



**HAL**  
open science

# Caractérisation fonctionnelle de deux formes de Déficience Intellectuelle associées à des mutations dans DYRK1A et PQBP1

Jérémie Courraud

► **To cite this version:**

Jérémie Courraud. Caractérisation fonctionnelle de deux formes de Déficience Intellectuelle associées à des mutations dans DYRK1A et PQBP1. Médecine humaine et pathologie. Université de Strasbourg, 2021. Français. NNT : 2021STRAJ114 . tel-03771327

**HAL Id: tel-03771327**

**<https://theses.hal.science/tel-03771327v1>**

Submitted on 7 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





*À mes parents*





# REMERCIEMENTS

Je tiens à remercier l'ensemble des membres de mon jury, Christel Depienne, Gaëlle Friocourt, Vincent Laugel, Yann Hérault et Jean-Louis Mandel. Un grand merci à vous pour avoir accepté de juger mon travail et de prendre le temps de partager vos expertises.

Je tiens à remercier ma directrice de thèse, Amélie Piton qui m'a soutenu depuis mon arrivée. Amélie je te suis d'abord extrêmement reconnaissant de cette confiance que tu m'as accordée dès le début. Un grand merci pour tout ce que tu as pu me transmettre, pour m'avoir permis de réaliser ces projets et pour avoir été présente tout en me laissant mon autonomie. Outre le plaisir que j'ai pu prendre au niveau scientifique, je suis heureux de me dire que ce que nous avons développé a pu avoir un impact positif et ça, c'est grâce à ton implication. Si c'était à refaire je re-signerais tous les jours à condition que je garde la même encadrante.

Je tiens également à remercier très chaleureusement le professeur Jean Louis Mandel, sans qui je ne serais tout simplement pas là aujourd'hui. Jean-Lou je te remercie infiniment pour m'avoir fait confiance toi aussi et bien sûr pour ton soutien. Merci pour ton implication et l'intérêt que tu as montré pour ces travaux, pour tes conseils et tes remarques qui tapent toujours juste. Merci aussi pour ces conversations, qu'elles aient été scientifiques ou non, en passant par les asperges, K. Kollwitz et allant jusqu'aux gènes hollywoodiens. Je ne me laisserais jamais de cette « DERNIERE » anecdote avant d'y aller.

Bien sûr je remercie également tous les membres de mon équipe étiquetés « former », à commencer par la docteure Angélique Quartier. Angélique merci de m'avoir encadré mais également de m'avoir intégré à l'équipe lorsque j'ai débarqué, surtout vu ce que tu étais en train de subir à ce moment-là. Merci pour ta bonne humeur et de m'avoir enseigné ce que tu savais, même après avoir quitté le labo (mais finalement est-ce qu'on peut vraiment dire que tu es partie ?). I would also like to thank dottore Francesca Mattioli. I'm pretty sure that you are fluent in French but you know, force of habits! Thank you for your help, especially when days were complicated and for being always here to trash talk these goddamn cells while eating chocolate. Un grand merci aux "non Piton" (tout le monde n'a pas cette chance, je sais que vous êtes toujours jaloux à ce jour), les susnommés docteurs Johan Gilet et Karima Habbas. Merci à

vous deux pour tous ces moments qu'ils aient été scientifiques ou non, je vous avoue que les déconnades me manquent. Merci, mille mercis à Gabrielle Rudolf pour avoir été là. Gabi je te suis très reconnaissant pour ton soutien, l'aide physique mais surtout psychologique que tu m'as offert, et bien sûr pour la relecture de ce manuscrit. Un très grand merci à Benjamin Durand, pour son aide plus que précieuse et son expertise. J'ai passé d'excellents moments, j'espère que l'occasion de travailler ensemble se représentera un jour. Je tiens également à remercier Imène et Loréline qui ont participé à l'un de ces projets. Je remercie également Éric Flatter, merci de m'avoir apporté ton expertise de la biomol, ces (longues) soirées n'auront pas été vaines. Pour finir dans les formers je tiens à remercier le professeur Jamel Chelly, pour ses remarques concernant mon projet mais également Vicky et Florent pour leur bonne humeur et les discussions que nous avons pu avoir et enfin Timothée, pour m'avoir accompagné dans cette sombre pièce pendant des mois, heureusement que tu étais là sinon le temps aurait été très long !

Après les formers, merci aux membres qui sont présents, à commencer par Nathalie qui m'a énormément aidé pour la réalisation de mes expériences. Merci à Hervé, pour tes remarques toujours pertinentes et pour m'avoir apporté ton aide à chaque fois que j'ai pu en avoir besoin. Merci à Oktay, Clarisse et Camille pour perpétrer cette bonne ambiance et pour l'aide que vous m'avez apportée, qu'elle soit physique ou psychologique. Thanks to Bloglarka, for the discussions and the relevant observations that actually helped me a lot to clarify some points. Nyilván használok a google fordítót, sajnos nem tudtam megtanulni magyarul. De tudom, hogy mindig kérhetek mosogatót (Én is megsérthetek valakit :D).

Je tiens également à remercier les membres des équipes de l'IGBMC (Godin, Hérault, Keyes) avec qui j'ai pu collaborer mais également passer de très bons moments. I would also like to thank Rosanna and Eric for this very nice collaboration, I actually learned a lot! Un merci particulier à Sirine, qui m'a apporté son soutien moral et son aide pour l'impression de ce manuscrit.

Je remercie également les plateformes de l'IGBMC (Karim en biologie moléculaire, Erwann et Elvire en microscopie, Anne au cell phenotyping et en particulier tout le service de culture cellulaire) qui ont pu m'aider pour la réalisation de ces projets. Merci à vous pour votre travail de qualité et pour m'avoir fait gagner un temps précieux. Mention spéciale pour Amélie Freismuth qui m'a même permis de ne plus congeler et décongeler mes NSCs, c'était super !

Je remercie la fondation Jérôme Lejeune pour le prix qu'elle m'a décerné, valorisant ainsi mon travail.

Enfin la vie au laboratoire est certes prenante mais il ne faut pas négliger tout ce qu'il peut y avoir autour. Je tiens donc à remercier mes amis pour avoir été présents et de ce fait soutenu même s'ils ne s'en sont pas forcément tous rendu compte. Et comme je vous connais j'ai utilisé un générateur d'ordre aléatoire, merci à Antoine, Leslie, Simon, Kévin, Christophe, Samuel, Kévin, Anthony, Amaury, Pol, Jonathan et Mélina.

Enfin, je tiens à remercier du fond du cœur mes parents et ma sœur sans ainsi que ma tante, ma cousine, mes grands-parents et bien sûr toute ma famille. Merci de m'avoir soutenu jusqu'ici, pour votre confiance sans faille depuis le début, pour tout ce que vous m'apportez, je sais que j'ai énormément de chance de vous avoir et je ne serai jamais arrivé jusqu'ici sans vous.

Pour finir, je remercie Marie, non seulement parce tu as pris, la peine, de relire, ce manuscrit, mais aussi parce que tu me supportes lorsque tu me poses naïvement une question et que tu dois par la suite subir mes explications sans fin. Mais surtout je veux te remercier parce que quelle que soit la journée, bonne ou mauvaise, tu es toujours là pour me soutenir et m'encourager.

# RÉSUMÉ

La déficience intellectuelle est un trouble du neurodéveloppement affectant 1 à 2 % de la population et la première raison de consultation génétique. Les causes génétiques de la DI sont variées, elles peuvent concerner des anomalies de nombre ou de structure des chromosomes, des variations du nombre de copies ou des anomalies ponctuelles dans un seul gène conduisant à une perte partielle ou totale de la fonction de la protéine (Ex : syndrome *DYRK1A*, syndrome de Renpenning). A l'heure actuelle, environ 1500 gènes ont pu être associés à des formes monogéniques de DI. Ces dernières années l'avènement du séquençage d'exome à visée diagnostique a permis d'améliorer de façon très importante l'identification de variants pathogènes chez les individus avec DI et d'augmenter le taux de diagnostic moléculaire. L'obtention d'un diagnostic étiologique précis et précoce est primordial pour les familles afin d'être en mesure de proposer une prise en charge adaptée de la DI chez le jeune enfant et éventuellement une approche thérapeutique s'il en existe. Cependant, deux problématiques majeures subsistent. La première concerne l'interprétation des nombreux variants identifiés. En effet, pour un certain nombre de variants, le faisceau d'arguments n'est pas suffisant pour conclure. Le variant est classé en tant que Variant de Signification Inconnue (VSI) et les familles concernées restent en errance diagnostique tant que ces variants ne sont pas reclassés. Pour améliorer la prise en charge des jeunes patients et envisager des approches thérapeutiques, il est donc primordial de mieux caractériser les conséquences de ces mutations lors du développement du cerveau. Les résultats obtenus ont permis de reclasser la plupart des variants identifiés dans ces gènes et de rendre un diagnostic aux familles concernées. Nous avons ainsi développé des approches robustes pour tester l'effet de n'importe quel variant qui sera identifié dans le futur dans ces gènes. La seconde difficulté découle du nombre de gènes impliqués, car chaque forme monogénique de DI représente une maladie ultra-rare, dont pour la plupart les mécanismes physiopathologiques ainsi que les voies cellulaires impliquées sont généralement mal connus. J'ai donc en parallèle investigué les conséquences moléculaires et cellulaires de l'inactivation de *DYRK1A* et *PQBPI* dans un modèle de progéniteurs neuronaux humains. J'ai ainsi pu identifier différents interacteurs de ces protéines ainsi que les gènes différenciellement exprimés suite à leur inactivation. La description de ces processus est un élément clef dans la compréhension des mécanismes physiopathologiques de ces troubles neurodéveloppementaux.

# SOMMAIRE

REMERCIEMENTS .....	5
RÉSUMÉ .....	8
SOMMAIRE .....	9
LISTE DES ABBRÉVIATIONS .....	12
INDEX DES FIGURES .....	14
INDEX DES TABLES .....	15
INDEX DES ANNEXES.....	16
<b>INTRODUCTION</b> .....	17
<b>I. Génétique de la Déficience intellectuelle et des maladies neurodéveloppementales</b> .....	19
1) Définition des troubles neurodéveloppementaux et de la déficience Intellectuelle.....	19
2) Etiologies .....	20
1. Causes environnementales .....	21
2. Causes génétiques.....	22
a. Anomalies chromosomiques .....	22
b. Anomalies du nombre de copies (CNV).....	22
c. Formes monogéniques.....	23
3) Identification des gènes impliqués dans la DI .....	24
1. La DI liée à l’X.....	24
2. Formes autosomiques dominantes.....	25
3. Formes autosomiques récessives.....	26
4) Diagnostic moléculaire de la DI et identification des variants .....	27
5) Interprétation des variants identifiés.....	27
1. Filtration des variants.....	28
2. Logiciels de prédiction.....	28
3. Classification ACMG.....	29
6) Variants de Signification Inconnue.....	30
7) Modèles d’études des VSI .....	31
1. Tests biochimiques.....	31
2. Analyses de cellules de patients.....	31
3. Analyse <i>in vitro</i> des protéines mutantes.....	32

4. Analyses des conséquences d'une altération de la fonction d'une protéine, exemple de la méthylation.....	32
8) Modèles d'étude des gènes de DI.....	33
1. Modèles animaux.....	33
2. Cellules souches neuronales humaines (hNSCs).....	34
<b>II. Une forme monogénique fréquente de DI, le syndrome DYRK1A.....</b>	<b>38</b>
1) De minibrain à <i>DYRK1A</i> .....	38
2) Découverte du syndrome DYRK1A.....	39
3) Caractéristiques du syndrome DYRK1A.....	40
4) Aspects moléculaires du syndrome DYRK1A.....	42
5) Gène et protéine DYRK1A.....	43
1. Généralités, orthologues et paralogues.....	43
2. Expression du gène.....	45
3. Expression de la protéine.....	49
4. Structure de la protéine et rôles des différents domaines.....	49
5. Régulation transcriptionnelle de <i>DYRK1A</i> .....	52
6. Activation de DYRK1A.....	53
7. Régulation d'activation de la protéine DYRK1A.....	54
6) Rôles de DYRK1A.....	55
1. Analyses omics.....	55
a. Chip-Seq et RNA-seq.....	57
b. Protéomique et phospho-protéomes.....	57
2. Régulation de la prolifération.....	59
a. Régulation du cycle cellulaire.....	59
b. Régulation de l'apoptose.....	60
c. Synthèse.....	61
3. Régulation de l'épissage alternatif.....	63
4. Régulation de la transcription.....	64
a. Remodelage de la chromatine.....	64
b. Marques épigénétiques.....	64
c. Facteurs de transcription.....	65
d. DYRK1A et ARN polymérase II.....	66
5. Autres rôles de DYRK1A.....	66
7) Modèles animaux de DYRK1A.....	67

1.	Drosophila melanogaster .....	67
2.	Mus Musculus.....	68
a.	Dans le cadre d'une surexpression de Dyrk1a .....	68
b.	Dans le cadre diminution d'expression de Dyrk1a.....	68
3.	Autres modèles : poulet, poisson zèbre et xénope.....	71
III.	<b>La DI liée à l'X causée par des mutations du gène PQBP1 .....</b>	<b>73</b>
1)	Historique et clinique .....	73
2)	Expression et domaines de la protéine PQBP1 .....	74
3)	PQBP1 et régulation de la transcription.....	77
4)	Rôle de PQBP1 dans la régulation de l'épissage .....	78
5)	PQBP1 et troubles neurodégénératifs .....	79
6)	Autres implications de PQBP1.....	80
	<b>OBJECTIFS DE LA THESE .....</b>	<b>81</b>
	<b>RESULTATS .....</b>	<b>85</b>
	<b>Article 1 : Integrative approach to interpret <i>DYRK1A</i> variants, leading to a frequent neurodevelopmental disorder .....</b>	<b>87</b>
	Discussion de l'article 1 .....	153
	<b>Article 2 : Role of <i>DYRK1A</i> in neural progenitor cells.....</b>	<b>159</b>
	Discussion de l'article 2 .....	211
	<b>Article 3 : Molecular consequences of variants in <i>PQBP1</i>, involved in the .....</b>	<b>219</b>
	X-linked Renpenning syndrome .....	219
	<b>DISCUSSION GÉNÉRALE .....</b>	<b>277</b>
	<b>ANNEXES .....</b>	<b>289</b>
	Annexe 1 : Novel mutations in <i>NLGN3</i> causing autism spectrum disorder and cognitive impairment .....	291
	Annexe 2 : Sex-specific impact of prenatal androgens on social brain default mode subsystems.....	297
	Annexe 3 : Conditional switching of <i>KIF2A</i> mutation provides new insights into cortical malformation pathogeny.....	298
	Annexe 4 : <i>De Novo</i> Frameshift Variants in the Neuronal Splicing Factor <i>NOVA2</i> Result in a Common C-Terminal Extension and Cause a Severe Form of Neurodevelopmental Disorder.....	299
	Annexe 5 : Biallelic <i>PDE2A</i> variants: a new cause of syndromic paroxysmal dyskinesia .....	300
	<b>RÉFÉRENCES .....</b>	<b>301</b>



# LISTE DES ABBREVIATIONS

aa : Acide Aminé

ACMG : American College of Medical Genetic and Geomics

ADHD : Attention Deficit Hyperactivity Disorders

ADN : Acide désoxyribo nucléique

AESH : Accompagnant d'Élève en Situation de Handicap

ARN : Acide Ribo Nucléique

ARNm : ARN messenger

ATP : Adénosine Tri Phosphate

Chip-Seq : Immunoprécipitation de la chromatine couplée à un séquençage

CNV : Variation du Nombre de Copies

CTD ou Cter : domaine Cterminal

DECIPHER : Database of Genomic variation and Phenotype in Human using Ensembl  
Ressources

DH : DYRK Homeobox

DI : Déficience Intellectuelle

DSCR : Down Syndrome Critical Region

DSM-V : Diagnostic and Statistical Manual of mental disorders, V<sup>th</sup> edition

DYRK1A : Dual specificity tyrosine-phosphorylation-regulated kinase 1A

ExAC : Exome Aggregation Consortium

FISH : Fluorescence In Situ Hybridization

FXS : Syndrome de l'X Fragile

GoF : Gain de Fonction

hESCs : Cellules Souches Embryonnaires humaines

hNSCs : Cellules Souches Neurales humaines

hNSCs-1 : Lignée hNSC SA001

hNSCs-2 : Lignée hNSC GMO1869  
kDa : Kilo Dalton  
IME : Institut Médico-Éducatif  
IPSCs : Cellules Souches Pluripotentes Induites  
IRM : Image par Résonance Magnétique  
ISH : In Situ Hybridization  
MAPT : Protéine de liaison aux microtubules (TAU)  
Mb : Méga base  
miR : micro ARN  
mnb : mini brain  
MO : Morpholino  
MRD7 : Retard Mental Dominant de type 7  
NAPA : N-terminal AutoPhosphorylation Accessory  
NGS : Séquençage de nouvelle génération  
NLS : Séquence de Localisation Nucléaire  
Nter : Domaine Nterminal  
PEST : domaine riche en acides aminés proline, acide glutamique, sérine et arginine  
PNDS : Protocole National de Diagnostic et de Soins  
PQBP1 : Poly Glutamine Binding Protein 1  
QI : Quotient Intellectuel  
RBP : Protéine de liaison à l'ARN  
RNAseq : séquençage de l'ARN  
SR : Syndrome de Renpenning  
RT-qPCR : Réverse Transcription- Réaction en Chaîne de Polymérase quantitative  
SD : Syndrome de Down  
shRNA : small hairpin RNA  
siRNA : small interfering RNA  
SNV : Variantion d'un seul nucleotide  
TND : Troubles neurodéveloppementaux  
TSA : Troubles du Spectre Autistique  
UTR : Région non traduite  
XLID : Déficience Intellectuelle Liée à l'X  
YAC : Chromosome artificiel de levure

# INDEX DES FIGURES

<b>Figure 1</b> : Distribution du score de QI dans la population générale.....	20
<b>Figure 2</b> : Répartition des étiologies de la DI. ....	21
<b>Figure 3</b> : Identification de gènes impliqués dans des formes monogéniques de DI en fonction de leur mode de transmission.....	24
<b>Figure 4</b> : Critères de pathogénicité des variants ACMG. ....	30
<b>Figure 5</b> : Provenance et différenciation des hNSCs. ....	36
<b>Figure 6</b> : Photographie d'un enfant porteur d'une variation perte de fonction dans <i>DYRK1A</i> . ....	41
<b>Figure 7</b> : Alignement des séquences protéiques des différentes DYRKs humaines et de quelques orthologues de <i>DYRK1A</i> . ....	44
<b>Figure 8</b> : Phylogénie des kinases du groupe CMGC .....	45
<b>Figure 9</b> : Pattern d'expression de <i>Dyrk1a</i> dans les progéniteurs neuronaux en différenciation chez la souris entre E12 et E15.....	46
<b>Figure 10</b> : Expression de <i>DYRK1A</i> dans différents tissus issus de prélèvement post-mortem chez l'Homme.....	48
<b>Figure 11</b> : Expression de l'ARN messager des protéines DYRKs humaines basée sur les données de la base de données GTEx. ....	49
<b>Figure 12</b> : Structures primaires des protéines DYRKs humaines.....	50
<b>Figure 13</b> : Représentation de la structure de <i>DYRK1A</i> en présence d'inhibiteurs. ....	52
<b>Figure 14</b> : Analyses d'enrichissement GOterms des principaux interacteurs de <i>DYRK1A</i> . ....	58
<b>Figure 15</b> : Représentation schématique de la protéine PQBP1 et de ses domaines.....	75

# INDEX DES TABLES

<b>Table 1</b> : Résumé des principales analyses OMICS impliquant DYRK1A .....	56
<b>Table 2</b> : Conséquences des variations d'expression de <i>DYRK1A</i> sur la prolifération cellulaire dans différents modèles d'études.. .....	62
<b>Tableau 3</b> : Récapitulatif des principaux modèles murins présentant une diminution d'expression de <i>Dyrk1a</i> .....	70
<b>Table 4</b> : Principaux modèles d'études de <i>PQBPI</i> .....	76

# INDEX DES ANNEXES

<b>Annexe 1</b> : Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment .....	291
<b>Annexe 2</b> : Sex-specific impact of prenatal androgens on social brain default mode subsystems.....	297
<b>Annexe 3</b> : Conditional switching of KIF2A mutation provides newinsights into cortical malformation pathogeny .....	298
<b>Annexe 4</b> : De Novo Frameshift Variants in the Neuronal Splicing Factor NOVA2 Result in a Common C-Terminal Extension and Cause a Severe Form of Neurodevelopmental Disorder.....	299
<b>Annexe 5</b> : Biallelic PDE2A variants: a new cause of syndromic paroxysmal dyskinesia...	300

# INTRODUCTION

---



# I. Génétique de la Déficience intellectuelle et des maladies neurodéveloppementales

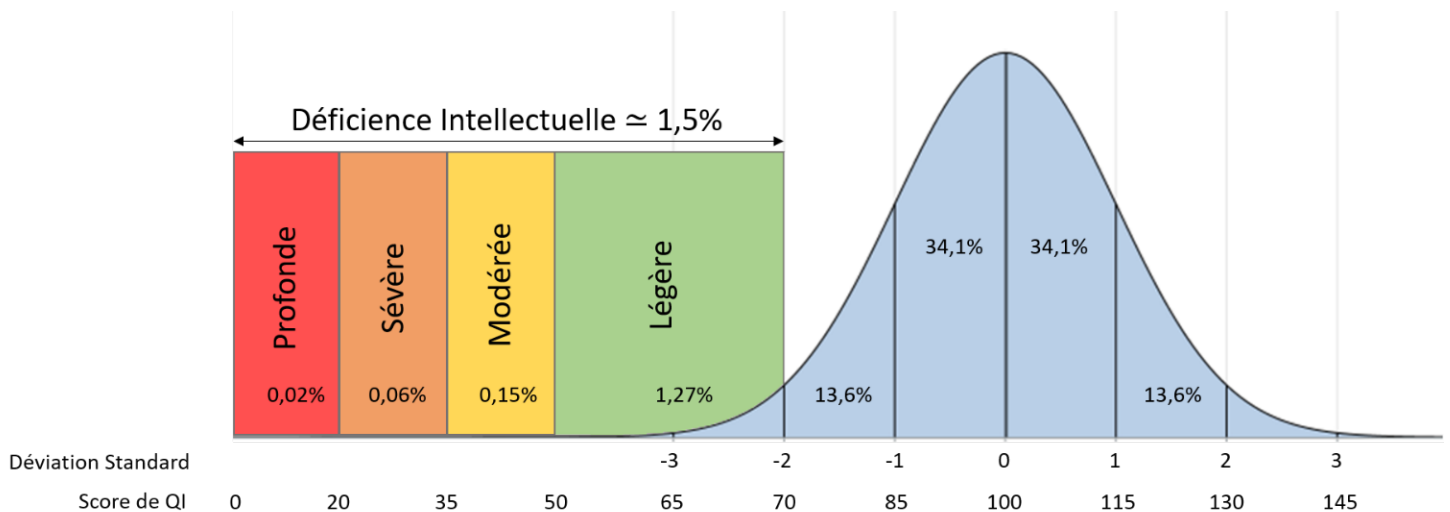
## 1) Définition des troubles neurodéveloppementaux et de la déficience Intellectuelle

Les troubles neurodéveloppementaux (TND) sont la manifestation clinique de conséquences d'anomalies survenues pendant le développement du cerveau. Au sein des TND on retrouve plusieurs troubles tels que : la Déficience Intellectuelle (DI) (appelée « retard mental » jusqu'en 2016 (Chiurazzi et al., 2016)), les troubles du Spectre Autistique (TSA), les Troubles du Déficit de l'Attention avec ou sans Hyperactivité (TDAH), la Schizophrénie ainsi que les troubles bipolaires (Diagnostic and Statistical Manual of Mental Disorders – 5ème édition 2013). La DI est la première raison de consultation génétique, on estime sa fréquence en France entre 1 et 2 % de la population. On remarquera une prévalence plus élevée chez les garçons que chez les filles avec un sexe ratio entre 1,3 et 1,9 selon les études (Leonard et al., 2002 ; Maulik et al., 2011 ; Buntinx et al., 2016). Plusieurs définitions provenant de différents organismes notamment l'American psychiatric association, l'American Association on Intellectual and Developmental Disabilities et l'Organisation Mondiale de la Santé sont proposées.

De manière globale, la DI se manifeste avant l'âge de 18 ans, généralement dès la petite enfance et est caractérisée par des limitations des capacités intellectuelles avec un QI (Quotient Intellectuel) inférieur à 70 ainsi qu'une limitation des comportements adaptatifs. L'évaluation des capacités intellectuelles peut être réalisée par le biais d'un QI standardisé et l'échelle de Wechsler (Burns et al., 2015) est le test le plus couramment utilisé. Cette échelle comprend 4 indexes évaluant : la compréhension verbale, le raisonnement perceptif, la mémoire de travail et la vitesse de traitement de l'information. On attribue au QI moyen de la population générale une valeur de 100 et le seuil limite de 70 correspond à une distance de deux écarts type ou plus du QI moyen. La valeur de QI permet également d'évaluer le degré de sévérité de DI qui est : légère dans 85% des cas (QI entre 50 et 70) ; modérée dans 10% des cas (QI entre 35 et 50) ; sévère (QI entre 20 et 35) dans 4% des cas ou profonde dans moins de 2% des cas (QI inférieur à 20) (Figure 1).



Au-delà de l'évaluation du QI, une évaluation des comportements adaptatifs est nécessaire pour préciser la DI. Cette évaluation est testée le plus souvent par l'échelle de la Vineland (Vineland Adaptive Behavior Scales) (Sparrow et al.,2005) qui permet de mesurer le fonctionnement adaptatif dans 4 différents domaines : les capacités de communication, les capacités de socialisation, les aptitudes du patient dans la vie de tous les jours et enfin la motricité. Malgré la standardisation de l'évaluation de la DI, ces tests sont peu réalisés en raison du temps requis à leur mise en place mais également de la nécessité de personnel accrédité (neuropsychologue).



**Figure : 1 Distribution du score de QI dans la population générale**

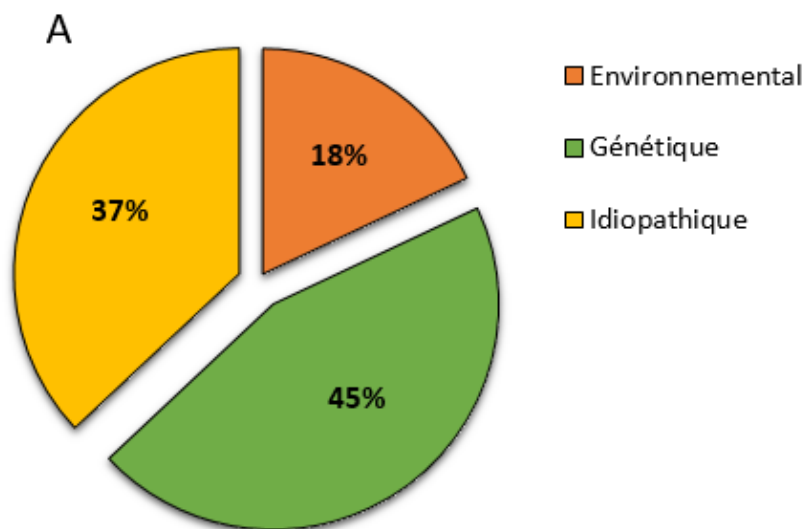
## 2) Etiologies

Les étiologies de la DI sont hétérogènes et variées mais restent d'origine génétique ou environnementale. En 2016, l'INSERM a estimé à 45% la fréquence des DI génétiques diagnostiquées, 18% la fréquence des DI liées à l'environnement incluant causes pré-natales et embryo-fœtopathies et à 37% la fréquence des DI idiopathiques (Expertise collective « Déficiences Intellectuelles » INSERM 2016). (Figure 2)

## 1. Causes environnementales

La prévalence des DI liées à des causes environnementales reste aujourd'hui difficile à estimer. En effet, les facteurs environnementaux peuvent avoir des pénétrances variables, être difficiles à associer, ou être simplement non décrits à l'heure actuelle. Parmi ces facteurs on distingue 3 grands groupes :

- Les causes infectieuses : certains virus comme le CMV (cytomégalovirus) ou parasites comme l'agent de la toxoplasmose (*Toxoplasma gondii*) peuvent avoir un effet cytopathogène durant l'embryogénèse et de ce fait entraver le bon développement du cerveau.
- Les causes toxiques : elles sont reliées à l'exposition à un agent chimique toxique durant la période pré ou post-natale et de très nombreuses substances ont aujourd'hui été décrites (tabac, arsenic, mercure, amphétamines, médicaments, ...). La cause toxique la plus fréquente reste l'alcool qui altère le développement de tous les organes (notamment le cerveau) et entraîne le syndrome d'alcoolisation fœtal.
- Les traumatismes : des chocs physiques directs (traumatismes crâniens, complications pendant la naissance, ...) ou physiologiques (hypoxie, anévrismes, malnutrition, ...) peuvent entraîner des lésions directes du cerveau.



**Figure 2 : Répartition des étiologies de la DI, d'après les chiffres de « Expertise collective Déficiences Intellectuelles » (Buntinx et al., 2016).**

## 2. Causes génétiques

Les causes génétiques de DI sont hétérogènes et peuvent aller de larges évènements chromosomiques à de simples substitutions. Presque tous les modes de transmission peuvent être observés : autosomal dominant (généralement *de novo*), autosomal récessif et lié à l’X.

### a. Anomalies chromosomiques

Les anomalies chromosomiques sont les causes les plus fréquentes et représentent plus de 20% des DI génétiques (Buntinx et al., 2016). Parmi les anomalies chromosomiques on retrouve les aneuploïdies (addition ou perte totale d’un ou plusieurs chromosomes) mais également les évènements de grandes anomalies structurales telles que les translocations, duplications délétions ou inversions de taille supérieure à 5Mb. Les anomalies chromosomiques furent les premières causes génétiques associées à la DI avec notamment le syndrome de Down (trisomie 21), mis en évidence grâce à des analyses de caryotypes par Marthe Gauthier et Jérôme Lejeune (Lejeune et al., 1959). Le SD reste aujourd’hui la cause de DI génétique la plus fréquente (Ropers et al., 2010) avec une fréquence avoisinant 1 sur 800 naissances. Les anomalies structurales sont plus rares et le plus souvent *de novo*.

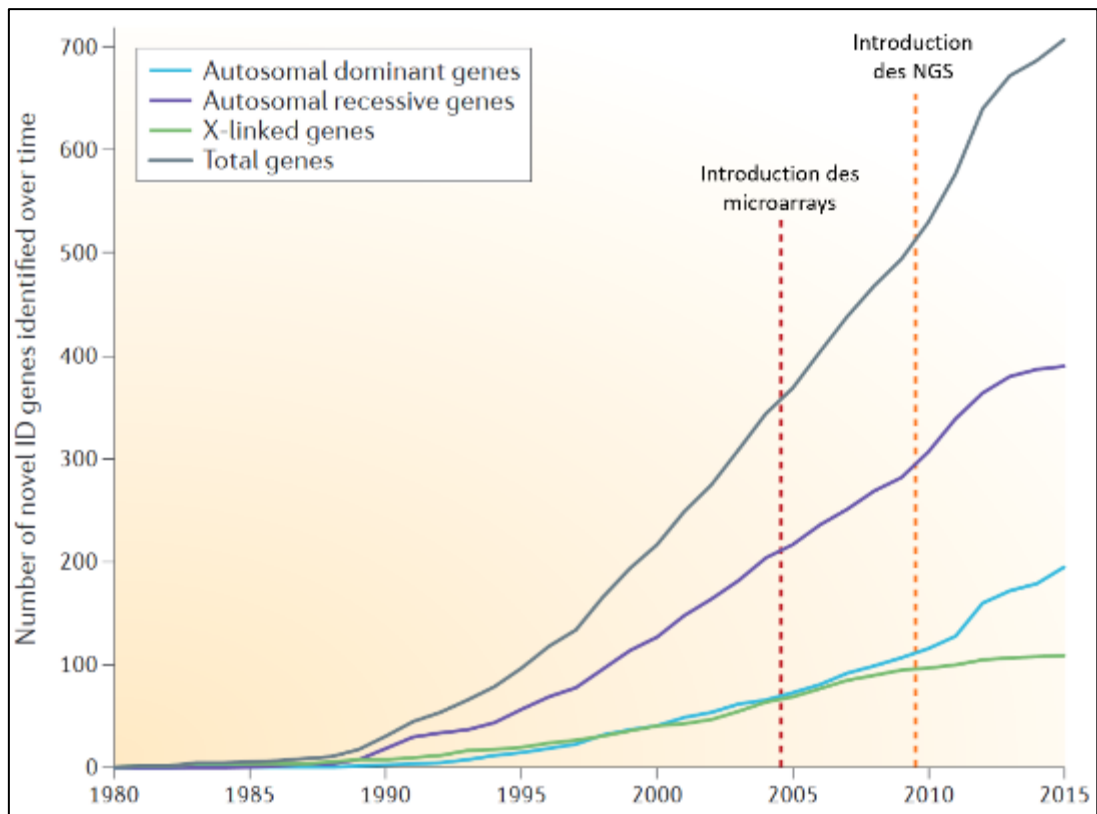
### b. Anomalies du nombre de copies (CNV)

Les CNV (Copy Number Variants) regroupent des anomalies structurales de tailles moyennes (quelques centaines de kb à 5Mb) comprenant délétions, insertions et duplications. Leur mise en évidence se fait la plupart du temps par Analyse Chromosomique par Puce à ADN de type « CGH » ou « SNP array » (ACPA). Les CNV représentent 10 à 15% des causes de DI génétiques (données du réseau français pour le diagnostic de la DI). Certains CNV sont récurrents, présentant des points de cassure similaires et entraînant la duplication ou délétion d’un même groupe de gènes. L’hypothèse la plus probable pouvant expliquer la récurrence de certains CNV concerne la présence de régions de forte homologie bordant le CNV, entraînant de ce fait des recombinaisons illégitimes (Pober et al., 2010). Par exemple, la délétion 7q11.13 conduit à une hétérozygotie pour environ 27 gènes dont celui de l’élastine (*ELN*) et est associée au syndrome de Williams Beuren, un syndrome rare associant DI et cardiopathie congénitale ainsi qu’un faciès typique et un comportement hypersocial caractéristique. L’élastine est une protéine fibreuse sécrétée dans la matrice extracellulaire et sa présence à l’état hétérozygote dans le syndrome de William-Beuren est associée entre autres aux atteintes cardiovasculaires. D’autres gènes présents dans la délétion sont suspectés de participer aux phénotypes cognitifs

comme les gènes *GTF21*, *LIMK1*, *CYLN2* et *GTD2IRD1*. Certains CNV peuvent affecter plusieurs gènes parmi lesquels un seul peut conduire à une DI quand il est muté, comme c'est le cas pour 3 individus rapportés dans ce travail présentant une délétion de plusieurs gènes sur le chromosome 21, dont *DYRK1A* (ind#1, ind#2 et ind #40 ; Courraud et al., 2021). Dans ce cas, le spectre phénotypique est en grande partie expliqué par la perte d'une des deux copies du gène *DYRK1A* mais il reste difficile de conclure quant aux conséquences issues de la perte des gènes présents dans ces régions comme *KCNJ2*.

### *c. Formes monogéniques*

Contrairement aux anomalies chromosomiques et aux CNV, les formes monogéniques concernent des variations affectant un seul gène, qu'elles soient de taille moyenne (plusieurs dizaines de kb voire Mb), qu'il s'agisse de mutations ponctuelles (SNV : Single Nucleotide Variant) qui ne concernent qu'un unique nucléotide, ou de petites insertions ou délétions de plusieurs nucléotides (indels). Les formes monogéniques peuvent être classées selon leur mode de transmission : dominant, récessif ou lié à l'X. Les premières formes monogéniques de DI furent identifiées dans de grandes familles comportant plusieurs individus atteints et concernent des syndromes facilement reconnaissables. L'avènement des NGS (Séquençage de Nouvelle Génération) a permis une accélération de l'identification des gènes impliqués dans des formes monogéniques de DI quel que soit leur mode de transmission (Figure 3). A l'heure actuelle, 1500 gènes ont pu être associés à une forme monogénique de DI (base de données SysID, Kochinke et al., 2016).



**Figure 3 : Identification de gènes impliqués dans des formes monogéniques de DI en fonction de leur mode de transmission (Visser et al., 2015).**

### 3) Identification des gènes impliqués dans la DI

#### 1. La DI liée à l’X

Le chromosome X comporte plus de 900 gènes codants et représente environ 5% de la totalité du génome humain mais aujourd’hui 127 gènes ont pu être validés comme associés à de la DI (Kochinke et al., 2016) représentant ainsi 10% des gènes totaux de DI. De plus, 55 gènes supplémentaires situés sur le chromosome X sont aujourd’hui considérés comme gènes candidats c’est-à-dire, potentiellement impliqués. Cet enrichissement fait partie des hypothèses expliquant en partie le biais de sexe observé pour la DI, avec un excès de près de 40% d’hommes rapporté. Ce mode de transmission fut mis en évidence dans les années 1940 grâce à la description de grandes familles avec un grand nombre d’hommes atteints. Par exemple, les premières descriptions de patients atteints du syndrome de l’X fragile (FXS) furent décrites en 1943 par Martin et Bell qui rapportent une dysmorphie faciale associée à un « retard mental à transmission familiale ». Ce n’est qu’en 1991 que le mécanisme d’inactivation de *FMRI* par

expansion de triplets CGG dans sa région promotrice conduisant à son inactivation transcriptionnelle sera mis en évidence (Oberlé et al., 1991). Le syndrome FXS est la forme la plus fréquente de DI héritée et sa fréquence est estimée à 1 sur 5000 pour les garçons et 1 sur 8000 pour les filles. D'autres gènes ont pu être identifiés par analyse de liaison dans de grandes familles avec seulement des garçons atteints, tels que le gène *RPS6KA3* (Trivier et al., 1996) responsable du syndrome de Coffin-Lowry, ou *PQBPI* responsable du syndrome de Renpenning (Kalscheueur et al., 2003). Le phénotype présenté par les femmes porteuses hétérozygotes de variations pathogènes dans un gène du X peut varier de l'absence de phénotype à la présence d'une DI (Exemple : syndrome de l'X fragile) en passant par des phénotypes plus légers comme des troubles d'apprentissage. Pour certains gènes, on retrouvera chez les mères porteuses un biais d'inactivation du X en faveur du X non porteur de la variation, comme par exemple pour le gène *UPF3B* (Tarpey et al., 2007) alors que pour d'autres il n'existe pas de corrélation entre les phénotypes des femmes porteuses et le profil d'inactivation du X dans le sang.

