L, Bienstock RJ, Keriel A, Stefanini M, Van Houten B and Egly JM (2003). "Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients." Mol Cell **11**(6): 1635-1646.

Durkin M (2002). "The epidemiology of developmental disabilities in low-income countries." Ment Retard Dev Disabil Res Rev **8**(3): 206-211.

Durkin MS, Hasan ZM and Hasan KZ (1998). "Prevalence and correlates of mental retardation among children in Karachi, Pakistan." Am J Epidemiol **147**(3): 281-288.

Ehninger D (2013). "From genes to cognition in tuberous sclerosis: implications for mTOR inhibitor-based treatment approaches." Neuropharmacology **68**: 97-105.

Ellison JW, Rosenfeld JA and Shaffer LG (2013). "Genetic basis of intellectual disability." Annu Rev Med **64**: 441-450.

Elsayed SM, Heller R, Thoenes M, Zaki MS, Swan D, Elsobky E, Zuhlke C, Ebermann I, Nurnberg G, Nurnberg P and Bolz HJ (2014). "Autosomal dominant SCA5 and autosomal recessive infantile SCA are allelic conditions resulting from SPTBN2 mutations." Eur J Hum Genet **22**(2): 286-288.

Elsner S, Simian D, Iosefson O, Marom M and Azem A (2009). "The mitochondrial protein translocation motor: structural conservation between the human and yeast Tim14/Pam18-Tim16/Pam16 co-chaperones." Int J Mol Sci **10**(5): 2041-2053.

Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, Milh M, Kortum F, Fritsch A, Pientka FK, Hellenbroich Y, Kalscheuer VM, Kohlhase J, Moog U, Rappold G, Rauch A, Ropers HH, Von Spiczak S, Tonnies H, Villeneuve N, Villard L, Zabel B, Zenker M, Laube B, Reis A, Wieczorek D, Van Maldergem L and Kutsche K (2010). "Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes." Nat Genet **42**(11): 1021-1026.

European Chromosome 16 Tuberous Sclerosis C (1993). "Identification and characterization of the tuberous sclerosis gene on chromosome 16." Cell **75**(7): 1305-1315.

Fetcho JR (1992). "The spinal motor system in early vertebrates and some of its evolutionary changes." Brain Behav Evol **40**(2-3): 82-97.

Fidler DJ, Bailey JN and Smalley SL (2000). "Macrocephaly in autism and other pervasive developmental disorders." Dev Med Child Neurol **42**(11): 737-740.

Field M, Tarpey PS, Smith R, Edkins S, O'meara S, Stevens C, Tofts C, Teague J, Butler A, Dicks E, Barthorpe S, Buck G, Cole J, Gray K, Halliday K, Hills K, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Varian J, West S, Widaa S, Mallya U, Wooster R, Moon J, Luo Y, Hughes H, Shaw M, Friend KL, Corbett M, Turner G, Partington M, Mulley J, Bobrow M, Schwartz C, Stevenson R, Gecz J, Stratton MR, Futreal PA and Raymond FL (2007). "Mutations in the BRWD3 gene cause X-linked mental retardation associated with macrocephaly." Am J Hum Genet **81**(2): 367-374.

Frazier AE, Dudek J, Guiard B, Voos W, Li Y, Lind M, Meisinger C, Geissler A, Sickmann A, Meyer HE, Bilanchone V, Cumsky MG, Truscott KN, Pfanner N and Rehling P (2004). "Pam16 has an essential role in the mitochondrial protein import motor." Nat Struct Mol Biol **11**(3): 226-233.

Garami A, Zwartkruis FJ, Nobukuni T, Joaquin M, Roccio M, Stocker H, Kozma SC, Hafen E, Bos JL and Thomas G (2003). "Insulin activation of Rheb, a mediator of mTOR/S6K/4E-BP signaling, is inhibited by TSC1 and 2." Mol Cell **11**(6): 1457-1466.

Gauthier J, Siddiqui TJ, Huashan P, Yokomaku D, Hamdan FF, Champagne N, Lapointe M, Spiegelman D, Noreau A, Lafreniere RG, Fathalli F, Joober R, Krebs MO, Delisi LE, Mottron L, Fombonne E, Michaud JL, Drapeau P, Carbonetto S, Craig AM and Rouleau GA (2011). "Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia." Hum Genet **130**(4): 563-573.

Ge D, Ruzzo EK, Shianna KV, He M, Pelak K, Heinzen EL, Need AC, Cirulli ET, Maia JM, Dickson SP, Zhu M, Singh A, Allen AS and Goldstein DB (2011). "SVA: software

for annotating and visualizing sequenced human genomes." Bioinformatics **27**(14): 1998-2000.

Gecz J, Shoubridge C and Corbett M (2009). "The genetic landscape of intellectual disability arising from chromosome X." Trends Genet **25**(7): 308-316.

Giannandrea M, Bianchi V, Mignogna ML, Sirri A, Carrabino S, D'elia E, Vecellio M, Russo S, Cogliati F, Larizza L, Ropers HH, Tzschach A, Kalscheuer V, Oehl-Jaschkowitz B, Skinner C, Schwartz CE, Gecz J, Van Esch H, Raynaud M, Chelly J, De Brouwer AP, Toniolo D and D'adamo P (2010). "Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly." Am J Hum Genet **86**(2): 185-195.

Gilissen C, Hehir-Kwa JY, Thung DT, Van De Vorst M, Van Bon BW, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A, Leach R, Klein R, Tearle R, Bo T, Pfundt R, Yntema HG, De Vries BB, Kleefstra T, Brunner HG, Vissers LE and Veltman JA (2014). "Genome sequencing identifies major causes of severe intellectual disability." Nature.

Gilissen C, Hoischen A, Brunner HG and Veltman JA (2011). "Unlocking Mendelian disease using exome sequencing." Genome Biol **12**(9): 228.

Gilissen C, Hoischen A, Brunner HG and Veltman JA (2012). "Disease gene identification strategies for exome sequencing." Eur J Hum Genet **20**(5): 490-497.

Glazov EA, Zankl A, Donskoi M, Kenna TJ, Thomas GP, Clark GR, Duncan EL and Brown MA (2011). "Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia." PLoS Genet **7**(3): e1002027.

Gnirke A, Melnikov A, Maguire J, Rogov P, Leproust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES and Nusbaum C (2009). "Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing." Nat Biotechnol **27**(2): 182-189.

Goloubinoff P, Mogk A, Zvi AP, Tomoyasu T and Bukau B (1999). "Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network." Proc Natl Acad Sci U S A **96**(24): 13732-13737.

Govek EE, Newey SE, Akerman CJ, Cross JR, Van Der Veken L and Van Aelst L (2004). "The X-linked mental retardation protein oligophrenin-1 is required for dendritic spine morphogenesis." Nat Neurosci **7**(4): 364-372.

Greco A, Gong SS, Ittmann M and Basilico C (1989). "Organization and expression of the cell cycle gene, ts11, that encodes asparagine synthetase." Mol Cell Biol **9**(6): 2350-2359.

Griffiths PD, Gardner SA, Smith M, Rittey C and Powell T (1998). "Hemimegalencephaly and focal megalencephaly in tuberous sclerosis complex." AJNR Am J Neuroradiol **19**(10): 1935-1938.

Guernsey DL, Jiang H, Hussin J, Arnold M, Bouyakdan K, Perry S, Babineau-Sturk T, Beis J, Dumas N, Evans SC, Ferguson M, Matsuoka M, Macgillivray C, Nightingale M, Patry L, Rideout AL, Thomas A, Orr A, Hoffmann I, Michaud JL, Awadalla P, Meek DC, Ludman M and Samuels ME (2010). "Mutations in centrosomal protein CEP152 in primary microcephaly families linked to MCPH4." Am J Hum Genet **87**(1): 40-51.

Haberle J, Gorg B, Rutsch F, Schmidt E, Toutain A, Benoist JF, Gelot A, Suc AL, Hohne W, Schliess F, Haussinger D and Koch HG (2005). "Congenital glutamine deficiency with glutamine synthetase mutations." N Engl J Med **353**(18): 1926-1933.

Haberle J, Shahbeck N, Ibrahim K, Schmitt B, Scheer I, O'gorman R, Chaudhry FA and Ben-Omran T (2012). "Glutamine supplementation in a child with inherited GS deficiency improves the clinical status and partially corrects the peripheral and central amino acid imbalance." Orphanet J Rare Dis **7**: 48.

Hamamy H, Antonarakis SE, Cavalli-Sforza LL, Temtamy S, Romeo G, Kate LP, Bennett RL, Shaw A, Megarbane A, Van Duijn C, Bathija H, Fokstuen S, Engel E, Zlotogora J, Dermitzakis E, Bottani A, Dahoun S, Morris MA, Arsenault S, Aglan MS, Ajaz M, Alkalamchi A, Alnaqeb D, Alwasiyah MK, Anwer N, Awwad R, Bonnefin M, Corry P, Gwanmesia L, Karbani GA, Mostafavi M, Pippucci T, Ranza-Boscardin E, Reversade B, Sharif SM, Teeuw ME and Bittles AH (2011). "Consanguineous marriages, pearls and perils: Geneva International Consanguinity Workshop Report." Genet Med **13**(9): 841-847.