Enfin, il existe également des DI liées à l'X à caractère dominant comme le syndrome de Rett, syndrome touchant majoritairement les filles. Le syndrome de Rett est causé par des variations perte de fonction le plus souvent *de novo* dans le gène *MECP2*, un gène codant pour une protéine de méthylation de l'ADN jouant un rôle dans la régulation transcriptionnelle (Lyst and Bird, 2015). Un autre exemple de forme féminine de DI liée à l'X impliquant des variations *de novo* est le gène *DDX3X*, récemment identifié lors d'analyse d'exome entier en trio (Snijders Blok et al., 2015), et maintenant connu pour être la forme la plus fréquente de DI chez les filles. Enfin, on sait par exemple que des variations perte de fonction dans le gène *PDCH19* codant pour une protocadhérine, conduisent à une forme d'encéphalopathie épileptique ne touchant que les femmes (Dibbens et al., 2008). De manière intéressante, les hommes porteurs de variations perte de fonction dans ce gène restent asymptomatiques mais un homme porteur d'une variation mosaïque fut finalement identifié (Depienne et al., 2009). Ces observations ont conduit à la caractérisation du mécanisme physiopathologique « d'interférence cellulaire » associé à ces variations, dans lequel l'interaction d'une protéine mutante avec une protéine sauvage conduit à un défaut du tissu et donc, à la pathologie.

## 2. Formes autosomiques dominantes

Les gènes de DI à transmission autosomique dominante représentent environ 35% des gènes de DI totaux (Kochinke et al., 2016) et contribuent beaucoup plus largement à la DI dans les

populations non consanguines (entre 70 et 80% ; données du laboratoire diagnostic génétique de Strasbourg). Dans le cadre de la DI, les variations à l'origine de formes autosomiques dominantes sont le plus souvent *de novo* en raison de la faible probabilité d'une transmission d'un parent atteint à son enfant, surtout dans le cas des formes de DI sévères. Les études de liaison génétique n'étant pas possibles pour identifier les gènes impliqués dans ces formes sporadiques de DI, la plupart des gènes identifiés avant le développement du séquençage haut-débit avaient été retrouvés dans des délétions ou des points de cassures de translocation. C'est le cas par exemple du gène *TCF4*, pour lequel des délétions avaient été rapportées chez des patients atteints du syndrome de Pitt-Hopkins (Amiel et al., 2007, Zweier et al., 2007), et pour lequel un séquençage sanger des régions codantes chez d'autres patients avec phénotype similaire a permis d'identifier plusieurs mutations ponctuelles *de novo*, confirmant définitivement l'implication du gène *TCF4* dans le syndrome de Pitt-Hopkins (Amiel et al., 2007).

Les mutations du gène *DYRK1A* ont été identifiées de la même manière (Courcet et al., 2012). Par la suite, des études de séquençage Sanger de gènes candidats à grande échelle ont permis d'identifier d'autres gènes comme le gène *SYNGAP1* (Hamdan et al., 2011). En 2010, l'utilisation du séquençage d'exome en trio a permis de souligner l'importance de la part des formes autosomiques dominantes avec variations *de novo* dans les cas isolés de DI et d'identifier plusieurs nouveaux gènes candidats (Vissers et al., 2010). Des études de séquençage d'exomes en trio ont par la suite été réalisées sur des cohortes plus larges et ont conduit à l'identification d'un grand nombre de gènes (De Ligt et al., 2012 ; Rauch et al., 2012 ; Chiurazzi et al., 2016).

### 3. Formes autosomiques récessives

Les DI génétiques à transmission autosomique récessive représentent une part plus minoritaire des DI génétiques, surtout dans des populations non-consanguines. La majeure partie de ces DI concernent des troubles métaboliques causés par l'absence d'une enzyme à l'état fonctionnel et on observera généralement des atteintes multi-systémiques. La phénylcétonurie, causée par des variations dans le gène *PAH* (Woo et al., 1983) qui code pour une phénylalanine hydroxylase, reste à ce jour la plus connue, dépistée de façon systématique à la naissance. On peut souligner que comme pour toutes les maladies suivant ce mode de transmission, les formes autosomiques récessives de DI présentent une prévalence plus importante dans des sous populations consanguines. Ainsi, si l'on reprend l'exemple de la phénylcétonurie, sa fréquence est d'environ 1 sur 10 000 naissances en Europe contre 1 sur

4000 dans les populations turques (Tadmouri et al., 2009). D'autres gènes impliqués dans des formes non métaboliques de DI ont été mis en évidence principalement dans des familles consanguines, par cartographie d'homozygotie ou séquençage d'exome (Najmabadi et al., 2011 ; Hu et al., 2019) comme le gène *THOC6* (Beaulieu et al., 2013) ou le gène *MED23* (Hashimoto et al., 2011).

#### 4) Diagnostic moléculaire de la DI et identification des variants

Avant le déploiement du NGS dans le cadre du diagnostic de la déficience intellectuelle, la recherche des causes génétiques constituait principalement à rechercher un syndrome de l'X fragile, une anomalie chromosomique (au départ par une analyse de caryotype, remplacé par la suite par une ACPA) ou un CNV (ACPA). En cas d'orientation diagnostique fondée sur des critères cliniques, une analyse de la région chromosomique pouvait être réalisée par FISH (Fluorescence In Situ Hybridization), ou un séquençage Sanger des régions codantes du gène suspecté d'être impliqué.

Aujourd'hui, la technologie de séquençage haut débit est largement utilisée pour le diagnostic des DI génétiques et a permis de faire fortement progresser le taux de diagnostic génétique. Le TS (Séquençage ciblé) permet d'investiguer un panel de gènes de DI pour un rendement diagnostique d'environ 25% (Redin et al., 2014). Le WES (séquençage d'exome en trio) conduit lui à un taux diagnostique d'environ 30% et est actuellement la méthode la plus utilisée par les laboratoires de diagnostic à travers l'Europe (De Ligt et al., 2012 ; Rauch et al., 2012). Enfin, le WGS (séquençage de génome complet) est indéniablement l'approche la plus complète mais est resté jusqu'à récemment peu utilisé en diagnostic clinique en raison de son coût, du volume de données à stocker mais aussi des difficultés à interpréter les variants dans les régions non codantes. Le WGS conduit à l'identification de variants causaux dans environ 40% des cas (Glissen et al., 2014). Le WGS est maintenant déployé en France pour la DI dans le cadre du Plan France Médecine Génomique (projet pilote DEFIDIAG, préindication qui s'ouvre sur les plateformes).

#### 5) Interprétation des variants identifiés

L'utilisation en routine des technologies de NGS dans les laboratoires de diagnostic génétique aboutissant à l'identification d'un grand nombre de variants, des étapes de filtration



puis d'analyse des variants ont été mises en place et des recommandations ont été proposées pour leur interprétation

## 1. Filtration des variants

L'identification d'un variant causal nécessite d'une part l'application de critères permettant de réduire le nombre de variations initialement identifiées par la technologie NGS et, d'autre part, de les trier en fonction de leur pertinence diagnostique. Malgré l'absence d'un consensus officiel concernant les moyens ou seuils utilisés pour le tri et l'analyse des variants, les critères utilisés par les différents laboratoires diagnostiques gravitent autour des mêmes principes. Parmi eux on comptera l'utilisation de grandes bases de données de séquençage. Ces bases permettent de faciliter la classification et l'interprétation des variations en renseignant sur la fréquence du variant d'intérêt au sein d'un échantillon de la population générale. La plus utilisée reste la base gnomAD (Karczewski et al., 2019) qui compile à l'heure actuelle environ 125 000 données d'exome et 15 000 de génomes, réalisés sur individus non apparentés et ne comportant pas de TND sévères. D'autres bases de données permettent également de référencer les anomalies génétiques retrouvées chez les patients avec dans certains cas les traits phénotypiques majeurs associés (DECIPHER, ClinVar, de novo-db). Lorsque les circonstances le permettent et en fonction du mode de transmission suspecté, la ségrégation du variant peut également être étudiée. L'étude de ségrégation peut être réalisée en amont de l'analyse NGS (analyse en trio) afin de réduire considérablement la quantité de variants traités ou à posteriori par séquençage Sanger, pour valider l'absence ou la présence du variant chez d'autres individus sains ou atteints de la même famille. Enfin, plusieurs paramètres techniques concernant la qualité du séquençage comme la profondeur et la couverture du séquençage sont un pré requis à la bonne évaluation des variants.

## 2. Logiciels de prédiction

Le développement d'outils informatiques et notamment de logiciels de prédictions est également devenu un élément incontournable dans l'interprétation des variations. Ces outils permettent de prédire dans une certaine mesure l'impact d'un variant à différents niveaux que ce soit par exemple sur l'épissage (SpliceSiteFinder) ou au niveau de la protéine (SIFT, Polyphen2, CADD). Le CADD (Combined Annotation Dependent Depletion) est à l'heure actuelle le score de prédiction le plus utilisé pour évaluer l'effet délétère des variants car il combine l'information provenant de plusieurs programmes. La génération de ce score est basée

sur un modèle de machine learning considérant 60 annotations différentes (conservation, encombrement, propriétés physico-chimiques, séquences nucléotidiques environnantes, etc) entraîné sur un échantillon de près de 9 millions de SNV (Kircher et al., 2014). Le score CADD est compris entre 0 et 35 et permet d'évaluer la probabilité d'un effet délétère du variant. On considère généralement pour la plupart des gènes : un effet délétère peu probable pour un score CADD inférieur à 20 ; un effet délétère potentiel pour un score CADD entre 20 et 25 ; un effet délétère très probable pour un score CADD >25. Cependant, les seuils utilisés dans l'interprétation des scores CADD peuvent varier selon les gènes. Ainsi, on considère qu'un score CADD supérieur à 20 concerne environ 1% des variants, un score supérieur à 25 représente environ 0,5% des variants et un score supérieur à 30 correspond à moins de 0,1% des variants. Aujourd'hui, plusieurs autres outils ont pu être développés dans différents laboratoires.

### 3. Classification ACMG

La classification de l'ACMG (American College of Medical Genetic and Genomics) a été proposée en 2015 pour codifier ces différents arguments nécessaires à l'interprétation des variants. Cette analyse est basée sur une liste d'arguments pondérés permettant de classer les variants identifiés dans 5 classes différentes : Bénin (classe 1), probablement bénin (classe 2), de signification inconnue (classe 3), probablement pathogène (classe 4) et pathogène (classe 5) (Richards et al., 2015 ; Amendola et al., 2016). Les critères utilisés pour cette classification sont regroupés dans différentes catégories dont certaines ont été évoquées précédemment (Figure 4) :

- Les données épidémiologiques : elles renseignent sur la présence ou non du variant dans la population générale
- Les données prédictives et structurales : elles renseignent sur l'impact potentiel du variant sur l'expression, la fonction et l'intégrité de la protéine
- Les données cliniques : elles prennent en compte la redondance de symptomatologie du patient avec celle d'autres patients diagnostiqués.
- La ségrégation familiale : elle renseigne sur la présence ou non du variant chez d'autres individus atteints ou sains dans la famille.
- Les données fonctionnelles : elles renseignent (lorsqu'elles existent) sur l'effet du variant dans un système *in vivo* ou *in vitro*.

Ce système de classement propose une certaine harmonisation dans l'interprétation des variants mais certains éléments restent subjectifs ou libre d'interprétation. Ainsi, un même variant pourra parfois être classé de manière différente avec une classe d'écart en fonction du laboratoire ayant réalisé l'analyse

		Benign		Pathogenic			
		Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population générale	Fréquence allélique forte			Absent des bases de données		Prévalence plus forte chez les individus affectés	
Données Prédictives et Computationnelles		Variant silencieux	Pas d'effet délétère prédis	Effet délétère prédis	Nouveau variant à la même position qu'un autre variant pathogène	Même changement d'aa qu'un variant établi pathogène	Variant non-sens dans un gène connu
Données fonctionnelles	Etudes fonctionnelles démontrant l'absence d'effet			Faux-sens dans un gène avec peu de faux-sens bénins	Hot-spot mutationnel dans un domaine fonctionnel	Etudes fonctionnelles démontrant l'effet délétère	
Ségrégation	Absence de ségrégation avec la pathologie			Co-ségrégation avec plusieurs individus affectés	Augmentation données de ségrégation		
<i>De novo</i>					<i>De novo</i> (absence de données parentales)	<i>De novo</i> (absent chez les parents)	
Données cliniques		Trouvées dans d'autres cas à cause alternative		Phénotype hautement spécifique d'un gène			

**Figure 4 : Critères de pathogénicité des variants ACMG (Adapté de Richards et al., 2015).** Le tableau organise les différents critères en fonction de leur significativité bénigne (côté gauche) ou pathogène (côté droit).

## 6) Variants de Signification Inconnue

Lors de l'analyse, si le faisceau d'argument reste trop pauvre et empêche la discrimination entre variant bénin et variant pathogène, la variation est alors décrite comme Variant de signification inconnue (VSI) ou variant de classe 3. Plusieurs paramètres peuvent en effet conduire à une classification VSI tels que : l'absence d'accès à la ségrégation du variant, une

clinique non évocatrice, une faible fréquence allélique, une incertitude ou discordance des logiciels de prédictions quant à ses conséquences au niveau de l'ARNm (stabilité, épissage, exon facultatif) ou de la protéine (site catalytique, repliement, stabilité, interaction avec les partenaires), mais également la présence du variant dans un gène encore non décrit en pathologie ou encore la pauvreté de la littérature actuelle concernant le gène en question. À l'heure actuelle une partie importante des variants faux-sens, en particulier ceux jamais rapportés, reste classée VSI. Les VSI sont devenus un enjeu majeur dans tous les domaines de la génétique médicale, ils concernent plus de 10% des DI génétiques (données du réseau français pour le diagnostic de la DI). Faute de tests fonctionnels, ils peuvent conduire à une situation d'errance diagnostique et empêcher une prise en charge efficace du patient. De ce fait, il reste nécessaire de réanalyser régulièrement les données face à des connaissances actualisées.

## 7) Modèles d'études des VSI

Les tests fonctionnels peuvent permettre de confirmer ou d'infirmer la pathogénicité d'un variant et l'objectif reste le plus souvent de démontrer la perte ou le gain de fonction entraîné par le VSI. Les validations fonctionnelles sont généralement réalisées *in vitro* et leurs mises au point dépend de la nature du variant ainsi que de son effet envisagé sur l'ARNm ou la protéine.

### 1. Tests biochimiques

Dans certains cas et notamment dans le cas des maladies métaboliques, il est possible de réaliser un test biochimique sur un prélèvement issu du patient. Par exemple, suite à l'identification d'un faux-sens chez un garçon avec une déficience intellectuelle associée à des traits autistiques dans le gène *MAOA* (situé sur le chromosome X et codant la monoamine oxydase A, une enzyme impliquée dans le métabolisme des catécholamines), un excès de substrats de cette enzyme a pu être mise en évidence dans les urines du patient, confirmant ainsi le caractère pathogène du variant faux-sens (Piton et al., 2014).

### 2. Analyses de cellules de patients

Dans le cas où elles sont disponibles, la validation fonctionnelle peut être réalisée sur des cellules issues directement du patient, telles que des cultures primaires de fibroblastes ou des cellules lymphoblastoïdes immortalisées à partir de leucocytes (LCL). Dans ce cas de figure il est possible d'étudier les conséquences du variant sur l'ARNm ou sur la protéine endogène. De plus, les résultats obtenus prennent en compte les paramètres spécifiques de l'individu, comme le fond génétique. Cependant, l'utilisation de ces cellules présente également un certain nombre

d'inconvénients. Dans le cas des pathologies suivant un mode de transmission autosomique dominant, la présence de l'allèle sauvage peut également rendre difficile l'interprétation des résultats. Enfin, ces types cellulaires correspondent à des tissus faciles d'accès (sang, peau, muscles) et peuvent ne pas être représentatifs de l'organe concerné par la pathologie, comme le cerveau pour la DI (régulation tissu spécifique, gène exprimé uniquement dans le cerveau ou seulement pendant le développement embryonnaire, etc).

### 3. Analyse *in vitro* des protéines mutantes

Dans le cas où les cellules issues du patient ne sont pas disponibles ou n'expriment pas la protéine d'intérêt, d'autres stratégies peuvent être utilisées. De nombreuses lignées cellulaires sont utilisées pour l'étude de l'effet de VSI (Exemple : HeLa, HEK293, SH-SY5Y, COS1, N2A). Différentes stratégies peuvent être mises en place dans le but d'exprimer la protéine porteuse de la variation d'intérêt, comme la surexpression via transfection d'un plasmide codant pour la protéine en question, ou l'édition du génome permettant de générer des lignées porteuses de variants étudiés. Selon la fonction de la protéine, plusieurs paramètres peuvent être investigués afin de comprendre l'effet du variant : le niveau d'expression et la stabilité protéique, la localisation cellulaire, l'activité enzymatique, l'interaction avec des partenaires ou encore la capacité des cellules à proliférer ou se différencier. Ces modèles cellulaires restent peu représentatifs des contextes physiopathologiques étudiés. C'est pourquoi l'utilisation de lignées iPSCs (Cellules Souches Pluripotentes induites) générées à partir de cellules de patients se démocratise, mais reste plutôt restreinte au cadre de la recherche. En effet, ces cellules directement issues du patient et donc porteuses de la variation ainsi que du fond génétique, sont largement polyvalentes car elles peuvent être différenciées en différents types cellulaires, dont les neurones, et sont de ce fait plus adaptés aux organes touchés par la pathologie.

### 4. Analyses des conséquences d'une altération de la fonction d'une protéine, exemple de la méthylation

Il a été récemment démontré que des variations de la méthylation de l'ADN peuvent être observées dans le cas d'inactivation de gènes régulant les processus épigénétiques (Tatton-Brown et al., 2017). Plusieurs signatures épigénétiques ont déjà pu être identifiées à l'heure actuelle, associées à des gènes tels que *NSD1*, *EZH2*, *CHD8* ou *SMARCA2*. Ainsi, les marques de méthylation de l'ADN furent utilisées dans le cadre de caractérisation de VSI. Par exemple, Une signature comprenant 229 CpG différenciellement méthylés fut identifiée chez des patients

porteurs de variations perte de fonction dans le gène *EZH2* (Choufani et al.,2020). Le gène *EZH2* code pour un composant du « Polycomb Repressive Complex 2 » (PRC2), un complexe régulant la structure de la chromatine par méthylation du résidu lysine 27 de l'histone H3. Les auteurs rapportent un variant *EZH2* présentant une hyperméthylation de CpG opposée à la sous méthylation observée chez les patients porteurs des variations perte de fonction. De plus, ce patient présentait un phénotype de retard de croissance, opposé au phénotype de croissance excessive observé chez autres patients. Finalement, ce variant fut validé comme gain de fonction grâce à une analyse enzymatique démontrant une activité augmentée. Ainsi, l'analyse de méthylation de l'ADN est un outil efficace permettant de reclasser les VSI identifiés dans ces gènes. Cette application nécessite cependant une implication directe ou indirecte du gène d'intérêt dans la régulation de ces processus de méthylation.

## 8) Modèles d'étude des gènes de DI

Les nombreux gènes impliqués dans la DI codent pour des protéines impliquées dans des fonctions moléculaires variées. De ce fait, les mécanismes physiopathologiques associés aux mutations de ces gènes sont pour la plupart généralement mal connus, alors même que leur identification est primordiale pour envisager des approches thérapeutiques. Au-delà de leur utilisation pour interpréter les variants dans des gènes connus, les tests réalisés sur cellules de patients ou sur d'autres types de cellules où le gène est inactivé ou les protéines mutantes exprimées, ainsi que la génération de modèles animaux mutés peuvent être utiles pour confirmer l'implication d'un gène dans la DI et pour comprendre les conséquences des mutations.

### 1. Modèles animaux

Si les modèles animaux sont un moyen d'analyse puissant dans la compréhension des mécanismes physiopathologiques, ils ne peuvent cependant pas être utilisés à grande échelle dans un contexte tel que l'étude des VSI hormis en présence de mutation récurrentes ou dans le cas de preuves de concept. Le génome du modèle animal peut être aujourd'hui facilement édité, notamment grâce à l'outil CRISPR/Cas9 mais peut également permettre une inactivation tissulaire spécifique ou temporelle dans le cas des modèles Cre/Lox. Le modèle doit être à même de récapituler le phénotype observé chez l'Homme et être choisi en fonction de la question posée. Outre le rôle du modèle animal dans la compréhension des mécanismes physiopathologique, il est également essentiel dans le développement et la mise en place de perspectives thérapeutiques. La souris et le rat sont des modèles de choix dans le cadre d'études

comportementales ou lors de l'étude du développement du cerveau. Par exemple, l'équipe de Yann Héroult a utilisé un modèle murin inactivé pour le gène *Kctd13* (Martin Lorenzo et al., 2021), un gène du locus 16p11.2, une région génomique sujette à des variations de type CNV associées entre autres à de la DI, pour mieux comprendre le rôle de ce gène dans la DI. Les auteurs observent chez les animaux mutant *Kctd13* +/- plusieurs phénotypes cognitifs également retrouvés chez les animaux porteurs de la délétion 16p11.2. Enfin, l'équipe rapporte que l'administration de la molécule « Fasudil » (un inhibiteur de la Rho-kinase) aux animaux *Kctd13* mutant permet d'améliorer certains phénotypes cognitifs de ces souris comme la mémoire de reconnaissance de l'objet.

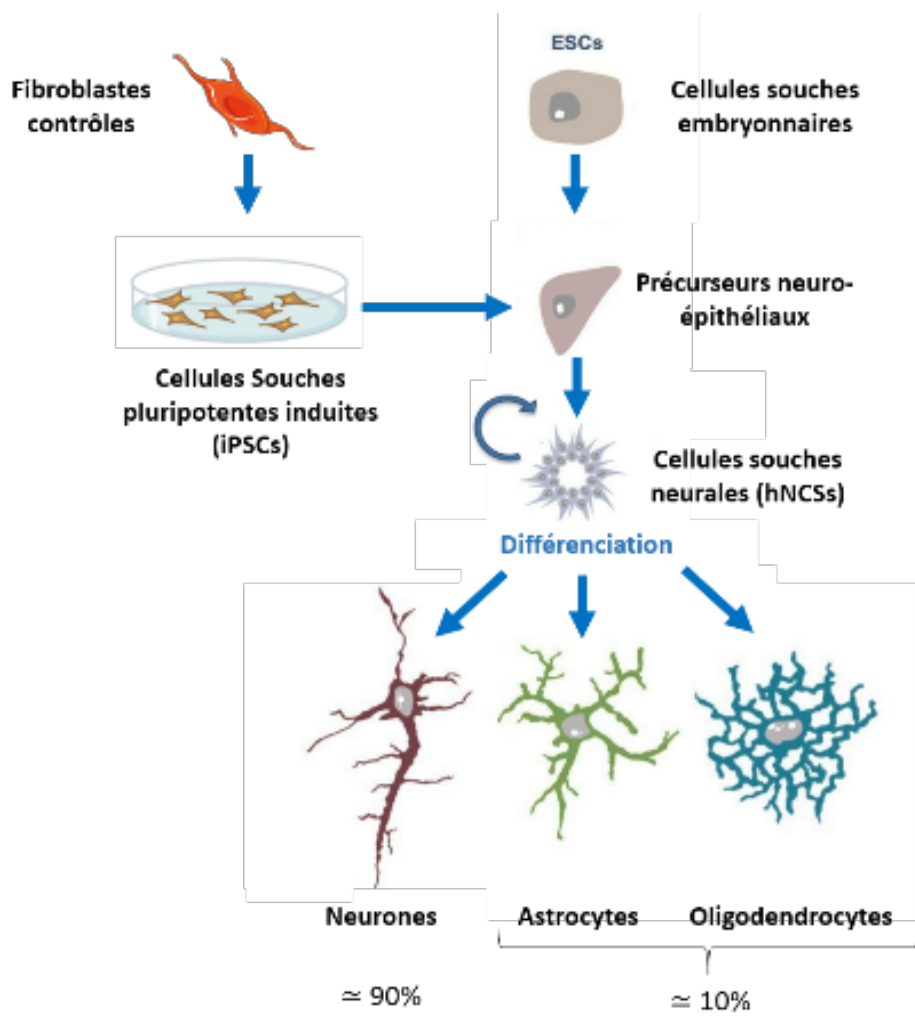
Dans le cas d'un phénotype robuste ou en raison de la facilité d'accès et de visualisation de l'organe, des modèles moins onéreux et générés plus rapidement pourront être utilisés, comme la drosophile, le poisson zèbre ou encore le vers *C. Elegans* et le xénope. Dans certains cas, c'est l'étude du modèle Drosophile qui a permis d'impliquer un gène dans la DI. Par exemple, l'équipe de Christiane Zweier, après avoir identifié des mutations faux-sens *de novo* dans *RHOBTB2* (codant une Rho GTPase) un gène initialement non connu en pathologie humaine, chez plusieurs patients avec DI et épilepsie sévère, a montré que la surexpression de ce gène chez la drosophile conduisait à une susceptibilité accrue à l'épilepsie ainsi qu'à des défauts locomoteurs tandis qu'un Knock Down du gène était associé avec une diminution de la taille et de la complexité de l'arborisation dendritique (Straub et al., 2018). Le modèle poisson zèbre est également utilisé en raison de son potentiel à récapituler le phénotype de différents TND, comme la réduction de la taille de la tête. Par exemple, les poissons inactivés pour le gène *Auts2* (un régulateur transcriptionnel) présentent une microcéphalie ainsi une diminution du nombre de neurones, phénotypes également retrouvés chez les patients (Oksenber et al., 2013) et ce phénotype peut être sauvé par injection d'un transcrit codant pour le cDNA *AUTS2* humain (Beunders et al., 2013).

## 2. Cellules souches neuronales humaines (hNSCs)

Comme vu précédemment, les modèles cellulaires sont largement utilisés dans le cadre de l'étude des TND. Cependant, les lignées « classiques » utilisées en laboratoires restent souvent peu pertinentes pour l'étude du développement du cerveau en raison de l'espèce (COS-1, N2A, PC12), des tissus desquels elles sont issues (HEK293, fibroblastes) ou encore des dérives physiologiques liées au caractère cancéreux de ces lignées (HeLa, U2OS, MCF-7). Cependant, la culture des cellules souches embryonnaires humaines (hESCs) ou la génération d'iPSCs (Takahashi et al., 2007) ont permis la création de modèles cellulaires humains plus pertinents

pour l'étude des TND. Ainsi, les hNSCs sont une lignée de cellules stables, capables d'auto renouvellement et correspondant à un stade intermédiaire de différenciation. En effet, les hNSCs disposent uniquement du caractère multipotent et correspondent à un stade de prédifférenciation en aval des précurseurs neuroépithéliaux mais toujours capables de donner les différents types de cellules neurales (Boissard et al., 2013) (Figure 5). Les hNSCs sont cultivées dans un milieu particulier en présence de trois facteurs de croissance (EFG, FGF-2 et BDNF) dont le retrait entraîne l'entrée des cellules en différenciation. Le processus de différenciation dure 21 jours et conduit à l'obtention d'une culture présentant principalement des neurones des couches corticales supérieures ( $\approx 90\%$ ) ainsi qu'une population d'astrocytes et d'oligodendrocytes ( $\approx 10\%$ ). Le laboratoire dispose grâce à une collaboration avec Alexandra Benchoua (institut ISTEM, Paris, France) de deux lignées de hNSCs (SA001, GM1869) (Boissard et al., 2013). Récemment, notre équipe a pu utiliser ces cellules pour étudier les conséquences d'une diminution d'expression de *NOVA2*, une protéine d'épissage alternatif associée à un syndrome de DI (Mattioli et al., 2020). Ainsi, nous avons par exemple pu mettre en évidence 41 gènes différentiellement exprimés dans les hNSCs partiellement inactivées pour *NOVA2*. Ces deux lignées sont cultivées dans les mêmes conditions et ne présentent aucune variation potentiellement délétère dans les gènes d'intérêt de mes études (*DYRK1A* et *PQBPI*) ainsi un caryotype normal (46, XY). Les hNSCs présentent cependant l'inconvénient d'être très difficilement transfectables. Il n'existe en effet aujourd'hui aucun protocole permettant l'entrée de matériel plasmidique dans ces lignées. En revanche, il reste toutefois possible de transfecter de petites molécules telles que les siRNA. C'est pourquoi lors de ce projet ce modèle fut uniquement utilisé pour étudier les conséquences de l'inactivation des gènes *DYRK1A* et *PQBPI* et non pour caractériser l'effet des différents variants d'intérêt. Enfin, la technologie d'interférence ARN ne conduisant qu'à une inactivation transitoire, il est également possible de générer des lignées hNSCs stables porteuses de variations dans les gènes d'intérêt grâce à la technologie CRISPR. De la même manière, la génération de telles lignées peut être réalisée par le biais d'iPSCs générées à partir de cellules de patients (comme des fibroblastes) et par différenciation de ces iPSCs en hNSCs. Ainsi, ce type de lignées stables peut par exemple permettre d'étudier les conséquences d'inactivation de gènes d'intérêt sur les processus de différenciation en neurones.





**Figure 5 : Provenance et différenciation des hNSCs.**



## II. Une forme monogénique fréquente de DI, le syndrome DYRK1A

Le gène *DYRK1A* code pour la protéine « Dual specificity tyrosine-phosphorylation-regulated kinase 1A » (kinase 1A à double spécificité Tyrosine et Serine/Thréonine). *DYRK1A* est situé chez l'Homme sur le chromosome 21 en position 21q22.13. *DYRK1A* fut impliqué en pathologie humaine et notamment dans deux syndromes neurodéveloppementaux (Syndrome de Down (SD) et syndrome DYRK1A) mais est également suspecté de jouer un rôle important dans certains syndromes neurodégénératifs tels que la maladie d'Alzheimer ou de Parkinson, ainsi que dans d'autres pathologies comme le cancer ou le diabète.

### 1) De minibrain à *DYRK1A*

*DYRK1A* fut décrit pour la toute première fois chez *Drosophila Melanogaster* par K.F Fishbach et M. Heisenber en 1984 (K.F Fishbach et M. Heisenber 1984). Dans cet article, les auteurs décrivent les mouches mutantes comme présentant une complexité des lobes optiques diminuée mais surtout une réduction de la taille du cerveau de 40 à 50% avec une réduction drastique du nombre de cellules, ce qui vaudra à l'homologue de *DYRK1A* l'appellation de *mnb* pour « mini-brain » (situé sur le chromosome X16E4-16E4). Quelques années plus tard, Tejedor et al. 1995 mettra en évidence une diminution de la prolifération des neuroblastes chez les mouches mutantes *mnb* ainsi que l'appartenance de *mnb* à une nouvelle famille de protéines kinases. Le rôle de *mnb* dans le développement du cerveau fut exploré dans un premier temps chez la mouche, puis ses homologues furent mis en évidence chez le rat (kentrup et al., 1996) et chez l'Homme (Song 1996 ; Guimerá et al., 1996). L'identification des différents homologues de *DYRK1A* et leurs cartographies sur le génome ont pu mettre en évidence que *DYRK1A* était extrêmement conservé au cours de l'évolution et localisé chez l'Homme dans la DSCR-1 (Down Syndrome Critical Region 1), renforçant ainsi l'intérêt des chercheurs pour ce gène. La DSCR-1 est décrite comme une portion du chromosome 21 d'environ 2,5Mb dont la duplication résume la majorité des traits phénotypiques des patients atteints du SD (Syndrome de Down). Dans les années qui suivirent, les modèles d'études animaux et cellulaires furent largement développés afin d'investiguer les conséquences de la surexpression de *DYRK1A*. Ainsi, l'implication de *DYRK1A* dans les déficits cognitifs du SD a pu être largement mise en évidence. Grâce à des modèles murins de surexpression par YAC (Yeast Artificial Chromosome) de portions variables de 0,5 à 2Mb de la région 21q22.2 contenant entre autres

*DYRK1A* (Desmond Smith et al., 1997). Les auteurs ont pu notamment souligner le déficit de mémoire d'apprentissage chez les animaux mutants.

Les modèles murins ont largement contribué à la compréhension du rôle de *DYRK1A* dans le développement du cerveau que ce soit dans un contexte d'haploinsuffisance (une seule copie de *Dyrk1a* fonctionnelle) ou dans un contexte de duplication/surexpression de *Dyrk1a*. L'implication de *DYRK1A* en pathologie humaine et notamment les conséquences de sa surexpression dans le développement des déficits cognitifs dans le SD fut admise durant la seconde moitié des années 90. Plus tard, *DYRK1A* sera également associé à un autre syndrome neurodéveloppemental causé par son haploinsuffisance (syndrome DYRK1A) ainsi qu'à d'autres pathologies comme le cancer, le diabète ou la maladie d'Alzheimer.

## 2) Découverte du syndrome DYRK1A

En 1997, une équipe identifie une micro délétion de la région 21q22.2 chez un patient atteint d'un retard de croissance utérin et d'une microcéphalie primaire (circonférence du crâne < 3 Standard déviation à la naissance) (Matsumoto et al., 1997). Parmi la quinzaine de gènes emportés par cette délétion, *DYRK1A* sera proposé comme celui potentiellement impliqué dans la microcéphalie. Ce n'est qu'en 2008 que les premières mutations impliquant *DYRK1A* seul sont identifiées chez des patients avec microcéphalie (Møller et al., 2008). Les auteurs rapportent 2 patients porteurs de grandes translocations dont les points de cassure sont situés respectivement dans les introns 2 et 6 du gène *DYRK1A*, entraînant de perte de fonction de l'un des allèles. Les deux enfants présentent notamment un retard de croissance intra-utérin, une microcéphalie primaire (périmètre crânien < 3 DS à la naissance) ainsi qu'une DI sévère. Deux années plus tard, la réanalyse de puces SNP réalisées sur 3009 patients atteints de DI a permis de mettre en évidence une microdélétion entraînant la perte des trois derniers exons de *DYRK1A* chez une autre patiente présentant des traits phénotypiques similaires (Van bon et al., 2011). Dans les années qui suivirent, de nombreuses autres délétions et variations ponctuelles furent mise en évidence dans *DYRK1A*, d'abord par séquençage Sanger puis par séquençage de panels de gènes ou d'exome. Ainsi, l'implication de *DYRK1A* dans la DI fut confirmée par la découverte successive de : 2 patients porteurs de décalages du cadre de lecture identifiés par séquençage sanger et qPCR sur 105 patients étudiés (Courcet et al., 2012) ; 10 patients porteurs soit de délétion (1), de décalages du cadre de lecture (4), de variants non-sens (3) ou faux-sens (2) identifiés par puces à ADN et séquençage ciblé ou séquençage d'exome (Bronicki et al. 2015). D'autres mutations ponctuelles furent rapportées par la suite (Ji et al., 2015 ; Van Bon

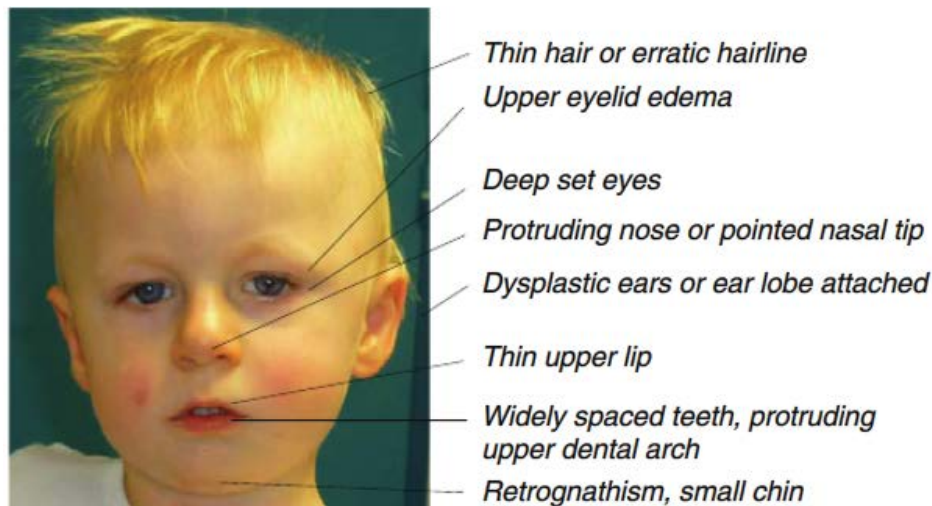
2016 ; Luco et al., 2016). En parallèle, *DYRK1A* fut également considéré comme un gène impliqué dans les troubles du spectre autistique (TSA) suite à la découverte de mutations *de novo* lors des études de séquençage d'exome réalisées sur des cohortes de patients présentant des TSA (O'Roak et al., 2012, Iossifov et al., 2012 ; Iossifov et al.,2014 ; Earl et al., 2017). Cependant, l'ensemble de ces patients TSA présenteraient également une DI.

A l'heure actuelle, le syndrome *DYRK1A* (OMIM #614104), causé par des délétions ou des mutations ponctuelles à l'état hétérozygote dans ce gène conduisant à une haploinsuffisance, est considéré comme un syndrome neurodéveloppemental fréquent, caractérisé par une DI associée à d'autres traits phénotypiques récurrents.

### 3) Caractéristiques du syndrome *DYRK1A*

Le syndrome *DYRK1A* est également référencé sous le terme « MRD7 » (Mental Retardation autosomal Dominant 7 ; OMIM #614104). Le phénotype des patients fut relativement bien décrit, notamment grâce aux cohortes réalisées sur les dernières années (Courcet et al., 2012 ; O'Roak et al., 2012 ; Bronicki et al., 2015 ; Ruaud et al., 2015 ; Ji et al., 2015 ; van Bon et al., 2016 ; Luco et al., 2016 ; Earl et al., 2017 ; Blackburn et al., 2019 ; Lee et al., 2020 ; Meissner et al., 2020 ; Courraud et al.,2021). Les premiers signes cliniques peuvent apparaître dès la période anténatale, avec un retard de croissance intra-utérin et une microcéphalie notamment. Le phénotype des patients à la période néonatale et la petite enfance est marqué par des troubles alimentaires avec des reflux gastro-intestinaux ainsi que des difficultés de succion et de déglutition. Durant la petite enfance, les patients présentent un retard global de développement ainsi que des troubles moteurs avec une hypertonie. Cette période est également marquée par un retard global de développement, avec un retard de langage important voir une absence de langage (le langage expressif étant plus affecté que le langage réceptif). Des troubles moteurs sont aussi présents avec une hypertonie et une démarche ataxique. La DI est constante dans le syndrome, dans la majorité des cas modérée à sévère. Des troubles du comportement sont aussi fréquemment rapportés chez les patients avec une hyperactivité, de l'anxiété et des troubles du sommeil. De plus, des traits autistiques ou des TSA sont décrits chez plus de la moitié des patients (Earl et al., 2017). Au niveau neurologique, des crises convulsives sont rapportées chez plus de la moitié des patients et des anomalies à l'IRM cérébrale peuvent également être mises en évidence (dilatation ou hypertrophie des ventricules, hypoplasie ou agénésie du corps calleux, retard de myélinisation de la substance blanche). D'autres signes cliniques non systématiques ont pu être relevés comme des anomalies de la

sphère urogénitale (hypospades, cryptorchidie) (Blackburn et al., 2019), des anomalies ophtalmologiques (myopie, strabisme, albinisme oculocutané) (Ji 2015 et al., ; Bronicki et al., 2015 ; Ernst et al., 2020) ou cardiaques (Van Bon et al., 2015). Enfin, dans la plupart des cas on notera une dysmorphie faciale assez caractéristique avec entre autres : un œdème de la paupière supérieure, des yeux enfoncés, une lèvre supérieure fine ainsi qu'un rétrognathisme (Figure 6).



**Figure 6 : Photographie d'un enfant porteur d'une variation perte de fonction dans *DYRK1A* montrant certaines caractéristiques morphologiques récurrentes au niveau du visage.** (Courraud et al., 2021).

Cependant, la clinique des patients n'est pas toujours évocatrice, la plupart des signes récurrents retrouvés chez les patients avec syndrome *DYRK1A* (comme la microcéphalie ou les TSA) étant loin d'être spécifiques du syndrome et pouvant être présents dans un grand nombre d'autres syndromes de DI. Par exemple, plusieurs patients *DYRK1A* avaient initialement été orientés vers un diagnostic pour le syndrome de Cockayne, un autre syndrome de DI causé généralement par des mutations dans les gènes *ERCC6* ou *ERCC8*, deux gènes codant pour des protéines impliquées dans le mécanisme de réparation NER (Nucleotide Excision Repair) de l'ADN (Boujelbene et al., en préparation). Les patients Cockayne présentent notamment : une microcéphalie, un retard de croissance ainsi que des anomalies de peau.

Il n'existe à l'heure actuelle aucune thérapie génique ni aucun traitement approuvé permettant de traiter le syndrome *DYRK1A*. La prise en charge des patients reste donc

symptomatique et nécessite l'intervention de différents professionnels de santé (gastro-entérologues, neuropédiatres, orthophonistes, kinésithérapeutes, pédopsychiatres) en fonction des difficultés de chaque individu. Plus largement, les patients bénéficient d'un suivi régulier par les professionnels de santé et peuvent être accompagnés par un AESH (Accompagnant d'Elève en Situation de Handicap) ou encore pris en charge dans un IME (Institut Médico-Educatif) ou autres établissements spécialisés. Afin de faciliter et d'améliorer le diagnostic et la prise en charge des patients, un PNDS (Protocole National de Diagnostic et de Soins) DYRK1A est en cours de rédaction.

#### 4) Aspects moléculaires du syndrome DYRK1A

Le syndrome DYRK1A est causé par une variation génétique entraînant la perte de fonction d'un des deux allèles *DYRK1A*. Cette modification peut être chromosomique (réarrangement chromosomique ; délétion/insertion ; CNV) ou ponctuelle (petite insertion/délétion causant un décalage du cadre de lecture, non-sens, faux-sens, variation affectant l'épissage) ( Møller et al., 2008 ; Van bon et al., 2011 ; Courcet et al., 2012 ; O'Roak et al., 2012 ; Redin et al., 2014 ; De Rubeis et al., 2014 ; Iglesias et al., 2014 ; Van Bon et al., 2016 ; Bronicki et al., 2015 ; Ruaud et al., 2015 ; Ji et al., 2015 ; Okamoto et al., 2015 ; Luco et al., 2016 ; Earl et al., 2017 ; Trujillano et al., 2017 ; Dang et al., 2018 ; Blackburn et al., 2019 ; Courraud et al., 2021). Les variants intragéniques entraînent généralement la synthèse d'une protéine catalytiquement inactive, souvent instable et dégradée par le protéasome. De nombreux variants faux-sens ont également été décrits et la quasi-totalité d'entre eux sont localisés dans le domaine kinase. En étudiant les positions de 171 variants faux-sens retrouvés dans la population générale (base de données gnomAD, anciennement ExAC), une équipe a pu mettre en évidence que la fréquence de ces variants était drastiquement diminuée au niveau du domaine kinase, supportant l'hypothèse que le domaine kinase de la protéine est extrêmement intolérant aux changements d'acides aminés (Evers et al., 2017). Les modélisations de structures réalisées par les auteurs suggèrent que ces variations interfèrent avec l'activité catalytique de DYRK1A, entraînant une perte de fonction. De récentes études fonctionnelles ont pu confirmer que ces variants affectaient en effet la protéine au niveau moléculaire que ce soit : en diminuant son expression ou en inhibant sa capacité d'autophosphorylation ou son activité enzymatique (Widowati et al., 2018 ; Arranz et al., 2019). Même si certains variants semblent n'entraîner qu'une perte partielle de stabilité de la protéine ou de son activité kinase (Arranz et al., 2019), il n'existe à l'heure actuelle aucune évidence d'une possible relation génotype-phénotype chez les patients.

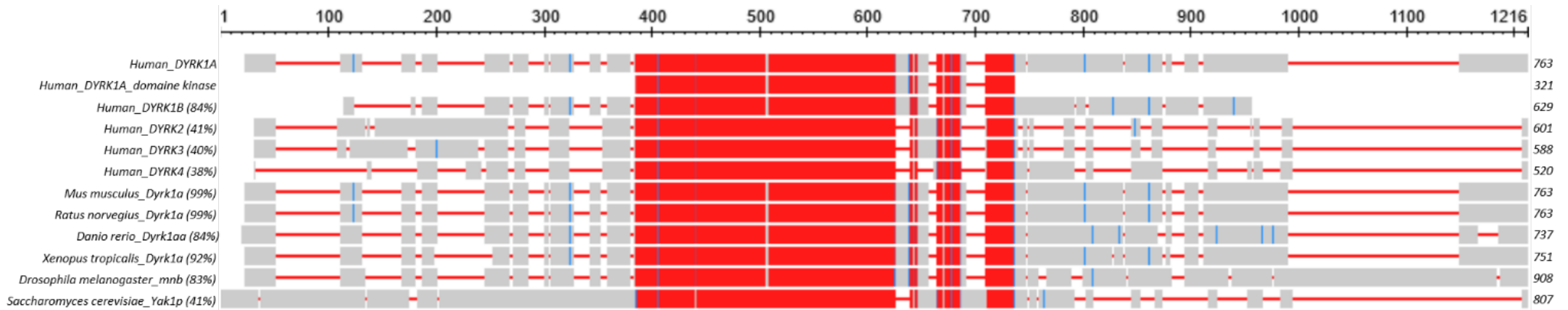
En effet, des patients porteurs d'une délétion complète du gène peuvent présenter des signes cliniques moins sévères que d'autres porteurs d'un variant faux-sens. Ces éléments semblent indiquer qu'une activité kinase « seuil » est requise au bon déroulement du neurodéveloppement. Il reste cependant très difficile d'estimer le pourcentage d'activité nécessaire à un développement normal du cerveau car des variants entraînant une diminution partielle de l'activité enzymatique et n'entraînant pas de DI ne sont pas détectés.

## 5) Gène et protéine DYRK1A

### 1. Généralités, orthologues et paralogues

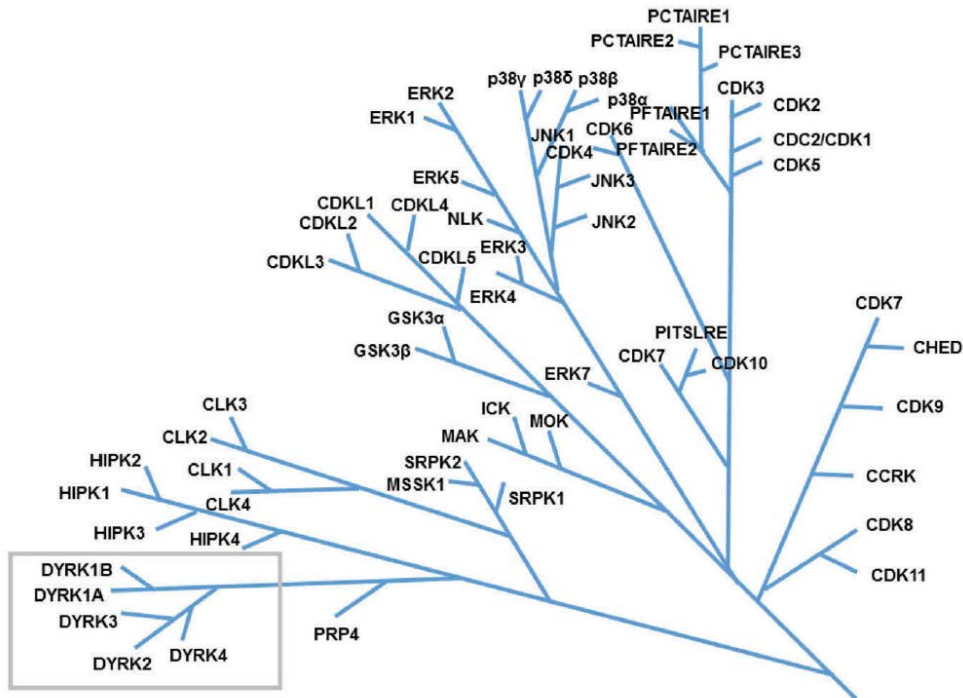
Le gène *DYRK1A* est situé chez l'Homme sur le chromosome 21 dans la région DSCR1- (Down syndrome Critical Region 1) et s'étend sur environ 160kb. *DYRK1A* code pour la protéine « *Dual specificity tyrosine-phosphorylation-regulated kinase IA* ». A l'heure actuelle, 9 transcrits d'épissage ont été décrits pour *DYRK1A*. Le plus grand de ces transcrits est généralement considéré comme transcrit de référence (NM\_001396.4) et comporte 11 exons codants pour 12 exons au total. Ce transcrit permet la synthèse d'une protéine de 763 acides aminés et d'une masse de 85kDa sans considération de modifications post-traductionnelles (The UniProt Consortium, 2020). DYRK1A fait partie de la famille des protéines DYRKs (Dual Specificity regulated Kinases) et est extrêmement conservé au cours de l'évolution : de la levure *Schizosaccharomyces pombe* à la souris (99% d'identité protéique avec *Mus Musculus*) (Figure 7).





**Figure 7 : Alignement des séquences protéiques des différentes DYRKs humaines et de quelques orthologues de DYRK1A.** Les alignements ont été réalisés avec la fonction d'alignement multiple de BLAST, le pourcentage d'identité est indiqué entre parenthèses et le nombre d'acides aminés composant les protéines est indiqué à droite.

Les protéines DYRKs sont incluses dans le groupe des CMGC qui comprend 5 autres familles de protéines kinases : les HIPKs (Homeodomain Interacting Protein Kiase), les MAPKs (Mitogen activated Protein Kinases), Les GSKs (Glycogen synthase kinases) et les CLKs (CCD2 like kinases) (Figure 8).



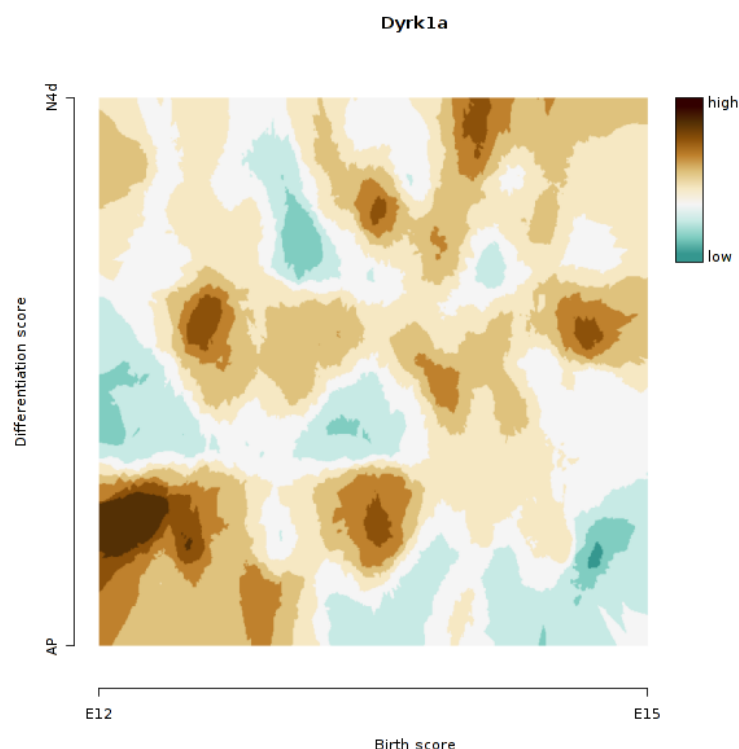
**Figure 8: Phylogénie des kinases du groupe CMGC** (Nguyen 2017 et al., 2017)

Chez l’Homme on retrouve 4 paralogues de *DYRK1A* : *DYRK1B*, *DYRK2*, *DYRK3* et *DYRK4* (ENSEMBL), *DYRK1B* étant le plus proche avec 74% d’identité protéique. Bien que ces 4 paralogues possèdent un certain pourcentage d’identité dans leur séquence avec *DYRK1A* (Figure 7), aucune étude n’a pour le moment fait état de leur potentielle implication dans un syndrome neurodéveloppemental. En effet, *DYRK1B* a été relié au syndrome d’obésité métabolique abdominal 3 ainsi que dans la prédisposition à certains cancers du foie, *DYRK2* et *DYRK4* ont également été associés à certains types de cancers et enfin *DYRK3* a été décrit comme fortement exprimé chez des patients anémiques (Laham et al., 2021).

## 2. Expression du gène

*DYRK1A* est exprimé de manière ubiquitaire mais son expression n’est pas totalement homogène en fonction des tissus considérés et des stades de développement. La présence de l’ARNm de *DYRK1A* a pu être mise en évidence dans le SNC (Système Nerveux Central) chez

la souris, l'embryon de poulet mais également dans le cerveau de patients atteints du SD (Guimera et al., 1996 ; Marti et al., 2003 ; Hämmerle et al., 2002 ; Wegiel et al., 2004). Un pic d'expression de *Dyrk1a* sera décrit dans le cerveau du rat à la naissance, suivi d'une diminution d'expression maintenue durant l'âge adulte (Okui et al., 1999). Chez l'Homme adulte, l'ARNm *DYRK1A* est également exprimé de manière ubiquitaire. De plus, des analyses de Northern Blot réalisées sur des extraits d'embryons humains ont pu montrer que *DYRK1A* était exprimé dans le cerveau, le foie, les poumons et les reins (Guimera et al., 1999). En 2008 une équipe a réalisé une analyse de l'expression de *Dyrk1a* durant le développement du cerveau chez la souris par méthode d'ISH, qui a permis de mettre en évidence une régulation spatio-temporelle complexe de l'expression, qui diffère selon les stades de développement et les zones du cerveau (Hämmerle et al., 2008). Ces résultats corrélerent avec des bases de données plus récentes qui rendent compte de l'expression des niveaux d'expression dans les progéniteurs neuronaux chez la souris en fonction du niveau de différenciation entre E12 et E15 (Telley et al., 2019) qui montrent une expression forte de *Dyrk1a* dans les progéniteurs à E12 avant l'entrée en différenciation, puis une expression variable après l'entrée des cellules en différenciation (Figure 9).



**Figure 9 : Pattern d'expression de *Dyrk1a* dans les progéniteurs neuronaux en différenciation** chez la souris. Les échelles représentent les stades embryonnaire E12 à E15 et l'état de différenciation AP (Progéniteurs apicaux) à N4d (neurones de 4 jours). L'échelle de couleur représente le niveau d'expression de l'ARNm *Dyrk1a*.

En 2013 les données transcriptomiques générées par le Consortium GTEx par séquençage d'ARN en cellules uniques sur 54 tissus différents provenant de près de 980 échantillons post-mortem d'humains adultes ont confirmé une expression ubiquitaire de *DYRK1A* avec au niveau du cerveau, un enrichissement dans le cervelet. (Figure 10).

Bulk tissue gene expression for DYRK1A (ENSG00000157540.19)

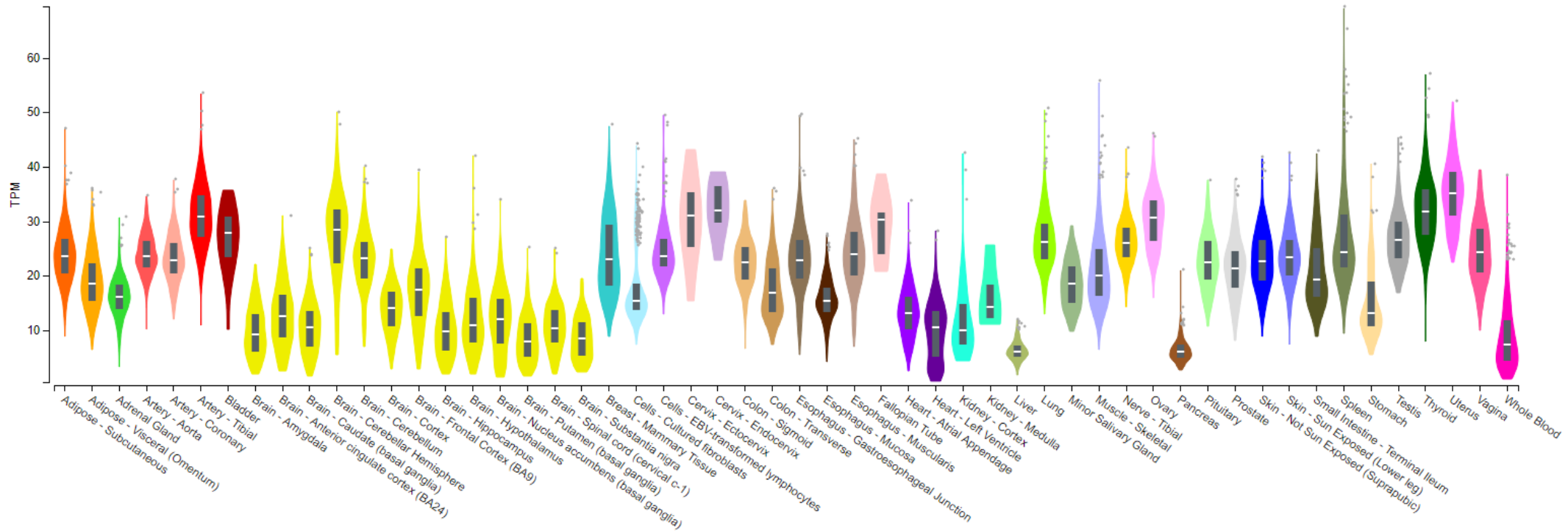
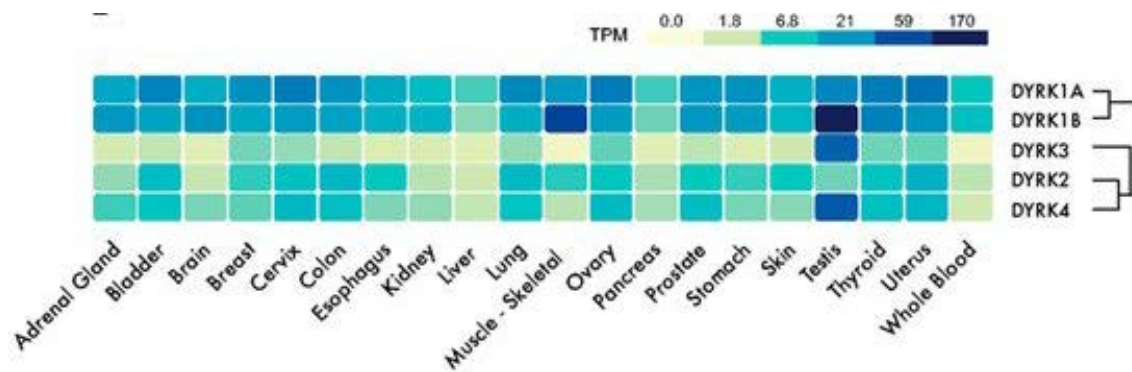


Figure 10 : Expression de DYRK1A dans différents tissus issus de prélèvement post-mortem chez l'Homme (GTEx Consortium, 2013).

Cette base de données nous renseigne également sur les expressions communes des différents gènes de la famille des DYRKs (Figure 11).



**Figure 11 : Expression de l'ARN messager des protéines DYRKs humaines basée sur les données de la base de données GTEx.** Les données sont représentées comme TPMs (Transcrits par million) médians. (Boni et al., 2020).

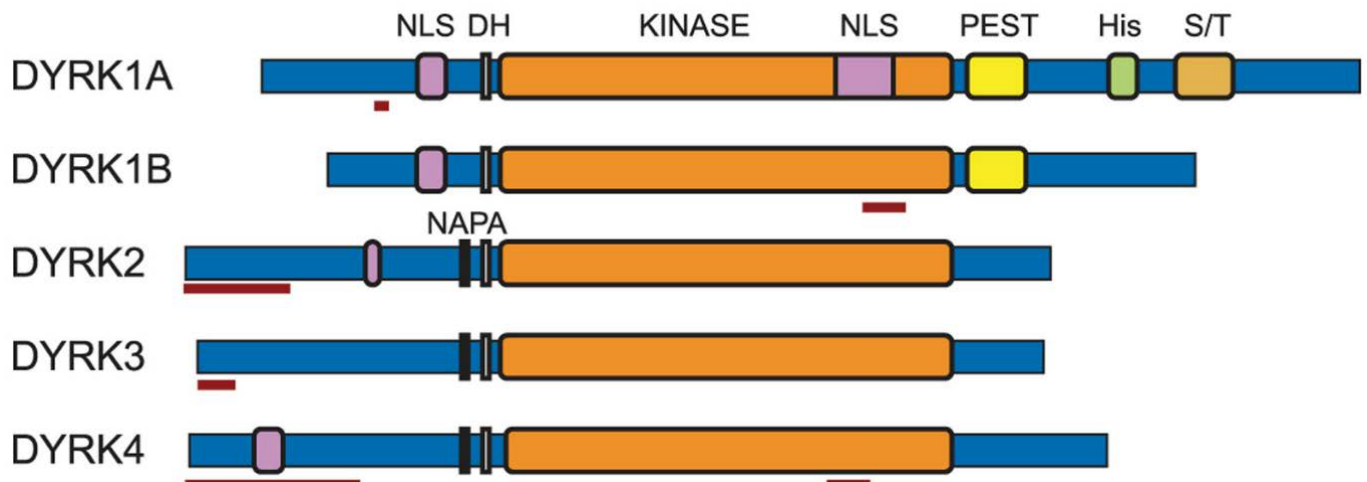
### 3. Expression de la protéine

Au niveau protéique, l'expression de DYRK1A est également ubiquitaire (Human protein Atlas version 20.1). Malgré la présence de deux séquences de localisation nucléaire (NLS), la localisation intracellulaire de DYRK1A semble varier selon les stades de développement et les types cellulaires. Des expériences de surexpression en modèles cellulaires font état d'une localisation majoritairement nucléaire avec une potentielle accumulation de la protéine dans les « nuclear speckles » (Alvarez et al., 2003). Chez la souris, *Dyrk1a* est exprimé durant la neurogénèse au niveau du tube neural, dans plusieurs zones de l'encéphale (Bulbe olfactif, hippocampe, thalamus et hypothalamus) et dans le cervelet. La protéine est présente dans le compartiment nucléaire mais également cytoplasmique dans les dendrites des cellules nerveuses (Hämmerle et al., 2008) et est détectée dans le soma des cellules en différenciation (Hämmerle et al., 2003). Chez l'Homme, DYRK1A semble localisé plus majoritairement dans le cytoplasme des cellules nerveuses (Marti et al., 2003) et est détecté dans la majorité des zones du cerveau (Hippocampe, cervelet, thalamus et hypothalamus) (Weigiel et al., 2011, Review).

### 4. Structure de la protéine et rôles des différents domaines

La protéine DYRK1A de référence (NM\_001396.4) est longue de 763 aa. Elle présente plusieurs domaines bien caractérisés. Deux éléments sont communs aux 5 membres de la famille : la boîte DH (DYRK Homeobox) (aa 137-153) nécessaire à l'auto activation de la protéine (Widowati et al., 2018), ainsi que le domaine kinase au centre de la protéine. Le

domaine Kinase de DYRK1A (aa 159-479) correspond à la région de la protéine la plus conservée au cours de l'évolution et confère à l'enzyme son activité catalytique (Figure 12).



**Figure 12: Structures primaires des protéines DYRKs humaines**

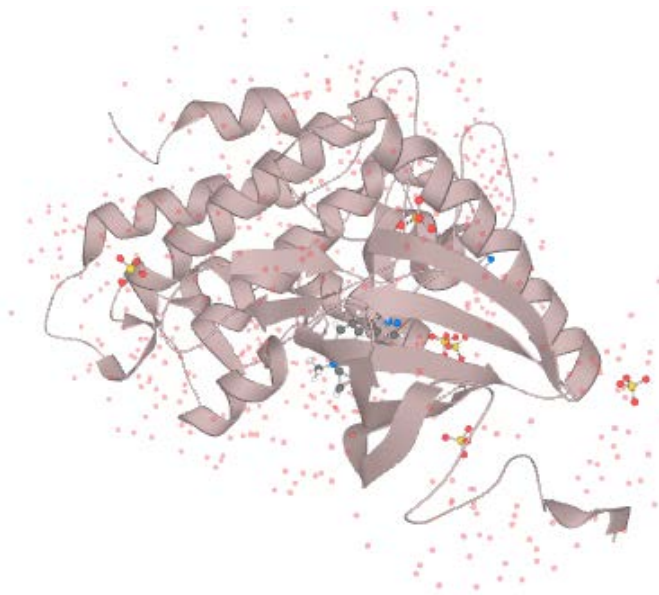
(Aranda et al., 2010, Review). Représentation de la structure primaire des différentes protéines de la famille des DYRKs : NLS (Séquence de Localisation Nucléaire) ; DH (DYRK Homology box) ; KINASE (Domaine Catalytique) ; PEST (Séquence peptidique riche en proline, acide glutamique, serine et thréonine) ; His (domaine riche en histidine) et S/T (Répétitions de résidus serines et thréonines) ; NAPP (région N-terminale nécessaire à l'autophosphorylation de la protéine) ; les lignes rouges indiquent les régions affectées par des événements d'épissage alternatif.

La réaction de phosphorylation correspond dans le cas des kinases au transfert d'un groupement phosphate provenant d'une molécule d'ATP (Adénosine TriPhosphate) sur un substrat qui peut être une protéine, un ose ou un lipide. DYRK1A permet le transfert de l'ion phosphate sur le groupement hydroxyle (-OH) de la chaîne latérale d'un aa sérine, thréonine ou tyrosine. En 2000, une séquence peptidique consensus de DYRK1A fut proposée (Himpel et al., 2000) : R-P-x(1,3)S/T-P (où R est une arginine, P une proline, S une sérine, T une thréonine et x(1,3) est une suite de 1 à 3 aa variables). Il est importante de noter que le domaine kinase de DYRK1A comporte un résidu tyrosine en position 321 particulièrement important dans la régulation de son activation.

DYRK1A présente également deux séquences NLS. La première (NLS1) est une séquence dite « bi-partite » située dans la région N-terminale de la protéine (aa 117-134) et est commune à DYRK1A, DYRK1B, DYRK2 et DYRK4. La seconde (NLS2) est présente dans le domaine kinase de la protéine (aa 389-395), est spécifique de DYRK1A et fut mise en évidence par la surexpression de différentes protéines tronquées (Alvarez et al., 2003). DYRK1A comprend également dans sa région C-terminale un domaine PEST (séquence riche en proline, acide glutamique, sérine et arginine) (aa 497 à 565) retrouvé aussi chez DYRK1B. Les domaines PEST sont généralement présents sur des protéines à demi-vie courte comme par exemple la cycline D1. Mais le fonctionnement moléculaire des PEST n'a pas encore été complètement élucidé. On sait que cette séquence peptidique régule notamment la stabilité de la protéine en favorisant sa dégradation par la voie du protéasome (Chakraborty et al., 2011) et également potentiellement son clivage par des calpaïnes dans le cas de DYRK1A (Jin et al., 2015). La région C-terminale de DYRK1A lui est spécifique, elle comporte deux domaines. Le premier est une région His (aa 599 à 619) composée de deux répétitions de respectivement 4 (aa599-602) et 13 (aa 607 à 619) résidus histidines. Cette région est proposée comme permettant l'adressage de DYRK1A dans les « nuclear speckles » afin d'y exercer un rôle de régulateur d'épissage alternatif (Alvarez et al., 2003). Le deuxième domaine de la région C-terminale de DYRK1A correspond à une répétition de résidus sérines et thréonines (aa 640-672) mais son rôle reste actuellement inconnu. En effet, l'expression de protéines mutantes présentant une délétion de cette région n'a présenté aucun phénotype particulier, notamment au niveau de la localisation de DYRK1A dans les « nuclear speckles » (Alvarez et al., 2003). Contrairement à DYRK2, DYRK3 et DYRK4, DYRK1A ne possède pas de domaine NAPA (N-terminal AutoPhosphorylation Accessory). Le domaine NAPA est nécessaire à la réaction d'activation par autophosphorylation de ces protéines DYRK et son activité est qualifiée de « chaperone like » (Kinstrie et al., 2010).

En ce qui concerne la structure cristallographique de DYRK1A, elle n'a pas encore été résolue dans son intégralité et seul le domaine catalytique bordé de quelques aa est actuellement disponible. Le dépôt le plus récent est celui de Sorrel et al. en 2019 et comporte une résolution de 1,05 Å sur la région des aa 127 à 485 (Figure 13).





**Figure 13 : Représentation de la structure de DYRK1A en présence d'inhibiteurs** réalisée par cristallographie aux rayons X (Sorrel et al. en 2019) (résolution 1,05 Å, positions aa 127-485, PDB entry 6S14).

## 5. Régulation transcriptionnelle de *DYRK1A*

Comme nous l'avons vu précédemment *DYRK1A* est exprimé de manière ubiquitaire et son niveau d'expression doit être finement régulé. La régulation transcriptionnelle de *DYRK1A* est peu décrite mais certains acteurs ont pu être identifiés. L'AP4 (Activator protein 4) est une protéine facilitant le recrutement de FT (Facteurs de Transcriptions) jouant un rôle de répresseur dans les cellules non neuronales et est un inhibiteur transcriptionnel de *DYRK1A* (Kim et al., 2006). En 2008, un second site d'initiation de transcription de *DYRK1A* est mis en évidence dans une lignée de cellules PC3 (Prostate Carcinoma) conduisant à la synthèse d'un transcrit comportant une région 5'UTR différente du transcrit initialement décrit (Maenz et al., 2008). Dans la même étude, les auteurs observent que la transcription de *DYRK1A* est positivement régulée par le facteur E2F1 en cellules HEK confirmant les résultats similaires de deux autres études réalisées en cellules U2OS (Ostéosarcome) et NIH3T3 (fibroblastes murins). On sait également que REST1 (RE1 Silencing Transcription factor 1) peut activer la transcription de *DYRK1A* dans des cellules de mammifères (HEK293 et HeLa) en se liant sur une séquence NRSE (neuron-restrictive silencer element) située dans son promoteur (Lu et al., 2011). La corrélation d'expression entre *Dyrk1a* et *Rest* fut validée sur coupes de cerveaux de souris

adultes. De plus, les auteurs ont également proposé un mécanisme dans lequel DYRK1A phosphoryle REST et entraîne sa dégradation par le protéasome, REST régulant la transcription de *DYRK1A* dans une boucle de rétrocontrôle négatif. Dans un modèle de cellules T98G (glioblastome humain), la même équipe a pu souligner que MEF2D (Myocyte-specific enhance factor 2) active la transcription d'une isoforme de *DYRK1A* plus courte (NM\_130438.2) que l'isoforme de référence en se fixant sur un élément de réponse dans un second promoteur de *DYRK1A* (Wang et al., 2017). Cette forme courte est composée de 529aa (contre 763 pour l'isoforme de référence) et diffère par sa partie terminale. En effet l'absence de l'exon 11 dans l'isoforme courte entraîne un décalage du cadre de lecture de l'exon 12 (dernier exon codant des autres isoformes) et ainsi la synthèse d'une protéine plus petite qui diffère de la forme de référence à partir de l'aa 515 et ne comporte de ce fait pas les domaines PEST, His et S/R. La protéine courte comporte les deux séquences de localisation nucléaire ainsi qu'un domaine kinase fonctionnel mais même si son rôle reste inconnu, les auteurs mettent en évidence une expression d'ARNm ainsi qu'une activité kinase de DYRK1A corrélées à l'expression de MEF2D. MEF2D est un facteur de transcription notamment impliqué dans le développement et la différenciation neuronal et musculaire, il est exprimé dans les neurones post mitotiques et permet leur survie. Etant donné l'importance du dosage de *DYRK1A* durant le neurodéveloppement et l'expression tissu spécifique d'isoformes d'épissages, cette régulation pourrait être importante dans la régulation de certaines voies cellulaires. Enfin, deux miR (micro ARN) sont proposés comme ciblant l'ARNm *DYRK1A* : miR1246 et miR-199b (Arbones et al., 2019, Review).

## 6. Activation de DYRK1A

Les protéines DYRKs ont la propriété de réaliser une autophosphorylation sur un résidu tyrosine. Cette autophosphorylation est considérée comme une auto-activation de l'enzyme qui lui confère son activité catalytique. Le motif nécessaire à l'activation de l'enzyme correspond à une séquence « Y-X-Y » (où Y sont des résidus tyrosines et X un aa aléatoire). Ce processus est appelé boucle d'auto-activation et c'est l'aa tyrosine 321 (Tyr<sup>321</sup>) qui est autophosphorylé par DYRK1A (Himpel et al., 2001).

Le fonctionnement moléculaire de ce processus est décrit comme une activation en *cis* de la protéine, réalisée directement durant la phase de traduction. En effet, DYRK1A adopterait dans un premier temps une forme de transition, capable de réaliser la réaction de phosphorylation sur la Tyr<sup>321</sup>. Une fois la réaction réalisée, la protéine acquerrait alors sa forme

mature (Himpel et al., 2001 ; Lochhead et al., 2005). On sait cependant que ce mécanisme ne nécessite aucun facteur protéique supplémentaire (chaperonnes comprises) et que DYRK1A est capable de s'autophosphoryler spontanément en présence d'ATP (Lochhead et al., 2005). La forme mature de DYRK1A correspond à la stabilisation de la forme active de la kinase, dont la conformation permet une interaction adéquate entre le site catalytique et le substrat (Nolen et al., 2004). Les DYRKs sont caractérisées comme capables de phosphoryler les trois types de résidus (sérine, thréonine et tyrosine) mais à l'heure actuelle, aucune tyrosine hormis la Tyr<sup>321</sup> n'a été démontrée comme phosphorylée par DYRK1A *in vivo*.

## 7. Régulation d'activation de la protéine DYRK1A

A l'heure actuelle, 58 sites de modifications post-traductionnelles sont référencées sur la base de données BioGRID (Stark et al., 2006). Le niveau de régulation le plus important de DYRK1A correspond à sa capacité d'autophosphorylation qui conditionne la présence de l'activité catalytique et est réalisée directement à la fin voire durant la traduction. Mais l'activité de DYRK1A semble également régulée par la présence d'interacteurs protéiques, permettant de moduler l'activité de l'enzyme dans le temps mais également en fonction du type cellulaire. En 2004, la protéine 14-3-3 est identifiée par méthode de double hybride chez la levure comme interagissant avec la région N-terminale de Dyrk1a (Kim et al., 2004). Cette interaction est proposée comme stimulant l'activité enzymatique de DYRK1A. Plus tard, une seconde équipe précisera qu'une phosphorylation au niveau du résidu sérine 520 de DYRK1A permet l'interaction avec le facteur 14-3-3, stimulant l'activité kinase de l'enzyme (Alvarez et al., 2007). Les protéines SPRED1 et SPRED2 (Sprouty-Related Protein with an EVH1 domain) initialement décrites comme inhibant la voie Ras/ERK semblent également diminuer l'activité kinase de DYRK1A dans la lignée HEK293 en interagissant directement avec le domaine kinase et modifiant sa capacité à reconnaître certains substrats (Li et al., 2010). Les modifications post-traductionnelles de DYRK1A sont actuellement peu décrites. La protéine LATS2 (une sérine/thréonine kinase) est proposée comme phosphorylant DYRK1A et stimulant son activité de phosphorylation de LIN52, un composant du complexe DREAM (Tschöp et al., 2011). Il est également proposé que la calpaïne-1 (une protéase), cliverait une partie du pool de protéine DYRK1A cellulaire dans sa partie C-terminale, procurant ainsi aux produits de clivage une activité kinase augmentée et conduirait notamment à une hyperphosphorylation de la protéine MAPT (Microtubule Associated Protein Tau) (Jin et al., 2015). Jusqu'à cette année, aucun mécanisme de régulation de l'activité de DYRK1A par action

directe d'une phosphatase n'avait été mis en évidence. C'est une équipe coréenne qui a identifié un inhibiteur protéique potentiel de DYRK1A : la phosphatase PPM1B (Lee et al., 2021). En effet, les auteurs utilisent la phosphorylation de MAPT (TAU) sur son résidu thréonine 212 (T212) comme rapporteur de l'activité kinase de DYRK1A. Ils mettent en évidence que PPM1B interagit directement avec DYRK1A et est capable de retirer le groupement phosphate présent sur le résidu sérine 258, inhibant ainsi son activité de phosphorylation et réduisant indirectement le taux de phosphorylation de la thréonine 212 de MAPT dans un modèle de cellules de mammifères en culture (HEK293).