Hamdan FF, Gauthier J, Araki Y, Lin DT, Yoshizawa Y, Higashi K, Park AR, Spiegelman D, Dobrzeniecka S, Piton A, Tomitori H, Daoud H, Massicotte C, Henrion E, Diallo O, Group SD, Shekarabi M, Marineau C, Shevell M, Maranda B, Mitchell G, Nadeau A, D'anjou G, Vanasse M, Srour M, Lafreniere RG, Drapeau P, Lacaille JC, Kim E, Lee JR, Igarashi K, Huganir RL, Rouleau GA and Michaud JL (2011). "Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability." Am J Hum Genet **88**(3): 306-316.

Hamdan FF, Gauthier J, Spiegelman D, Noreau A, Yang Y, Pellerin S, Dobrzeniecka S, Cote M, Perreau-Linck E, Carmant L, D'anjou G, Fombonne E, Addington AM, Rapoport JL, Delisi LE, Krebs MO, Mouaffak F, Joober R, Mottron L, Drapeau P, Marineau C, Lafreniere RG, Lacaille JC, Rouleau GA, Michaud JL and Synapse to Disease G (2009a). "Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation." N Engl J Med **360**(6): 599-605.

Hamdan FF, Piton A, Gauthier J, Lortie A, Dubeau F, Dobrzeniecka S, Spiegelman D, Noreau A, Pellerin S, Cote M, Henrion E, Fombonne E, Mottron L, Marineau C, Drapeau P, Lafreniere RG, Lacaille JC, Rouleau GA and Michaud JL (2009b). "De novo STXBP1 mutations in mental retardation and nonsyndromic epilepsy." Ann Neurol **65**(6): 748-753.

Hanson PI and Whiteheart SW (2005). "AAA+ proteins: have engine, will work." Nat Rev Mol Cell Biol **6**(7): 519-529.

Hart CE, Race V, Achouri Y, Wiame E, Sharrard M, Olpin SE, Watkinson J, Bonham JR, Jaeken J, Matthijs G and Van Schaftingen E (2007). "Phosphoserine aminotransferase deficiency: a novel disorder of the serine biosynthesis pathway." Am J Hum Genet **80**(5): 931-937.

Harvey RJ, Carta E, Pearce BR, Chung SK, Supplisson S, Rees MI and Harvey K (2008a). "A critical role for glycine transporters in hyperexcitability disorders." Front Mol Neurosci **1**: 1.

Harvey RJ, Topf M, Harvey K and Rees MI (2008b). "The genetics of hyperekplexia: more than startle!" Trends Genet **24**(9): 439-447.

Hashmi MA (1997). "Frequency of consanguinity and its effect on congenital malformation--a hospital based study." J Pak Med Assoc **47**(3): 75-78.

Hearn T, Spalluto C, Phillips VJ, Renforth GL, Copin N, Hanley NA and Wilson DI (2005). "Subcellular localization of ALMS1 supports involvement of centrosome and basal body dysfunction in the pathogenesis of obesity, insulin resistance, and type 2 diabetes." Diabetes **54**(5): 1581-1587.

Hirata H, Saint-Amant L, Downes GB, Cui WW, Zhou W, Granato M and Kuwada JY (2005). "Zebrafish bandoneon mutants display behavioral defects due to a mutation in the glycine receptor beta-subunit." Proc Natl Acad Sci U S A **102**(23): 8345-8350.

Hirota Y and Lahti JM (2000). "Characterization of the enzymatic activity of hChlR1, a novel human DNA helicase." Nucleic Acids Res **28**(4): 917-924.

Hodson S, Marshall JJ and Burston SG (2012). "Mapping the road to recovery: the ClpB/Hsp104 molecular chaperone." J Struct Biol **179**(2): 161-171.

Hongo S, Chiyo T and Takeda M (1996). "Cloning of cDNA for asparagine synthetase from rat Sertoli cell." Biochem Mol Biol Int **38**(1): 189-196.

Horowitz B, Madras BK, Meister A, Old LJ, Boyes EA and Stockert E (1968). "Asparagine synthetase activity of mouse leukemias." Science **160**(3827): 533-535.

Huang N, Lee I, Marcotte EM and Hurles ME (2010). "Characterising and predicting haploinsufficiency in the human genome." PLoS Genet **6**(10): e1001154.

Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, Coates G, Fairley S, Fitzgerald S, Fernandez-Banet J, Gordon L, Graf S, Haider S, Hammond M, Holland R, Howe K, Jenkinson A, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Rios D, Schuster M, Slater G, Smedley D, Spooner W, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wilder S, Zadissa A, Birney E, Cunningham F, Curwen V, Durbin R, Fernandez-Suarez XM, Herrero J, Kasprzyk A, Proctor

G, Smith J, Searle S and Flicek P (2009). "Ensembl 2009." Nucleic Acids Res **37**(Database issue): D690-697.

Huber C, Faqeih EA, Bartholdi D, Bole-Feysot C, Borochowitz Z, Cavalcanti DP, Frigo A, Nitschke P, Roume J, Santos HG, Shalev SA, Superti-Furga A, Delezoide AL, Le Merrer M, Munnich A and Cormier-Daire V (2013). "Exome sequencing identifies INPPL1 mutations as a cause of opsismodysplasia." Am J Hum Genet **92**(1): 144-149.

Hunkapiller T, Kaiser RJ, Koop BF and Hood L (1991). "Large-scale and automated DNA sequence determination." Science **254**(5028): 59-67.

Ikegawa S (2006). "Genetic analysis of skeletal dysplasia: recent advances and perspectives in the post-genome-sequence era." J Hum Genet **51**(7): 581-586.

Jackson AP, Eastwood H, Bell SM, Adu J, Toomes C, Carr IM, Roberts E, Hampshire DJ, Crow YJ, Mighell AJ, Karbani G, Jafri H, Rashid Y, Mueller RF, Markham AF and Woods CG (2002). "Identification of microcephalin, a protein implicated in determining the size of the human brain." Am J Hum Genet **71**(1): 136-142.

Jacquemont S, Coe BP, Hersch M, Duyzend MH, Krumm N, Bergmann S, Beckmann JS, Rosenfeld JA and Eichler EE (2014). "A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders." Am J Hum Genet **94**(3): 415-425.

Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, Bourgeron T and Paris Autism Research International Sibpair S (2003). "Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism." Nat Genet **34**(1): 27-29.

Jaworski J (2007). "ARF6 in the nervous system." Eur J Cell Biol **86**(9): 513-524.

Jaworski J, Spangler S, Seeburg DP, Hoogenraad CC and Sheng M (2005). "Control of dendritic arborization by the phosphoinositide-3'-kinase-Akt-mammalian target of rapamycin pathway." J Neurosci **25**(49): 11300-11312.

Jensen A (1983). "Effects of inbreeding on mental ability factors." Personality and individual differences **4**(1): 71-87.

Jubinsky PT, Messer A, Bender J, Morris RE, Ciraolo GM, Witte DP, Hawley RG and Short MK (2001). "Identification and characterization of Magmas, a novel mitochondria-associated protein involved in granulocyte-macrophage colony-stimulating factor signal transduction." Exp Hematol **29**(12): 1392-1402.

Jubinsky PT, Short MK, Mutema G and Witte DP (2003). "Developmental expression of Magmas in murine tissues and its co-expression with the GM-CSF receptor." J Histochem Cytochem **51**(5): 585-596.

Kabashi E, Lin L, Tradewell ML, Dion PA, Bercier V, Bourgouin P, Rochefort D, Bel Hadj S, Durham HD, Vande Velde C, Rouleau GA and Drapeau P (2010). "Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo." Hum Mol Genet **19**(4): 671-683.

Kaindl AM, Passemard S, Kumar P, Kraemer N, Issa L, Zwirner A, Gerard B, Verloes A, Mani S and Gressens P (2010). "Many roads lead to primary autosomal recessive microcephaly." Prog Neurobiol **90**(3): 363-383.

Kalscheuer VM, Tao J, Donnelly A, Hollway G, Schwinger E, Kubart S, Menzel C, Hoeltzenbein M, Tommerup N, Eyre H, Harbord M, Haan E, Sutherland GR, Ropers HH and Gecz J (2003). "Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation." Am J Hum Genet **72**(6): 1401-1411.

Kamada F, Aoki Y, Narisawa A, Abe Y, Komatsuzaki S, Kikuchi A, Kanno J, Niihori T, Ono M, Ishii N, Owada Y, Fujimura M, Mashimo Y, Suzuki Y, Hata A, Tsuchiya S, Tominaga T, Matsubara Y and Kure S (2011). "A genome-wide association study identifies RNF213 as the first Moyamoya disease gene." J Hum Genet **56**(1): 34-40.

Kang JQ, Shen W and Macdonald RL (2009). "Two molecular pathways (NMD and ERAD) contribute to a genetic epilepsy associated with the GABA(A) receptor GABRA1 PTC mutation, 975delC, S326fs328X." J Neurosci **29**(9): 2833-2844.

Karsenty G, Kronenberg HM and Settembre C (2009). "Genetic control of bone formation." Annu Rev Cell Dev Biol **25**: 629-648.

Kaufman L, Ayub M and Vincent JB (2010). "The genetic basis of non-syndromic intellectual disability: a review." J Neurodev Disord **2**(4): 182-209.