## 6) Rôles de DYRK1A

La protéine DYRK1A est un élément cellulaire clef et dispose d'un intérêt et d'une littérature extrêmement riche, notamment en raison de sa très forte conservation au cours de l'évolution et de son implication en pathologie humaine à plusieurs niveaux (SD, syndrome DYRK1A, neurodégénérescence, cancer, ...). Aujourd'hui, 409 interacteurs de DYRK1A sont recensés sur la base de données BioGRID (Stark et al., 2006) et plusieurs dizaines de substrats mis en évidence, reliant la protéine de manière directe ou indirecte à de nombreux processus cellulaires. De plus, l'importance de son dosage dans la cellule ajoute un niveau de difficulté supplémentaire à la compréhension des rôles de DYRK1A, les conséquences d'une diminution d'expression ou de sa sur-activation étant différentes et ne présentant pas toujours un effet « miroir ». Les exemples de cibles et d'interacteurs présentés dans ce travail restent non exhaustifs.

### 1. Analyses omics

Le terme OMIC désigne une analyse computationnelle massive de données d'un certain type (protéines, ADN, ARN, lipides, ...). Les OMICs permettent une investigation puissante, globale, sans a priori et sont devenus des outils incontournables dans la découverte de nouveaux mécanismes mais présentent cependant plusieurs inconvénients. Elles restent souvent onéreuses et comportent un grand nombre de variabilités qui dépendent de paramètres tels que : le type cellulaire étudié, la fenêtre temporelle, les variations inter-individuelles ou les biais techniques. Les OMICs ont apporté un niveau de complexité additionnel et permis de considérer les rôles et interactions multi-systémiques que peuvent avoir un gène ou une protéine, y compris pour *DYRK1A* (Table 1).

<b>Auteurs</b>	<b>OMICs</b>	<b>Modèle</b>	<b>Implications de DYRK1A / Observations</b>
Malcher 2011	Microarray	<i>Saccharomyces cerevisiae</i>	Formation de biofilm et résistance au stress
Varjosalo 2013	Protéome	T-REx™-293	Identifications d'interacteurs
Jang 2014	Human exon array analysis	HeLa/T98G	Marques épigénétiques et immunité
Miyata 2014	Protéome	COS7	Identifications d'interacteurs
Di Vona 2015	Chip-Seq	HeLa/T98G	Régulation de la traduction, épissage, cycle cellulaire
Huttlin 2015	Protéome	HEK293	Identifications d'interacteurs
Thompson 2015	RNAseq	Souris <i>Dyrk1a</i> +/-, cellules immunitaires (pre B et T)	Cycle cellulaire et transcription médiée par E2F
Huttlin 2017	Protéome	HEK293T, HCT116, ou MCF10A	Identifications d'interacteurs
Nguyen 2018	P-protéome	Cerveaux de souris TgDYRK1A, Ts65Dn	Transmission synaptique, cytosquelette, projection des neurones
Li 2018	Protéome et ChipSeq	HEK293	Cycle cellulaire, traduction, transcription, adhésion cellulaire
Roewenstrunk 2019	Protéome	HeLa/T98G	Cycle cellulaire, apoptose, transcription, cytosquelette
Eguard 2019	Protéome	HeLa	Anaphase Promoting Complex, DNA repair, transcription
Menon 2019	Protéome	T98G	Métabolisme des ARN, cycle cellulaire, jonctions cellulaires
De Toma 2019	Protéome et p-protéome	Souris TgDYRK1A <sup>2</sup>	Identification de hubs régulés : SMAD3, PRNP, HRAS, PTEN
Cho 2019	RNA-seq	Poisson zèbre <i>dyrk1a</i> +/- (48 hpf)	Processus d'oxydo-réduction, adhésion cellulaire, développement
Willsey 2020	RNA-seq	Xénope <i>Dyrk1a</i> +/- (cerveaux au stade 46)	Réplication de l'ADN, cycle cellulaire
Recasens 2021	Protéome et p-protéome	U251 ; glioblastomes humains	Cycle cellulaire, réplication de l'ADN

**Table 1 : Résumé des principales analyses OMICS impliquant DYRK1A**

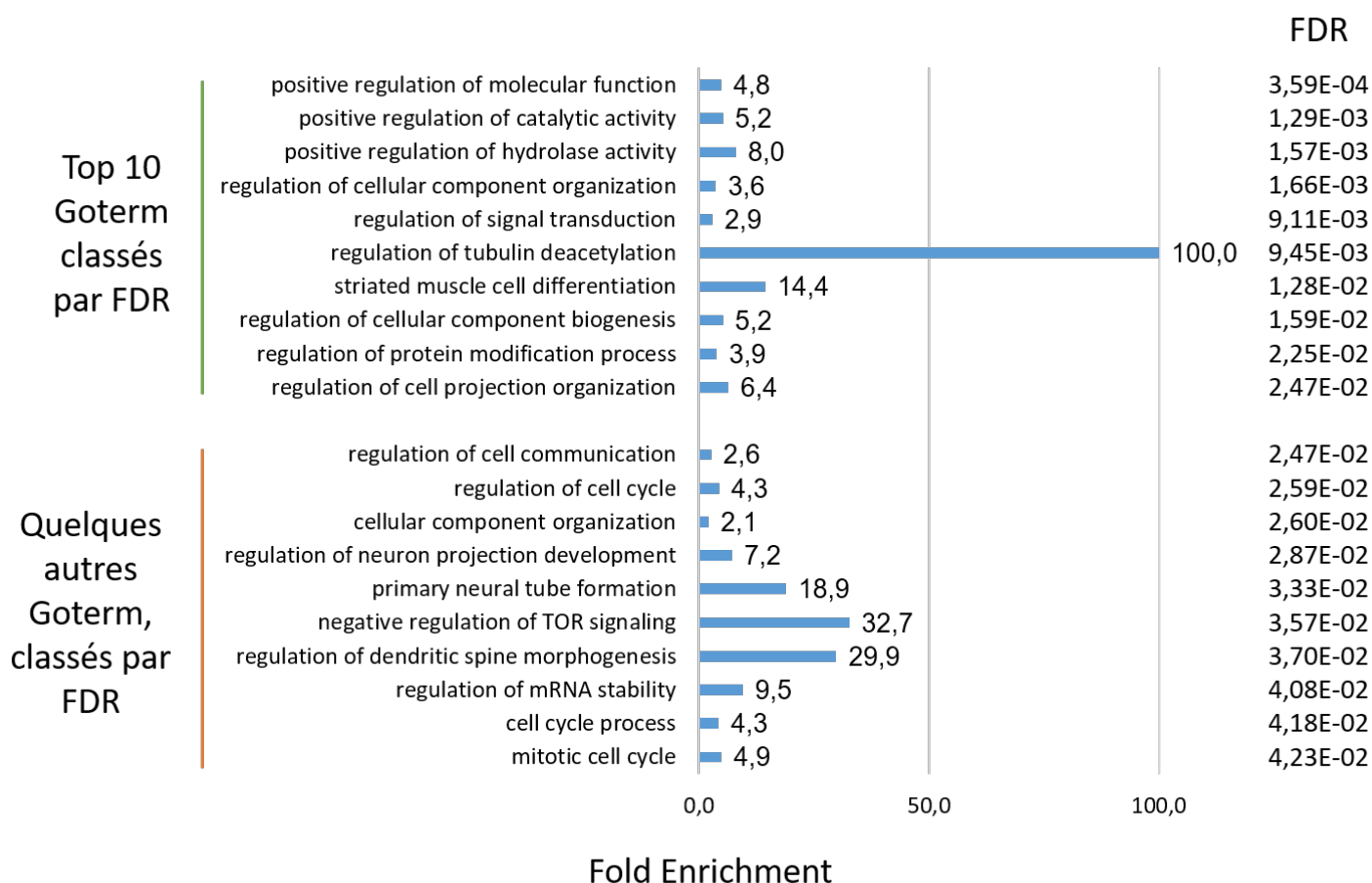
T-REx™-293 : lignée embryonnaire de rein humain exprimant de façon stable la protéine répresseur de tétracycline ; HeLa : lignée du cancer du cancer du col de l'utérus humain ; COS7 : lignée de fibroblastes rénaux de singe immortalisés par le virus SV-40 ; T98G et U251 : lignées de glioblastome humain ; HEK293 : lignée embryonnaire de rein humain ; HCT116 : lignée du cancer du côlon humain ; MCF10A : lignée épithéliale non maligne de sein humain ; TgDYRK1A : lignée murine comportant une copie supplémentaire du gène *DYRK1A* humain ; TgDYRK1A<sup>2</sup> : lignée murine comportant une ou plusieurs copies du transgène *sMT-1a/Dyrk1A* ; Ts65Dn : lignée murine de trisomie partielle pour 132 gènes de la partie distale des chromosomes MMU16 et MMU17 ; hpf : heures post-fécondation.

### *a. Chip-Seq et RNA-seq*

Ce n'est que tardivement que le rôle nucléaire de DYRK1A, notamment dans la régulation de la transcription, fut investigué. Une première analyse de Chip-Seq fut réalisée sur DYRK1A en cellules HeLa et T98G (Di Vona et al., 2015 ; Di Vona et al., 2021, Preprint BioRxiv), révélant une interaction de la kinase avec l'ARN polymérase II et un enrichissement significatif sur les séquences « TCTCGCGAGA ». Plus récemment, une autre expérience de ce type en cellules HEK293 a mis en évidence une co-localisation de DYRK1A avec CBP et p300, deux acétyltransférases, au niveau de régions enhancers (Li et al., 2018). Cependant, en comparant leurs résultats à ceux obtenus précédemment par Di Vona, les auteurs mettent en évidence un faible chevauchement entre les régions génomiques occupées par DYRK1A dans les lignées T98G et HEK293, indiquant que la présence de DYRK1A sur les séquences enhancers semble être tissu spécifique. La comparaison des analyses GOterms de ces deux études permet cependant de relever des implications similaires de DYRK1A dans les deux types cellulaires, notamment concernant la régulation de la traduction, de la réparation de l'ADN et du cycle cellulaire. Dans le but d'identifier les gènes dont la transcription est régulée de manière directe ou indirecte par DYRK1A, plusieurs équipes ont pu réaliser un transcriptome et ce, dans différents modèles. La comparaison de ces différentes études met en évidence un très faible recoupement des gènes différentiellement exprimés en cas de diminution d'expression de *DYRK1A*. En effet, il ne semble pas exister de groupe de gènes dont l'expression est modulée par DYRK1A commun à l'ensemble des études. Cette observation met en évidence le caractère tissu spécifique de la régulation transcriptionnelle exercée par DYRK1A. On notera également que ces études ont été réalisées dans des espèces différentes (souris, xénope, poisson zèbre, cellules humaines) et que malgré le degré de conservation de cette kinase, son rôle peut également partiellement différer selon les modèles utilisés. En plus d'observer une hétérogénéité des gènes régulés, DYRK1A semble pouvoir adopter un rôle activateur ou inhibiteur de la transcription qui peut être modulé selon les modèles utilisés.

### *b. Protéomique et phospho-protéomes*

Un très grand nombre de partenaires de DYRK1A ont aujourd'hui pu être décrits, notamment grâce à la technique de spectrométrie de masse. Ces analyses protéomiques réalisées le plus souvent à partir de cellules de mammifères en culture ont permis d'élargir l'intéactome de DYRK1A (Table 1) mais également de confirmer son implication dans un grand nombre de processus cellulaires (Figure 14).



**Figure 14 : Analyses d'enrichissement GOterms des principaux interacteurs de DYRK1A.**

Les principaux interacteurs correspondent à l'ensemble des 44 protéines interagissant avec la protéine DYRK1A humaine et retrouvés dans un minimum de deux études dans la base de données BioGRID. Les analyses d'enrichissement ont été réalisées grâce à l'outil « GO Enrichment Analysis » (Mi et al., 2018) release 2021-06.

Les technologies de protéomique, notamment la spectrométrie de masse, ont joué un rôle important dans l'identification du grand nombre d'interacteurs de DYRK1A décrits à l'heure actuelle. Cependant, il faut également garder à l'esprit que ces techniques ultra-sensibles sont susceptibles de générer un certain nombre de faux-positifs. Aussi, il reste primordial de valider les interactions détectées par ces méthodes dites « high-throughput » par des moyens plus directs comme la co-immunoprécipitation suivie de Western Blot. De la même manière que pour les analyses transcriptomiques, la recherche d'interacteurs de DYRK1A fut réalisée dans différents modèles et types cellulaires. Les premières expériences protéomiques ayant identifié certains interacteurs de DYRK1A étaient des criblages de grande ampleur, visant à étudier de manière

globale l'interactome des protéines kinases dans des modèles de cellules humaines en culture (notamment HEK293) (Varjosalo et al., 2013 ; Huttlin et al., 2015). Par la suite, plusieurs études de spectrométrie de masse sur éluas de co-immunoprécipitation furent réalisées en modèles cellulaires humains mais également sur cellules et tissus murin (Table 1). Là aussi, la comparaison de ces différentes études révèle une variabilité de résultats réelle. De la même manière que pour l'étude des rôles de DYRK1A au niveau transcriptomique, ces expériences furent réalisées dans des modèles différents mais également dans des contextes différents (KO hétérozygote murin, surexpression de DYRK1A en modèle cellulaire par construction plasmidique), expliquant au moins partiellement cette variabilité. Certains facteurs présentent tout de même une affinité remarquable avec la DYRK1A comme la protéine DCAF7, retrouvée comme interacteur dans l'ensemble des expériences protéomiques réalisées à ce jour. Plus récemment, des expériences de phosphoprotéome ont pu renseigner sur les variations de phosphorylation pouvant être à l'origine d'une augmentation ou diminution d'expression de DYRK1A (Nguyen et al., 2018 ; De Toma et al., 2019 ; Recasens et al., 2021). En comparant différents tissus cérébraux (hippocampe, cortex et cervelet) de souris sur ou sous exprimant DYRK1A (Nguyen et al., 2018), l'équipe de Yann Héroult a pu mettre en évidence différentes cibles directes et indirectes de DYRK1A pouvant différer selon les régions du cerveau étudiées. L'équipe a notamment mis en évidence une action directe de DYRK1A phosphorylant la protéine SYN1 (synapsine1) sur son résidu sérine 551 et que le niveau de phosphorylation de ce résidu était proportionnel à la quantité de protéine DYRK1A présente. Plus tard, une seconde étude soulignera également une altération de la synapse dans le cas d'une surexpression de DYRK1A en modèle murin (De Toma et al., 2019). Enfin, très récemment, une troisième équipe a étudié le phospho-protéome associé à une inhibition de DYRK1A en utilisant l'inhibiteur de DYRK1A L41 (Recasens et al., 2021). Les auteurs rapportent ici un rôle prépondérant de DYRK1A dans la régulation de la prolifération passant notamment par sa capacité à phosphoryler la protéine CDK1.

## 2. Régulation de la prolifération

### a. Régulation du cycle cellulaire

Le tout premier phénotype attribué à DYRK1A fut une diminution du nombre de cellules dans le cerveau des mouches mutantes *mnb* (K.F Fishbach et M. Heisenber, 1984). Par la suite, plusieurs substrats de DYRK1A impliqués dans la régulation de la prolifération ou du cycle cellulaire furent décrits et la cycline D1 reste l'exemple le plus connu. La cycline D1 est un



régulateur majeur de la transition G1/S, la protéine est exprimée dans une fenêtre de temps très courte et se dimérise avec le complexe CDK4/6 pour permettre l'entrée de la cellule en phase S. Ainsi, DYRK1A phosphoryle la cycline D1 sur son résidu thréonine 286 et entraînant sa dégradation par le protéasome, poussant la cellule à stopper son cycle (Chen et al., 2013). DYRK1A a également été décrit comme phosphorylant la cycline D2 sur son résidu Thr280, entraînant là encore la dégradation de sa cible par le protéasome et *in fine*, une diminution de l'avancée des cardiomyocytes dans le cycle cellulaire (Hille et al., 2016). Comme vu précédemment, DYRK1A phosphoryle également la protéine LIN52, un facteur d'assemblage du complexe DREAM. Par ce biais, DYRK1A régule indirectement l'activité de DREAM et donc l'expression de gènes régulant notamment les transitions G1/S et G2/M dans la cellule (Arbones et al., 2019, Review). Plus récemment, un séquençage d'ARN totaux réalisé après inactivation de *Dyrk1a* chez *Xenopus Tropicalis* a révélé une augmentation de la transcription des ARNm codant pour les Cyclines B1 et B2 (Willsey et al., 2020), dont les expressions sont maximales durant la phase M du cycle cellulaire. En observant les embryons au stade 10 (stade d'initiation de la gastrula) traités à l'harmine (un inhibiteur pharmacologique de DYRK1A), l'équipe a également observé une proportion plus importante de cellules dans les phases S (cellules exprimant PCNA) et M (cellules exprimant la phospho-Histone 3).

#### *b. Régulation de l'apoptose*

L'apoptose est définie comme un mécanisme de mort cellulaire programmée et est parfois qualifiée de « suicide cellulaire ». Elle joue notamment un rôle prépondérant lors du développement embryonnaire. *DYRK1A* fut relié à l'apoptose pour la première fois en 2008 et identifié comme phosphorylant la CASP9 (Caspase-9) sur son résidu thréonine 125 *in vitro* dans des cellules U2OS (Seifert et al., 2008). Quelques mois plus tard, ces résultats furent confirmés par un second groupe dans les progéniteurs de la rétine chez la drosophile et la souris (Laguna et al., 2008). Cette phosphorylation permet d'inhiber le clivage de la CASP9 empêchant l'activation de cette dernière et inhibant ainsi le clivage de la CASP3 (Caspase-3) et donc la mort cellulaire par la voie apoptotique. Parmi les cibles de DYRK1A, on retrouve le suppresseur de tumeur SIRT1 (Sirtuin 1), une NAD désacétylase. La phosphorylation de SIRT1 sur son résidu thréonine 522 stimule son activité et conduit à la désacétylation de p53 et permet une inhibition de l'apoptose. De ce fait, une diminution d'expression ou d'activité de DYRK1A conduit à une mort cellulaire par la voie p53 (Guo et al., 2010). Plus récemment et de manière surprenante, une équipe a surexprimé DYRK1A par transduction dans des lignées  $\beta$  pancréatiques et a pu observer une apoptose plus importante de ces cellules. De la même

manière, l'inactivation partielle de *Dyrk1a* par ARN interférence a révélé une atténuation de l'activité des caspases-3 et 8 dans ces cellules  $\beta$ , suggérant cette fois-ci que DYRK1A stimulerait l'apoptose (Lu et al., 2019). Par immunoprécipitation de DYRK1A et un traitement à l'harmine des cellules pancréatiques, les auteurs ont pu montrer que DYRK1A interagissait avec le facteur IRS2 (insuline Receptor-Substrate 2), augmentait son niveau phosphorylation, conduisant à une dégradation de IRS2 par le protéasome. Cette dégradation menant à une entrée en apoptose des cellules  $\beta$ .

### c. Synthèse

Au vu du phénotype de microcéphalie présent chez les patients mais également retrouvé dans différents modèles animaux porteurs de mutations dans *Dyrk1a* (Drosophile, souris, rat, poisson zèbre, xénope), la protéine fut impliquée très tôt dans les phénomènes d'activation ou de maintien de la prolifération. Cependant, une vue plus globale de la littérature concernant l'implication de *DYRK1A* dans la prolifération cellulaire révèle un rôle complexe de la protéine dépendant des organismes, types cellulaires et de la modification de son dosage. Les effets d'une modification du dosage de DYRK1A dépendent du type cellulaire utilisé, du fond génétique et de l'environnement de culture des cellules (Table 2). Par exemple, une augmentation de l'expression de DYRK1A conduit à une stimulation de la prolifération des cellules cardiaques murines (Hille et al., 2016) et DYRK1A fut retrouvé comme surexprimé dans différentes lignées cancéreuses et son inhibition peut conduire à une réduction de la capacité proliférative de ces cellules (Radhakrishnan et al., 2016)

L'inactivation partielle de *Dyrk1a* conduit à l'inverse à une diminution de la capacité proliférative dans d'autres modèles (progéniteurs hippocampaux de rat, neuroblastome murins N2A) (Yabut et al., 2010 ; Park et al., 2010). On notera cependant que dans la majorité des cas, l'inactivation partielle de DYRK1A conduit à une diminution de la capacité proliférative.

<b>Auteurs</b>	<b>Expression</b>	<b>Conséquences sur la prolifération</b>	<b>Modèles</b>
Chakrabarti 2007	+	DYRK1A + inhibe prolifération	Souris Ts65Dn
Contestabile 2007	+	DYRK1A + inhibe prolifération	Souris Ts65Dn ; fœtus humains
Park 2010	+	DYRK1A + inhibe prolifération	Lignée H19-7
Ishihara 2010	+	DYRK1A + inhibe prolifération	Souris Ts1Cje et Ts2Cje
Yabut 2010	+	DYRK1A + inhibe prolifération	Lignée N2A ; cerveaux de souris électroporés in utero
Soppa 2014	+	DYRK1A + inhibe prolifération	Lignées SH-SY5Y
Lee 2016	+	DYRK1A et DYRK1B + inhibe prolifération	Lignées de glioblastomes humain, NSC murines
Shen 2015	- et +	DYRK1A + inhibe prolifération / DYRK1A - stimule prolifération	CP CβP de rat ; îlots CβP primaires humaines
Litovchick 2011	- et +	DYRK1A + inhibe prolifération / DYRK1A – stimule l'entrée en phase S	NIH3T3, T98G, U2OS, DW1990
Hämmerle 2011	- et +	DYRK1A + active prolifération / DYRK1A - inhibe prolifération	PC12, ME poulet, télencéphale murin (Harmine ou EGCG)
Pozo 2013	+	DYRK1A + stimule prolifération	Lignée de glioblastomes issues de patients
Radhakrishnan 2016	+	DYRK1A + stimule prolifération	Lignée HNSCC
Hille 2016	+	DYRK1A + stimule prolifération	Souris Dyrk1A MHC
Kim 2021	/	DYRK1A + stimule prolifération	Cellules B de patients DS leucémiques
Chen 2013	+	DYRK1A + et - inhibe prolifération	Lignées fibroblastes de patients Down Syndrome
Fischback & Heisenberg 1984	-	DYRK1A – diminue le pool de cellules	Drosophile mutante mnb +/-
Tejedor 1995	-	DYRK1A - inhibe prolifération	Drosophile mutante mnb +/-
Laguna 2008	- et +	DYRK1A - inhibe prolifération (apoptose)	PRM murins (Dyrk1A +/- et tgYAC152f7)
Ferron 2010	-	DYRK1A - inhibe prolifération	Cultures de neurosphère murines Dyrk1a +/-
Rachdi 2014	-	DYRK1A - inhibe prolifération	Souris DYRK1A +/- ; CP de cellules pancréatiques murines
Barallobre 2014	- et +	DYRK1A - inhibe prolifération (mort cellulaire)	Souris Dyrk1a +/- et mBACtgDyrk1a
Wang 2015	-	DYRK1A - stimule prolifération	Lignée CβP humaine traitée à l'aminopyrazine
Chen 2021	-	DYRK1A - stimule prolifération	Lignée MMK1 traitée par siRNA DYRK1A
Recasens 2021	-	DYRK1A - (60%) stimule prolifération ; DYRK1A - arrête prolifération	U251 traités par shRNA DYRK1A ou avec la L41

**Table 2 : Conséquences des variations d'expression de DYRK1A sur la prolifération cellulaire dans différents modèles d'études.** Ts65Dn : lignée murine de trisomie partielle pour 132 gènes de la partie distale des chromosomes MMU16 et MMU17 ; Ts1Cje et TS2Cje: lignées murine de trisomie partielle dont *Dyrk1a* ; tgYAC152f7 : lignée murine de trisomie pour les gènes *TTC3*, *DYRK1A* et *KCNJ6* ; lignée murine mBACtgDyrk1a : comporte trois copies du gène *Dyrk1a* ; lignée murine Dyrk1A MHC : surexpression conditionnelle *Dyrk1a* dans le tissu cardiaque ; H19-7 lignée de progéniteurs hippocampaux immortalisés de rat ; N2A : lignée de neuroblastome de souris ; PC12 : lignée de phéochromocytome de rat ; SH-SY5Y : lignée de neuroblastome humain ; NIH3T3 : lignée de fibroblastes embryonnaires murins ; T98G, MMK1 et U251 : lignées de glioblastome humain ; U2OS : lignée d'ostéosarcome humain ; SW1990 : lignée de cancer du pancréas humain ; HNSCC : lignée de carcinome épidermoïde de la tête et du cou humain ; NSC : cellules souches neurales ; PRM : Progéniteurs de Rétine Murins ; C $\beta$ P : Cellules  $\beta$  pancréatiques ; DS : Syndrome de Down ; mnb : mini brain ; ME : Moëlle Epinière ; CP : Cultures Primaires ; MO : Morpholinos ; cKO : Knock Out conditionnel ; Harmine, EGCG, L41 : Inhibiteurs pharmacologiques de DYRK1A.

### 3. Régulation de l'épissage alternatif

Parmi les cibles et interacteurs de DYRK1A, on retrouve également plusieurs RBP (RNA Binding Protein) régulant l'épissage alternatif. La première implication de DYRK1A dans ce processus fut mise en évidence en 2006. Par surexpression de DYRK1A en cellules COS-7 via un plasmide, l'équipe rapporte que le facteur SF3B1 est une cible directe de DYRK1A, phosphorylé sur son résidu thréonine 434 (Graaf et al., 2006). En plus de phosphoryler la protéine MAPT (TAU), DYRK1A inhibe indirectement l'inclusion de l'exon 10 de *MAPT* par quatre mécanismes : 1. en phosphorylant le facteur d'épissage ASF sur ses résidus sérine 227, 234 et 238 (Shi et al., 2008) ; 2. En phosphorylant le facteur SC-35 (SRFS2) (Qian et al., 2013) ; 3. En phosphorylant le facteur SRp55 conduisant à une incapacité de ce dernier à se localiser dans les « nuclear speckles » (Yin et al., 2012) ; 4. En phosphorylant le facteur 9G8 conduisant à sa localisation nucléaire (Ding et al., 2012). Cette inhibition d'inclusion de l'exon 10 de *TAU* conduit à la régulation de la balance d'expression entre les isoformes 3R et 4R de MAPT dans le cerveau, balance dont la dérégulation est suspectée de conduire à différentes Tauopathies.

## 4. Régulation de la transcription

Le rôle de DYRK1A dans la régulation de la transcription est également avéré et différents modes d'action ont pu être découverts à ce jour. En effet, DYRK1A présente un grand nombre de cibles pouvant réguler l'expression génique à différents niveaux en agissant par exemple sur la structure de la chromatine, en modifiant l'activité de facteurs de transcription, en interagissant avec l'ARN polymérase II ou encore en modifiant certaines marques épigénétiques.

### a. Remodelage de la chromatine

L'ensemble des régulations ou modifications au niveau de la chromatine opérées par DYRK1A sont indirectes. Le premier facteur de remodelage interagissant avec DYRK1A mis en évidence est la protéine Arip4 (androgen receptor-inter-acting protein 4), une ADN hélicase codée par un gène paralogue d'*ATRX* (une autre hélicase ADN associée à un syndrome de DI liée à l'X) et notamment impliqué dans la régulation de l'expression de l'AR (Récepteur aux Androgènes) (Sitz et al., 2004). De manière intéressante, DYRK1A est capable d'induire la cascade transcriptionnelle en activant l'AR via son interaction avec ARIP4 mais également indépendamment de son activité kinase. Plus tard, en utilisant un modèle murin comportant un YAC exprimant les versions humaines des gènes *TTC3*, *KCNJ6* et *DYRK1A*, une équipe met en évidence d'une part : une dérégulation de l'expression de gènes régulés par les facteurs REST/NRSF (REST/neuron-restrictive silencer factor) et, d'autre part, une liaison au complexe SWI-SNF (un complexe de remodelage de la chromatine connu pour interagir avec REST/NRSF) (Lepagnol Bestel et al., 2009). Les auteurs souligneront que la surexpression du gène *DYRK1A* humain conduit dans le modèle murin à une augmentation du nombre de transcrits *Rest/Nrsf*, modifiant ainsi l'équilibre d'expression des gènes cibles et entraînant une diminution de la taille des neurites chez les animaux.

### b. Marques épigénétiques

L'implication de DYRK1A dans la régulation de marques épigénétiques est récente mais deux mécanismes ont aujourd'hui été démontrés. Le premier concerne la phosphorylation de l'histone H3 par DYRK1A (Jang et al., 2014). Les auteurs utilisent un système de surexpression de DYRK1A dans un modèle de fibroblastes murins mettant en évidence la phosphorylation de l'histone H3 sur ses résidus thréonine 45 et sérine 57 par DYRK1A dans des régions promotrices de gènes régulés par la protéine HP1 (heterochromatin protein 1, un

répresseur transcriptionnel). Ces phosphorylations inhibent le recrutement de HP1 au niveau de ces séquences et entraînent une ouverture de la chromatine conduisant *in fine* à une activation transcriptionnelle des gènes concernés dont plusieurs sont impliqués dans la réponse immunitaire. Récemment, DYRK1A fut relié à l'action de deux histones acétyl transférases : p300 et CBP (ou CREBBP) (Li et al., 2018). Les auteurs mettent en évidence une interaction directe, ainsi qu'une hyperphosphorylation de ces deux protéines en cas de surexpression de DYRK1A dans une lignée HEK293. De plus, un Chip-Seq réalisé dans une lignée T98G démontre la présence simultanée des trois protéines sur 232 régions du génome. Enfin, l'équipe montre que l'interaction de DYRK1A avec p300 et CBP permet la régulation transcriptionnelle de plusieurs gènes (*DNAJC12*, *WNT11*, *DYNCH1*, ...) via acétylation de la lysine 27 de l'histone H3 (H3K27).

### *c. Facteurs de transcription*

Même si DYRK1A est impliqué dans des mécanismes de remodelage de la chromatine et de modification épigénétique, son rôle de régulateur transcriptionnel passe également par son interaction avec différents facteurs de transcription. DYRK1A a notamment été montré comme phosphorylant CREB1 (cAMP Response Element-Binding Protein 1) et stimule son activité transcriptionnelle, induisant la différenciation de progéniteurs hippocampaux (H19-7) (Yang et al., 2001). DYRK1A active également l'activité transcriptionnelle de GLI1 en entraînant sa rétention dans le noyau (Mao et al., 2002). Ce mécanisme passe par la phosphorylation de GLI1 sur son résidu sérine 408 situé dans une région NLS, provoquant sa rétention dans le noyau (Ehe et al., 2017). GLI1 est un effecteur agissant en aval de la voie Hedgehog, une voie de régulation clef du développement embryonnaire notamment connue pour son implication dans le développement des membres dans laquelle le rôle de GLI1 est de réguler l'expression de gènes impliqués dans la prolifération et dans le déterminisme de différents types cellulaires. Un autre facteur de transcription dont l'interaction avec DYRK1A est bien caractérisée est la protéine NFAT (Nuclear factor of activated T-cells). En 2006 une première équipe suggère que la surexpression de *DYRK1A* bloque la voie calcineurine-NFAT (Arron et al., 2006) en réduisant l'activité du facteur de transcription et cette inhibition par phosphorylation sera confirmée en 2009 (Lee et al., 2009 ; Kuhn et al., 2009). Aujourd'hui, plusieurs protéines NFAT codées par le même gène diffèrent par leur partie C-terminale (NFATc1, NFATc2, NFATc3 et NFATc4) et DYRK1A augmente la translocation nucléaire de NFATc1 ainsi que sa stabilité en diminuant son ubiquitination par phosphorylation sur ses résidus sérines 261, 278, 403 et 409 mais possède l'effet inverse sur de NFATc2 (Liu et al., 2017). Initialement décrit comme

activateur de la transcription dans les cellules T, le rôle de NFAT dans le neurodéveloppement par la voie calcineurine est notamment nécessaire à la croissance axonale et à la formation des synapses (Nguyen et Giovanni, 2008). Le facteur REST est également un interacteur connu de DYRK1A. La protéine REST (RE1 Silencing Transcription Factor) se fixe sur les sites RE-1 (Responsive Element) et joue un rôle de répresseur transcriptionnel, modulant l'expression de gènes codants pour des protéines synaptiques, des neurotransmetteurs ou encore des canaux ioniques. REST est un activateur transcriptionnel de *DYRK1A*, mais DYRK1A régule également négativement la stabilité du facteur REST en le phosphorylant et conduisant à sa dégradation par le protéasome (Lu et al., 2011).

#### *d. DYRK1A et ARN polymérase II*

En 2015, une étude rapporte pour la première fois une interaction directe entre le domaine His de DYRK1A et l'ARN polymérase II (Di Vona et al., 2015). Dans cet article les auteurs proposent une phosphorylation du domaine CTD de l'ARN pol II sur ses résidus sérines 2 et 5 (aa régulant l'activation de l'ARN pol II) par DYRK1A. Après analyse d'un Chip-seq réalisé en cellules HeLa et T98G, l'équipe propose également que la protéine serait recrutée sur des séquences promotrices comportant une séquence « TCTCGCGAGA ». Très récemment, dans un preprint très récent (Di Vona et al., 2021, BioRxiv), cette même équipe suggère après une analyse plus en profondeur des données Chip-Seq précédentes, un rôle majeur de DYRK1A dans la régulation de la transcription de plusieurs protéines ribosomales et donc, un impact indirect sur la traduction cellulaire globale. A l'heure actuelle, aucune autre étude n'a validé ces régulations transcriptionnelles, sans doute en raison des types cellulaires utilisés lors des expériences de Chip-Seq. L'interaction directe de DYRK1A avec l'ARN pol II fut également rapportée par analyse de spectrométrie de masse et Immunoprécipitations en cellules HeLa (Yu et al., 2019). Les auteurs proposent un modèle dans lequel DCAF7 stabilise DYRK1A sous la forme d'un complexe nucléaire permettant aux deux protéines de se fixer à l'ARN pol II et de promouvoir une activité transcriptionnelle conduisant à la différenciation d'une lignée de myoblastes (C2C12) par le biais de l'expression de gènes myogéniques.

## 5. Autres rôles de DYRK1A

Les actions de DYRK1A dans les processus cellulaires sont diverses et variées. En plus de ses implications dans la régulation du cycle cellulaire, de l'apoptose, de l'épissage alternatif, du remodelage de la chromatine, des marques épigénétiques et de l'activité transcriptionnelle,

DYRK1A possède également des cibles et interacteurs impliqués dans d'autres processus tels que : le métabolisme car DYRK1A phosphoryle par exemple la protéine CRY2 régulant partiellement les oscillations circadiennes moléculaires (Kurabayashi et al., 2010) ; Le cytosquelette neuronal avec comme vu précédemment une régulation de la balance des isoformes 3R et 4R de MAPT (TAU), mais également une phosphorylation de la  $\beta$ -tubuline (Mc Kenney et al., 2016), de SEPT4 (Septine 4)(Sitz et al., 2008) et d'APP (Ryoo et al., 2008) ; La fonction synaptique par phosphorylation de DNM1 (Dynamine 1) (Huang et al., 2004), de l'AMPH1 (Amphiphysine-1) (Murakami et al., 2006) de SYNT1 (Synaptojanine 1) (Chen et al., 2014), du récepteur au glutamate GluN2A (Grau et al., 2014) ou encore la régulation du transport axonal de certaines protéines comme APP (Bessone et al., 2021, Preprint BioRxiv).

## 7) Modèles animaux de DYRK1A

Au vu du rôle majeur de DYRK1A dans la cellule mais également en raison de son implication en pathologie humaine, plusieurs modèles animaux furent générés dans le but d'étudier les conséquences de son inactivation ou de sa surexpression.

### 1. *Drosophila melanogaster*

La première mouche (*Drosophila melanogaster*) mutante porteuse d'une inactivation d'un des allèles du gène *mnb* codant pour l'homologue de DYRK1A fut générée en 1984 et présentait une diminution de la complexité des lobes optiques ainsi que de la taille globale du cerveau (K.F Fishbach et M. Heisenber 1984). L'inactivation des deux allèles *mnb* chez la mouche est létale avant le stade « pupes ». Ces observations primaires seront validées plus tard par d'autres équipes qui affineront également le phénotype de ces mutants en mettant par exemple en évidence une diminution du nombre de neurones due à une baisse de prolifération des cellules progénitrices (Tejedor et al., 1995). Le phénotype sera également complété par une diminution globale de la taille du corps au moins partiellement due au rôle de régulation transcriptionnelle de *mnb* dans la voie Yki (Degoutin et al., 2013), une voie majeure d'activation de la transcription chez la mouche conduisant à l'expression de gènes de croissance et de prolifération cellulaire. Enfin, une étude plus récente met en évidence que dans les précurseurs neuronaux, *mnb* a pour rôle d'entraîner une sortie du cycle cellulaire en régulant le facteur Dacapo (l'homologue de *CDKN1B*), facilitant la différenciation des cellules en neurones (Shaikh et al., 2016). Les auteurs rapportent chez les mouches mutantes *mnb* une prolifération



mais également une augmentation de la mort cellulaire, expliquant la diminution de la taille du cerveau.