Kawauchi T, Sekine K, Shikanai M, Chihama K, Tomita K, Kubo K, Nakajima K, Nabeshima Y and Hoshino M (2010). "Rab GTPases-dependent endocytic pathways regulate neuronal migration and maturation through N-cadherin trafficking." Neuron **67**(4): 588-602.

Kelley RI, Robinson D, Puffenberger EG, Strauss KA and Morton DH (2002). "Amish lethal microcephaly: a new metabolic disorder with severe congenital microcephaly and 2-ketoglutaric aciduria." Am J Med Genet **112**(4): 318-326.

Kim MJ, Dunah AW, Wang YT and Sheng M (2005). "Differential roles of NR2A- and NR2B-containing NMDA receptors in Ras-ERK signaling and AMPA receptor trafficking." Neuron **46**(5): 745-760.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B and Schilling TF (1995). "Stages of embryonic development of the zebrafish." Dev Dyn **203**(3): 253-310.

Kimura Y, Okamura Y and Higashijima S (2006). "alx, a zebrafish homolog of Chx10, marks ipsilateral descending excitatory interneurons that participate in the regulation of spinal locomotor circuits." J Neurosci **26**(21): 5684-5697.

Kirin M, Mcquillan R, Franklin CS, Campbell H, Mckeigue PM and Wilson JF (2010). "Genomic runs of homozygosity record population history and consanguinity." PLoS One **5**(11): e13996.

Kleefstra T and Hamel BC (2005). "X-linked mental retardation: further lumping, splitting and emerging phenotypes." Clin Genet **67**(6): 451-467.

Kleefstra T, Yntema HG, Oudakker AR, Banning MJ, Kalscheuer VM, Chelly J, Moraine C, Ropers HH, Fryns JP, Janssen IM, Sistermans EA, Nillesen WN, De Vries LB, Hamel BC and Van Bokhoven H (2004). "Zinc finger 81 (ZNF81) mutations associated with X-linked mental retardation." J Med Genet **41**(5): 394-399.

Knorz VJ, Spalluto C, Lessard M, Purvis TL, Adigun FF, Collin GB, Hanley NA, Wilson DI and Hearn T (2010). "Centriolar association of ALMS1 and likely centrosomal functions of the ALMS motif-containing proteins C10orf90 and KIAA1731." Mol Biol Cell **21**(21): 3617-3629.

Kobayashi H, Yamazaki S, Takashima S, Liu W, Okuda H, Yan J, Fujii Y, Hitomi T, Harada KH, Habu T and Koizumi A (2013). "Ablation of Rnf213 retards progression of diabetes in the Akita mouse." Biochem Biophys Res Commun **432**(3): 519-525.

Kumar A, Girimaji SC, Duvvari MR and Blanton SH (2009a). "Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly." Am J Hum Genet **84**(2): 286-290.

Kumar P, Henikoff S and Ng PC (2009b). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-1081.

Kumar V, Zhang MX, Swank MW, Kunz J and Wu GY (2005). "Regulation of dendritic morphogenesis by Ras-PI3K-Akt-mTOR and Ras-MAPK signaling pathways." J Neurosci **25**(49): 11288-11299.

Kuper J, Wolski SC, Michels G and Kisker C (2012). "Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation." EMBO J **31**(2): 494-502.

Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Hasegawa T, Chinen Y, Tomita Ha HA, Kinoshita A, Mizuguchi T, Yoshiura Ki K, Ohta T, Kishino T, Fukushima Y, Niikawa N and Matsumoto N (2002). "Haploinsufficiency of NSD1 causes Sotos syndrome." Nat Genet **30**(4): 365-366.

Kutsche K, Yntema H, Brandt A, Jantke I, Nothwang HG, Orth U, Boavida MG, David D, Chelly J, Fryns JP, Moraine C, Ropers HH, Hamel BC, Van Bokhoven H and Gal A (2000). "Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation." Nat Genet **26**(2): 247-250.

Lander ES and Botstein D (1987). "Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children." Science **236**(4808): 1567-1570.

Langmead B, Trapnell C, Pop M and Salzberg SL (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Larsen TM, Boehlein SK, Schuster SM, Richards NG, Thoden JB, Holden HM and Rayment I (1999). "Three-dimensional structure of Escherichia coli asparagine synthetase B: a short journey from substrate to product." Biochemistry **38**(49): 16146-16157.

Lee JH, Huynh M, Silhavy JL, Kim S, Dixon-Salazar T, Heiberg A, Scott E, Bafna V, Hill KJ, Collazo A, Funari V, Russ C, Gabriel SB, Mathern GW and Gleeson JG (2012). "De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly." Nat Genet **44**(8): 941-945.

Lehmann AR (2001). "The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases." Genes Dev **15**(1): 15-23.

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**(3): 117-134.

Lerche H, Shah M, Beck H, Noebels J, Johnston D and Vincent A (2013). "Ion channels in genetic and acquired forms of epilepsy." J Physiol **591**(Pt 4): 753-764.

Lerche H, Weber YG, Jurkat-Rott K and Lehmann-Horn F (2005). "Ion channel defects in idiopathic epilepsies." Curr Pharm Des **11**(21): 2737-2752.

Levitus M, Waisfisz Q, Godthelp BC, De Vries Y, Hussain S, Wiegant WW, Elghalbzouri-Maghrani E, Steltenpool J, Rooimans MA, Pals G, Arwert F, Mathew CG, Zdzienicka MZ, Hiom K, De Winter JP and Joenje H (2005). "The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J." Nat Genet **37**(9): 934-935.

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, Macdonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, Mcintosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL and Venter JC (2007). "The diploid genome sequence of an individual human." PLoS Biol **5**(10): e254.

Li BS, Gu LJ, Luo CY, Li WS, Jiang LM, Shen SH, Jiang H, Shen SH, Zhang B, Chen J, Xue HL and Tang JY (2006). "The downregulation of asparagine synthetase expression can increase the sensitivity of cells resistant to l-asparaginase." Leukemia **20**(12): 2199-2201.

Li H and Durbin R (2009a). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and Genome Project Data Processing S (2009b). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Liao JC and Fetcho JR (2008). "Shared versus specialized glycinergic spinal interneurons in axial motor circuits of larval zebrafish." J Neurosci **28**(48): 12982-12992.

Liaw D, Marsh DJ, Li J, Dahia PL, Wang SI, Zheng Z, Bose S, Call KM, Tsou HC, Peacocke M, Eng C and Parsons R (1997). "Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome." Nat Genet **16**(1): 64-67.

Ligeti E, Welti S and Scheffzek K (2012). "Inhibition and termination of physiological responses by GTPase activating proteins." Physiol Rev **92**(1): 237-272.

Liu W, Morito D, Takashima S, Mineharu Y, Kobayashi H, Hitomi T, Hashikata H, Matsuura N, Yamazaki S, Toyoda A, Kikuta K, Takagi Y, Harada KH, Fujiyama A, Herzig R, Krischek B, Zou L, Kim JE, Kitakaze M, Miyamoto S, Nagata K, Hashimoto N and Koizumi A (2011). "Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development." PLoS One **6**(7): e22542.

Lizard G, Rouaud O, Demarquoy J, Cherkaoui-Malki M and Iuliano L (2012). "Potential roles of peroxisomes in Alzheimer's disease and in dementia of the Alzheimer's type." J Alzheimers Dis **29**(2): 241-254.

Lo Nigro C, Chong CS, Smith AC, Dobyns WB, Carrozzo R and Ledbetter DH (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): 157-164.

Lu Q, Wierzbicki S, Krasilnikov AS and Schmitt ME (2010). "Comparison of mitochondrial and nucleolar RNase MRP reveals identical RNA components with distinct enzymatic activities and protein components." RNA **16**(3): 529-537.

Lugtenberg D, Yntema HG, Banning MJ, Oudakker AR, Firth HV, Willatt L, Raynaud M, Kleefstra T, Fryns JP, Ropers HH, Chelly J, Moraine C, Gecz J, Van Reeuwijk J, Nabuurs SB, De Vries BB, Hamel BC, De Brouwer AP and Van Bokhoven H (2006). "ZNF674: a new kruppel-associated box-containing zinc-finger gene involved in nonsyndromic X-linked mental retardation." Am J Hum Genet **78**(2): 265-278.

Madhavan T and Narayan J (1991). "Consanguinity and mental retardation." J Ment Defic Res **35 (Pt 2)**: 133-139.

Mahmood S, Ahmad W and Hassan MJ (2011). "Autosomal Recessive Primary Microcephaly (MCPH): clinical manifestations, genetic heterogeneity and mutation continuum." Orphanet J Rare Dis **6**: 39.

Majnemer A and Shevell MI (1995). "Diagnostic yield of the neurologic assessment of the developmentally delayed child." J Pediatr **127**(2): 193-199.

Maloof J, Sledz K, Hogg JF, Bodensteiner JB, Schwartz T and Schochet SS (1994). "Unilateral megalencephaly and tuberous sclerosis: related disorders?" J Child Neurol **9**(4): 443-446.

Mandel JL and Chelly J (2004). "Monogenic X-linked mental retardation: is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations." Eur J Hum Genet **12**(9): 689-693.

Manolio TA (2010). "Genomewide association studies and assessment of the risk of disease." N Engl J Med **363**(2): 166-176.

Margulies EH, Vinson JP, Program NCS, Miller W, Jaffe DB, Lindblad-Toh K, Chang JL, Green ED, Lander ES, Mullikin JC and Clamp M (2005). "An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing." Proc Natl Acad Sci U S A **102**(13): 4795-4800.

Mari F, Azimonti S, Bertani I, Bolognese F, Colombo E, Caselli R, Scala E, Longo I, Grosso S, Pescucci C, Ariani F, Hayek G, Balestri P, Bergo A, Badaracco G, Zappella M, Broccoli V, Renieri A, Kilstrup-Nielsen C and Landsberger N (2005). "CDKL5 belongs to the same molecular pathway of MeCP2 and it is responsible for the early-onset seizure variant of Rett syndrome." Hum Mol Genet **14**(14): 1935-1946.

Matsuura Y and Stewart M (2005). "Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling." EMBO J **24**(21): 3681-3689.

May PA and Gossage JP (2001). "Estimating the prevalence of fetal alcohol syndrome. A summary." Alcohol Res Health **25**(3): 159-167.

Maydan G, Noyman I, Har-Zahav A, Neriah ZB, Pasmanik-Chor M, Yeheskel A, Albin-Kaplanski A, Maya I, Magal N, Birk E, Simon AJ, Halevy A, Rechavi G, Shohat M, Straussberg R and Basel-Vanagaite L (2011). "Multiple congenital anomalies-hypotonia-seizures syndrome is caused by a mutation in PIGN." J Med Genet **48**(6): 383-389.

Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M and Depristo MA (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

Mckusick VA (1998). Mendelian inheritance in man : a catalog of human genes and genetic disorders. Baltimore, Johns Hopkins University Press.

Mclaren J and Bryson SE (1987). "Review of recent epidemiological studies of mental retardation: prevalence, associated disorders, and etiology." Am J Ment Retard **92**(3): 243-254.

Mefford HC, Batshaw ML and Hoffman EP (2012). "Genomics, intellectual disability, and autism." N Engl J Med **366**(8): 733-743.

Megarbane A, Dagher R and Melki I (2008a). "Sib pair with previously unreported skeletal dysplasia." Am J Med Genet A **146A**(22): 2916-2919.

Megarbane A, Mehawej C, El Zahr A, Haddad S and Cormier-Daire V (2014). "A second family with autosomal recessive spondylometaphyseal dysplasia and early death." Am J Med Genet A **164A**(4): 1010-1014.

Megarbane A, Samaras L, Chedid R, Chouery E, Chretien D, Caillaud C, Abou-Ghoch J and Jalkh N (2008b). "Developmental delay, dysmorphic features, neonatal spontaneous fractures, wrinkled skin, and hepatic failure: a new metabolic syndrome?" Am J Med Genet A **146A**(24): 3198-3201.

Mehawej C, Delahodde A, Legeai-Mallet L, Delague V, Kaci N, Desvignes JP, Kibar Z, Capo-Chichi JM, Chouery E, Munnich A, Cormier-Daire V and Megarbane A (2014). "The impairment of MAGMAS function in human is responsible for a severe skeletal dysplasia." PLoS Genet **10**(5): e1004311.

Meng Y, Zhang Y, Tregoubov V, Janus C, Cruz L, Jackson M, Lu WY, Macdonald JF, Wang JY, Falls DL and Jia Z (2002). "Abnormal spine morphology and enhanced LTP in LIMK-1 knockout mice." Neuron **35**(1): 121-133.

Merienne K, Jacquot S, Pannetier S, Zeniou M, Bankier A, Gecz J, Mandel JL, Mulley J, Sassone-Corsi P and Hanauer A (1999). "A missense mutation in RPS6KA3 (RSK2) responsible for non-specific mental retardation." Nat Genet **22**(1): 13-14.

Metzker ML (2010). "Sequencing technologies - the next generation." Nat Rev Genet **11**(1): 31-46.

Miller SA, Dykes DD and Polesky HF (1988). "A simple salting out procedure for extracting DNA from human nucleated cells." Nucleic Acids Res **16**(3): 1215.

Min KW, Hwang JW, Lee JS, Park Y, Tamura TA and Yoon JB (2003). "TIP120A associates with cullins and modulates ubiquitin ligase activity." J Biol Chem **278**(18): 15905-15910.

Mitsui T, Kawai H, Sakoda S, Miyata M and Saito S (1994). "Hereditary parkinsonism with multiple system degeneration: beneficial effect of anticholinergics, but not of levodopa." J Neurol Sci **125**(2): 153-157.

Mnatzakanian GN, Lohi H, Munteanu I, Alfred SE, Yamada T, Macleod PJ, Jones JR, Scherer SW, Schanen NC, Friez MJ, Vincent JB and Minassian BA (2004). "A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome." Nat Genet **36**(4): 339-341.

Mochida GH, Mahajnah M, Hill AD, Basel-Vanagaite L, Gleason D, Hill RS, Bodell A, Crosier M, Straussberg R and Walsh CA (2009). "A truncating mutation of TRAPPC9 is

associated with autosomal-recessive intellectual disability and postnatal microcephaly." Am J Hum Genet **85**(6): 897-902.

Moczko M, Schonfisch B, Voos W, Pfanner N and Rassow J (1995). "The mitochondrial ClpB homolog Hsp78 cooperates with matrix Hsp70 in maintenance of mitochondrial function." J Mol Biol **254**(4): 538-543.

Modell B and Darr A (2002). "Science and society: genetic counselling and customary consanguineous marriage." Nat Rev Genet **3**(3): 225-229.

Mogk A, Schlieker C, Strub C, Rist W, Weibezahn J and Bukau B (2003). "Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity." J Biol Chem **278**(20): 17615-17624.

Morton NE (1978). "Effect of inbreeding on IQ and mental retardation." Proc Natl Acad Sci U S A **75**(8): 3906-3908.

Motazacker MM, Rost BR, Hucho T, Garshasbi M, Kahrizi K, Ullmann R, Abedini SS, Nieh SE, Amini SH, Goswami C, Tzschach A, Jensen LR, Schmitz D, Ropers HH, Najmabadi H and Kuss AW (2007). "A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation." Am J Hum Genet **81**(4): 792-798.

Murdock DG, Boone BE, Esposito LA and Wallace DC (1999). "Up-regulation of nuclear and mitochondrial genes in the skeletal muscle of mice lacking the heart/muscle isoform of the adenine nucleotide translocator." J Biol Chem **274**(20): 14429-14433.

Musante L and Ropers HH (2014). "Genetics of recessive cognitive disorders." Trends Genet **30**(1): 32-39.

Nachman MW and Crowell SL (2000). "Estimate of the mutation rate per nucleotide in humans." Genetics **156**(1): 297-304.

Nadif Kasri N, Nakano-Kobayashi A, Malinow R, Li B and Van Aelst L (2009). "The Rho-linked mental retardation protein oligophrenin-1 controls synapse maturation and plasticity by stabilizing AMPA receptors." Genes Dev **23**(11): 1289-1302.

Najm J, Horn D, Wimplinger I, Golden JA, Chizhikov VV, Sudi J, Christian SL, Ullmann R, Kuechler A, Haas CA, Flubacher A, Charnas LR, Uyanik G, Frank U, Klopocki E, Dobyns WB and Kutsche K (2008). "Mutations of CASK cause an X-linked brain malformation phenotype with microcephaly and hypoplasia of the brainstem and cerebellum." Nat Genet **40**(9): 1065-1067.

Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P, Zecha A, Mohseni M, Puttmann L, Vahid LN, Jensen C, Moheb LA, Bienek M, Larti F, Mueller I, Weissmann R, Darvish H, Wrogemann K, Hadavi V, Lipkowitz B, Esmaeeli-Nieh S, Wieczorek D, Kariminejad R, Firouzabadi SG, Cohen M,

Fattahi Z, Rost I, Mojahedi F, Hertzberg C, Dehghan A, Rajab A, Banavandi MJ, Hoffer J, Falah M, Musante L, Kalscheuer V, Ullmann R, Kuss AW, Tzschach A, Kahrizi K and Ropers HH (2011). "Deep sequencing reveals 50 novel genes for recessive cognitive disorders." Nature **478**(7367): 57-63.

Nakashima A, Yoshino K, Miyamoto T, Eguchi S, Oshiro N, Kikkawa U and Yonezawa K (2007). "Identification of TBC7 having TBC domain as a novel binding protein to TSC1-TSC2 complex." Biochem Biophys Res Commun **361**(1): 218-223.

Neuwald AF, Aravind L, Spouge JL and Koonin EV (1999). "AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes." Genome Res **9**(1): 27-43.

Newschaffer CJ, Croen LA, Daniels J, Giarelli E, Grether JK, Levy SE, Mandell DS, Miller LA, Pinto-Martin J, Reaven J, Reynolds AM, Rice CE, Schendel D and Windham GC (2007). "The epidemiology of autism spectrum disorders." Annu Rev Public Health **28**: 235-258.