## 2. Mus Musculus

### a. Dans le cadre d'une surexpression de *Dyrk1a*

Le modèle souris reste le modèle animal le plus utilisé dans l'étude du rôle de *DYRK1A*. Il fut utilisé dans un premier temps pour investiguer les conséquences de sa surexpression dans le contexte du SD dont les patients sont porteurs de trois copies fonctionnelles de *DYRK1A*. Plusieurs modèles murins porteurs d'insertions de tailles variables furent générés. Certains comportent en effet la totalité des gènes homologues humains de la DSCR-1 (modèle Ts65Dn), d'autres comportent en trois exemplaires quelques gènes seulement comme le modèle Tg(CEPHY152F7)12Hgc créé par insertion d'un YAC contenant les gènes *TTC3*, *DYRK1A* et *KCNJ6* humains. Enfin, le modèle Tg(*Dyrk1a*)189N3Yah généré à l'ICS (Institut Clinique de la Souris), comporte uniquement l'insertion d'une troisième copie du gène *Dyrk1a* murin (Guedj et al., 2012). De manière très synthétique, la surexpression de *Dyrk1a* dans le modèle murin entraîne des anomalies au niveau cérébral (pas de micro ou macrocéphalie, ventricules cérébraux et hippocampe plus grands), au niveau moléculaire (dégénérescence des neurones cholinergiques, niveaux d'APP et de MAPT (TAU) hyperphosphorylé augmentés), mais également au niveau comportemental (troubles de l'apprentissage, troubles de la mémoire à long terme et de la mémoire spatiale, troubles cognitifs) récapitulant certains phénotypes cognitifs retrouvés chez les patients DS et soulignant l'importance de *DYRK1A* dans ce syndrome et comme cible thérapeutique potentielle. En effet, plusieurs études ont pu mettre en évidence la correction de différents phénotypes des souris Ts65Dn en inhibant *DYRK1A* (Laguna et al., 2013 ; Jiang et al., 2015). Une des plus récentes a pu montrer une correction partielle des phénotypes cognitifs et moléculaires des souris Ts65Dn grâce à l'utilisation d'une molécule pharmacologique inhibant l'activité kinase de *DYRK1A* : le F-DANDY (hydroxy derivatives of 3,5-diaryl-7-azaindoles) (Neumann et al., 2018).

### b. Dans le cadre diminution d'expression de *Dyrk1a*

Des modèles murins présentant une diminution d'expression de *Dyrk1a* furent également générés (Table 3) pour mieux comprendre le rôle de la kinase, notamment sur le développement du cerveau. On sait que les souris double mutantes (*Dyrk1A* *-/-*) présentent des

retards de développement et de croissance avec une réduction globale de la taille du corps. Cette double mutation est également létale à l'état embryonnaire entre E10,5 et E13,5 en raison des défauts de maturation du système nerveux (Fotaki et al., 2002). Les souris porteuses d'une mutation perte de fonction dans *Dyrk1A* à l'état hétérozygote (*Dyrk1A +/-*) sont biologiquement viables et présentent plusieurs phénotypes cellulaires. Au niveau macroscopique, on retrouve chez les souris mutantes une microcéphalie primaire ainsi qu'un retard de croissance post-natal systématique, des troubles dans les fonctions neuromusculaires chez les souriceaux (Fotaki et al., 2002 ; Arranz et al., 2019) et une susceptibilité accrue à l'épilepsie induite par le froid ou le PTZ (Souchet et al., 2014 ; Raveau et al., 2018 ; Arranz et al., 2019). Les analyses d'imagerie et histologiques mettent en évidence une augmentation de la taille des ventricules et une diminution ou absence du corps calleux (Levy et al., 2021). On notera que ces atteintes macroscopiques résument la symptomatologie retrouvée chez les patients.

Au niveau histologique ces souris hétérozygotes montrent une diminution de l'arborisation dendritique dans les neurones pyramidaux (Benavides-Piccine et al., 2005) et de la formation globale de dendrites et d'épines dendritiques (Dang et al., 2018). Une altération de la balance Glutamate/GABA fut observée, avec une augmentation de la proportion de neurones gabaergiques et de synapses excitatrices dans le néocortex (Arranz et al., 2019). Ces observations semblent confirmer l'effet protecteur de *DYRK1A* dans la susceptibilité à l'épilepsie, *DYRK1A* inhibant le développement de neurones glutamatergiques. Enfin, de nombreux phénotypes comportementaux ont également pu être mis en évidence tels qu'une hypoactivité générale (Fotaki et al., 2004), une diminution de la mémoire de travail (Arque et al., 2008), des changements dans les vocalisations (Scattoni et al., 2008) ainsi qu'une baisse de la sociabilité (Raveau et al., 2018) rappelant certains éléments retrouvés chez les patients *DYRK1A* atteints de TSA.

<b>Auteurs</b>	<b>Génotypes</b>	<b>Phénotype</b>
Fotaki 2002	Dyrk1a +/- et -/- constitutif	Létalité embryonnaire (-/-) ; microcéphalie (+/-) ; comportement anxieux (+/-)
Fotaki 2004	Dyrk1a +/- constitutif	Défauts du développement moteur, hypoactivité
Benavides-Piccione 2005	Dyrk1a +/- constitutif	Neurones pyramidaux plus petits et moins branchés
Arque 2008	Dyrk1a +/- constitutif	Altérations de la mémoire spatiale
Arque 2009	Dyrk1a +/- constitutif	Altérations musculaires (traction, réflexe préhensile), diminution du CA1-CA3
Ferrón 2010	Dyrk1a +/- constitutif	Diminution de la prolifération, augmentation du nombre de cellules apoptotiques
Guedj 2012	Dyrk1a +/- constitutif	Microcéphalie, défauts dans les structures cérébrales et couches corticales
Barallobre 2014	Dyrk1a +/- constitutif	Microcéphalie, moins de neurones, apoptose des neurones
Rachdi 2014	Dyrk1a +/- constitutif	Intolérance au glucose, diminution de la prolifération des cellules Beta
Souchet 2014	Dyrk1a +/- constitutif	Susceptibilité à l'épilepsie
Thompson 2015	Dyrk1a +/- constitutif	Promotion des cellules en quiescence
Raveau 2018	Dyrk1a +/- constitutif	Défauts de cognition, vocalisation, sociabilité ; susceptibilité à l'épilepsie
Arranz 2019	Dyrk1a +/- constitutif	ASD, épilepsie, augmentation de la densité des neurones, défaut dans la balance synapses inhibitrices/excitatrices
Levy 2021	Dyrk1a +/- et -/- cKO cortex	Microcéphalie, diminution taille des soma, ASD, hypophosphorylations (mTOR, S6, ERK1, ERK2)

**Table 3 : Récapitulatif des principaux modèles murins présentant une diminution d'expression de Dyrk1a.**

### 3. Autres modèles : poulet, poisson zèbre et xénope

Le rôle de DYRK1A fut également étudié dans d'autres modèles de vertébrés. Par méthode d'électroporation *in ovo* en 2011. Hämmerle et al. ont étudié les conséquences d'une surexpression ou inhibition de *Dyrk1a* sur la prolifération cellulaire dans la moelle épinière d'embryons de poulet (*Galus galus*). En surexprimant DYRK1A par le biais d'une construction plasmidique, les auteurs rapportent un effet antiprolifératif et un nombre de cellules apoptotiques inchangé par rapport à la condition sauvage. En revanche, lors de l'utilisation d'harmine, les auteurs constatent un effet hyperprolifératif ainsi qu'une augmentation importante du nombre de cellules apoptotiques. Cette étude rapporte également une sortie forcée du cycle cellulaire par induction de *p27* (ou *CDKN1B*) en cas de surexpression ou d'inhibition de DYRK1A, conduisant à une entrée précoce dans le processus de différenciation neuronale (Hämmerle et al., 2011).

Plus récemment, un modèle poisson zèbre (*Danio rerio*) porteur d'une inactivation de l'orthologue de *DYRK1A* (*Dyrk1aa*) à l'état homozygote fut généré (Kim et al., 2017). Les poissons mutants présentent notamment une microcéphalie mais les auteurs rapportent un certain nombre d'altérations comportementales telles que de l'anxiété et une diminution de l'interaction et de la cohésion sociale. Enfin, par méthode d'ISH, Kim et al. démontre une diminution de *c-fos* (un marqueur de l'activation neuronale). Finalement, en 2020, un modèle xénope (*Xenopus tropicalis*) porteur d'une perte de fonction de *dyrk1a* à l'état hétérozygote fut généré (Willsey et al., 2020). Les auteurs rapportent une expression forte de l'ARNm *dyrk1a* dans les cellules en prolifération durant la neurogénèse. Les animaux mutants présentent une microcéphalie et notamment une diminution de la taille du télencéphale. Leur étude a permis de mettre en évidence une ciliogénèse altérée chez les mutants, potentiellement causée par un défaut de la dynamique des microtubules. De plus, les auteurs ont également réalisé une étude transcriptomique en comparant les animaux sauvages aux mutant *dyrk1a*, cette expérience a révélé 294 gènes différentiellement exprimés avec un enrichissement en GOterms pour la réplication de l'ADN et le cycle cellulaire. Ces implications sont redondantes avec les fonctions déjà connues de DYRK1A et confirment son rôle prépondérant dans le développement chez les vertébrés.



### III. La DI liée à l’X causée par des mutations du gène PQBP1

Le gène *PQBP1* code pour la « Polyglutamine Binding Protein 1 » et est situé chez l’Homme sur le chromosome X en position Xp11.23. *PQBP1* est impliqué en pathologie humaine notamment dans un syndrome neurodéveloppemental : le Syndrome de Renpenning. *PQBP1* est également suspecté de jouer un rôle dans certains syndromes neurodégénératifs tels que la maladie d’Alzheimer ou la Chorée de Huntington.

#### 1) Historique et clinique

Le syndrome de Renpenning (SR) (ou syndrome : Sutherland-Haan ; Hamel ; Porteous ; Golabi-Ito-Hall ; MIM300463) est une forme de DI syndromique mise évidence par Hans Renpenning en 1962 (Renpenning et al., 1962). Les auteurs rapportent dans cette étude 20 patients issus d’une grande famille avec un retard mental « confiné aux hommes et transmis par des femmes normales », suggérant un mode de transmission lié à l’X. Les causes moléculaires du SR seront mises en évidence 40 ans plus tard par l’identification de variations perte de fonction dans le gène *PQBP1* (Kalscheuer et al., 2003). Dans cet article, les auteurs ont pu décrire 21 patients de sexe masculin répartis dans 5 familles et tous porteurs d’une variation de type insertion ou délétion d’un dinucléotide AG entraînant un décalage du cadre de lecture. Ils mettent de ce fait en évidence un hotspot mutationnel correspondant à une répétition de 6 dinucléotides AG (AG<sub>6</sub>) dans l’exon 4. Par la suite, d’autres variants conduisant à des décalages du cadre de lecture situés en dehors du hotspot AG<sub>6</sub> seront décrits (Lenski et al., 2004 ; Stevenson et al., 2005 ; Cossée et al., 2006 ; Lubs et al., 2006 ; Martínez-Garay et al., 2007 ; Germanaud et al., 2011 ; Hu et al., 2016). Jusqu’à récemment, un seul variant faux-sens (Tyr65Cys) était rapporté chez 3 patients d’une même famille (Lubs et al., 2006). La modification de ce résidu tyrosine très conservé dans le domaine WW de PQBP1 entraîne la perte d’interaction entre PQBP1 et un facteur d’épissage (WBP11 alias SIPP1) ainsi que les potentielles régulations médiées par ce domaine (Mizugochi et al., 2016). Plus récemment, deux autres variants faux-sens ont pu être mis en évidence (Arg243Trp et Pro244Leu) cette fois-ci dans le domaine C-terminal de la protéine (Redin et al., 2014 ; Hu et al., 2016). Si les prédictions *in silico* plaident en faveur d’une pathogénicité de ces variants, seule la pathogénicité du variant p.Pro244Leu fut démontrée. En effet, le changement Pro244Leu entraîne incapacité de liaison entre la protéine PQBP1 et le facteur U5-15kDa (Liu et al 2020).

Un patient porteur d'une duplication de 4,7Mb comprenant plusieurs dizaines de gènes dont *PQBP1* fut également rapporté en 2011 (Flynn et al., 2010) et il existe à l'heure actuelle près d'une quarantaine de cas recensés sur la base DECIPHER (Firth et al., 2009) dont beaucoup associés à des TND mais aucun comportant une duplication de *PQBP1* seul. Les conséquences d'une surexpression de *PQBP1* chez l'Homme restent de ce fait aujourd'hui inconnues. Le SR est un syndrome ultra rare avec seulement une trentaine de familles recensées dans la littérature. Les patients atteints du SR présentent des manifestations cliniques communes avec notamment : une microcéphalie primaire, un retard de croissance, une dysmorphie faciale (face longue et triangulaire, sourcils épars, oreilles décollées, ...), une maigreur et une petite taille. On notera également la présence de traits phénotypiques variables comme une atrophie musculaire, des dysfonctions du palais ainsi que des anomalies génitales (Germanaud et al., 2011 ; Abdled-Salam et al., 2018 ; Mameesh et al., 2019 ; Masih et al., 2020). Malheureusement ces caractéristiques cliniques restent variables et sont retrouvées dans un très grand nombre de syndrome de DI génétiques. Elles ne permettent pas de ce fait d'orienter le diagnostic lors de la consultation.

## 2) Expression et domaines de la protéine PQBP1

Le gène *PQBP1* code pour la Poly Glutamine Binding Protein 1 (aussi connue sous le nom de NPW38) et comprend 6 exons codants et 10 variants d'épissage ont actuellement été décrits. Le transcrit de référence (NM\_001032381.1) code pour une protéine longue de 265aa pour une masse moléculaire d'environ 38kDa. La protéine comporte un domaine C-terminal conservé chez les vertébrés mais sa séquence est unique et il n'existe pas d'homologues décrits (Okazawa et al., 2001). *PQBP1* est exprimé de manière ubiquitaire (Waragai et al., 1999) avec une prépondérance pour les tissus nerveux, en particulier le cortex et l'hippocampe. *PQBP1* est en particulier fortement exprimé dans les progéniteurs corticaux murins (ref LI 2013). Une équipe chinoise a pu démontrer une régulation transcriptionnelle de *Pqbp1* murin par le facteur SOX2 (Li et al., 2013). En effet, plusieurs séquences de fixation de SOX2 sont présentes dans la région en aval du gène *Pqbp1* et cette régulation a pu être validée par des expériences de gènes rapporteurs codant pour la luciférase. Concernant sa localisation, la protéine est décrite comme largement nucléaire après IF en culture primaire de neurones murins (Wang et al., 2013) mais semble également localisée au niveau des granules de stress cytoplasmiques des rhabdomères chez la drosophile (Wan et al., 2015) et dans les corps nucléaires des cellules de mammifères (Okazawa et al., 2002). Plusieurs domaines de PQBP1 ont pu être décrits (Figure 11). La protéine longue de 265aa comporte un domaine WW (aa 48-82) qui permet l'interaction avec

des régions riches en prolines présentes dans de nombreuses protéines (Sudol et al., 1996 ; Macias et al., 2002). PQBP1 comporte également un domaine PRD (Polar Riche aa Domain ; aa 104-163) (Okazawa et al., 2001) dont le rôle est d'interagir avec les régions polyglutamine d'autres protéines (Imafuku et al., 1998 ; Okazawa et al., 2002). Une séquence NLS (aa 175-192), facilitant sa translocation dans le noyau et reconnue par le récepteur karyopherin $\beta$ 2 (Kap $\beta$ 2) est également présente (Lee et al., 2006). Enfin, la région C-terminale de PQBP1 baptisée CTD (C-terminal Domaine ; aa 192-265) confère à la protéine une capacité d'interaction avec d'autres facteurs. D'autres cibles et évènements de régulation ont pu être reliés à PQBP1 et actuellement, 65 interacteurs sont décrits sur la base de données BioGRID. Les rôles de PQBP1 concernent tout particulièrement la régulation de la transcription et de l'épissage alternatif et plusieurs modèles d'études ont à ce jour été utilisés pour comprendre ses implications (Table 4).



**Figure 15: Représentation schématique de la protéine PQBP1 et de ses domaines.** Les positions en acides aminés sont indiquées en dessous.



<b>Auteurs</b>	<b>Modèles</b>	<b>Conséquences phénotypiques et moléculaires</b>
Waragai 2000	Double hybride, levure	PQBP1 interagit via son domaine CTD avec U5-15kDa
Okazawa 2002	HEK293	PQBP1 interagit avec l'ARN pol II en cours de transcription
Llorian 2004	COS-1, HeLa, HEK293	SPP1 interagit avec le domaine WW de PQBP1
Qi 2005	Souris	<i>Pqbp1</i> est exprimé dans le SNC avec un pic d'expression à la naissance
Ito 2009	Souris cKO <i>Pqbp1</i> +/-	Augmentation d'expression de c-fos et de l'acétylation de l'histone H3
Takahashi 2009	<i>C. elegans</i> (shRNA)	PQBP1 est impliqué dans le métabolisme lipidique
Tamura 2010	<i>D. Melanogaster</i> <i>dPQBP1</i> +/-	Diminution de la durée de vie et des capacités d'apprentissage
Wang 2013	CP neurones murins ; shRNA <i>PQBP1</i>	<i>Pqbp1</i> régule l'apoptose par épissage de Bcl-X via son interaction avec SF3B1
Li 2013	NSPCs murins	Sox2 régule partiellement l'expression de <i>Pqbp1</i>
Iwasaki 2014	<i>Xenopus laevis</i>	Diminution d'expression de <i>sox2</i> ; Défauts de neurodéveloppement
Ito 2015	Souris <i>Pqbp1</i> +/- cKO (nestin-Cre)	Microcéphalie, allongement du cycle cellulaire des NSPC
Yang 2020	Souris <i>Pqbp1</i> +/- cKO (nestin-Cre)	Diminution de : taille, masse osseuse, nombre de chondrocytes
Shen 2020	SK-N-BE, CH murines cKO <i>Pqbp1</i> +/-	PQBP1 régule positivement la synthèse protéique en interagissant avec eEF2
Liu 2020	HeLa	PQBP1 transporte U5-15kDa au noyau et le variant P244L inhibe cette interaction

**Table 4 : Principaux modèles d'études de PQBP1**

COS-1 : lignée de fibroblastes rénaux de singe immortalisés par le virus SV-40 ; HeLa : lignée du cancer du cancer du col de l'utérus humain ; HEK293 : lignée embryonnaire de rein humain ; NSPCs : cellules souches progénitrices neurales ; SK-N-BE : lignée de neuroblastome murin ; CP : Culture Primaire ; CH : Coupes Hippocampales.

### 3) PQBP1 et régulation de la transcription

On sait que PQBP1 est capable de se lier à l'ARN Pol II et semble avoir une affinité plus importante lorsque cette dernière est phosphorylée, indiquant une interaction préférentielle en cours d'élongation de transcription (Okazawa et al., 2002). Les auteurs ont également rapporté que l'interaction de PQBP1 via son domaine CTD avec l'Ataxin1 (Atxn1) diminue le niveau de phosphorylation de l'ARN Pol II. Les auteurs rapportent également que le complexe PQBP1/Atxn1-mutée (plus de 40 aa glutamine à la suite) conduit à une réduction du taux de phosphorylation de l'ARN Pol II et, *in fine*, à une diminution globale de son activité transcriptionnelle. Chez la drosophile, la perte de fonction partielle de l'homologue de *PQBP1* (insertion d'un transposon dans le premier exon de *dPQBP1* conduit à une diminution de l'espérance de vie des mouches mutantes, une diminution d'expression des gènes codant pour la sous unité NR1 du récepteur NMDA dans les neurones de projection ainsi qu'une baisse des capacités d'apprentissage (Tamura et al., 2010).

La souris *Pqbp1* KO hétérozygote est létale (Iwasaki et al., 2014) et le premier modèle présentant une diminution d'expression est une souris transgénique exprimant de manière endogène et ubiquitaire un shRNA conduisant à une dégradation d'environ 50% des ARNm *Pqbp1* (Ito et al., 2009). L'équipe a pu mettre en évidence de légères différences comportementales comportant anxiété et cognition. Les auteurs soulignent une absence d'augmentation d'expression c-fos (un marqueur de l'activité neuronale voyant son expression augmentée en cas d'activité des neurones) chez les souris *Pqbp1* mutantes. Deux années plus tard, deux modèles de KO conditionnels furent générés. Ces modèles utilisent l'expression d'une Cre sous le contrôle de deux promoteurs respectifs : le promoteur Nestin, permettant une inactivation dans les cellules progénitrices et le promoteur et Synapsin conduisant à une inactivation dans les neurones différenciés (Ito et al., 2015). Seul le modèle Nestin-Cre présente une forte microcéphalie indiquant que l'altération de *Pqbp1* dans les cellules progénitrices conduit au phénotype de microcéphalie. On notera que les deux modèles présentent des structures corticales inchangées. Les auteurs rapportent un allongement de la phase M au niveau du cycle des NSC murines KO pour *Pqbp1*. Cet allongement du cycle cellulaire conduit à une diminution du pool total de précurseurs au moment de la neurogénèse et de ce fait, à une diminution du nombre de neurones. Les auteurs ont également développé un AAV codant pour l'ARNm de *PQBP1* humain, qui injecté à E10 conduit à un sauvetage du phénotype de

microcéphalie et des atteintes comportementales. Une analyse transcriptomique réalisée sur cultures de neurones a révélé un rôle essentiel de PQBP1 dans la stabilité de l'ARNm *APC4*, un composant du complexe de promotion de l'anaphase (Anaphase Promoting Complex) aux rôles multiples, notamment très impliqué dans la régulation du cycle cellulaire (Skaar et al., 2009 ; Sivakumar et al., 2015). En utilisant les mêmes modèles KO conditionnels, cette équipe a rapporté récemment que les souris *Pqbp1* cKOs présentent une petite taille ainsi qu'une réduction de la masse osseuse liée notamment à des défauts de différenciation des ostéoclastes (Yang et al., 2020).

#### 4) Rôle de PQBP1 dans la régulation de l'épissage

La protéine PQBP1 est connue pour réguler la transcription des gènes mais également leur épissage, des processus cruciaux dans le bon déroulement du neurodéveloppement. On sait que le domaine CTD de PQBP1 interagit notamment avec la protéine U5-15kDa (ou TXNL4A ; snRNP15) du spliceosome via un motif YxxPxxVL (Waragai et al., 2000 ; Mizuguchi et al., 2014 ; Mizuguchi et al., 2016 ; Liu et al., 2020). Dans ces travaux les auteurs décrivent PQBP1 comme un facteur adaptateur, permettant de faciliter le transport de U5-15kDa au noyau mais également comme un potentiel facteur de recrutement ou de stabilisation d'autres protéines d'épissage (Okazawa et al., 2002.). En plus d'interagir avec U5-15kDa, PQBP1 régule l'épissage alternatif de plusieurs cibles via une interaction avec le facteur SF1B3 (un composant du complexe snRNP-U2) (Wang et al., 2013). Les auteurs ont démontré que PQBP1 régule l'épissage alternatif d'ARNm codant pour des facteurs apoptotiques comme le facteur Bcl-X, une protéine transmembranaire présente dans la membrane mitochondriale. Ce dernier existe sous deux isoformes, la plus longue (Bcl-xL) permet d'inhiber le relargage de molécules telles que le cytochrome C et possède donc un rôle anti apoptotique. La plus courte (Bcl-xS) à l'inverse possède un rôle inhibiteur de Bcl-xL et a donc un rôle pro-apoptotique. De ce fait, les auteurs décrivent l'inactivation de *PQBP1* comme favorisant la forme pro-apoptotique Bcl-xS soulignant son implication dans la régulation du processus d'apoptose. Cette interaction entre PQBP1 et SF1B3 semble nécessaire à la reconnaissance spécifique de sites d'épissage et entraîne notamment en cas de déplétion de *Pqbp1* dans des cultures primaires de neurones murins (shRNA transduits à l'aide de lentivirus) une diminution de la croissance dendritique. Dans les embryons de *X. Laevis*, une équipe a pu confirmer l'interaction de PQBP1 avec WBP11, un facteur d'épissage exprimé durant le développement embryonnaire de l'animal (Iwasaki et al., 2014). De plus, l'inactivation partielle de *pqbp1a* et *pqbp1b* par MO mais également leur surexpression durant le développement du xénope conduit à une diminution

d'expression de *sox2* ainsi qu'à des malformations de la plaque neurale et des défauts de gastrulation et neurulation. Cette étude a pu démontrer que PQBP1 diminuait la réponse au FGF en forçant l'inclusion de l'exon 8a lors de l'épissage de l'ARNm codant pour la protéine FGF-2. Cette régulation d'épissage a pour conséquence la synthèse de deux récepteurs différents FGFR2IIIb et FGFR2IIIc, répondant à des ligands différents : respectivement FGF1-3 -7-10 et 22 pour FGFR2IIIb et FGF2-4-6-8-9-17 et 18 pour FGFR2IIIc (Gong et al., 2014). La perte d'équilibre de cette balance entre les deux récepteurs est ainsi suspectée de contribuer à l'apparition de phénotypes développementaux. Plus récemment, une équipe a étudié chez la drosophile les conséquences du décalage du cadre de lecture causée par une insertion ou délétion au niveau du hotspot mutationnel dans *PQBP1* (Zhang et al., 2017). Les auteurs ont pu mettre en évidence que le nouvel épitope C-terminal synthétisé interagissait préférentiellement avec la protéine FMRP non phosphorylée, conduisant à son ubiquitination et stimulant de ce fait sa dégradation. Très récemment, une équipe a pu montrer que la conformation des protéines PQBP1 tronquées était plus compacte que la protéine sauvage conduisant également à leur dimérisation (Chen et al., 2021). La protéine PQBP1 est décrite comme intrinsèquement désordonnée et la présence de dimères semble s'expliquer par une fonction auto inhibitrice de dimérisation du domaine CTD qui est absent chez ces mutants. Les protéines PQBP1 perdent de ce fait leurs capacités régulatrices, notamment liées à l'interaction avec l'ARN Pol II et les différents facteurs d'épissage.

## 5) PQBP1 et troubles neurodégénératifs

La protéine PQBP1 est largement décrite dans le cadre de son rôle dans le SR notamment de par ses activités de régulation de la transcription et de l'épissage. Plusieurs études rapportent une implication potentielle de PQBP1 dans d'autres contextes comme les troubles neurodégénératifs. En effet, certaines protéines impliquées dans des troubles neurodégénératifs ont été montrées comme interagissant avec PQBP1. Par exemple, la protéine Atxn1 (codée par le gène *SCA1* et impliquée dans l'ataxie cérébelleuse de type 1) interagit directement avec PQBP1 au niveau des corps d'inclusion nucléaires dans les neurones et pourrait jouer un rôle physiopathologique dans le cas d'un patient exprimant une Atxn1 mutée (Okazawa et al., 2002). On sait également que PQBP1 est capable d'interagir avec HTT, la protéine impliquée dans la maladie d'Huntington (Waragai et al., 1999). Enfin, PQBP1 interagit avec la protéine SRRM2 dans les cultures primaires de neurones murins au niveau des nuclear speckles (Tanaka et al., 2018). La perte d'interaction entre ces deux protéines dans les modèles d'étude de la maladie d'Alzheimer murins (souris exprimant des versions mutées des gènes *APP* et *PSEN1*) entraîne

une perte de stabilité de PQBP1 conduisant à des défauts d'épissages alternatifs dans la synapse. Les auteurs ont également pu démontrer que l'injection d'AAV-*PQBP1* permettait de restaurer le phénotype de diminution de densité des dendrites observé au préalable chez les animaux malades. On peut également noter qu'une souris transgénique surexprimant *PQBP1* humain fut également générée et a conduit à la mise en évidence d'un phénotype neurodégénératif (Okuda et al., 2003). Ces souris ont en effet présenté une atteinte tardive et graduelle des moto neurones ainsi qu'une diminution du nombre de cellules de Purkinje dans le cervelet. Ces résultats vont donc également dans le sens d'une implication de PQBP1 dans certains troubles neurodégénératifs.

## 6) Autres implications de PQBP1

Chez *C. Elegans* l'expression de *pqbp-1.1* est ubiquitaire durant toute la vie de l'animal mais majoritaire dans le tractus intestinal. Une équipe a pu montrer que la répression de *pqbp-1.1* par shRNA conduit à des défauts d'incorporation des triglycérides, suggérant l'implication de la protéine dans le métabolisme lipidique. Ces défauts d'incorporation ont également été validés par la même équipe dans des cultures d'adipocytes murins, suggérant une explication au moins partielle au retard de croissance général et au corps fin des patients (Takahashi et al., 2009). Une autre étude rapporte une potentielle interaction du domaine CTD de PQBP1 avec l'ADNc du virus VIH (Virus d'Immunodéficience Humaine) qui déclencherait une réaction immunitaire médiée par la protéine IRF3 dans les cellules dendritiques (Yoh et al., 2015). PQBP1 est également suspecté d'être impliqué dans certaines maladies coronaires (Talukar et al., 2016).

On sait également que PQBP1 interagit avec la dynamine 2 et inhibe sa capacité de GTPase. De ce fait, l'inactivation de *Pqbp1* dans des cultures primaires de neurones hippocampiques de rat conduit à des anomalies de morphologie du cil primaire, également retrouvées dans le cortex de souris mutantes *in vivo* (Ikeuchi et al., 2013). Il est intéressant de noter que ce groupe n'a également pas vu de différence significative au niveau du nombre de cellules entrant en apoptose (marquage TUNEL), que ce soit dans les zones ventriculaires ou dans la zone corticale durant le développement.

# **OBJECTIFS DE LA THESE**

---



La déficience intellectuelle (DI) est un trouble du neurodéveloppement (TND) affectant 1 à 2 % de la population et la première raison de consultation génétique. Les causes génétiques de la DI sont variées et extrêmement hétérogènes, elles peuvent concerner des anomalies de nombre ou de structure des chromosomes, des variations du nombre de copies ou des anomalies ponctuelles affectant un seul gène et conduisant à une perte partielle ou totale de la fonction de la protéine. A l'heure actuelle plus de 1000 gènes ont pu être associés à un syndrome de DI. L'avènement du séquençage d'exome à visée diagnostique a permis ces dernières années d'améliorer de façon très importante l'identification de variants pathogènes chez les individus avec DI et d'augmenter le taux de diagnostic moléculaire. L'obtention d'un diagnostic étiologique précis et précoce est primordial pour les familles afin d'être en mesure de proposer une prise en charge adaptée de la DI chez le jeune enfant et éventuellement une approche thérapeutique s'il en existe. Malgré ces progrès importants en matière de diagnostic moléculaire, deux problématiques majeures subsistent :

- 1) L'interprétation des nombreux variants identifiés, pour un certain nombre de variants (pour beaucoup des variants faux-sens), le faisceau d'arguments n'est pas suffisant pour conclure. Le variant est classé en tant que Variant de Signification Inconnue (VSI) et les familles concernées restent en errance diagnostique tant que ces variants ne sont pas reclassés.
- 2) Du fait du nombre de gènes impliqués, chaque forme monogénique de DI représente une maladie ultra-rare, dont pour la plupart les mécanismes physiopathologiques ainsi que les voies cellulaires impliquées sont généralement mal connus. Pour améliorer la prise en charge des jeunes patients et envisager des approches thérapeutiques, il est donc primordial de mieux caractériser les conséquences de ces mutations lors du développement du cerveau.

Mon travail de doctorat s'est partagé entre ces deux problématiques. Mes objectifs étaient :

- De développer des approches expérimentales robustes afin de prouver le caractère pathogène ou bénin de VSI identifiés dans un des gènes les plus fréquemment muté dans la DI, le gène *DYRK1A* mais également dans *PQBPI*, un second gène impliqué dans des formes de DI liées à l'X.
- De caractériser les conséquences moléculaires et cellulaires de l'inactivation de *DYRK1A* et *PQBPI* par interférence ARN dans un modèle de progéniteurs neuronaux humains afin de mieux comprendre les mécanismes physiopathologiques impliqués.
- De contribuer à la finalisation d'autres projets tels que la validation de l'effet de VSI dans le gène *NLGN3*

Enfin, j'ai également eu la chance de pouvoir participer à plusieurs projets au sein de mon équipe mais également en dehors.





# RESULTATS

---



**Article 1 : Integrative approach to interpret *DYRK1A* variants, leading to a frequent neurodevelopmental disorder**

---



## Résumé

Le syndrome DYRK1A est une des formes de DI les plus fréquentes et représente près de 0,5% des patients DI. *DYRK1A* code pour une kinase à double spécificité Tyrosine et Serrine/Thréonine extrêmement conservée au cours de l'évolution. Les premières variations de fonction identifiées chez des patients DYRK1A furent des événements de type translocations et microdélétions. En plus de révolutionner l'identification de nouveaux gènes impliqués dans la DI, le séquençage d'exome a permis de mettre en évidence plusieurs dizaines de variations ponctuelles (décalages du cadre de lecture, non-sens, petites insertions/délétions, variants d'épissage et faux-sens) dans le gène *DYRK1A*. De manière globale, l'interprétation de ces variants est devenue un enjeu majeur pour les laboratoires de diagnostic génétique, notamment concernant les variants faux-sens. Malgré le développement d'outils informatiques permettant de prédire l'impact délétère des variants sur la fonction de la protéine, plus de 10% de ces variants restent à l'heure actuelle classés comme VSI. Le développement de tests fonctionnels permettant de caractériser l'effet des VSI permet d'une part de mettre fin au phénomène d'errance diagnostic des familles et d'autre part, de mieux comprendre les mécanismes physiopathologiques associés aux pathologies.

L'objectif de ce travail était de développer une approche permettant de tester l'impact de n'importe quel variant identifié dans *DYRK1A* afin de pouvoir conclure à son effet pathogène ou son caractère bénin. Ce travail a débuté par le développement d'un score clinique spécifique au syndrome DYRK1A ( $CS_{DYRK1A}$ ) par Amélie Piton, Benjamin Durand, Marie Vincent et Marjolaine Willems, grâce à une étude approfondie des traits phénotypiques retrouvés chez 32 individus atteints. Ce score permet d'évaluer si le phénotype d'un patient est évocateur ou non du syndrome DYRK1A. J'ai pu en parallèle étudier l'impact fonctionnel de 17 variants identifiés chez des patients DYRK1A, en modèle cellulaire. J'ai dans un premier temps observé l'effet de ces variants au niveau de l'ARNm muté dans des cultures primaires de fibroblastes de patients et ai pu observer une dégradation de certains par le système NMD, des effets sur l'épissage alternatif conduisant à des décalages du cadre de lecture ou encore une absence d'effet sur la stabilité de l'ARNm pour les variants faux-sens. J'ai dans un second temps surexprimé les protéines DYRK1A sauvage et porteuses des variants étudiés dans différents modèles cellulaires (HeLa, COS-1 et HEK293) afin d'étudier leurs impacts à différents niveaux. En effet, j'ai pu observer pour une majorité d'entre eux une diminution du niveau

d'expression des protéines mutantes, causée par une dégradation prématurée par le protéasome. J'ai également pu montrer que la plupart des variants n'avaient pas d'effet sur la capacité de la protéine à se localiser dans le compartiment nucléaire de cellules HeLa mais que l'un d'entre eux conduisait à un phénotype d'agrégation dans les cellules. Par la suite, je me suis intéressé à l'impact de ces variants sur la capacité des protéines mutantes à réaliser le processus d'autophosphorylation sur le résidu Tyr321, prérequis indispensable permettant à l'enzyme de passer sous sa forme de kinase active. J'ai de ce fait pu montrer que la plupart des variants inhibaient ce processus. J'ai également pu valider que l'absence d'autophosphorylation des protéines mutantes était bien corrélée avec une perte totale de l'activité kinase, en étudiant la capacité des protéines mutantes à phosphoryler MAPT (TAU), une cible bien connue de DYRK1A. Pour finir et étant donné l'implication de DYRK1A dans la régulation de phénomènes épigénétiques, nous avons réalisé une collaboration avec Rosanna Weksberg et Éric Chater-Diehl dans le but d'identifier une potentielle signature épigénétique associée au syndrome DYRK1A. Après analyse du méthylome d'ADN extrait à partir d'échantillons de sang de patients contrôles ou porteurs de variations pathogènes *DYRK1A*, nos collaborateurs ont pu mettre en évidence une signature épigénétique forte comprenant près de 402 CpG différenciellement méthylés chez les patients DYRK1A. De ce fait, ils ont dans un second temps étudié la ségrégation des ADN extraits à partir de sang de patients porteurs de VSI dans *DYRK1A* et ont évalué l'effet délétère des VSI en fonction de leur clusterisation avec les échantillons contrôles ou ceux des patients DYRK1A. Nous avons pu montrer que la classification des variants réalisée par analyse du méthylome correspondait aux observations et conclusions réalisées sur les modèles cellulaires. Cependant, suite aux analyses épigénétiques, un VSI (p.Gly486Asp) fut proposé comme potentiel gain de fonction. Étant donné l'absence d'effet de ce variant sur la capacité de la protéine mutante à phosphoryler MAPT (TAU) en modèle cellulaire, l'effet de ce variant reste à l'heure actuelle indéterminé.

Pour conclure, nous avons pu développer au cours de ce projet une approche multidisciplinaire permettant de mieux interpréter les variants identifiés dans le gène *DYRK1A*. De ce fait, nous avons pu à la suite de ce travail conclure quant à la pathogénicité ou au caractère bénin de 12 VSI sur 13 testés initialement.