Ng D, Thakker N, Corcoran CM, Donnai D, Perveen R, Schneider A, Hadley DW, Tifft C, Zhang L, Wilkie AO, Van Der Smagt JJ, Gorlin RJ, Burgess SM, Bardwell VJ, Black GC and Biesecker LG (2004). "Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in BCOR." Nat Genet **36**(4): 411-416.

Ng PC and Henikoff S (2003). "SIFT: Predicting amino acid changes that affect protein function." Nucleic Acids Res **31**(13): 3812-3814.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA and Shendure J (2009). "Targeted capture and massively parallel sequencing of 12 human exomes." Nature **461**(7261): 272-276.

Niccols A (2007). "Fetal alcohol syndrome and the developing socio-emotional brain." Brain Cogn **65**(1): 135-142.

Nicholas AK, Khurshid M, Desir J, Carvalho OP, Cox JJ, Thornton G, Kausar R, Ansar M, Ahmad W, Verloes A, Passemard S, Misson JP, Lindsay S, Gergely F, Dobyns WB, Roberts E, Abramowicz M and Woods CG (2010). "WDR62 is associated with the spindle pole and is mutated in human microcephaly." Nat Genet **42**(11): 1010-1014.

Nishimura F, Nishihara M, Mori M, Torii K and Takahashi M (1995). "Excitability of neurons in the ventromedial nucleus in rat hypothalamic slices: modulation by amino acids at cerebrospinal fluid levels." Brain Res **691**(1-2): 217-222.

Ogawa Y, Miyamoto Y, Asally M, Oka M, Yasuda Y and Yoneda Y (2010). "Two isoforms of Npap60 (Nup50) differentially regulate nuclear protein import." Mol Biol Cell **21**(4): 630-638.

Ogura T, Whiteheart SW and Wilkinson AJ (2004). "Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases." J Struct Biol **146**(1-2): 106-112.

Ogura T and Wilkinson AJ (2001). "AAA+ superfamily ATPases: common structure--diverse function." Genes Cells **6**(7): 575-597.

Osborne LR, Martindale D, Scherer SW, Shi XM, Huizenga J, Heng HH, Costa T, Pober B, Lew L, Brinkman J, Rommens J, Koop B and Tsui LC (1996). "Identification of genes from a 500-kb region at 7q11.23 that is commonly deleted in Williams syndrome patients." Genomics **36**(2): 328-336.

Otte C and Rauch A (2013). "[Intellectual disability - a frequent reason for referral to medical genetics]." Praxis (Bern 1994) **102**(24): 1467-1473.

Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, Walford GA, Sugiana C, Boneh A, Chen WK, Hill DE, Vidal M, Evans JG, Thorburn DR, Carr SA and Mootha VK (2008). "A mitochondrial protein compendium elucidates complex I disease biology." Cell **134**(1): 112-123.

Parish JL, Rosa J, Wang X, Lahti JM, Doxsey SJ and Androphy EJ (2006). "The DNA helicase ChlR1 is required for sister chromatid cohesion in mammalian cells." J Cell Sci **119**(Pt 23): 4857-4865.

Parsell DA, Kowal AS, Singer MA and Lindquist S (1994). "Protein disaggregation mediated by heat-shock protein Hsp104." Nature **372**(6505): 475-478.

Pavlowsky A, Chelly J and Billuart P (2012). "Emerging major synaptic signaling pathways involved in intellectual disability." Mol Psychiatry **17**(7): 682-693.

Peng J, Huang CH, Short MK and Jubinsky PT (2005). "Magmas gene structure and evolution." In Silico Biol **5**(3): 251-263.

Petrij F, Dorsman JC, Dauwerse HG, Giles RH, Peeters T, Hennekam RC, Breuning MH and Peters DJ (2000). "Rubinstein-Taybi syndrome caused by a De Novo reciprocal translocation t(2;16)(q36.3;p13.3)." Am J Med Genet **92**(1): 47-52.

Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, Searle S, Farrell CM, Loveland JE, Ruef BJ, Hart E, Suner MM, Landrum MJ, Aken B, Ayling S, Baertsch R, Fernandez-Banet J, Cherry JL, Curwen V, Dicuccio M, Kellis M, Lee J, Lin MF, Schuster M, Shkeda A, Amid C, Brown G, Dukhanina O, Frankish A, Hart J, Maidak BL, Mudge J, Murphy MR, Murphy T, Rajan J, Rajput B, Riddick LD, Snow C, Steward C, Webb D, Weber JA, Wilming L, Wu W, Birney E, Haussler D, Hubbard T, Ostell J, Durbin R and Lipman D (2009). "The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes." Genome Res **19**(7): 1316-1323.

Pueschel SM, Louis S and Mcknight P (1991). "Seizure disorders in Down syndrome." Arch Neurol **48**(3): 318-320.

Puffenberger EG, Strauss KA, Ramsey KE, Craig DW, Stephan DA, Robinson DL, Hendrickson CL, Gottlieb S, Ramsay DA, Siu VM, Heuer GG, Crino PB and Morton DH (2007). "Polyhydramnios, megalencephaly and symptomatic epilepsy caused by a homozygous 7-kilobase deletion in LYK5." Brain **130**(Pt 7): 1929-1941.

Pugh RA, Wu CG and Spies M (2012). "Regulation of translocation polarity by helicase domain 1 in SF2B helicases." EMBO J **31**(2): 503-514.

Pulvers JN, Bryk J, Fish JL, Wilsch-Brauninger M, Arai Y, Schreier D, Naumann R, Helppi J, Habermann B, Vogt J, Nitsch R, Toth A, Enard W, Paabo S and Huttner WB (2010). "Mutations in mouse Aspm (abnormal spindle-like microcephaly associated) cause not only microcephaly but also major defects in the germline." Proc Natl Acad Sci U S A **107**(38): 16595-16600.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ and Sham PC (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-575.

Puri R, Verma I and Bhargava I (1978). "Effects of consanguinity in a community in Pondicherry." Medical Genetics in India **2**: 129-139.

Pussegoda KA (2010). "Exome sequencing: locating causative genes in rare disorders." Clin Genet **78**(1): 32-33.

Qin Y, Zhu Y, Baumgart JP, Stornetta RL, Seidenman K, Mack V, Van Aelst L and Zhu JJ (2005). "State-dependent Ras signaling and AMPA receptor trafficking." Genes Dev **19**(17): 2000-2015.

Quinlan AR and Hall IM (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Ramakers GJ (2002). "Rho proteins, mental retardation and the cellular basis of cognition." Trends Neurosci **25**(4): 191-199.

Rampon C, Tang YP, Goodhouse J, Shimizu E, Kyin M and Tsien JZ (2000). "Enrichment induces structural changes and recovery from nonspatial memory deficits in CA1 NMDAR1-knockout mice." Nat Neurosci **3**(3): 238-244.

Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, Zenker M, Huffmeier U, Thiel C, Ruschendorf F, Nurnberg P, Reis A and Trautmann U (2006). "Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation." Am J Med Genet A **140**(19): 2063-2074.

Reed W and Reed S (1965). Mental retardation : A family study., Philadelphia: Saunders.

Reiner O, Carrozzo R, Shen Y, Wehnert M, Faustinella F, Dobyns WB, Caskey CT and Ledbetter DH (1993). "Isolation of a Miller-Dieker lissencephaly gene containing G protein beta-subunit-like repeats." Nature **364**(6439): 717-721.

Rex CS, Chen LY, Sharma A, Liu J, Babayan AH, Gall CM and Lynch G (2009). "Different Rho GTPase-dependent signaling pathways initiate sequential steps in the consolidation of long-term potentiation." J Cell Biol **186**(1): 85-97.

Richards NG and Kilberg MS (2006). "Asparagine synthetase chemotherapy." Annu Rev Biochem **75**: 629-654.

Ridanpaa M, Van Eenennaam H, Pelin K, Chadwick R, Johnson C, Yuan B, Vanvenrooij W, Pruijn G, Salmela R, Rockas S, Makitie O, Kaitila I and De La Chapelle A (2001). "Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia." Cell **104**(2): 195-203.

Riviere JB, Mirzaa GM, O'roak BJ, Beddaoui M, Alcantara D, Conway RL, St-Onge J, Schwartzentruber JA, Gripp KW, Nikkel SM, Worthylake T, Sullivan CT, Ward TR, Butler HE, Kramer NA, Albrecht B, Armour CM, Armstrong L, Caluseriu O, Cytrynbaum C, Drolet BA, Innes AM, Lauzon JL, Lin AE, Mancini GM, Meschino WS, Reggin JD, Saggar AK, Lerman-Sagie T, Uyanik G, Weksberg R, Zirn B, Beaulieu CL, Finding of Rare Disease Genes Canada C, Majewski J, Bulman DE, O'driscoll M, Shendure J, Graham JM, Jr., Boycott KM and Dobyns WB (2012). "De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes." Nat Genet **44**(8): 934-940.

Rizzo JM and Buck MJ (2012). "Key principles and clinical applications of "next-generation" DNA sequencing." Cancer Prev Res (Phila) **5**(7): 887-900.

Roeleveld N, Zielhuis GA and Gabreels F (1997). "The prevalence of mental retardation: a critical review of recent literature." Dev Med Child Neurol **39**(2): 125-132.

Rohena L, Neidich J, Truitt Cho M, Gonzalez KD, Tang S, Devinsky O and Chung WK (2013). "Mutation in SNAP25 as a novel genetic cause of epilepsy and intellectual disability." Rare Dis **1**: e26314.

Romano C, Tine A, Fazio G, Rizzo R, Colognola RM, Sorge G, Bergonzi P and Pavone L (1990). "Seizures in patients with trisomy 21." Am J Med Genet Suppl **7**: 298-300.

Romeo Y, Moreau J, Zindy PJ, Saba-El-Leil M, Lavoie G, Dandachi F, Baptissart M, Borden KL, Meloche S and Roux PP (2013). "RSK regulates activated BRAF signalling to mTORC1 and promotes melanoma growth." Oncogene **32**(24): 2917-2926.

Rooryck C, Diaz-Font A, Osborn DP, Chabchoub E, Hernandez-Hernandez V, Shamseldin H, Kenny J, Waters A, Jenkins D, Kaissi AA, Leal GF, Dallapiccola B, Carnevale F, Bitner-Glindzicz M, Lees M, Hennekam R, Stanier P, Burns AJ, Peeters H, Alkuraya FS and Beales PL (2011). "Mutations in lectin complement pathway genes COLEC11 and MASP1 cause 3MC syndrome." Nat Genet **43**(3): 197-203.

Ropers HH (2006). "X-linked mental retardation: many genes for a complex disorder." Curr Opin Genet Dev **16**(3): 260-269.

Ropers HH (2008). "Genetics of intellectual disability." Curr Opin Genet Dev **18**(3): 241-250.

Ropers HH (2010). "Genetics of early onset cognitive impairment." Annu Rev Genomics Hum Genet **11**: 161-187.

Ropers HH and Hamel BC (2005). "X-linked mental retardation." Nat Rev Genet **6**(1): 46-57.

Roy S, Short MK, Stanley ER and Jubinsky PT (2012). "Essential role of Drosophila black-pearl is mediated by its effects on mitochondrial respiration." FASEB J **26**(9): 3822-3833.

Rudolf J, Makrantoni V, Ingledew WJ, Stark MJ and White MF (2006). "The DNA repair helicases XPD and FancJ have essential iron-sulfur domains." Mol Cell **23**(6): 801-808.

Rumenapp U, Freichel-Blomquist A, Wittinghofer B, Jakobs KH and Wieland T (2002). "A mammalian Rho-specific guanine-nucleotide exchange factor (p164-RhoGEF) without a pleckstrin homology domain." Biochem J **366**(Pt 3): 721-728.

Ruzzo EK, Capo-Chichi JM, Ben-Zeev B, Chitayat D, Mao H, Pappas AL, Hitomi Y, Lu YF, Yao X, Hamdan FF, Pelak K, Reznik-Wolf H, Bar-Joseph I, Oz-Levi D, Lev D, Lerman-Sagie T, Leshinsky-Silver E, Anikster Y, Ben-Asher E, Olender T, Colleaux L, Decarie JC, Blaser S, Banwell B, Joshi RB, He XP, Patry L, Silver RJ, Dobrzeniecka S, Islam MS, Hasnat A, Samuels ME, Aryal DK, Rodriguiz RM, Jiang YH, Wetsel WC, Mcnamara JO, Rouleau GA, Silver DL, Lancet D, Pras E, Mitchell GA, Michaud JL and Goldstein DB (2013). "Deficiency of asparagine synthetase causes congenital microcephaly and a progressive form of encephalopathy." Neuron **80**(2): 429-441.

Saint-Amant L (2006). "Development of motor networks in zebrafish embryos." Zebrafish **3**(2): 173-190.

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM and Smith M (1977). "Nucleotide sequence of bacteriophage phi X174 DNA." Nature **265**(5596): 687-695.

Sarnat HB and Curatolo P (2008). Malformations of the nervous system. Edinburgh ; New York, Elsevier.

Schatz G and Dobberstein B (1996). "Common principles of protein translocation across membranes." Science **271**(5255): 1519-1526.

Schiestl RH and Gietz RD (1989). "High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier." Curr Genet **16**(5-6): 339-346.

Schneider CA, Rasband WS and Eliceiri KW (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Methods **9**(7): 671-675.

Scholl-Burgi S, Haberlandt E, Heinz-Erian P, Deisenhammer F, Albrecht U, Sigl SB, Rauchenzauner M, Ulmer H and Karall D (2008). "Amino acid cerebrospinal fluid/plasma ratios in children: influence of age, gender, and antiepileptic medication." Pediatrics **121**(4): e920-926.

Schwartz CE, Gillessen-Kaesbach G, May M, Cappa M, Gorski J, Steindl K and Neri G (2000). "Two novel mutations confirm FGD1 is responsible for the Aarskog syndrome." Eur J Hum Genet **8**(11): 869-874.

Scott RM and Smith ER (2009). "Moyamoya disease and moyamoya syndrome." N Engl J Med **360**(12): 1226-1237.

Seidman LJ, Buka SL, Goldstein JM, Horton NJ, Rieder RO and Tsuang MT (2000). "The relationship of prenatal and perinatal complications to cognitive functioning at age 7 in the New England Cohorts of the National Collaborative Perinatal Project." Schizophr Bull **26**(2): 309-321.

Shen J, Gilmore EC, Marshall CA, Haddadin M, Reynolds JJ, Eyaid W, Bodell A, Barry B, Gleason D, Allen K, Ganesh VS, Chang BS, Grix A, Hill RS, Topcu M, Caldecott KW, Barkovich AJ and Walsh CA (2010). "Mutations in PNKP cause microcephaly, seizures and defects in DNA repair." Nat Genet **42**(3): 245-249.

Shendure J and Ji H (2008). "Next-generation DNA sequencing." Nat Biotechnol **26**(10): 1135-1145.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM and Sirotkin K (2001). "dbSNP: the NCBI database of genetic variation." Nucleic Acids Res **29**(1): 308-311.

Shiraishi S, Zhou C, Aoki T, Sato N, Chiba T, Tanaka K, Yoshida S, Nabeshima Y, Nabeshima Y and Tamura TA (2007). "TBP-interacting protein 120B (TIP120B)/cullin-associated and neddylation-dissociated 2 (CAND2) inhibits SCF-dependent ubiquitination of myogenin and accelerates myogenic differentiation." J Biol Chem **282**(12): 9017-9028.

Shoichet SA, Hoffmann K, Menzel C, Trautmann U, Moser B, Hoeltzenbein M, Echenne B, Partington M, Van Bokhoven H, Moraine C, Fryns JP, Chelly J, Rott HD, Ropers HH and Kalscheuer VM (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**(6): 1341-1354.

Short MK, Hallett JP, Tar K, Dange T, Schmidt M, Moir R, Willis IM and Jubinsky PT (2012). "The yeast magmas ortholog pam16 has an essential function in fermentative growth that involves sphingolipid metabolism." PLoS One **7**(7): e39428.

Shorvon SD (2000). General principles of treatment in epilepsy, London: Blackwell Science.

Shoubridge C, Tarpey PS, Abidi F, Ramsden SL, Rujirabanjerd S, Murphy JA, Boyle J, Shaw M, Gardner A, Proos A, Puusepp H, Raymond FL, Schwartz CE, Stevenson RE, Turner G, Field M, Walikonis RS, Harvey RJ, Hackett A, Futreal PA, Stratton MR and Gecz J (2010). "Mutations in the guanine nucleotide exchange factor gene IQSEC2 cause nonsyndromic intellectual disability." Nat Genet **42**(6): 486-488.

Sikorski RS and Boeke JD (1991). "In vitro mutagenesis and plasmid shuffling: from cloned gene to mutant yeast." Methods Enzymol **194**: 302-318.

Sinha D, Joshi N, Chittoor B, Samji P and D'silva P (2010). "Role of Magmas in protein transport and human mitochondria biogenesis." Hum Mol Genet **19**(7): 1248-1262.

Siomi H, Siomi MC, Nussbaum RL and Dreyfuss G (1993). "The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein." Cell **74**(2): 291-298.

Siu VM, Ratko S, Prasad AN, Prasad C and Rupar CA (2010). "Amish microcephaly: Long-term survival and biochemical characterization." Am J Med Genet A **152A**(7): 1747-1751.

Skibbens RV (2004). "Chl1p, a DNA helicase-like protein in budding yeast, functions in sister-chromatid cohesion." Genetics **166**(1): 33-42.

Skuse DH (2005). "X-linked genes and mental functioning." Hum Mol Genet **14 Spec No 1**: R27-32.

Smitherman M, Lee K, Swanger J, Kapur R and Clurman BE (2000). "Characterization and targeted disruption of murine Nup50, a p27(Kip1)-interacting component of the nuclear pore complex." Mol Cell Biol **20**(15): 5631-5642.

Sparkes R, Patton D and Bernier F (2007). "Cardiac features of a novel autosomal recessive dilated cardiomyopathic syndrome due to defective importation of mitochondrial protein." Cardiol Young **17**(2): 215-217.

Srivastava AK and Schwartz CE (2014). "Intellectual disability and autism spectrum disorders: Causal genes and molecular mechanisms." Neurosci Biobehav Rev **46P2**: 161-174.

Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeysinghe S, Krawczak M and Cooper DN (2003). "Human Gene Mutation Database (HGMD): 2003 update." Hum Mutat **21**(6): 577-581.