# Integrative approach to interpret DYRK1A variants, leading to a frequent neurodevelopmental disorder

Jeremie Courraud<sup>1,2,3,4</sup>, Eric Chater-Diehl<sup>5</sup>, Benjamin Durand<sup>1,2,3,4</sup>, Marie Vincent<sup>6</sup>, Maria del Mar Muniz Moreno<sup>1,2,3,4</sup>, Imène Boujelbene<sup>1,2,3,4,7</sup>, Nathalie Drouot<sup>1,2,3,4</sup>, Loréline Genschik<sup>1,2,3,4</sup>, Elise Schaefer<sup>8</sup>, Mathilde Nizon<sup>6</sup>, Bénédicte Gerard<sup>7</sup>, Marc Abramowicz<sup>9</sup>, Benjamin Cogné<sup>6</sup>, Lucas Bronicki<sup>10</sup>, Lydie Burglen<sup>11</sup>, Magalie Barth<sup>12</sup>, Perrine Charles<sup>13</sup>, Estelle Colin<sup>12</sup>, Christine Coubes<sup>14</sup>, Albert David<sup>6</sup>, Bruno Delobel<sup>15</sup>, Florence Demurger<sup>16</sup>, Sandrine Passemard<sup>17</sup>, Anne-Sophie Denomme<sup>18</sup>, Laurence Faivre<sup>18</sup>, Claire Feger<sup>7</sup>, Mélanie Fradin<sup>20</sup>, Christine Francannet<sup>21</sup>, David Genevieve<sup>14</sup>, Alice Goldenberg<sup>22</sup>, Anne-Marie Guerrot<sup>22</sup>, Bertrand Isidor<sup>6</sup>, Katrine M. Johannesen<sup>24,25</sup>, Boris Keren<sup>13</sup>, Maria Kibæk<sup>23</sup>, Paul Kuentz<sup>18</sup>, Michele Mathieu-Dramard<sup>26</sup>, Bénédicte Demeer<sup>26</sup>, Julia Metreau<sup>27</sup>, Rikke Steensbjerre Møller<sup>24,25</sup>, Sébastien Moutton<sup>18</sup>, Laurent Pasquier<sup>20</sup>, Kristina Pilekær Sørensen<sup>23</sup>, Laurence Perrin<sup>28</sup>, Mathilde Renaud<sup>29</sup>, Pascale Saugier<sup>22</sup>, Marlene Rio<sup>30</sup>, Joane Svane<sup>23</sup>, Julien Thevenon<sup>31</sup>, Frederic Tran Mau Them<sup>18</sup>, Cathrine Elisabeth Tronhjem<sup>23</sup>, Antonio Vitobello<sup>18</sup>, Valerie Layet<sup>32</sup>, Stéphane Auvin<sup>33</sup>, Khaoula Khachnaoui<sup>34</sup>, Marie-Christine Birling<sup>35</sup>, Severine Drunat<sup>36</sup>, Allan Bayat<sup>23</sup>, Christèle Dubourg<sup>37</sup>, Salima El Chehadeh<sup>7</sup>, Christina Fagerberg<sup>23</sup>, Cyril Mignot<sup>12</sup>, Michel Guipponi<sup>9</sup>, Thierry Bienvenu<sup>38</sup>, Yann Herault<sup>1,2,3,4</sup>, Julie Thompson<sup>39</sup>, Marjolaine Willems<sup>14</sup>, Jean-Louis Mandel<sup>1,2,3,4</sup>, Rosanna Weksberg<sup>5,40,41,42,43</sup>, \*Amélie Piton<sup>1,2,3,4,7,44</sup>

## AFFILIATIONS

<sup>1</sup> Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch 67400, France

<sup>2</sup> Centre National de la Recherche Scientifique, UMR7104, Illkirch 67400, France

<sup>3</sup> Institut National de la Santé et de la Recherche Médicale, U964, Illkirch 67400, France

<sup>4</sup> Université de Strasbourg, Illkirch 67400, France

<sup>5</sup> Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada

<sup>6</sup> Service de Génétique Médicale, CHU de Nantes & Inserm, CNRS, Université de Nantes, l'institut du thorax, 44000 Nantes, France

<sup>7</sup> Unité de Génétique Moléculaire, IGMA, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

<sup>8</sup> Service de Génétique Médicale, IGMA, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

<sup>9</sup> Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland

<sup>10</sup> Department of Genetics, CHEO, Ottawa, ON, Canada.

<sup>11</sup> Centre de référence des malformations et maladies congénitales du cervelet et Département de génétique et embryologie médicale, APHP, Sorbonne Université, Hôpital Armand Trousseau, 75012 Paris, France

<sup>12</sup> Pediatrics & Biochemistry and Genetics, Department, Angers Hospital, Angers, France.

<sup>13</sup> Genetic Department, University Hospital Pitié-Salpêtrière, AP-HP, Paris, France

<sup>14</sup> Département de Génétique Médicale maladies rares et médecine personnalisée, Centre de Référence Maladies Rares Anomalies du Développement, Hôpital Arnaud de Villeneuve, Université Montpellier, France

<sup>15</sup> Centre de Génétique Chromosomique, GHICL, Hôpital Saint Vincent de Paul, Lille, France

<sup>16</sup> Service de Génétique, CH Bretagne Atlantique- Vannes

<sup>17</sup> Département de Génétique, Hôpital Universitaire Robert Debré, APHP, Paris, France.

<sup>18</sup> Centre de Génétique et Centre de Référence Anomalies du développement et Syndromes malformatifs, Hôpital d'Enfants and INSERM UMR1231 GAD, FHU TRANSLAD, CHU de Dijon, Dijon, France

<sup>19</sup> Unité Fonctionnelle d'Innovation en Diagnostique Génomique des Maladies Rares, Pôle de Biologie, FHU-TRANSLAD, CHU Dijon Bourgogne, F-21000, Dijon, France

<sup>20</sup> Centre de Référence Maladies Rares, Unité Fonctionnelle de Génétique Médicale, CHU, Rennes, France

<sup>21</sup> Service de Génétique médicale, CHU de Clermont-Ferrand, Clermont-Ferrand, France

<sup>22</sup> Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics and Reference Center for Developmental Disorders, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France

<sup>23</sup> Department of Clinical Genetics, Odense Denmark Hospital, Odense University Hospital, Odense, Denmark

<sup>24</sup> Department of Epilepsy Genetics and Personalized Treatment, The Danish Epilepsy Centre, Dianalund, Denmark

<sup>25</sup> Institute for Regional Health Services, University of Southern Denmark, Odense Denmark

<sup>26</sup> Service de Génétique Clinique, Centre de référence maladies rares, CHU d'Amiens-site Sud, Amiens, France

<sup>27</sup> APHP, Service de neurologie pédiatrique, Hôpital Universitaire Bicetre, Le Kremlin-Bicetre, France

<sup>28</sup> Department of Genetics, Robert Debré Hospital, AP-HP, Paris, France

<sup>29</sup> Service de Génétique Clinique et de Neurologie, Hôpital Brabois Enfants, Nancy, France



<sup>30</sup> Department of medical genetics and reference centre for rare intellectual disabilities. INSERM UMR 1163, Paris Descartes-Sorbonne Paris Cité University, Imagine Institute, Necker Enfants Malades Hospital, Paris, France

<sup>31</sup> Department of Genetics and Reproduction, Centre Hospitalo-Universitaire Grenoble-Alpes, Grenoble, France

<sup>32</sup> Consultations de génétique, Groupe Hospitalier du Havre, Le Havre, France

<sup>33</sup> Center for rare epilepsies & epilepsy unit Robert-Debré Hospital, APHP, & INSERM NeuroDiderot, Université de Paris, Paris, France

<sup>34</sup> Université Côte d'Azur, Inserm U1081, CNRS UMR7284, IRCAN, CHU de Nice, Nice, France

<sup>35</sup> ICS, Mouse Clinical Institute, Illkirch-Graffenstaden, France

<sup>36</sup> Département de Génétique, Hôpital Universitaire Robert Debré, Paris

<sup>37</sup> Laboratoire de Génétique Moléculaire, CHU Pontchaillou, UMR 6290 CNRS, IGDR, Faculté de Médecine, Université de Rennes 1, Rennes, France

<sup>38</sup> Molecular Genetics Laboratory, Cochin Hospital, APHP.Centre-Université de Paris, and INSERM UMR 1266, Institut de Psychiatrie et de Neurosciences de Paris, 75014 Paris, France

<sup>39</sup> Complex Systems and Translational Bioinformatics (CSTB), ICube laboratory - CNRS, Fédération de Médecine Translationnelle de Strasbourg (FMTS), University of Strasbourg, Strasbourg, France

<sup>40</sup> Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada

<sup>41</sup> Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5S 1A1, Canada-RBW

<sup>42</sup> Department of Pediatrics, University of Toronto, Toronto, ON, Canada

<sup>43</sup> Institute of Medical Science, School of Graduate Studies, University of Toronto, Toronto, Ontario, Canada

<sup>44</sup> Institut Universitaire de France

The authors declare no conflict of interest.

\*Address for correspondence and material request:

Amélie Piton, PhD

Laboratoire "Mécanismes génétiques des maladies neurodéveloppementales", IGBMC, Illkirch, France

Tel : +33369551652

E-mail: [piton@igbmc.fr](mailto:piton@igbmc.fr)

## Competing interests

None.

## ABSTRACT

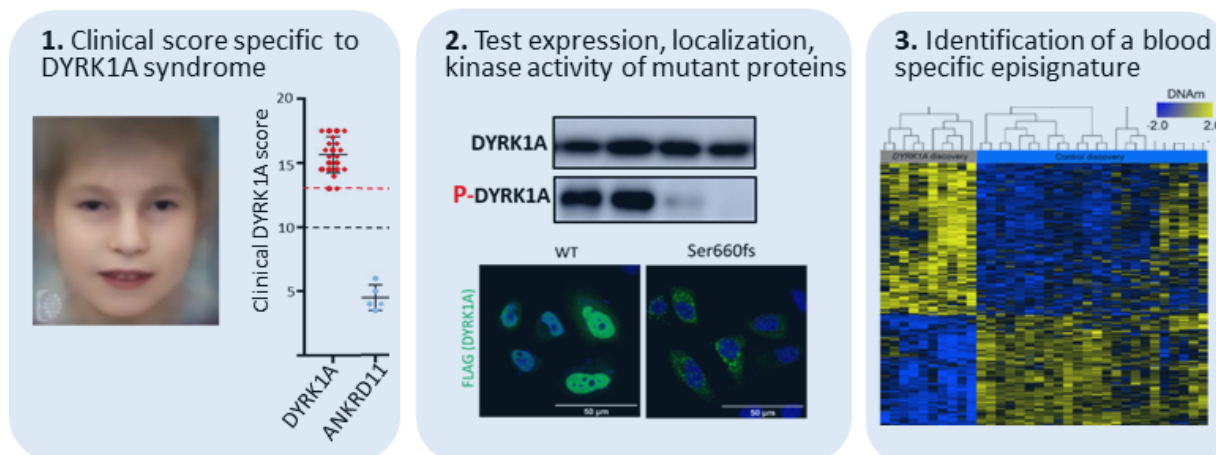
**PURPOSE:** *DYRK1A* syndrome is among the most frequent monogenic forms of intellectual disability (ID). We refined the molecular and clinical description of this disorder and developed tools to improve interpretation of missense variants, which remains a major challenge in human genetics.

**METHODS:** We reported clinical and molecular data for fifty individuals with ID harboring *DYRK1A* variants and developed i) a specific *DYRK1A* clinical score, ii) amino acid conservation data generated from one hundred of *DYRK1A* sequences across different taxa, iii) *in vitro* overexpression assays to study level, cellular localization, and kinase activity of *DYRK1A* mutant proteins, and iv) a specific blood DNA methylation signature.

**RESULTS:** This integrative approach was successful to reclassify several variants as pathogenic. However, we questioned the involvement of some others, such as p.Thr588Asn, still reported as likely pathogenic, and showed it does not cause an obvious phenotype in mice.

CONCLUSION: Our study demonstrated the need for caution when interpreting variants in *DYRK1A*, even those occurring *de novo*. The tools developed will be useful to interpret accurately the variants identified in the future in this gene.

## Integrative tools to interpret variants causing *DYRK1A* syndrome



**Conclusion:** Integration of the different tools to interpret variants in *DYRK1A*



## INTRODUCTION

Intellectual disability (ID) and autism spectrum disorder (ASD) are two highly heterogeneous groups of neurodevelopmental disorders (NDD) with substantial genetic contributions which overlap strongly both at the clinical and genetic levels. More than one thousand genes have been implicated in monogenic forms of NDD, with an important contribution of autosomal dominant forms caused by *de novo* variants<sup>1</sup>. One of these genes, *DYRK1A* (*dual-specificity-tyrosine-phosphorylation-regulated-kinase-1A*)<sup>2</sup>, located on chromosome 21, is among the genes the most frequently mutated in individuals with ID<sup>1</sup>.

The first *DYRK1A* disruptions were identified in individuals with intrauterine growth restriction (IUGR), primary microcephaly and epilepsy<sup>3</sup>. Few years later, the first frameshift variant was described in a patient with similar features<sup>4</sup>. The clinical spectrum associated with *DYRK1A* pathogenic variants (*MRD7 for Mental Retardation 7* in OMIM) was further refined with the publication of additional patients, presenting suggestive facial dysmorphism, severe speech impairment and feeding difficulty, while epilepsy and prenatal microcephaly were not always present<sup>5-9</sup>. Pathogenic variants were also identified in cohorts of individuals with ASD<sup>10</sup>, but all have ID<sup>11</sup>. The *DYRK1A* gene encodes a dual tyrosine-serine/threonine (Tyr-Ser/Thr) kinase composed of a central catalytic domain including Tyrosine-321, involved in *DYRK1A*

activation by autophosphorylation<sup>12</sup>, two nuclear localization signal sequence (NLS), and additional functional domains. *DYRK1A* is ubiquitously expressed during embryonic development and in adult tissues. Its location is both cytoplasmic and nuclear and varies by cell type and stage of development<sup>13</sup>. By the number and diversity of its proposed protein targets, *DYRK1A* regulates numerous cellular functions (reviews<sup>14,15</sup>), among them the MAPT (Tau) protein phosphorylated by *DYRK1A* on its Thr212 position<sup>16</sup>.

High Throughput Sequencing (HTS) has revolutionized the identification of genetic variants for diagnostic applications but a major challenge remains in the interpretation of the vast number of variants, especially for highly heterogeneous disease such as ID. A combination of genetic, clinical and functional approaches, summarized by the American College of Medical Genetics and the Association for Molecular Pathology (ACMG/AMP), are commonly used to interpret these variants<sup>17</sup>. A significant proportion of the variants, especially the missense variants, remain classified as variants of unknown significance (VUS, according to ACMG/AMP) after the primary analysis. For autosomal dominant diseases with complete penetrance such as *DYRK1A* syndrome, the *de novo* occurrence of a variant is a strong but not absolute argument for pathogenicity, and the parents' genotype is not always available. Clinical observations are useful but could also lead to misinterpretation, as numerous ID-associated manifestations are unspecific. Many tools have been developed to predict *in silico* the pathogenicity of missense variants but they remain predictions and protein-specific functional tests are useful to confirm variants effect. DNA methylation (DNAm) is also a powerful tool to test variant pathogenicity in disorders associated with epigenetic regulatory genes. We discovered that pathogenic variants in these genes can exhibit disorder-specific DNAm signatures comprised of consistent, multilocus DNAm alterations in peripheral blood, useful for classifying variants in these genes as pathogenic or benign<sup>18-21</sup>. *DYRK1A* has numerous targets, and while it is not well-described as an epigenetic regulator, it has been shown to phosphorylate histone H3 and histone acetyltransferase proteins<sup>22,23</sup>. Therefore, we hypothesized that pathogenic variants in *DYRK1A* might generate a specific DNAm signature in blood.

We reviewed the clinical signs present in 34 individuals never reported and carrying deletions or clearly pathogenic variants in *DYRK1A* to refine the clinical spectrum of *DYRK1A* syndrome. Based on these data, we developed a score to help to recognize affected individuals and to interpret variants (reverse phenotyping). In parallel, we developed *in silico* and *in vitro* approaches to assess variant effects on *DYRK1A* function. Finally, we defined a DNAm signature specific to *DYRK1A* syndrome in patient blood. We used this multifaceted approach to interpret 17 variants identified in patients with ID/NDD and demonstrated its efficiency in optimizing accurate interpretation of variants.

## METHODS

### *Patients and molecular analysis*

Variants in *DYRK1A* (NM\_001396.4) were identified during routine genetic analyses in individuals referred to clinical genetic services for intellectual disability in France, Denmark and Switzerland: CGH-array, *DYRK1A* coding regions sanger sequencing, targeted next generation sequencing of ID genes (TES)<sup>24-26</sup>, trio or simplex clinical exome sequencing (CES) or exome sequencing (ES), and confirmed by an additional method. Combined Annotation Dependent Depletion (CADD) and NnsplICE were used to predict the effect of missense and splice variants respectively. Fibroblasts were established from skin biopsies for Ind #1, #11, #22, #24, Bronicki#9 and Bronicki#10 and cultivated as previously described<sup>27</sup>. Paxgene blood

samples were collected for Ind #9, #18, #19 and #30. mRNA extraction, RNA sequencing, RT-PCR or qPCR (primers available on request) were performed as previously described<sup>28</sup>.

### ***Phenotypic analysis and clinical scoring***

Clinical information and photographs were provided by the referring clinicians for the 44 individuals reported here (**Table S1**) as well as for six individuals previously published (individuals Bronicki#2, #3, #8, #9 and #10 and Ruaud#2)<sup>6,7</sup>(n=50 individuals in total). Based on the most frequent signs and the morphometric characteristics presented by the 34 individuals with truncating variants in *DYRK1A*, a clinical score out of 20 was established (DYRK1A\_I, n=21 individuals with photographs available **Table S2**). Clinical scores were calculated for this initial cohort plus a second cohort (replication cohort, DYRK1A\_R, n=13) with individuals already described in previous publications<sup>5-7</sup> by experienced clinical geneticists, as well as for individuals with other frequent monogenic forms of ID, caused by pathogenic variants in *DDX3X* (n=5), *ANKRD11* (n=5), *ARID1B* (n=8), *KMT2A* (n=6), *MED13L* (n=5), *SHANK3* (n=6) or *TCF4* (n=6) genes.

### ***Definition of sets of missense variants and conservation analysis***

To evaluate which tools are pertinent to predict effect of missense variants on the DYRK1A protein, we used different sets of variants : 1) variants presumed to be benign (negative N-set, n=115) including missense variants reported “benign” in ClinVar plus variants found more than once in GnomAD (november 2019 release), 2) variants reported as “pathogenic”/“likely pathogenic” in Clinvar (positive P-set, n=16), and 3) other variants reported here, in literature, or as VUS in Clinvar (test T-set, n=44)(**Table S3**). 123 orthologous sequences of human DYRK1A were extracted from the OrthoInspector database version 3.0. and a multiple sequence alignment (MSA) was constructed using the Clustal Omega software. The MSA was then manually refined to correct local alignment errors using the Jalview MSA editor<sup>29</sup>. The refined MSA was used as input to the PROBE software<sup>30</sup>, in order to identify conserved regions in the sequences. The sequences in the MSA were divided into five separate clades: Vertebrates, Metazoans, Protists, Plants and Fungi.

### ***In vitro analysis of variant effect on DYRK1A protein***

*DYRK1A* expression plasmids (addgene #101770, Huen lab) were generated from the pMH-SFB-*DYRK1A* vector containing the N-terminal FLAG-tagged human *DYRK1A* cDNA sequence (NM\_001396.4). Variant sequences were obtained by site-directed mutagenesis as described<sup>31</sup>. Antibodies used are listed in **Supplementary Methods**. HeLa, HEK293 and COS1 cells were maintained and transfected for 24h with *DYRK1A* plasmids (for immunofluorescence) or *DYRK1A* plasmids plus pEGFP-N1 plasmid (for Western Blot) as previously done<sup>31,32</sup>. For autophosphorylation analysis and DCFA7 (WDR68) interaction, proteins were extracted from transfected HEK293 cells and immunoprecipitated with anti-FLAG antibody as described<sup>31</sup>. Phosphorylated Tyr321 DYRK1A was visualized as described<sup>33</sup> and normalized by the level of total DYRK1A protein. Kinase activity was investigated by co-transfecting *DYRK1A* and *MAPT* plasmids (MAPT\_OHu28029C\_pcDNA3.1(+)-C-HA, geneScript) in HEK293 cells, adapted from<sup>34</sup>. The level of DYRK1A, MAPT and pMAPT (Thr212) proteins were normalized with GAPDH.

### ***DNA methylation signature***

Methylation analysis was performed using blood DNA from individuals with *DYRK1A* LoF variants (n=16), split into signature discovery (n=10) and validation (n=6) cohorts, based primarily on whether age at time of blood collection was available, and age- and sex-matched neurotypical controls (n=24). Whole blood DNA samples were prepared, hybridized to the Illumina Infinium Human MethylationEPIC BeadChip and analyzed as previously described<sup>20</sup>, a total of n=774,590 probes were analyzed for differential methylation. Standard quality control metrics showed good data quality for all samples except Ind#20. Briefly, *limma* regression with covariates age, sex, and five predicted blood cell types identified a DNAm signature with a Benjamini-Hochberg adjusted *p*-value<0.05 and 10% methylation difference. Next, we developed a support vector machine (SVM) model with linear kernel trained on including non-redundant CpG sites<sup>20</sup> using the methylation values for the discovery cases vs. controls. The model generated scores ranging between 0 and 1 for tested samples, classifying samples as “positive” (score>0.5) or “negative” (score<0.5). Additional neurotypical controls (n=94) and *DYRK1A* LoF validation samples (n=6) were scored to test model specificity and sensitivity respectively, and samples with pathogenic *KMT2A* (n=8) and *ARID1B* (n=4) variants and *DYRK1A* missense and distal frameshift (n=11) variants were tested.

## RESULTS

### *Identification of genetic variants in DYRK1A in individuals with ID*

We collected molecular and clinical information from 50 individuals with ID (44 never reported and six previously reported<sup>6,7</sup>) carrying a variant in *DYRK1A* identified in clinical and diagnostic laboratories: structural variants deleting or interrupting *DYRK1A* and recurrent or novel nonsense, frameshift, splice and missense variants (**Table 1, Figure S1**). When blood or fibroblast samples were available, we characterized the consequences of these variants on *DYRK1A* mRNA by RNA-sequencing and RT-qPCR (**Figure S2, Supplementary Text**). For one variant, c.1978del, occurring in the last exon of the gene (**Ind #18**), the mutant transcripts escape to nonsense mRNA mediated decay (NMD) and result in a truncated protein p.Ser660fs (or p.Ser660Profs\*43) retaining its entire kinase domain (**Figure S2F**). The variants occurred *de novo* in most of the cases (36/44), one was inherited from a mosaic father and parental DNA was not available for the seven others.

### *Clinical manifestations in individuals with pathogenic variants in DYRK1A and definition of a clinical score*

We reviewed the clinical manifestations of the patients with truncating variants, except p.(Ser660fs)(**Table S1, Supplementary text**). Recurrent features include, consistently with what was reported<sup>5,6,8,11</sup>: moderate to severe ID, prenatal or postnatal progressive microcephaly, major speech impairment, feeding difficulties, seizures and especially history of febrile seizures, autistic traits and anxiety, delayed gross motor development with unstable gait, brain MRI abnormalities including dilated ventricles and corpus callosum hypoplasia and recurrent facial features (**Figure 1A, Figure S3**). We found genital abnormalities as reported<sup>35</sup> but no obvious renal anomalies. We noted the importance of skin manifestations and especially atopic dermatitis. We used recurrent features to establish a “*DYRK1A*- clinical score” (CS<sub>*DYRK1A*</sub>) on a 20 point scale (**Figure 1A**), which aims to reflect specificity rather than severity of the phenotype. High scores, ranging from 13 to 18.5 (mean=15.5), were obtained for the individuals having a pathogenic variant in *DYRK1A* described here (*DYRK1A\_I*) or previously (*DYRK1A\_R*) (**Table S2, Figure 1B**). The threshold of CS<sub>*DYRK1A*</sub> ≥13 appears to be discriminant between individuals with LoF variants in *DYRK1A* (all ≥13) and individuals suffering from another form of ID (all<13) and is considered as “highly

suggestive” ( $13 > CS_{DYRK1A} > 10$ : “intermediate”;  $CS_{DYRK1A} < 10$ : “poorly evocative”). A clinical score on 15 points without photograph is less discriminative (**Figure S4**).

### *In silico analysis of missense variant effects*

We evaluated the discriminative power of the CADD score, commonly used in medical genetics<sup>36</sup> to interpret missense variants in *DYRK1A*. If a significant difference in the CADD score distribution is observed between the “benign” variants (N-set, see **Methods**) and those reported as “pathogenic” (P-set, see **Methods**) ( $p$ -value  $< 0.0001$ ), a substantial proportion of the N-set variants still have a CADD score above thresholds (20 or 25) usually used to define pathogenicity (**Figure S5A**). This could be explained by the high degree of amino acid conservation of *DYRK1A* among vertebrates, and this could lead to over-interpretation of pathogenicity. We performed sequence alignment with orthologs from different taxon (**Figure S6**) and confirmed that using sequences from vertebrates only is not efficient to classify missense variants, as one third of the N-set variants affect amino acids conserved in all species (**Figure S5B**,  $V=100\%$ ). Considering conservation parameters going beyond vertebrates appears more discriminant (13/16 P-set and 1/115 N-set variants are conserved in 100% of vertebrates, 90% of metazoan and 80% of other animals) (**Figure S5B**).

### *In vitro characterization of consequences of missense variants on DYRK1A protein*

In order to test the consequences of the missense variants *in vitro*, we overexpressed wild-type (WT) and mutant *DYRK1A* proteins in three different cell lines (HEK293, HeLa, COS1) including a truncating pathogenic variant Arg413fs and a benign missense variant from gnomAD, Ala341Ser. A significant and drastic decrease in *DYRK1A* protein level (**Figure 2A**), due to a reduction of protein stability (**Figure S7A**), was observed for the truncating Arg413fs and missense Asp287Val, Ser311Phe, Arg467Gln, Gly168Asp and Ile305Arg variants in each cell type. None of the variants affects *DYRK1A* interaction with DCAF7 (WDR68) (**Figure S7B**). To be active, *DYRK1A* has to undergo an autophosphorylation on Tyrosine 321<sup>12</sup>. To measure the level of active *DYRK1A* protein, we detected phospho-*DYRK1A* (Tyr321) by immunoprecipitation followed by immunoblot using anti-phospho-HIPK2, as previously described<sup>33</sup> (**Figure 2B**). We confirmed that the three variants previously tested (Asp287Val, Ser311Phe and Arg467Gln) abolish autophosphorylation<sup>33,37</sup>, as do the Gly168Ap and Ile305Arg variants. The Ser324Arg *DYRK1A* variant showed only residual autophosphorylation. No effect on autophosphorylation was observed for Arg255Gln, Tyr462His, Gly486Asp and Thr588Asn. No effect was detected either for the Glu366Asp amino acid change, but the analysis of patient’s blood mRNA showed that this variant (c.1098G>T) affects splicing leading to a deletion of 49 amino acids p.Ile318\_Glu366del (**Figure S2G**). We used this strategy to test additional variants reported in databases and showed that Arg158His, affecting a highly conserved amino acid position but reported twice in GnomAD, does not affect *DYRK1A* protein. Ala277Pro, reported as pathogenic in ClinVar, as well as Gly171Arg, Leu241Pro and Pro290Arg, reported initially as VUS in ClinVar, affect both *DYRK1A* level and autophosphorylation (**Figure S8A-B**, **Table S3**, **Figure S5B**). None of the missense variants appears to affect *DYRK1A* cellular localization, contrary to Arg413fs variant or changes in NLS domains (**Supplementary Text**, **Figure S8C**). However, we observed an aggregation of *DYRK1A* proteins with the distal frameshift variant Ser660fs (**Figure 2C**), preventing a correct measure of the protein and autophosphorylation levels.

### *Identification of a DNAm signature associated with DYRK1A pathogenic variants*

To determine if *DYRK1A* is associated with specific changes in genome-wide DNA methylation (DNAm) in blood, we used methylation array analysis. We compared DNAm in blood for a subset (*discovery*) of our cohort carrying pathogenic LoF variants in *DYRK1A* with age- and sex-matched neurotypical controls and identified n=402 differentially methylated CpG sites corresponding to 165 RefSeq genes (**Table S4, Figure 3A-B**). The sensibility and specificity of the score (0-1) derived from this signature was validated using additional individuals with *DYRK1A* truncating variants (*validation*), additional controls and individuals with pathogenic variants in other genes (**Table S5; Figure 3C**). Next, we scored the samples with missense variants and found that six classified positively (p.Asp287Val, p.Ser311Phe, p.Arg467Gln, p.Gly168Asp, p.Ile305Arg and p.Ser324Arg) and three negatively (p.Arg255Gln, p.Tyr462His, p.Thr588Asn) (**Figure 3B-C, Table S5, Figure S9**). The sample with the distal frameshift variant p.Ser660fs classified as DNAm positive, with a high score (0.92). The sample with the p.Gly486Asp variant clustered out from both *DYRK1A* cases and controls and its methylation profile was even opposite to *DYRK1A* LoF cases (increased methylation at sites decreased in LoF cases and vice versa, **Figure S9, S11**), suggesting this variant might have a gain-of-function (GoF) effect. A notable feature of these GoF CpG sites is that they tended to cluster together, as for instance in the *HIST1H3E* promoter (**Table S4**).

### ***Integration of the different tools to reclassify variants***

We integrated the clinical score, *in silico* predictions, functional assays results and DNAm score to evaluate the pathogenicity of the variants and reclassify them according to ACMG/AMP categories (**Figure 4, Table S6**). We found that variants p.Gly168Asp, p.Asp287Val, p.Ile305Arg, p.Ser311Phe and p.Arg467Gln, identified in individuals with intermediate to high CS<sub>DYRK1A</sub> scores, led to reduced protein level as well as an absence of autophosphorylation activity, which was previously described for three of them<sup>33,37</sup>. All classified as DNAm-positive, definitively supporting their pathogenicity. For the p.Ser324Arg variant, identified *de novo* in a patient with an intermediate CS<sub>DYRK1A</sub> score, we observed only a slight decrease of *DYRK1A* stability and a partial decrease of its autophosphorylation ability. The binary nature of the DNAm signature, showing a positive score, definitively supports its pathogenic effect.

The p.Arg255Gln and p.Tyr462His variants were identified in individuals with low CS<sub>DYRK1A</sub> score, they had relatively high CADD score (24 and 29.6) but affect amino acids not highly conserved. They had no effect on protein level, autophosphorylation and cellular localization of *DYRK1A* and classified DNAm-negative, and were therefore both considered as likely benign. While parental DNA was not available to test the inheritance of p.Arg255Gln, the variant p.Tyr462His occurred *de novo*. This individual has an affected brother who does not carry the variant, and exome sequencing of the whole family failed to identify additional promising variants, even taken into account the possibility of two different origins for the brothers. This remains puzzling and WGS is ongoing for both to go further. Another *de novo* variant affecting the same position p.Tyr462Cys was identified in a girl with mild developmental delay, hypotonia and hypermobility without facial dysmorphism, who finally obtained another molecular diagnosis (personal communication Sander Stegmann, Maastricht University Medical Center). The p.Thr588Asn variant, previously reported as likely pathogenic<sup>6</sup>, affects a mildly conserved amino acid and appears to have no effect on mRNA or protein level and function, as described by others<sup>33,37</sup>. To go further, we tested the ability of the mutant Thr588Asn *DYRK1A* to phosphorylate MAPT on its Thr212 and confirmed it does not affect its kinase activity (**Figure S10**). Moreover, a knock-in mouse model generated for Thr588Asn failed to present any decrease of kinase activity and any obvious behavioral phenotype (**Supplementary Text, Figure S12**). In addition, if this variant was not reported in gnomAD, two other amino acid changes are reported at the same position: p.Thr588Pro, (44 times including once at the homozygote state) and p.Thr588Ala (once). All these arguments plus the fact that the patient's DNAm signature was negative were convincing enough to reclassify this variant as likely benign. The fact that it occurred *de novo* in a girl with a high CS<sub>DYRK1A</sub>

(15.5/20) remains puzzling, while no additional promising variants were identified in trio-exome sequencing data and no positive classification was found using ~20 DNAm signatures available. However, the girl also presents additional manifestations unusual for *DYRK1A* syndrome such as truncal obesity.

Only one sample showed a DNAm profile different from both controls and individuals with *DYRK1A* syndrome (**Figure 3A, Figure S9 and S11**). This individual has a low  $CS_{DYRK1A}$ , presenting relative macrocephaly and ASD without ID and carries a *de novo* p.Gly486Asp variant. This variant was previously reported in another individual with NDD<sup>38</sup>, but it was not possible to obtain DNA, clinical or inheritance information. No significant change in protein and autophosphorylation level was observed for this variant and analysis of MAPT Thr212 phosphorylation failed to confirm the potential GoF effect (**Figure S10**).

We characterized the consequences of a distal frameshift *de novo* variant, p.Ser660fs. Its overexpression leads to cytoplasmic aggregation of *DYRK1A*, which makes it difficult to quantify the real effect on protein level, autophosphorylation or kinase activity (**Figure S8 and S10**). However, its DNAm overlaps those of other individuals with truncating variants located further upstream in the protein, confirming its pathogenic effect (**Figure S9**). To test if these aggregations could be driven by the novel C-terminal extension added by the frameshift variant (43 amino acids), we introduced nonsense variants at the same position (Ser660\* and Ser661\*). As no aggregate was detected (**Figure 2C & S8D**), we concluded that the C-terminal extension is responsible for the self-aggregation of the mutant *DYRK1A* protein. Interestingly, the two truncating variants Ser660\* and Ser661\* did not affect *DYRK1A* level, localization or autophosphorylation (**Figure S8A-B**).

## DISCUSSION

Here we report clinical manifestations of 34 novel patients with clear loss-of-function (LoF) variants in *DYRK1A*, refining the clinical spectrum associated with *DYRK1A* syndrome. We used recurrent signs present in individuals to establish a clinical score, which may seem outdated in the era of pangenomic approaches but is in fact very useful to interpret variants of unknown significance identified by these approaches. Indeed, here we demonstrated that the combination of clinical data together with *in silico* and *in vitro* observations are essential to interpret variants accurately.

Since *DYRK1A* is a highly conserved gene in vertebrates, we assumed that *in silico* predictive tools using conservation calculated mainly from vertebrates might overestimate the potential pathogenicity of missense variants. We showed that deeper conservation analyses using additional taxa are useful to improve the predictions for missense variants. However, *in silico* analyses have their limitations, and functional assays are essential to assess variant effect conclusively. We therefore tested the effect of 17 variants and showed that ten of them decreased both *DYRK1A* protein level and *DYRK1A* autophosphorylation level. The remaining variants showed no effect on protein function (**Figure 4**). However, the absence of effects observed during series of functional tests does not totally exclude a potential effect.

Over the past five years, several studies have found patients with specific monogenic disorders involving genes encoding epigenetic regulatory proteins are associated with DNAm signatures in blood. The advantage of such signatures is the high rate of clear classification (positive/pathogenic vs negative/benign) they provide for variants. Considering the potential role played by *DYRK1A* in epigenetic regulation<sup>22,23,39</sup>, we tested whether *DYRK1A* LoF leads to such a DNAm profile and identified a *DYRK1A* DNAm signature with high sensitivity and specificity (**Figure 3**). We undertook this work in whole-blood (as most signature work is done) due its clinical availability and *in silico* tools to account for cell proportion differences. We expect many of these changes to be blood-specific in patients with *DYRK1A* pathogenic variants. However,



enough DNAm changes may overlap other tissues for the blood signature to have cross-tissue utility for variant classification, as we found for fibroblasts in Sotos Syndrome<sup>18</sup>. The combination of clinical score (CS<sub>DYRK1A</sub>), *in silico* predictions, functional assays and DNAm signature allow to reclassify ten missense variants as pathogenic. Three variants were considered as likely benign: a variant located in the catalytic domain whose inheritance was unknown, p.Arg255Gln, and two *de novo* variants located at the end or outside of this domain: p.Tyr462His and p.Thr588Asn.

Still based on methylation data, we suspected a gain-of-function (GoF) effect for another *de novo* variant located outside the catalytic domain: p.Gly486Asp. We have already shown that DNAm profiles at gene-specific signature sites provide a functional readout of each variants effect, GoF activity. Indeed, in previous work, we found the same pattern for a patient with a missense variant in *EZH2*, typically associated with Weaver syndrome. The patient, presenting undergrowth rather than overgrowth characteristic of Weaver syndrome, had an opposite DNAm profile to *EZH2* cases relative to controls and carried a missense variant which was demonstrated to increase *EZH2* activity<sup>19</sup>. In our case, we could not confirm the putative GoF effect of Gly486Asp by measuring MAPT-Thr212 phosphorylation. Arranz et al. observed on the contrary an increase of *DYRK1A* kinase activity<sup>37</sup> for this variant, but they reported significant increase for five additional variants, including one present four times in GnomAD (Arg528Trp), which might question the sensitivity of the test.

We identified a *de novo* distal frameshift variant in the last exon of *DYRK1A* leading to *DYRK1A* aggregation *in vitro*, which needs to be confirmed *in vivo*. Interestingly, nonsense changes introduced at this position (aa660 and 661) lead to the expression of a protein which seems to be stable, does not aggregate, and maintains its autophosphorylation capacity and ability to phosphorylate MAPT (**Figure 2C, S8C-D, S10**). Therefore, we think that distal truncating variants should be interpreted with caution, especially when they escape to NMD. Three additional such distal variants are reported in individuals with ID/NDD in ClinVar/literature (**Table S8**) for which DNAm analysis on blood samples would be interesting to perform.

In conclusion, we developed a combination of tools efficient to interpret variants identified in *DYRK1A*. We showed that missense variants located outside and inside the catalytic domain as well as variants leading to distal premature stop codon are not necessarily pathogenic. These results illustrate that variants in *DYRK1A*, as well as in other NDD causative genes, should be interpreted with caution, even if they occur *de novo*. In the future, we recommend performing DNAm analysis if blood DNA sample is available or, if not, *in vitro* testing of variant effect on *DYRK1A* autophosphorylation.

## DATA AVAILABILITY

Variants were submitted to ClinVar or to Decipher database (as indicated in Table 1). Additional data are available upon request.

## ACKNOWLEDGEMENTS

The authors would like to thank the families for their participation and support. The authors also thank the Agence de Biomédecine, Fondation APLM, Fondation Maladies Rares and Fondation Jérôme Lejeune for financial support. We also thank the Centre National de Génotypage, the diagnostic laboratories of Hôpitaux Universitaires de Strasbourg (HUS), clinical genetic residents, the GenomEast sequencing and the molecular biology platforms of IGBMC.

## CONFLICTS OF INTEREST

None

## ETHICS DECLARATION

Individuals were referred by clinical geneticists for genetic testings as part of routine clinical care. All patients enrolled and/or their legal representative have signed informed consent for research use and authorization for publication, including for photograph publication for those included in **Figure S3**. All the institutions received local IRB approval to use these data in research purpose. The main IRB approval was obtained from the Ethics Committee of the Strasbourg University Hospital (CCPPRB). For experiments performed in animal, they were all done in compliance with the ARRIVE Guidelines and were non-invasive procedures approved by Com’Eth, the local ethic committee.