Stevenson RE, Schwartz CE and Schroer RJ (2000). X-linked mental retardation. New York, Oxford University Press.

Stewart M (2006). "Structural basis for the nuclear protein import cycle." Biochem Soc Trans **34**(Pt 5): 701-704.

Stromme P, Mangelsdorf ME, Scheffer IE and Gecz J (2002a). "Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in Aristaless related homeobox gene, ARX." Brain Dev **24**(5): 266-268.

Stromme P, Mangelsdorf ME, Shaw MA, Lower KM, Lewis SM, Bruyere H, Lutcherath V, Gedeon AK, Wallace RH, Scheffer IE, Turner G, Partington M, Frints SG, Fryns JP, Sutherland GR, Mulley JC and Gecz J (2002b). "Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy." Nat Genet **30**(4): 441-445.

Sulonen AM, Ellonen P, Almusa H, Lepisto M, Eldfors S, Hannula S, Miettinen T, Tyynismaa H, Salo P, Heckman C, Joensuu H, Raivio T, Suomalainen A and Saarela J (2011). "Comparison of solution-based exome capture methods for next generation sequencing." Genome Biol **12**(9): R94.

Swerdlow H, Wu SL, Harke H and Dovichi NJ (1990). "Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette." J Chromatogr **516**(1): 61-67.

Tadmouri GO, Nair P, Obeid T, Al Ali MT, Al Khaja N and Hamamy HA (2009). "Consanguinity and reproductive health among Arabs." Reprod Health **6**: 17.

Tanaka M, Miyoshi J, Ishizaki H, Togawa A, Ohnishi K, Endo K, Matsubara K, Mizoguchi A, Nagano T, Sato M, Sasaki T and Takai Y (2001). "Role of Rab3 GDP/GTP exchange protein in synaptic vesicle trafficking at the mouse neuromuscular junction." Mol Biol Cell **12**(5): 1421-1430.

Tang SJ, Reis G, Kang H, Gingras AC, Sonenberg N and Schuman EM (2002). "A rapamycin-sensitive signaling pathway contributes to long-term synaptic plasticity in the hippocampus." Proc Natl Acad Sci U S A **99**(1): 467-472.

Tarpey P, Parnau J, Blow M, Woffendin H, Bignell G, Cox C, Cox J, Davies H, Edkins S, Holden S, Korny A, Mallya U, Moon J, O'meara S, Parker A, Stephens P, Stevens C, Teague J, Donnelly A, Mangelsdorf M, Mulley J, Partington M, Turner G, Stevenson R, Schwartz C, Young I, Easton D, Bobrow M, Futreal PA, Stratton MR, Gecz J, Wooster R and Raymond FL (2004). "Mutations in the DLG3 gene cause nonsyndromic X-linked mental retardation." Am J Hum Genet **75**(2): 318-324.

Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'meara S, Latimer C, Dicks E, Menzies A, Stephens P, Blow M, Greenman C, Xue Y, Tyler-Smith C, Thompson D, Gray K, Andrews J, Barthorpe S, Buck G, Cole J, Dunmore R, Jones D, Maddison M, Mironenko T, Turner R, Turrell K, Varian J, West S, Widaa S, Wray P, Teague J, Butler A,

Jenkinson A, Jia M, Richardson D, Shepherd R, Wooster R, Tejada MI, Martinez F, Carvill G, Goliath R, De Brouwer AP, Van Bokhoven H, Van Esch H, Chelly J, Raynaud M, Ropers HH, Abidi FE, Srivastava AK, Cox J, Luo Y, Mallya U, Moon J, Parnau J, Mohammed S, Tolmie JL, Shoubridge C, Corbett M, Gardner A, Haan E, Rujirabanjerd S, Shaw M, Vandeleur L, Fullston T, Easton DF, Boyle J, Partington M, Hackett A, Field M, Skinner C, Stevenson RE, Bobrow M, Turner G, Schwartz CE, Gecz J, Raymond FL, Futreal PA and Stratton MR (2009). "A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation." Nat Genet **41**(5): 535-543.

Tashiro A and Yuste R (2004). "Regulation of dendritic spine motility and stability by Rac1 and Rho kinase: evidence for two forms of spine motility." Mol Cell Neurosci **26**(3): 429-440.

Tatton-Brown K, Seal S, Ruark E, Harmer J, Ramsay E, Del Vecchio Duarte S, Zachariou A, Hanks S, O'brien E, Aksglaede L, Baralle D, Dabir T, Gener B, Goudie D, Homfray T, Kumar A, Pilz DT, Selicorni A, Temple IK, Van Maldergem L, Yachelevich N, Childhood Overgrowth C, Van Montfort R and Rahman N (2014). "Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability." Nat Genet **46**(4): 385-388.

Taylor DC, Falconer MA, Bruton CJ and Corsellis JA (1971). "Focal dysplasia of the cerebral cortex in epilepsy." J Neurol Neurosurg Psychiatry **34**(4): 369-387.

Tee AR, Manning BD, Roux PP, Cantley LC and Blenis J (2003). "Tuberous sclerosis complex gene products, Tuberin and Hamartin, control mTOR signaling by acting as a GTPase-activating protein complex toward Rheb." Curr Biol **13**(15): 1259-1268.

Thiel CT, Horn D, Zabel B, Ekici AB, Salinas K, Gebhart E, Ruschendorf F, Sticht H, Spranger J, Muller D, Zweier C, Schmitt ME, Reis A and Rauch A (2005). "Severely incapacitating mutations in patients with extreme short stature identify RNA-processing endoribonuclease RMRP as an essential cell growth regulator." Am J Hum Genet **77**(5): 795-806.

Thompson JD, Gibson TJ and Higgins DG (2002). "Multiple sequence alignment using ClustalW and ClustalX." Curr Protoc Bioinformatics **Chapter 2**: Unit 2 3.

Thornton GK and Woods CG (2009). "Primary microcephaly: do all roads lead to Rome?" Trends Genet **25**(11): 501-510.

Trivier E, De Cesare D, Jacquot S, Pannetier S, Zackai E, Young I, Mandel JL, Sassone-Corsi P and Hanauer A (1996). "Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome." Nature **384**(6609): 567-570.

Uziel G, Ghezzi D and Zeviani M (2011). "Infantile mitochondrial encephalopathy." Semin Fetal Neonatal Med **16**(4): 205-215.

Vale RD (2000). "AAA proteins. Lords of the ring." J Cell Biol **150**(1): F13-19.

Valenti D, De Bari L, De Filippis B, Henrion-Caude A and Vacca RA (2014). "Mitochondrial dysfunction as a central actor in intellectual disability-related diseases: An overview of Down syndrome, autism, Fragile X and Rett syndrome." Neurosci Biobehav Rev.

Van Der Crabben SN, Verhoeven-Duif NM, Brilstra EH, Van Maldergem L, Coskun T, Rubio-Gozalbo E, Berger R and De Koning TJ (2013). "An update on serine deficiency disorders." J Inherit Metab Dis **36**(4): 613-619.

Van Der Lelij P, Chrzanowska KH, Godthelp BC, Rooimans MA, Oostra AB, Stumm M, Zdzienicka MZ, Joenje H and De Winter JP (2010a). "Warsaw breakage syndrome, a cohesinopathy associated with mutations in the XPD helicase family member DDX11/ChlR1." Am J Hum Genet **86**(2): 262-266.

Van Der Lelij P, Oostra AB, Rooimans MA, Joenje H and De Winter JP (2010b). "Diagnostic Overlap between Fanconi Anemia and the Cohesinopathies: Roberts Syndrome and Warsaw Breakage Syndrome." Anemia **2010**: 565268.

Van Karnebeek CD, Shevell M, Zschocke J, Moeschler JB and Stockler S (2014). "The metabolic evaluation of the child with an intellectual developmental disorder: diagnostic algorithm for identification of treatable causes and new digital resource." Mol Genet Metab **111**(4): 428-438.

Van Karnebeek CD and Stockler S (2012). "Treatable inborn errors of metabolism causing intellectual disability: a systematic literature review." Mol Genet Metab **105**(3): 368-381.

Van Slegtenhorst M, De Hoogt R, Hermans C, Nellist M, Janssen B, Verhoef S, Lindhout D, Van Den Ouweland A, Halley D, Young J, Burley M, Jeremiah S, Woodward K, Nahmias J, Fox M, Ekong R, Osborne J, Wolfe J, Povey S, Snell RG, Cheadle JP, Jones AC, Tachataki M, Ravine D, Sampson JR, Reeve MP, Richardson P, Wilmer F, Munro C, Hawkins TL, Sepp T, Ali JB, Ward S, Green AJ, Yates JR, Kwiatkowska J, Henske EP, Short MP, Haines JH, Jozwiak S and Kwiatkowski DJ (1997). "Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34." Science **277**(5327): 805-808.

Veltman JA and Brunner HG (2012). "De novo mutations in human genetic disease." Nat Rev Genet **13**(8): 565-575.

Villard L, Fontes M, Ades LC and Gecz J (2000). "Identification of a mutation in the XNP/ATR-X gene in a family reported as Smith-Fineman-Myers syndrome." Am J Med Genet **91**(1): 83-85.