## WEB RESSOURCES

The URLs for online tools and data presented herein are:

CADD: <https://cadd.gs.washington.edu/>

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

Clustal Omega: <https://www.ebi.ac.uk/Tools/msa/clustalo/>

dbSNP: <http://www.ncbi.nlm.nih.gov/projects/SNP/>

Decipher: <https://decipher.sanger.ac.uk/>

GnomAD: <http://gnomad.broadinstitute.org/>

Integrative Genomics Viewer (IGV): <http://www.broadinstitute.org/igv/>

Mutation Nomenclature: <http://www.hgvs.org/mutnomen/recs.html>

NnsplICE: [https://www.fruitfly.org/seq\\_tools/splice.html](https://www.fruitfly.org/seq_tools/splice.html)

OMIM: <http://www.omim.org/>

UCSC: <http://genome.ucsc.edu/>

CADD score: <https://cadd.gs.washington.edu/>

OrthoInspector database: <https://www.lbgi.fr/orthoinspectorv3/databases>

## BIBLIOGRAPHY

1. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature*. 2017;542(7642):433-438. doi:10.1038/nature21062
2. Gonzalez-Mantilla AJ, Moreno-De-Luca A, Ledbetter DH, Martin CL. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *JAMA Psychiatry*. 2016;73(3):275-283. doi:10.1001/jamapsychiatry.2015.2692
3. Møller RS, Kübart S, Hoeltzenbein M, et al. Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. *Am J Hum Genet*. 2008;82(5):1165-1170. doi:10.1016/j.ajhg.2008.03.001
4. Courcet J-B, Faivre L, Malzac P, et al. The DYRK1A gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. *J Med Genet*. 2012;49(12):731-736. doi:10.1136/jmedgenet-2012-101251
5. van Bon BWM, Coe BP, Bernier R, et al. Disruptive de novo mutations of DYRK1A lead to a syndromic form of autism and ID. *Mol Psychiatry*. 2016;21(1):126-132. doi:10.1038/mp.2015.5
6. Bronicki LM, Redin C, Drunat S, et al. Ten new cases further delineate the syndromic intellectual disability phenotype caused by mutations in DYRK1A. *Eur J Hum Genet*. 2015;23(11):1482-1487. doi:10.1038/ejhg.2015.29
7. Ruaud L, Mignot C, Guët A, et al. DYRK1A mutations in two unrelated patients. *Eur J Med Genet*. 2015;58(3):168-174. doi:10.1016/j.ejmg.2014.12.014
8. Ji J, Lee H, Argiropoulos B, et al. DYRK1A haploinsufficiency causes a new recognizable syndrome with microcephaly, intellectual disability, speech impairment, and distinct facies. *Eur J Hum Genet*. 2015;23(11):1473-1481. doi:10.1038/ejhg.2015.71
9. Meissner LE, Macnamara EF, D’Souza P, et al. DYRK1A pathogenic variants in two patients with syndromic intellectual disability and a review of the literature. *Mol Genet Genomic Med*. Published online November 7, 2020:e1544. doi:10.1002/mgg3.1544
10. O’Roak BJ, Vives L, Fu W, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science*. 2012;338(6114):1619-1622. doi:10.1126/science.1227764

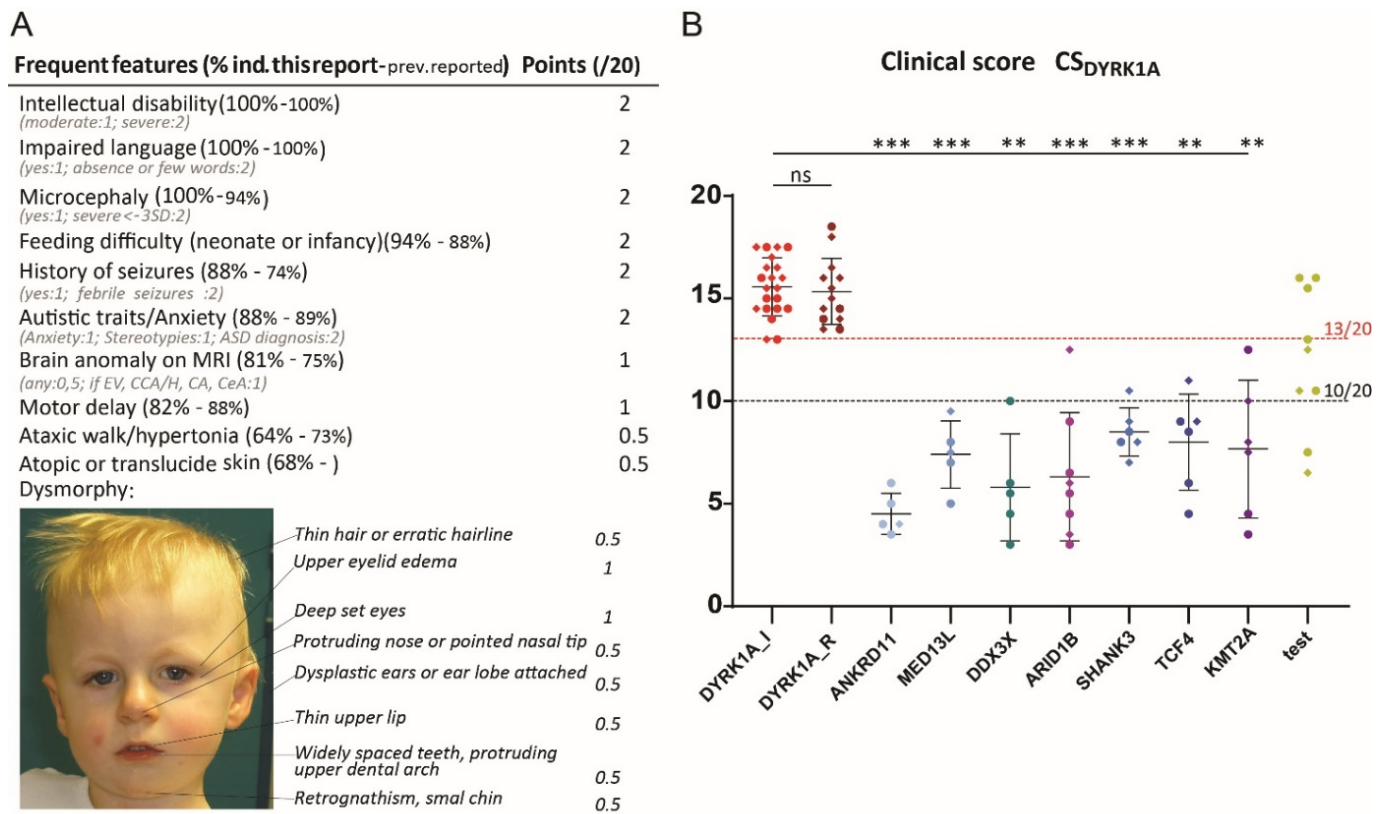
11. Earl RK, Turner TN, Mefford HC, et al. Clinical phenotype of ASD-associated DYRK1A haploinsufficiency. *Mol Autism*. 2017;8:54. doi:10.1186/s13229-017-0173-5
12. Himpel S, Panzer P, Eirnbter K, et al. Identification of the autophosphorylation sites and characterization of their effects in the protein kinase DYRK1A. *Biochem J*. 2001;359(Pt 3):497-505. doi:10.1042/0264-6021:3590497
13. Hämmerle B, Elizalde C, Tejedor FJ. The spatio-temporal and subcellular expression of the candidate Down syndrome gene Mnb/Dyrk1A in the developing mouse brain suggests distinct sequential roles in neuronal development. *Eur J Neurosci*. 2008;27(5):1061-1074. doi:10.1111/j.1460-9568.2008.06092.x
14. Tejedor FJ, Hämmerle B. MNB/DYRK1A as a multiple regulator of neuronal development. *FEBS J*. 2011;278(2):223-235. doi:10.1111/j.1742-4658.2010.07954.x
15. Duchon A, Herault Y. DYRK1A, a Dosage-Sensitive Gene Involved in Neurodevelopmental Disorders, Is a Target for Drug Development in Down Syndrome. *Front Behav Neurosci*. 2016;10:104. doi:10.3389/fnbeh.2016.00104
16. Woods YL, Cohen P, Becker W, et al. The kinase DYRK phosphorylates protein-synthesis initiation factor eIF2Bepsilon at Ser539 and the microtubule-associated protein tau at Thr212: potential role for DYRK as a glycogen synthase kinase 3-priming kinase. *Biochem J*. 2001;355(Pt 3):609-615. doi:10.1042/bj3550609
17. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi:10.1038/gim.2015.30
18. Choufani S, Cytrynbaum C, Chung BHY, et al. NSD1 mutations generate a genome-wide DNA methylation signature. *Nat Commun*. 2015;6:10207. doi:10.1038/ncomms10207
19. Choufani S, Gibson WT, Turinsky AL, et al. DNA Methylation Signature for EZH2 Functionally Classifies Sequence Variants in Three PRC2 Complex Genes. *Am J Hum Genet*. 2020;106(5):596-610. doi:10.1016/j.ajhg.2020.03.008
20. Chater-Diehl E, Ejaz R, Cytrynbaum C, et al. New insights into DNA methylation signatures: SMARCA2 variants in Nicolaides-Baraitser syndrome. *BMC Med Genomics*. 2019;12(1):105. doi:10.1186/s12920-019-0555-y
21. Aref-Eshghi E, Bend EG, Colaiacovo S, et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. *Am J Hum Genet*. 2019;104(4):685-700. doi:10.1016/j.ajhg.2019.03.008
22. Jang SM, Azebi S, Soubigou G, Muchardt C. DYRK1A phosphorylates histone H3 to differentially regulate the binding of HP1 isoforms and antagonize HP1-mediated transcriptional repression. *EMBO Rep*. 2014;15(6):686-694. doi:10.15252/embr.201338356
23. Li S, Xu C, Fu Y, et al. DYRK1A interacts with histone acetyl transferase p300 and CBP and localizes to enhancers. *Nucleic Acids Res*. 2018;46(21):11202-11213. doi:10.1093/nar/gky754
24. Redin C, Gérard B, Lauer J, et al. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *J Med Genet*. 2014;51(11):724-736. doi:10.1136/jmedgenet-2014-102554
25. Carion N, Briand A, Cuisset L, Pacot L, Afenjar A, Bienvenu T. Loss of the KH1 domain of FMR1 in humans due to a synonymous variant causes global developmental retardation. *Gene*. 2020;753:144793. doi:10.1016/j.gene.2020.144793
26. Nasser H, Vera L, Elmaleh-Bergès M, et al. CDK5RAP2 primary microcephaly is associated with hypothalamic, retinal and cochlear developmental defects. *J Med Genet*. 2020;57(6):389-399. doi:10.1136/jmedgenet-2019-106474
27. Balak C, Benard M, Schaefer E, et al. Rare De Novo Missense Variants in RNA Helicase DDX6 Cause Intellectual Disability and Dysmorphic Features and Lead to P-Body Defects and RNA Dysregulation. *Am J Hum Genet*. 2019;105(3):509-525. doi:10.1016/j.ajhg.2019.07.010
28. Quartier A, Chatrousse L, Redin C, et al. Genes and Pathways Regulated by Androgens in Human Neural Cells, Potential Candidates for the Male Excess in Autism Spectrum Disorder. *Biol Psychiatry*. Published online January 9, 2018. doi:10.1016/j.biopsych.2018.01.002
29. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189-1191. doi:10.1093/bioinformatics/btp033
30. Kress A, Lecompte O, Poch O, Thompson JD. PROBE: analysis and visualization of protein block-level evolution. *Bioinformatics*. 2018;34(19):3390-3392. doi:10.1093/bioinformatics/bty367
31. Mattioli F, Isidor B, Abdul-Rahman O, et al. Clinical and functional characterization of recurrent missense variants implicated in THOC6-related intellectual disability. *Hum Mol Genet*. 2019;28(6):952-960. doi:10.1093/hmg/ddy391
32. Quartier A, Courraud J, Thi Ha T, et al. Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment. *Hum Mutat*. 2019;40(11):2021-2032. doi:10.1002/humu.23836
33. Widowati EW, Ernst S, Hausmann R, Müller-Newen G, Becker W. Functional characterization of DYRK1A missense variants associated with a syndromic form of intellectual deficiency and autism. *Biol Open*. 2018;7(4). doi:10.1242/bio.032862
34. Lee K-S, Choi M, Kwon D-W, et al. A novel de novo heterozygous DYRK1A mutation causes complete loss of DYRK1A function and developmental delay. *Sci Rep*. 2020;10(1):9849. doi:10.1038/s41598-020-66750-y
35. Blackburn ATM, Bekheirnia N, Uma VC, et al. DYRK1A-related intellectual disability: a syndrome associated with congenital anomalies of the kidney and urinary tract. *Genet Med*. 2019;21(12):2755-2764. doi:10.1038/s41436-019-0576-0
36. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315. doi:10.1038/ng.2892
37. Arranz J, Balducci E, Arató K, et al. Impaired development of neocortical circuits contributes to the neurological alterations in DYRK1A haploinsufficiency syndrome. *Neurobiol Dis*. 2019;127:210-222. doi:10.1016/j.nbd.2019.02.022
38. Dang T, Duan WY, Yu B, et al. Autism-associated Dyrk1a truncation mutants impair neuronal dendritic and spine growth and interfere with postnatal cortical development. *Mol Psychiatry*. 2018;23(3):747-758. doi:10.1038/mp.2016.253

39. Lepagnol-Bestel A-M, Zvara A, Maussion G, et al. DYRK1A interacts with the REST/NRSF-SWI/SNF chromatin remodelling complex to deregulate gene clusters involved in the neuronal phenotypic traits of Down syndrome. *Hum Mol Genet.* 2009;18(8):1405-1414. doi:10.1093/hmg/ddp047

## FIGURE LEGENDS

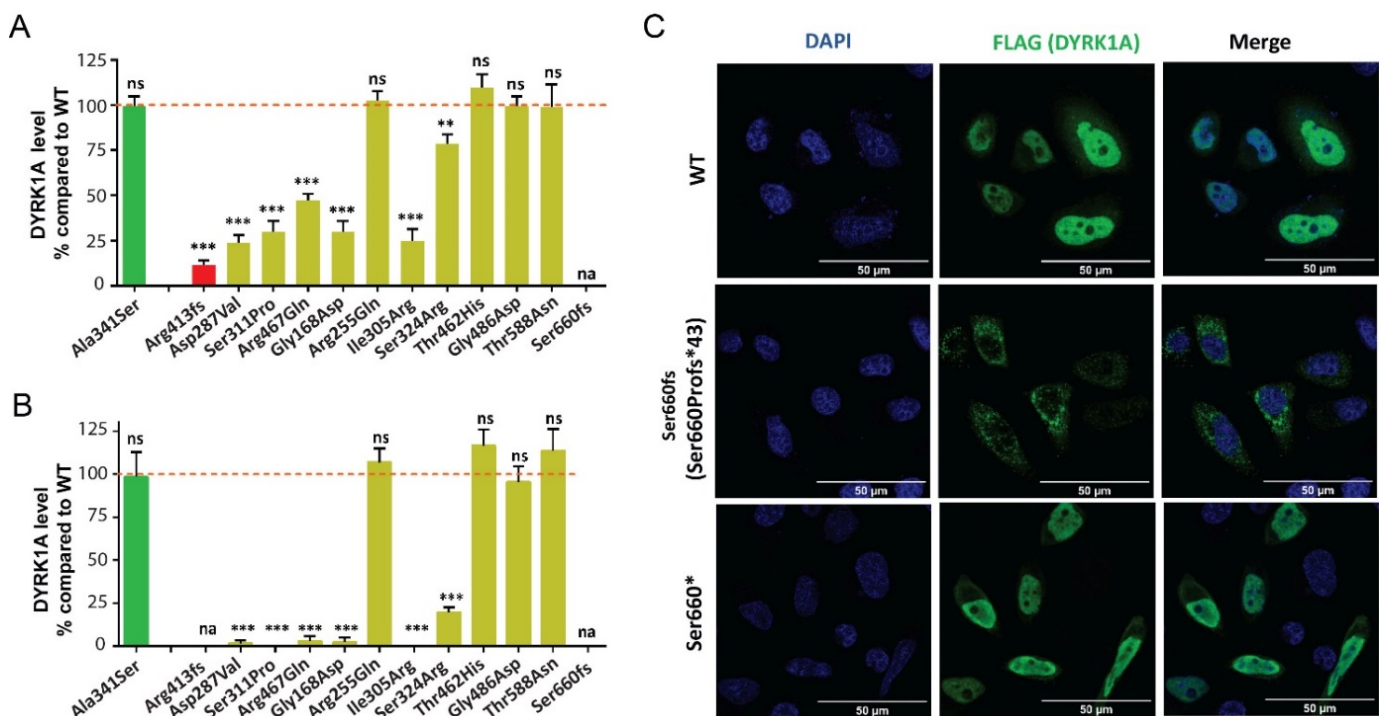
### Figure 1. Clinical score for Intellectual Disability associated to *DYRK1A* haploinsufficiency

(A) Clinical score out of 20 points established according to the most recurrent clinical features presented by patients (the weight assigned to each symptom being based on its recurrence): clinical symptoms are out of 15 points, while the facial appearance is out of 5 points. EV: enlarged ventricles; CCA/H: corpus callosum agenesis or hypoplasia, CA: cerebral atrophy, CeA: cerebellar atrophy (B) Clinical scores calculated for individuals carrying pathogenic variants in *DYRK1A* reported here and for whom photographs were available (n=21)(initial cohort, *DYRK1A\_I*, scores 13- 17.5 with a mean of 15.5), the previously published individuals (replication cohort, *DYRK1A\_R*, scores 13.5-18.5, mean=15.3) and the individuals affected with other frequent monogenic forms of ID, associated to variants in *ANKRD11*, *MED13L*, *DDX3X*, *ARID1B*, *SHANK3*, *TCF4* or *KMT2A* (scores 3-12.5, mean=7). The clinical score for the individuals carrying missense or distal frameshift variants are indicated in yellow (test); the threshold of  $CS_{DYRK1A} \geq 13$  appeared to be discriminant between individuals with LoF variants in *DYRK1A* (all  $\geq 13$ ) and individuals suffering from another form of ID (all  $< 13$ ). A score above this threshold was therefore considered “highly suggestive”. We classified individuals with  $CS_{DYRK1A} < 10$  as “poorly evokative” and individuals with a  $CS_{DYRK1A}$  comprised between 10 and 13 as “intermediate”. Brown-Forsythe and Welch ANOVA tests with Dunnett’s T 3 multiple comparisons test were performed. ns: not significant; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ , error bars represent SD.



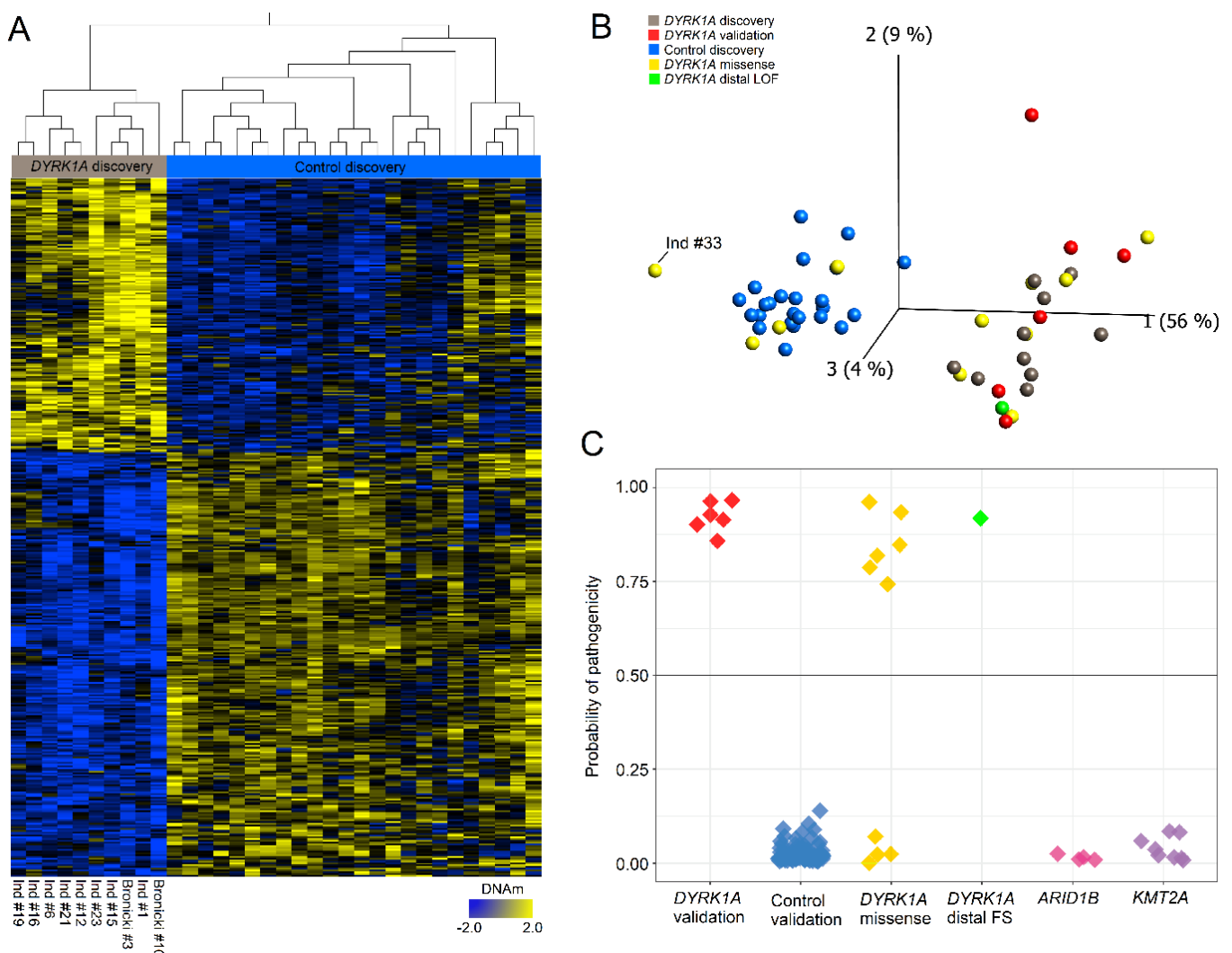
## Figure 2. Expression, localization and Tyr321 phosphorylation of DYRK1A mutant proteins

(A) Level of variant DYRK1A proteins expressed in HeLa, HEK293 and COS cells transiently transfected with DYRK1A constructs. Protein levels were normalized on the level of GFP proteins (expressed from a cotransfected pEGFP plasmid). Quantifications were performed on a total of  $n \geq 9$  series of cells ( $n \geq 3$  HeLa cells,  $n \geq 3$  HEK293 and  $n \geq 3$  COS cells) using ImageJ software. One-way ANOVA with multiple comparison test was performed to compare the level of variant DYRK1A proteins to the level of wild-type DYRK1A protein (orange dashes), applying Bonferroni's correction: ns: not significant; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; error bars represent SEM, standard error of the mean; in green, the variant from gnomAD, in red, a truncating variant and in gold, the variants tested in this study (B) DYRK1A's ability to autophosphorylate on Tyr321 was tested in HEK293 cells ( $n=3$ ) by immunoprecipitations with anti-DYRK1A followed by an immunoblot using anti-phospho-HIPK2 as described in Widowati et al. DYRK1A phospho-Tyr321 levels were normalized with DYRK1A total level (orange dashes). Variant DYRK1A phospho-Tyr321 levels were normalized with total DYRK1A protein levels and expressed as percentage of wild-type level. One-way ANOVA test was performed to compare variants to wild-type DYRK1A levels. ns: not significant; \*\*\* $p < 0.001$ ; error bars represent SEM, standard error of the mean (C) Immunofluorescence experiment showing that Ser660fs (alias Ser660Profs\*43) variant leads to DYRK1A protein aggregation when overexpressed in HeLa cells, using a FLAG-tagged DYRK1A proteins carrying Ser660Profs43. No aggregation was observed for the Ser660\* variant.



**Figure 3. DNA methylation signature of *DYRK1A* loss-of-function functionally classifies *DYRK1A* VUS.**

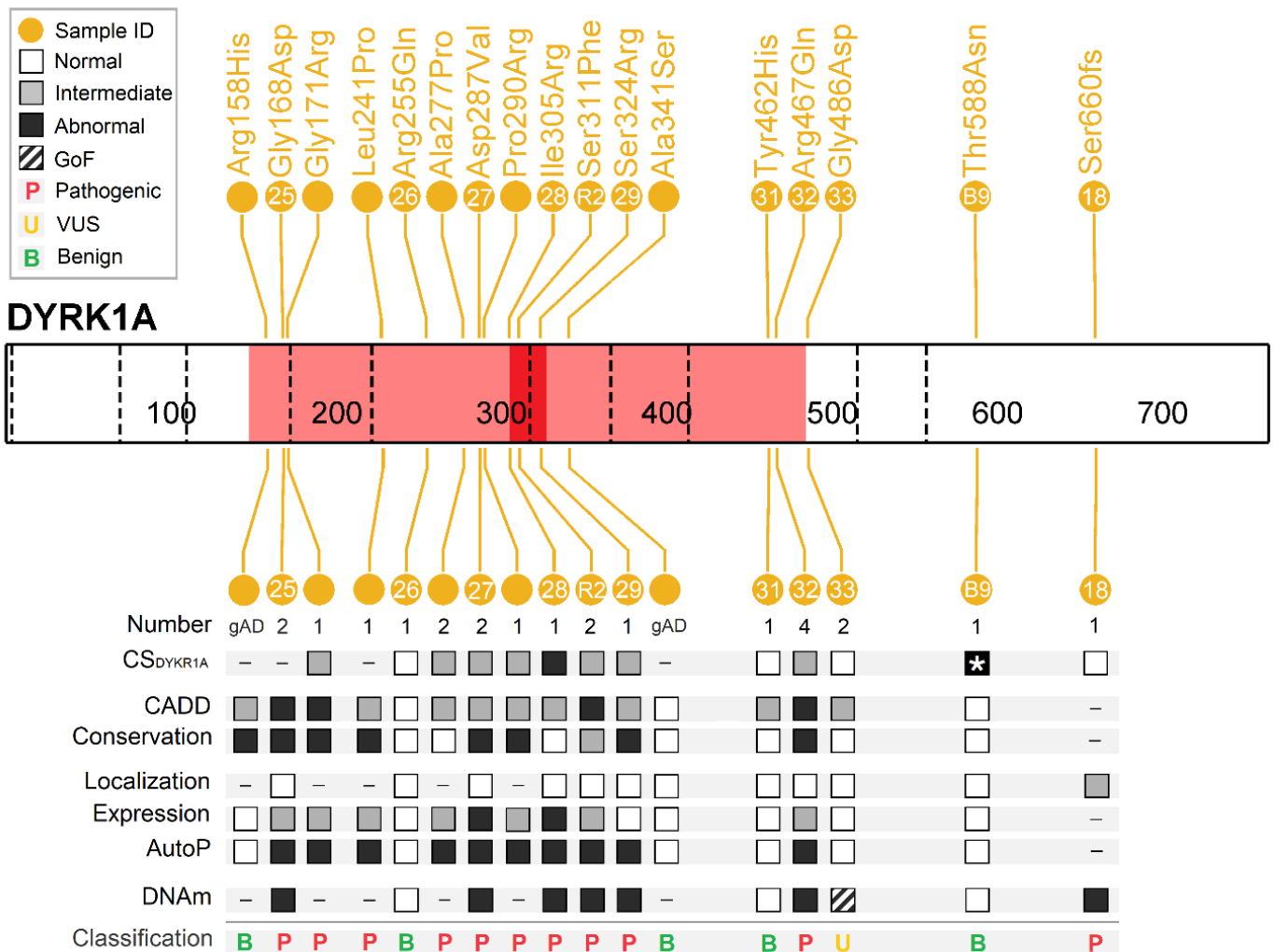
(A) Heatmap showing the hierarchical clustering of discovery *DYRK1A* LoF cases (n = 10) and age- and sex-matched neurotypical discovery controls (n = 24) used to identify the 402 differentially methylated signature sites shown. Each row corresponds to a CpG site differentially methylated (DM) and the color gradient represents the normalized DNA methylation value from -2.0 (blue) to 2.0 (yellow) at each site. DNA methylation at these sites clearly separate discovery cases (grey) from discovery controls (blue). Euclidian distance metric is used for the clustering dendrogram. (B) Principal components analysis (PCA) visualizing the DNAm profiles of the study cohort at the 402 signature sites. Validation of *DYRK1A* LoF cases (not used to define the signature sites; red) cluster with discovery cases, while missense (yellow) and distal LoF (green) variants cluster with either cases or controls. Ind #33 (Gly486Asp) has an opposite DNAm profile to *DYRK1A* LoF cases at these sites, suggesting a GoF. (C) Support vector machine (SVM) classification model based on the DNA methylation values in the discovery groups. Each sample is plotted based on its scoring by the model. All samples are clearly positive (>0.5) or negative (<0.5). All *DYRK1A* validation cases from our cohort (n=6) classified positively, all control validation cases (n=94) classified negatively. Missense variants classified clearly positively or negatively (yellow), the distal frameshift variant (Ind #18, c.1978del)(green), analyzed in duplicate, classified positively. Pathogenic *ARID1B* (Coffin-Siris syndrome) and *KMT2A* (Wiedemann Steiner syndrome) also classified negatively.





**Figure 4. Summary of the analysis performed to reclassify variants in *DYRK1A***

Representation of the *DYRK1A* protein (the kinase domain is indicated in red and the catalytic domain in dark red) with the positions of the different variants tested with the sample #ID of the individuals indicated inside the circles. Number: number of individuals with ID reported with the variant; gAD: variant reported in individuals from gnomAD; CS<sub>DYRK1A</sub> poorly (white), intermediate (grey) or highly (black) evocative, or unknown (-); the white star indicates that the individual presents a high CS<sub>DYRK1A</sub> score but also additional clinical manifestations unusual for *DYRK1A* syndrome; CADD below 25 (white), between 25 and 30 (grey) or above 30 (black); conservation: highly conserved V=100%, M>90%, O>80% (black), moderately V=100%, M>90%, O<80% (grey) or midly V=100% M<90%, O<80% (white); Expression or autophosphorylation being normal (white), intermediate decreased (grey), strongly decreased (black); Localization was normal (white), affected (grey) or not tested (-); DNAmethylation positive (black), negative (white), suggestive of a GoF effect (hashed) or not tested (-). Final classification: Pathogenic (P), Benign/Likely benign (B), Unknown significance (U).



## **Table 1. List of variants identified in *DYRK1A* in individuals with intellectual disability**

This list includes variants identified in the 44 individuals never reported as well as variants previously reported in six individuals<sup>6,7</sup> for whom we collected biological samples and additional clinical information. The truncating variants clearly pathogenic are represented in red; the missense variants and the distal frameshift variant to test are in gold. Variants are reported according to standardized nomenclature defined by the reference human genome GRCh37/hg19 and the *DYRK1A* isoform NM\_001396.4; del: deletion of the gene ; Trans.: translocation interrupting the gene ; Ns: nonsense ; Fs: frameshift, Spl: splice ; Ms: missense variants; Infr.: inframe deletion; TES: targeted exome sequencing of ID genes TES<sup>1</sup> : panel of ID genes from Carion et al.<sup>47</sup>; TES<sup>2</sup> : panel of ID gene adapted from Redin et al.<sup>46</sup>; TES<sup>3</sup> : panel of 44 ID genes; TES<sup>4</sup>: panel of microcephaly genes from Nasser et al.<sup>48</sup> ; CES: clinical exome sequencing, ES: exome sequencing, Sanger: Sanger sequencing, CGH-array: comparative genomic hybridization-array. <sup>a</sup>: the consequences of c.1098G>T is p.Ile318\_Glu366del instead of p.Glu366Asp; M : male ; F : female ; *DYRK1A\_I* : initial cohort used to establish *DYRK1A* clinical score (CS<sub>*DYRK1A*</sub>) ; *DYRK1A\_R* : replication cohort used to confirm the relevance of the CS<sub>*DYRK1A*</sub> ; Disc : discovery cohort used to establish DNA methylation signature (DNAm) ; Valid: validation cohort used to confirm DNA methylation signature (DNAm); test : variants tested for pathogenicity using DNAm.