Vincent AK, Noor A, Janson A, Minassian BA, Ayub M, Vincent JB and Morel CF (2012). "Identification of genomic deletions spanning the PCDH19 gene in two unrelated girls with intellectual disability and seizures." Clin Genet **82**(6): 540-545.

Visel A, Thaller C and Eichele G (2004). "GenePaint.org: an atlas of gene expression patterns in the mouse embryo." Nucleic Acids Res **32**(Database issue): D552-556.

Vissers LE, De Ligt J, Gilissen C, Janssen I, Steehouwer M, De Vries P, Van Lier B, Arts P, Wieskamp N, Del Rosario M, Van Bon BW, Hoischen A, De Vries BB, Brunner HG and Veltman JA (2010). "A de novo paradigm for mental retardation." Nat Genet **42**(12): 1109-1112.

Vissers LE, Van Ravenswaaij CM, Admiraal R, Hurst JA, De Vries BB, Janssen IM, Van Der Vliet WA, Huys EH, De Jong PJ, Hamel BC, Schoenmakers EF, Brunner HG, Veltman JA and Van Kessel AG (2004). "Mutations in a new member of the chromodomain gene family cause CHARGE syndrome." Nat Genet **36**(9): 955-957.

Wagner K, Mick DU and Rehling P (2009). "Protein transport machineries for precursor translocation across the inner mitochondrial membrane." Biochim Biophys Acta **1793**(1): 52-59.

Wang K, Li M and Hakonarson H (2010). "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." Nucleic Acids Res **38**(16): e164.

Wang X, Wang H, Cao M, Li Z, Chen X, Patenia C, Gore A, Abboud EB, Al-Rajhi AA, Lewis RA, Lupski JR, Mardon G, Zhang K, Muzny D, Gibbs RA and Chen R (2011). "Whole-exome sequencing identifies ALMS1, IQCB1, CNGA3, and MYO7A mutations in patients with Leber congenital amaurosis." Hum Mutat **32**(12): 1450-1459.

Warman ML, Cormier-Daire V, Hall C, Krakow D, Lachman R, Lemerrer M, Mortier G, Mundlos S, Nishimura G, Rimoin DL, Robertson S, Savarirayan R, Sillence D, Spranger J, Unger S, Zabel B and Superti-Furga A (2011). "Nosology and classification of genetic skeletal disorders: 2010 revision." Am J Med Genet A **155A**(5): 943-968.

Watson P, Black G, Ramsden S, Barrow M, Super M, Kerr B and Clayton-Smith J (2001). "Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein." J Med Genet **38**(4): 224-228.

Westermann B and Neupert W (2000). "Mitochondria-targeted green fluorescent proteins: convenient tools for the study of organelle biogenesis in Saccharomyces cerevisiae." Yeast **16**(15): 1421-1427.

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, Mcguire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA and Rothberg JM (2008). "The complete genome of an individual by massively parallel DNA sequencing." Nature **452**(7189): 872-876.

Williams SR, Aldred MA, Der Kaloustian VM, Halal F, Gowans G, Mcleod DR, Zondag S, Toriello HV, Magenis RE and Elsea SH (2010). "Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems." Am J Hum Genet **87**(2): 219-228.

Winnepenninckx B, Errijgers V, Hayez-Delatte F, Reyniers E and Frank Kooy R (2002). "Identification of a family with nonspecific mental retardation (MRX79) with the A140V mutation in the MECP2 gene: is there a need for routine screening?" Hum Mutat **20**(4): 249-252.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-Macdonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M and Davis RW (1999). "Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis." Science **285**(5429): 901-906.

Woodley MA (2009). "Inbreeding depression and IQ in a study of 72 countries." Intelligence **37**: 268-276.

World Health Organization. (1980). International classification of impairments, disabilities, and handicaps : a manual of classification relating to the consequences of disease. Geneva

Albany, N.Y., World Health Organization ;

sold by WHO Publications Centre USA.

Wortmann SB, Kluijtmans LA, Engelke UF, Wevers RA and Morava E (2012). "The 3-methylglutaconic acidurias: what's new?" J Inherit Metab Dis **35**(1): 13-22.

Wu Y, Arai AC, Rumbaugh G, Srivastava AK, Turner G, Hayashi T, Suzuki E, Jiang Y, Zhang L, Rodriguez J, Boyle J, Tarpey P, Raymond FL, Nevelsteen J, Froyen G, Stratton M, Futreal A, Gecz J, Stevenson R, Schwartz CE, Valle D, Huganir RL and Wang T (2007). "Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans." Proc Natl Acad Sci U S A **104**(46): 18163-18168.

Wu Y, Sommers JA, Khan I, De Winter JP and Brosh RM, Jr. (2012). "Biochemical characterization of Warsaw breakage syndrome helicase." J Biol Chem **287**(2): 1007-1021.

Wu Y, Sommers JA, Suhasini AN, Leonard T, Deakyne JS, Mazin AV, Shin-Ya K, Kitao H and Brosh RM, Jr. (2010). "Fanconi anemia group J mutation abolishes its DNA repair function by uncoupling DNA translocation from helicase activity or disruption of protein-DNA complexes." Blood **116**(19): 3780-3791.

Wu Y, Suhasini AN and Brosh RM, Jr. (2009). "Welcome the family of FANCJ-like helicases to the block of genome stability maintenance proteins." Cell Mol Life Sci **66**(7): 1209-1222.

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ and Viegas-Pequignot E (1999). "Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene." Nature **402**(6758): 187-191.

Yangngam S, Plong-On O, Sripo T, Roongpraiwan R, Hansakunachai T, Wirojanan J, Sombuntham T, Ruangdaraganon N and Limprasert P (2014). "Mutation screening of the neurexin 1 gene in thai patients with intellectual disability and autism spectrum disorder." Genet Test Mol Biomarkers **18**(7): 510-515.

Yoshimura S, Gerondopoulos A, Linford A, Rigden DJ and Barr FA (2010). "Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors." J Cell Biol **191**(2): 367-381.

Zahir F and Friedman JM (2007). "The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility." Clin Genet **72**(4): 271-287.

Zeymer C, Barends TR, Werbeck ND, Schlichting I and Reinstein J (2014). "Elements in nucleotide sensing and hydrolysis of the AAA+ disaggregation machine ClpB: a structure-based mechanistic dissection of a molecular motor." Acta Crystallogr D Biol Crystallogr **70**(Pt 2): 582-595.

Zheng S, Kim H and Verhaak RG (2014). "Silent mutations make some noise." Cell **156**(6): 1129-1131.

Zlotogora J and Shalev SA (2010). "The consequences of consanguinity on the rates of malformations and major medical conditions at birth and in early childhood in inbred populations." Am J Med Genet A **152A**(8): 2023-2028.

RESSOURCES ÉLECTRONIQUES :

1000 genomes : http://browser.1000genomes.org/index.html

Annovar: http://www.openbioinformatics.org/annovar

Allen Brain atlas: http://www.brain-map.org

BioEdit software: http://www.mbio.ncsu.edu/bioedit/bioedit.html

ClustalW: http://www.ebi.ac.uk/Tools/msa/clustalw2

Complete genomics: http://www.completegenomics.com/public-data/69-Genomes/

Database of Genomic Variants: http://projects.tcag.ca/variation/

dbSNP: http://www.ncbi.nlm.nih.gov

Discovery Studio program: http://accelrys.com/products/discovery-studio/

Ensembl Genome Browser: http://www.ensembl.org/index.html

Eurexpress: http://www.eurexpress.org/

ExAC Genome Browser: http://exac.broadinstitute.org/

Exome Variant Server: http://evs.gs.washington.edu/EVS/

Fast PHASE: http://stephenslab.uchicago.edu/software.html#fastphase

Histoserv: http://www.histoservinc.com/

HUGO Gene Nomenclature Committe: http://www.genenames.org/

Image J: http://rsbweb.nih.gov/ij/

Infrafrontier mouse disease model: www.infrafrontier.eu

Integrative Genomics Viewer: http://www.broadinstitute.org/igv/

International Classification of Diseases: http://www.cdc.gov

KEEG pathways database: http://www.genome.jp/kegg/pathway.html

Leiden Open Variation database: http://www.lovd.nl/3.0/home

MGI mouse knockout database: http://www.informatics.jax.org/

Mutation taster: http://www.mutationtaster.org

McGill University Genome Quebec Innovation Center: http://gqinnovationcenter.com

NCBI CCDS: http://www.ncbi.nlm.nih.gov/CCDS

NCBI Refseq: http://www.ncbi.nlm.nih.gov/refseq

NCBI OMIM: http://www.ncbi.nlm.nih.gov/omim

NCBI Homologene : http://www.ncbi.nlm.nih.gov/homologene

Orphanet: http://www.orpha.net

PLink: http://pngu.mgh.harvard.edu/~purcell/plink

Picard's MarkDuplicates: http://picard.sourceforge.net/index.shtml

Polyphen-2: http://genetics.bwh.harvard.edu/pph2

Primer 3: http://bioinfo.ut.ee/primer3-0.4.0/

Protein Data Bank: http://www.rcsb.org/pdb/home/home.do

Provean: http://provean.jcvi.org/index.php

Uniprot: http://www.uniprot.org

SIFT: http://sift.jcvi.org