Variant					Individual			Reporting		Analyses performed				
GRCh37 (Chr21)	NM_001396.4	NP_001387.2	Type	Method	Ind. number	Sex	Inheritance	this individual	Additional ind. (ClinVar)	CS <sub>DYRK1A</sub>	<i>In silico</i>	mRNA	<i>in vitro</i>	DNAm
g:38481804_40190458del (DYRK1A ; >10 other genes)	NA	NA	del.	CGH-array	Ind #1	M	de novo	NA	NA	DYRK1A_I	-	yes	-	Disc.
g:38722881_39426450del (DYRK1A; KCNJ2)	NA	NA	del.	CGH-array	Ind #2	F	de novo	NA	NA	DYRK1A_I	-	-	-	-
g:38302140_40041414del (DYRK1A ; >5 other genes)	NA	NA	del.	CGH-array	Ind #40	M	de novo	NA	NA	DYRK1A_I	-	-	-	-
t(9;21)(p12;q22)	between exon 2 & 3	NA	trans.	CGH-array	Ind #3	M	de novo	NA	NA	DYRK1A_I	-	-	-	-
g.38852961C>T	c.349C>T	p.Arg117*	Ns	ES	Ind #4	M	NA	NA	(5x) vcv000373087	DYRK1A_I	-	-	-	-
				ES	Ind #5	F	father mosaic	NA		DYRK1A_I	-	-	-	-
				ES	Ind #6	M	de novo	NA		DYRK1A_I	-	-	-	Disc.
g.38858865C>T	c.613C>T	p.Arg205*	Ns	TES <sup>2</sup>	Bronicki_#2	M	de novo	PMID: 25920557 (ClinVar_SCV000281731=SCV000196058)	(x7) vcv000162153	DYRK1A_R	-	-	-	-
				TES <sup>4</sup>	Ind #44	M	NA	ClinVar_scv001432338		DYRK1A_I	-	-	-	-
g.38862575C>T	c.763C>T	p.Arg255*	Ns	TES <sup>1</sup>	Ind #7	M	de novo	ClinVar_scv001712097	(5x) vcv000162152	DYRK1A_I	-	-	-	Valid.
				NA	Ind #43	F	NA	ClinVar_scv001432353		DYRK1A_I	-	-	-	-
g.38862611C>T	c.799C>T	p.Gln267*	Ns	ES	Ind# 35	M	de novo	Clinvar_SUB9815736	no	DYRK1A_I	-	-	-	-
g.38862748T>A	c.936T>A	p.Cys312*	Ns	CES	Ind #37	F	de novo	Clinvar_scv001712105	no	DYRK1A_I	-	-	-	-
g.38877655C>T	c.1309C>T	p.Arg437*	Ns	TES <sup>4</sup>	Ind #34	F	not in mother	NA	(7x) vcv000162158	DYRK1A_I	-	-	-	-
g.38877745C>T	c.1399C>T	p.Arg467*	Ns	TES <sup>4</sup>	Ind#42	F	de novo	ClinVar_scv001432351	(3x) vcv000204005	DYRK1A_I	-	-	-	-
g.38850510dup	c.235dup	p.Arg79fs	Fs	TES <sup>4</sup>	Ind #41	F	de novo	ClinVar_scv001432470	no	DYRK1A_I	-	-	-	-
g.38850565_38850566del	c.290_291 del	p.Ser97fs	Fs	CES	Ind #8	M	de novo	ClinVar_scv001712106	(2x) vcv000418949	DYRK1A_I	-	-	-	-
g.38850572_38850576del	c.297_301del	p.Leu100fs	Fs	NA	Ind #9	M	de novo	ClinVar_scv000485020	no	DYRK1A_I	-	-	-	-
g.38853089del	c.477del	p.Tyr159*	Fs	ES	Ind #10	F	de novo	Clinvar_scv001712094	no	DYRK1A_I	-	-	-	-
g.38862514_38862515del	c.702_703del	p.Cys235fs	Fs	ES	Ind #11	M	de novo	ClinVar_scv000965742	no	DYRK1A_I	-	yes	-	Valid.
g.38858873_38858876delinsGAA	c.621_624 delinsGAA	p.Glu208fs	Fs	TES <sup>2</sup>	Bronicki_#3	M	de novo	PMID: 25920557 (ClinVar_SCV000281736 = SCV000196059)	no	DYRK1A_R	-	-	-	Disc.
g.38862594del	c.782del	p.Leu261fs	Fs	TES <sup>2</sup>	Ind #12	F	de novo	ClinVar_scv001437790	no	DYRK1A_I	-	-	-	Disc.
g.38862656dup	c.844dup	p.Ser282fs	Fs	TES <sup>1</sup>	Bronicki_#8	M	de novo	PMID: 25920557 (ClinVar_SCV000196064)	no	DYRK1A_R	-	-	-	-
g.38865371del	c.1004del	p.Gly335fs	Fs	TES <sup>1</sup>	Ind #13	F	not in mother	Clinvar_scv001712087	no	DYRK1A_I	-	-	-	Valid.
g.38865375dup	c.1008dup	p.Pro337fs	Fs	ES	Ind #14	F	de novo	Decipher_351807	no	DYRK1A_I	-	-	-	-
g.38865400del	c.1033del	p.Trp345fs	Fs	CES	Ind #15	F	de novo	Clinvar_scv001712108	no	DYRK1A_I	-	-	-	Disc.

g.38868553dup	c.1232dup	p.Arg413fs	Fs	TES <sup>2</sup>	Bronicki_#10	F	de novo	PMID: 25920557 (ClinVar_SCV000196066)	no	DYRK1A_R	-	yes	yes	Disc.
g.38877616del	c.1270del	p.His424fs	Fs	ES	Ind #39	F	de novo	Clinvar_scv001712090	no	DYRK1A_I	-	-	-	-
g.38877679dup	c.1333dup	p.Thr445fs	Fs	ES	Ind #16	F	de novo	Clinvar_scv000778256	no	DYRK1A_I	-	-	-	Disc.
g.38877837del	c.1491delC	p.Ala498fs	Fs	ES	Ind #17	F	de novo	Clinvar_scv000494645	SCV000056592	DYRK1A_I	-	-	-	Valid.
g.38884520del	c.1978del	p.Ser660fs= p.Ser660Profs*43	Fs	TES <sup>1</sup>	Ind #18	F	de novo	Clinvar_scv001712088	no	test	-	yes	yes	test
g.38852939G>T	c.328-1G>T	p.?	Spl.	TES <sup>2</sup>	Ind #19	F	de novo	Clinvar_scv001437768	no	DYRK1A_I	-	yes	-	Disc.
g.38862475A>G	c.665-2A>G	p.?	Spl.	Sanger	Ind #20	M	de novo	Clinvar_scv001712086	SCV000492145	DYRK1A_I	-	-	-	Valid.
g.38862468_38862472del	c.665-9_665-5del	p.?	Spl.	Sanger	Ind #21	M	de novo	Clinvar_scv001437772	SCV000677027	DYRK1A_I	-	-	-	Disc.
			Spl.	TES <sup>2</sup>	Ind #36	M	de novo			DYRK1A_I	-	-	-	-
g.38862764G>C	951+1G>C	p.?	Spl.	TES <sup>1</sup>	Ind #38	M	de novo	Clinvar_scv001712091	no	DYRK1A_I	-	-	-	Valid.
g.38862767_38862770del	c.951+4_951+7del	p.?	Spl.	ES	Ind #22	M	de novo	Clinvar_scv000965731	SCV000709803	DYRK1A_I	-	yes	-	-
g.38877584A>G	c.1240-2A>G	p?	Spl.	ES	Ind #23	M	de novo	Clinvar_scv000966166	no	DYRK1A_I	-	-	-	Disc.
g.38877585_38877586insTAA	c.1240-1_1240insTAA	p.Glu414*	Spl.	TES <sup>4</sup>	Ind #24	F	de novo	Clinvar_scv001432455	no	DYRK1A_I	-	yes	-	-
g.38853115G>A	c.503G>A	p.Gly168Asp	Mis.	TES <sup>1</sup>	Ind #25	F	de novo	Clinvar_scv001712092	SCV000573105	test	yes	-	yes	test
g.38862576G>A	c.764G>A	p.Arg255Gln	Mis.	TES <sup>1</sup>	Ind #26	F	NA	Clinvar_scv001712093	no	test	yes	-	yes	test
g.38862672A>T	c.860A>T	p.Asp287Val	Mis.	ES	Ind #27	M	de novo	Clinvar_scv000598121	SCV001446739	test	yes	-	yes	test
g.38862726T>G	c.914T>G	p.Ile305Arg	Mis.	Sanger	Ind #28	F	de novo	Clinvar_scv001712110	no	test	yes	-	yes	test
g.38862744C>T	c.932C>T	p.Ser311Phe	Mis	NA	Ruauud_#2	M	de novo	PMID: 25641759 (ClinVar_SCV000586742)	SCV000520979	test	yes	-	yes	test
g.38865339T>A	c.972T>A	p.Ser324Arg	Mis.	TES <sup>3</sup>	Ind #29	M	de novo	Clinvar_scv000902439	no	test	yes	-	yes	test
g.38865465G>T	c.1098G>T <sup>a</sup>	p.Ile318_Glu366del	Infr.	ES	Ind #30	F	de novo	Decipher_434484	no	test	yes	yes	yes	-
g.38877730T>C	c.1384T>C	p.Tyr462His	Mis.	TES <sup>2</sup>	Ind #31	M	de novo	Clinvar_scv001437769	no	test	yes	-	yes	test
g.38877746G>A	c.1400G>A	p.Arg467Gln	Mis.	TES <sup>2</sup>	Ind #32	F	de novo	Clinvar_scv001437771	(3x) vcv000209150	test	yes	-	yes	test
g.38877803G>A	c.1457G>A	p.Gly486Asp	Mis.	ES	Ind #33	M	de novo	Clinvar_scv000747759	no	test	yes	-	yes	test
g.38884305C>A	c.1763C>A	p.Thr588Asn	Mis	ES	Bronicki_#9	F	de novo	PMID: 25920557 (ClinVar_SCV000965705= SCV000196065)	no	test	yes	yes	yes	test

## SUPPLEMENTARIES

### Supplementary methods - Antibodies

DYRK1A protein quantification (Western Blot)	mouse anti-FLAG antibody (1:1:000; Sigma Aldrich #F1804)
	mouse anti-GFP antibody (in house)
DYRK1A localization (Immunocytochemistry)	mouse anti-FLAG antibody (1:1:000; Sigma Aldrich #F1804)
Immunoprecipitation	rabbit anti-DYRK1A antibody (1:1000; Cohesion Biosciences #CPA1357)
Interaction with DCAF7	anti-WDR68 antibody (1:2500; abcam ab138490)
DYRK1A autophosphorylation	rabbit anti-phospho-HIPK2 antibody (1:1000, Thermofisher #PA5-13045)
	rabbit anti-DYRK1A antibody (1:1000; Cohesion Biosciences #CPA1357)
MAPT phosphorylation	mouse anti-TAU-5 antibody (Thermofisher #MA5-12808)
	rabbit anti-pTAU-T212 antibody (Thermofisher #44-740G)
	mouse anti-FLAG antibody (1:1:000; Sigma Aldrich #F1804)
	Mouse anti-GAPDH (Sigma #MAB374)

### Supplementary text

#### Details on splice and frameshift variants and their consequences on *DYRK1A* mRNA

Deletions encompassing *DYRK1A* and chromosomal rearrangement t(9;21)(p12;q22) interrupting the gene were reported in four individuals (**Ind #1-3, 40**). We identified recurrent nonsense variants p.Arg117\* (**Ind #4, #5, #6**), p.Arg205\* (**Ind #44**), p.Arg255\* (**Ind #7, #43**), p.Arg437\* (**Ind#34**) and p. p.Thr467\* (**Ind#42**) as well as novel nonsense variants (**Ind#35, #37**) or small indels (**Ind #8-17, #39, #41**), one of them occurring in the last exon affecting the distal region of the protein: p.Ser660fs (**Ind #18**). Seven of the variants identified are predicted to affect splice sites, half of them previously reported elsewhere (**Ind #19-24, #36, #38**). The remaining individuals (**Ind #25-33**) carry missense variants, three of them already reported: p.Gly168Asp, p.Asp287Val and p.Arg467Gln. We tested the consequences of splice variants identified on *DYRK1A* mRNA (**Figure S2**) when possible (blood or fibroblasts available) by RNA-sequencing (**Ind #19, 22, 24**) or RT-qPCR (**Ind #18**). We found that c.328-1G>T (**Ind #19**) leads to abnormal splicing events between exons 4 and 5 including intron 4 retention and use of alternative cryptic acceptor sites in exon 5 (**Figure S2A**). The c.951+4\_951+7 del (**Ind #22**) leads to retention of intron 7 or skipping of exon 7, both resulting in a premature truncation (**Figure S2B**). RNA sequencing performed on mRNA extracted from **Ind #24** fibroblasts revealed that the TAA insertion at the beginning of exon 10 was included in the transcripts, leading to a stop codon p.Glu414\*. *DYRK1A* mRNA levels were only slightly decreased in **Ind #22** and **#24** when compared to individual carrying truncating variant (**Ind #11**), suggesting that the aberrant transcripts escape nonsense mediated mRNA decay (NMD), at least partially (**Figure S2D**). The variant previously reported c.1232dup, p.Arg413fs (**Bronicki #10**)<sup>1</sup> is subjected to NMD as illustrated by the decrease of *DYRK1A* mRNA level in patient's fibroblasts, and the low level of mutant allele in cells, restored by NMD blocking agent (**Figure S2E**). However, it is worth noting that mutant transcripts carrying the distal frameshift c.1978del (**Ind #18**),

located in the last exon of the gene, escape to NMD and therefore result in a truncated protein p.Ser660fs having its entire kinase domain (**Figure S2F**).

### **Clinical manifestations in individuals with pathogenic variant in *DYRK1A***

All the individuals with clearly loss-of-function variants in *DYRK1A* present with moderate to severe ID except two (**Ind #3, #34**). Language was affected in all individuals, severely in most (no speech, or only few words or short sentences). Individuals tend to present failure to thrive, even sometimes from the uterine stage (20/32), especially on the weight gain (from -1 to -4SD). The size is less affected comprised between -0.5 to -2SD. Microcephaly is however a constant trait, although not always present from birth. All individuals except two (31/33) had feeding difficulties during the neonatal period (poor sucking, gastrostomy, etc), which can be very severe during infancy and can persist during childhood and even into adulthood (selective, smashed food only, etc). The large majority of individuals present a history of seizures (29/33) mainly including febrile episodes (n=18). Hypotonia was noted in only half of the individuals (16/29), but a majority had motor delay with a walk acquired after 18 months of age (27/33)(mean = 23 months), and the gait could continue to be unstable and ataxic (9). Hyperreflexia and hypertonia have been observed in some patients (12). Sleep disorders were reported in some patients (10). Behavioral manifestations observed in patients included anxiety (13) and autistic traits with stereotyped behaviours (24). A diagnosis of autism spectrum disorder (ASD) has been established in only four patients, but few have had the appropriate tests. Other behavioral manifestations such as water fascination (5), absence of fear (5) were noted. MRI revealed ventricular dilation (9), thin corpus callosum (9), as well as cortical or cerebellar atrophy (9). Other manifestations include gastrointestinal manifestations such as constipation (14) or gastroesophageal reflux (10), intestinal anomalies such as inguinal hernia (4) and urogenital anomalies already reported to be frequent<sup>2</sup>, including cryptorchidism (4). No obvious kidney anomalies have been reported but not all the individuals underwent a renal ultrasound. We observed a thin skin in a high frequency of patients (15), often associated with dermatitis or atopic skin after birth or during infancy (11), sometimes very pronounced. After reanalysis of literature, we found that atopic dermatitis was also reported in additional individuals. Vision anomalies include myopia (6), hypermetropia (8) and astigmatism (6). An optic nerve hypoplasia was noticed in at least three individuals. A papillary pallor was reported in some individuals (6). Individuals shared common facial appearance including erratic hairline with thin hair, deep set eyes with upper eyelid edema, protruding nose or pointed nasal tip, dysplastic ears, thin upper lip, widely spaced teeth with protruding upper dental arch, retro/micrognathism (**Figure 1, Figure S3**).

## Effect of missense variants on DYRK1A subcellular localization

The cellular localization of overexpressed WT and variant DYRK1A proteins in HeLa cells, studied by immunostaining, revealed three types of cellular distributions: 1) mainly nuclear (N), 2) nuclear and cytoplasmic (N+C) and 3) mainly cytoplasmic (C) (**Figure S8C**). WT DYRK1A protein is mainly localized in the nucleus (80% N; 20% N+C), and this localization is affected when we mutated, separately or combined, the two nuclear localization signals NLS1 (aa 117-134) and NLS2 (aa 389-395) confirming that each NLS contributes to the nuclear localization of DYRK1A, as previously reported<sup>3</sup>. We observed a significant decrease of nuclear localization for Arg413fs, but none of the missense variants tested seems to affect DYRK1A localization.

## Generating and phenotyping analysis of the DYRK1A Thr588Asn Mutant mouse line

The *Dyrk1a*<sup>T588N</sup> mutant mouse line was established for YH at the Institut Clinique de la Souris-(PHENOMIN-ICS, Illkirch, France; <http://www.phenomin.fr>) and breed in our animal facility at the PHENOMIN-ICS (Agreement C67-218-40). The targeting vector was constructed as follows. A 3.4 kb fragment encompassing part of intron 11 and corresponding to the 5' homology arm was amplified by PCR on C57BL/6N ES cell genomic DNA and subcloned in an MCI proprietary vector containing two multiple cloning sites, three repeated SV40 polyA sequences as well as a flipped Neomycin resistance cassette surrounded by 2 LoxP sites. The variant (ACC > AAC) leading to the threonine to asparagine change at position 588 (T588N) was introduced in a second cloning step by fusion of 2 PCR products. Finally, a 3.6 kb fragment corresponding to the 3' homology was cloned in a third cloning step to obtain the final targeting construct. The linearized construct was electroporated in C57BL/6N mouse embryonic stem (ES) cells (ICS proprietary line S3). After G418 selection, targeted clones were identified by long-range PCR and further confirmed by Southern blot with an internal (Neo) probe and a 3' external probe. Two positive ES clones were validated by karyotype spreading and microinjected into BALB/cN blastocysts. Resulting male chimeras were bred with Flp deleter females showing maternal contribution<sup>4</sup>. Germline transmission with the direct excision of the selection cassette was achieved in the first litter. The mouse line was bred on a pure C57BL/6N genetic background in a specific pathogen free environment at the PHENOMIN-ICS animal mouse, with poplar wood granulate bedding (SAFE, Augis, France) and access to normal diet (D03 and D04; SAFE, Augis, France) and tap water treated with Chlorine dioxide (0,8ppm) ad libitum under a classical 12-12 light dark cycle (dark 7pm to 7am during phenotyping). We found a close to mendelian transmission ratio for the segregation of the variant both in heterozygotes and homozygotes. The female to male ratio was as expected.

The expression of the Thr588Asn allele of *Dyrk1a* in heterozygous was studied by Western blot at 7 weeks of age in 5 wild type and 5 *Dyrk1a*<sup>T588N/+</sup> animals. Twenty micrograms of proteins per sample were separated by classical electrophoresis and transferred on a membrane (BIO-RAD, Schiltigheim, Fr). Incubation with primary antibody Anti-Dyrk1a (Abnova, 1/1000) was followed with an incubation with a secondary antibody Anti-mouse (Abnova; 1/5000). We used B-Actin as an internal control that was detected with an anti-mouse B-actin-HRP (Sigma, 1/150000). For revelation, we used the Clarity™ Western ECL Substrate (BIO-RAD-#170-5061). No statistically meaningful difference was observed between the genotypes (Student t.test=0.33; **Figure S12B**). We performed a functional characterization of DYRK1A kinase activity following the protocol already published<sup>5</sup>, using 13 *Dyrk1a*<sup>T588N/+</sup> and 17 control littermate brains (**Figure S12B**). No difference in DYRK1A kinase activity was found regarding the sex or genotype. We raised a cohort with mutant and control littermates from both sex (12 wt male, 12 *Dyrk1a*<sup>T588N/+</sup> male, 12 wt females and 11 *Dyrk1a*<sup>T588N/+</sup> females) to test the cognition, memory, locomotor activity, and assess the anxiety and autism like stereotypies on these animals. We performed the battery of tests as shown in (**Figure S12D**) following the pipeline and protocols previously described<sup>6-9</sup>. All the animal experiments were done in compliance with the ARRIVE Guidelines<sup>10,11</sup> in accordance with the Directive of the European Parliament: 2010/63/EU, revising/replacing Directive 86/609/EEC and with French Law (Decret n° 2013-118 01 and its supporting annexes entered into legislation 01February 2013) relative to the protection of animals used in scientific experimentation and supervised by the Com'eth our local ethical committee. The tests were administered in the following order: open field, novel object recognition (noted NOR, performed 24h after the open field), Y-maze, repetitive behaviour, sociability 3 chambers test and reciprocal tests. In the open field test we assessed the locomotor activity, exploratory drive and anxiety; and we have taken into account randomization of the animals, blinding of the experimenter during the animal research. With the EthoVision system (Noldus, the Netherlands), we measured the total distance, time spend on each of the three areas centre, periphery and walls. Sex was not affecting any of these parameters, so the downstream statistical assessments was done pooling all the animals together divided just by genotype and not considering the sex, increasing the sampling size. We found no difference or effect between the control and mutant *Dyrk1a*<sup>T588N/+</sup> mice in the total distance travelled or time spend on each area. All the animals performed the test and did not show anxiety as they spent quite a high amount of time travelling over all the arena. The novel object recognition test (NOR) was used to study the memory of the animals to discriminate and explore novel objects over familiar ones. We did not identify a deficit in object recognition between the two genotypes (**Figure S12E**). We evaluated the motor activity and the working memory in the Y maze. There was no difference in the percentage of visits done to each arm, no defect in the spontaneous alternation (**Figure S12F**) or delay leaving the initial arm. Repetitive behaviour is one of the 3 main clinical manifestations of autism, together with deficits in social interaction and communication. Thus, we analysed in a 10 min test the number and time spend climbing, digging, and rearing by the mutant animals and control littermates but we did not observe any special stereotypic of repetitive behaviour in the *Dyrk1a*<sup>T588N/+</sup> mice except a slight increase in climbing frequency (**Figure S12G**). In the three chambers test, we analysed both the time spent in proximity of the empty cage during the presentation, or for the familiar and then the new

congener in the discrimination phase. No phenotype was detected in the presentation phase and a significant difference in exploration was observed for the Thr588Asn mutant in the discrimination phase (**Figure S12H**). For the reciprocal sociability test where the social interactions between pairs of mice is analyzed, always using congeners that were not cage mates and in the case of both mutants and control mice in the cohort adding an unknown wild-type animal of the same sex and similar age and size to assess the interactions. We analysed both the time spend in proximity nose to nose or proximity nose to tail but no difference in interaction was observed for *Dyrk1a*<sup>T588N/+</sup>. Although is worthy to mention that the Gardner-Altman effect size plots show a slight tendency on the *Dyrk1a*<sup>T588N/+</sup> mice to decrease the number of contacts and increase the distance with the congener (**Figure S12I**).

### Supplementary acknowledgements

We would like to thank the Centre National de Génotypage for their participation in library preparation and DNA sequencing. We thank all the people from the GenomEast sequencing platform, IGBMC cloning platform for their technical and bioinformatics supports. They thank people for the diagnostic laboratory of Hôpitaux Universitaire de Strasbourg (HUS) for performing follow-up of variants and giving diagnosis to family as well as additional clinicians not included in the author list for their clinical contributions.

### Supplementary Figures

#### Figure S1. Variants identified in DYRK1A in individuals with ID

Schematic representation of DYRK1A protein secondary structure with its different domains: Nuclear localization sequence 1 and 2 (NLS1 and NLS2); DYRK Homology box (DH); catalytic domain; PEST domain (PEST); His rich domain (His) and Serine Threonine (S/T) repeat domain. Variants identified in the cohort are represented with the number of individuals (Ind#) carrying then **(A)** truncating variants (nonsense, frameshift, splice variants) **(B)** missense variants. The variants identified tested in this study are indicated in gold.

#### Figure S2. Consequences of variants on *DYRK1A* mRNA

Sashimi plot from Integrative Genome Viewer (IGV) showing consequences of the splice variants identified by RNA-Seq in mRNA extracted from **(A) Ind #19** blood, showing that c.328-1G>T generate different abnormal transcripts with a) intron 4 retention, leading to a premature stop codon (p.Tyr111Argfs\*6), b) use of an alternative acceptor site 18bps or c) 21pbs downstream the regular one, leading to deletion of few amino

acids (p.Val110\_Lys115del or p.Val110Lys116del) **(B)** **Ind #22** fibroblasts, showing that c.951+4\_951+7del variant leads to a skipping of exon 7 in half of the mutated transcripts (p.Val222Aspfs\*22) and a retention of intron 7 in the other half (p.Ile318fs\*10); **(C)** IGV view of RNA-seq obtained from **Ind #24** fibroblasts, showing c.1240-1\_1240insTAA causes the insertion of these 3 nucleotides in the mRNA at the beginning of exon 10 leading to a premature stop codon p.(Glu414\*) **(D)** Quantitative expression of *DYRK1A* mRNA (normalized by the expression of two reference genes, *GAPDH* and *YWHAZ*) **(E)** Sequencing of *DYRK1A* mRNA in Individual Bronicki#10 cells (c.1232dup, p.(Arg413fs)) treated or not with a NMD blocking agent (emetine). Sequencing of blood *DYRK1A* cDNA in **(F)** Ind #18 showing an equal amount of transcripts carrying the c.1978del variant compared to wild-type allele, suggesting that mutated transcripts escape to NMD, and in **(G)** Ind #30, carrying the variant c.1098G>T, showing the skipping of exon 8 induced by this variant (c.952\_1098del) leading to the deletion of 49 amino acids p.Ile318\_Glu366del instead of one amino acid substitution p.Glu366Asp as first predicted (the probability of using exon 8 donor splice site was decreased by the variant: MaxEnt: -72.7%; NNSPLICE: -54.1%).

### **Figure S3. Photographs of individuals with variant in *DYRK1A***

Photographs of individuals with variants in *DYRK1A*, showing that they share similar facial features.

### **Figure S4. Distribution of the clinical scores calculated without photograph (on 15 points)**

Clinical scores calculated without photograph for individuals carrying pathogenic variants in *DYRK1A* from the initial cohort (*DYRK1A\_I*), the replication cohort (*DYRK1A\_R*) and the individuals affected with other frequent monogenic forms of ID, associated to pathogenic variants in *ANKRD11*, *MED13L*, *DDX3X*, *ARID1B*, *SHANK3*, *TCF4* or *KMT2A*. Brown-Forsythe and Welch ANOVA tests with Dunnett's T 3 multiple comparisons test were performed. ns: not significant; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ , error bars represent SD.

### **Figure S5. Predictions of effect of missense variants and conservation of *DYRK1A* protein**

**(A)** Distribution of the CADD score for missense variants a) not presumed to be not disease-causing (negative set, N-set,  $n = 115$ , see **Methods**), b) presumed to be pathogenic (positive set, P-set,  $n = 16$ ) and c) other missense variants (test set, T-set,  $n = 41$ ). A CADD score  $\geq 20$  means that the variant belongs to the top 1% of variants predicted to be the most deleterious,  $\geq 25$  that the variant belongs to the top 0.3% of variants predicted the most deleterious  $\geq 30$  the variant belongs to the top 0.1% of variants predicted the most deleterious. Brown-Forsythe and Welch ANOVA tests with Dunnett's T 3 multiple comparisons test were performed. ns: not significant; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ , error bars represent SD **(B)** Percentage of variants from the N-set



(presumably benign), P-set (presumably pathogenic) and T-set (to test) having a CADD score  $\geq 20$ ,  $\geq 25$  or  $\geq 30$ , or conserved across 100% of Vertebrate species (V=100%), conserved in at least 90% of Metazoan species (M $\geq 90\%$ ), and at least 80% of other animals (O $\geq 80\%$ ). <sup>a</sup> p.Arg158His variant (V=100% ; M=96% ; O=92%); <sup>b</sup> all variants except p.Ser311Phe variant (100% ; 100% ; 71%), p.Thr588Asn (V= 85%) and p.Ala277Pro (100%; 12%; 27%); variants ; <sup>c</sup> p.Gly171Arg (100%; 100% ; 86%), p.Lys188Arg (100%; 96%; 97%), p.Leu207Pro (100% ; 93% ; 93%), p.Leu241Pro (100% ; 100% ; 97%), p.Leu245Arg (100% ; 100% ; 97%), p.Asp287Tyr (100% ; 100% ; 97%), p.Asp287Asn (100% ; 100% ; 97%), p.Pro290Arg (100%; 100%; 97%), p.Arg328Trp (100%; 100%; 95%), p.Leu347Arg (100% ; 96% ; 84%).

### **Figure S6. Conservation of DYRK1A proteins across the different taxons**

Schematic view of the Multiple Sequence Alignment (MSA) of DYRK1A protein orthologs from Vertebrates, Metazoans, Protists, Fungi and Plants. MSA positions are numbered according to the human protein and colored by a red gradient according to the level of amino acid conservation in each group.

### **Figure S7. Effect of variants on DYRK1A half-life and interaction with DCAF7**

Investigation of variants consequences on DYRK1A protein stability and interaction with its partner DCAF7. (A) Stability of DYRK1A proteins in HEK293 transfected with DYRK1A plasmids and treated with cycloheximide 40 $\mu$ g/mL (stopped at 1, 2, 4 and 8 hours after treatment). Constructs was detected by SDS-PAGE by immunoblotting whole cell lysates using an anti-FLAG antibody and quantification realized with GAPDH (B) Co-immunoprecipitation of DYRK1A in HEK293 cells transfected with *DYRK1A* plasmids using an anti-FLAG antibody, without antibody (Empty) or with Mouse against Rabbit antibody (MAR) as negative and species isotype controls. DCAF7 (WDR68) interaction was detected by SDS-page immunoblotting with a specific antibody (1:2500; abcam anti-WDR68 antibody ab138490).

### **Figure S8. *In vitro* effects of additional variants on DYRK1A proteins and cellular localization of variant proteins**

(A-B) Effect of additional variants on DYRK1A level and ability to autophosphorylate: one variant from the N-set (reported twice in gnomAD) but affecting an highly conserved amino acid position, Arg158His, one variant from the P-set but affecting a position poorly conserved after vertebrates, Ala277Pro, and three variants initially reported as VUS in ClinVar (the last one was reclassified as Likely Pathogenic during the course of this study) and affecting highly conserved positions : p.Gly171Arg, p.Leu241Pro and p.Pro290Arg. We also tested the functional effet of two nonsense changes we created for the need of this study : Ser660\* and Ser661\*. (A) Level of variant DYRK1A proteins expressed in HEK293 (n=3 series) cells transiently

transfected with DYRK1A constructs. Protein levels were normalized on the level of GFP proteins (expressed from a cotransfected pEGFP plasmid). One-way ANOVA with multiple comparison test was performed to compare the level of variant DYRK1A proteins to the level of wild-type DYRK1A protein (orange dashes), applying Bonferroni's correction: ns: not significant; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; error bars represent SEM, standard error of the mean, (B) DYRK1A's ability to autophosphorylate on Tyr321 was tested in HEK293 cells ( $n=3$ ) by immunoprecipitations with anti-DYRK1A followed by an immunoblot using an anti-phospho-HIPK2 as described in Widowati et al. DYRK1A phospho-Tyr321 levels were normalized with DYRK1A total DYRK1A protein levels (orange dashes), and expressed as percentage of WT level. One-way ANOVA test was performed to compare variants to wild-type DYRK1A levels. ns: not significant; \*\*\* $p < 0.001$ ; error bars represent SEM, standard error of the mean (C) Cellular localization of DYRK1A variant proteins observed in HeLa cells after overexpression of FLAG-tagged wild-type and variant DYRK1A constructs ( $n=3$  series of HeLa cells; 50 cells minimum counted per series; scale bars of illustrative images correspond to  $10\mu\text{m}$ ). Three types of localization of DYRK1A protein in cells have been observed: DYRK1A located mostly in the nucleus (N), in both nucleus and cytoplasm (N+C), or mostly in the cytoplasm (C). Scale bars:  $50\mu\text{m}$ . WT protein is mainly localized in the nucleus (N/N+C/C: 80/20/0%). The shift into cytoplasmic localization of DYRK1A is stronger when NLS1 is disrupted (0/63/37%) compared to NLS2 (22/71/7%), and even more drastic when both are mutated, with DYRK1A stacked in the cytoplasm in a large majority of cells (0/26/74%). Chi-square test was performed to compare localization of variant DYRK1A proteins to wild-type DYRK1A protein ns: not significant; \*\*\* $p < 0.001$ ; error bars represent SEM, standard error of the mean. (D) Immunofluorescence experiment showing that the Ser661\* variant does not lead to DYRK1A protein aggregation when overexpressed in HeLa cells.

### Figure S9. DNAm profile of all DYRK1A cases DYRK1A DNAm signature sites.

Heatmap showing the hierarchical clustering of discovery *DYRK1A* LoF cases ( $n = 10$ ; grey) and age- and sex-matched neurotypical discovery controls ( $n = 24$ ; blue) used to identify the 402 differentially methylated (DM) signature sites shown (each row corresponds to a CpG site). The DNAm values for *DYRK1A* LoF validation cases ( $n=6$ , red), missense variants ( $n=10$ , yellow) and distal LoF variant ( $n=1$ , green) are shown. The colour gradient represents the normalized DNA methylation value from -2.0 (blue) to 2.0 (yellow) at each site. Ind #33 (Gly486Asp) has an opposite DNAm values to other *DYRK1A* cases at these sites, also clustering out from controls. Euclidean distance metric is use for the clustering dendrogram.

### Figure S10. Effect of variants on DYRK1A kinase activity regarding MAPT (TAU) Thr212 phosphorylation levels.

MAPT Thr212 phosphorylation assays performed to test variants with particular profiles : Ser324Arg (partial autophosphorylation), Gly486Asp (potential GoF), Thr588Asn (potential phenocopy), Ser660fs (protein aggregates) as well as Ser660\* and Ser661\*. Immunoblots on total protein extracts from HEK293 cells transfected with DYRK1A and MAPT plasmids using anti-FLAG, anti TAU-5 and anti-pT212-TAU antibodies (A) Transfection of different quantities of DYRK1A WT plasmid and 1µg of MAPT plasmid (B) Transfection of DYRK1A proteins harboring the distal truncating variant Ser660fs or the two nonsense changes Ser660\* and Ser661\* and 1µg of MAPT plasmid (C) Transfection of other DYRK1A variants : Ser311Phe (known to abolish kinase activity), Ser324Arg, Gly486Asp and Thr588Asn and 1µg of MAPT plasmid.

**Figure S11. DNAm profile of Ind #33 with DYRK1A Gly486Asp variant displays an opposite DNAm profile at individual signature CpGs.**

DNA methylation values are shown at the 25 most hypermethylated and hypomethylated CpG sites in the signature. Red bars show the average delta beta value at each site in *DYRK1A* LoF discovery cases (n=10), error bar represents the standard error of the mean (SEM). Yellow bars represent the delta beta value Ind #33 at these sites (the beta value of Ind #33 minus the average beta of discovery controls). At most sites, the delta beta for Ind #33 is opposite that of *DYRK1A* cases, indicating DNAm levels in the opposite direction of *DYRK1A* LoF cases relative to controls. Importantly, this is a different phenomenon than Ind #33 having control-like DNAm values, in which case the bars would be near 0.

**Figure S12. Generation, characterisation and behavioural analysis of the heterozygous mouse mutant line carrying a Thr588Asn variant in *Dyrk1a*.** (A) Generating the Thr588Asn (T588N) allele. The T588N allele phenotyped is named ‘T588N with a WT potential allele’ and was obtained by homologous recombination in embryonic stem cells using a targeting vector. LoxP and Flp sites are indicated respectively in green and grey (B) No difference in western blot quantification of DYRK1A level, normalized to b-actin level, in 5 hippocampal protein extracts isolated from *Dyrk1a*<sup>T588N/+</sup> individuals and 5 control littermates (C) Similarly no changes was observed in the phosphorylation of a specific DYRK1A-phosphorylation peptide (-tyde) using brain extract isolated from the hippocampi of *Dyrk1a*<sup>T588N/+</sup> and control littermates (respectively n=13 and 17) (D) schematic representation of the behavioural pipeline used to study 23 *Dyrk1a*<sup>T588N/+</sup> and 24 wild-type (wt) littermates from both sexes at the starting age of 10 weeks (E) In the novel object recognition, a preference was found in the percentage of time exploring the novel object (NO) compared to the familiar object (FO) in the *Dyrk1a*<sup>T588N/+</sup> (F) Spontaneous alternation in the Y maze was not affected by the variant (one single t test with 50% chance level). (G) A significant difference was observed in specific repetitive

behaviour climbing, but not in digging or rearing activities analysed during the repetitive behavioral test (**H**) In the three-chamber sociability test, the percentage of time exploring the familiar stranger (**F**) versus the novel stranger (**N**) was significantly different in the mutant T588N heterozygotes (**I**) During freely moving social interaction we did not detect any anomaly. The statistical significance is noted as followed \*adj P.value= $\leq 0.05$ , \*\*  $0.05 < \text{adj P.value} < 0.01$ , \*\*\* adj P.value  $< 0.001$

**Table S1. Clinical information and summary of recurrent clinical signs observed in individuals with pathogenic variants in *DYRK1A***

M: male; F: female; IUGR: Intrauterine growth restriction; CMV: research for cytomegalovirus infection; BW: birth weight; BS: birth size; BOFC: birth occipitofrontal head circumference; A/FW: absence/few words; ID: intellectual disability, ASD: Autism spectrum disorder; M/S: moderate/severe; mo; months old; SD: standard deviation; AtSD: atrial septal defect; VSD: ventricular septal defect; GER: Gastroesophageal reflux; MRI: Magnetic resonance imaging; EV: enlarged ventricles, CCA/H: corpus callosum agenesis/hypoplasia, CA: cortical or subcortical atrophy, CeA: cerebellar atrophy; NA: not available; <sup>a</sup> Previously reported: Individuals with truncating variants in *DYRK1A* reported in literature (n=80)<sup>1,2,12-36</sup>

**Table S2. Calculation of the clinical score  $CS_{DYRK1A}$  in the different cohorts**

NA: not available; *DYRK1A\_I*: Initial cohort of individuals with variants disrupting *DYRK1A* (nonsense, frameshift, splice, deletions or translocations) reported in this study, used to set up the clinical score ; *DYRK1A\_R*: Cohort of replication including individuals with variants disrupting *DYRK1A* who were previously described with enough clinical information and photographs available (n=12)<sup>1,12,18</sup> \* mild macrocephaly

**Table S3. List of *DYRK1A* missense variants reported in databases (gnomAD, ClinVar, Decipher) and literature and identified in this report (N-set, P-set and T-set)**

training sets: N-set: variants not presumed to be disease-causing, i.e missense variants annotated as “benign”/“likely benign” in ClinVar as well as variants reported more than once in GnomAD (november 2019 release) (n=115); P-set: missense variants reported as “pathogenic”/“likely pathogenic” in Clinvar (n=16) T-set: missense variants reported here, in literature, or as VUS in Clinvar (n=44); FreqRefV, FreqRefM, FreqRefO, FreqRefP, FreqRefF: frequency of the reference amino acid among Vertebrates, Metazoans, Other protist animals, Plants or Fungi; FreqSubV, FreqSubM, FreqSubO, FreqSubP, FreqSubF : frequency of the novel amino acid among Vertebrates, Metazoans, Other animals, Plants or Fungi;

#### **Table S4. List of CpG sites included in DYRK1A DNAm signature**

GoF: gain-of-function. One third of the signature sites (134/402) the  $\beta$  value for p.Gly486Asp was outside the range observed for that of all discovery controls.

#### **Table S5. SVM score associated with the different variants**

SVM: Support Vector Machine classification and scores

#### **Table S6. Summary of the analysis performed to reclassify variants in *DYRK1A***

<sup>1</sup>Highly conserved : V=100% M $\geq$ 90% O $\geq$ 80% \* The consequences of c.1098G>T is p.Ile318\_Glu366del instead of p.Glu366Asp ; \*\* same individual submitted twice \*\*\* : this amino acid change is not reported in gnomAD, but Thr588Pro is reported 44 times (including in one homozygote) and Thr588Ala is reported once ; DD: developmental delay, ID: intellectual disability ; clinical information reported in ClinVar and/or in littérature for the individual <sup>a</sup>: "Abnormality of the frontal hairline, abnormality of the skin, cataract, cerebellar atrophy, feeding difficulties in infancy, intrauterine growth retardation, microcephaly, proportionate short stature, single transverse palmar crease, specific learning disability, ventriculomegaly" ; <sup>b</sup>: « Short chin, Truncal obesity » ; <sup>c</sup>: « Developmental regression, intellectual disability, hypotonia, mild ataxia, intention tremor, dysmorphisms, primary microcephaly, failure to thrive, demyelination, sun sensitivity, incontinence and anxiety » ; <sup>d</sup>: « ASD, learning disorder and macrocephaly » ; <sup>e</sup>: « Abnormal facial shape, Down-sloping shoulders, Genu valgum, Global developmental delay, Hypoplastic toenails » in an individual carrying an additional nonsense variant in *DYRK1A* ; <sup>f</sup>: « Decreased facial expression, Global developmental delay, Hypoplastic left heart, Micrognathia, Postnatal microcephaly, Retinal dystrophy » ; <sup>g</sup>: *DYRK1A*-related ; <sup>h</sup>: « MR/ID/DD; Seizures; Brain MRI positive, Dysmorphic features, Cardiovascular; Craniofacial; Hematologic (child onset), Gastrointestinal (child onset); Musculoskeletal/Structural; Neurologic (child onset), Ophthalmologic, Renal conditions » ; <sup>i</sup>: « Brachycephaly, Intellectual disability, Seizures, Muscular hypotonia, Global developmental delay » ; dn : *de novo* ; ACMG/AMP Criteria : PS2 : *De novo* confirmed ; PS3 : *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product; PM2 : Absent from controls ; in Exome Sequencing Project, 1000 Genomes or ExAC ; PP3 : *in silico* analysis support a deleterious effect on the gene product ; PP4 : Patient's phenotype is highly specific for gene ; BS1: Allele frequency is greater than expected for disorder ; BS3 : Well-established *in vitro* or *in vivo* functional studies shows no damaging effect on protein function or splicing ; BP4 : *in silico* analysis suggest no impact on gene or gene product.

## **Table S7. Summary of functional studies previously performed for additional missense variants in *DYRK1A***

<sup>1</sup>Highly conserved : V=100% M $\geq$ 90% O $\geq$ 80% \* The consequences of c.1098G>T is p.Ile318\_Glu366del instead of p.Glu366Asp ; DD: developmental delay, ID: intellectual disability ; Clinical information reported in ClinVar and/or in literature : <sup>a</sup>: “speech delay, motor delay, moto coordination disorder, seizures and mild physical dysmorphism”; <sup>b</sup>: “mild ID, infantile spasms, speech delay, social interaction and repetitive behaviors, anxiety”; <sup>c</sup>: “Abnormality of the palmar creases, abnormality of the skeletal system, amblyopia, astigmatism, cleft soft palate, constipation, delayed speech and language development, global developmental delay, intrauterine growth retardation, microcephaly”; <sup>d</sup>: “IUGR, DD, microcephaly, ID, severe speech delay”; <sup>e</sup>: “ASD+ID, anxiety, perseveration and additional upsets and aggressive behavior”; <sup>f</sup>: “Severe ID, seizures febrile + generalized tonico-clonic”; <sup>g</sup>: “ID, seizures and microcephaly”; <sup>h</sup>: “Intellectual disability, microcephaly, feeding difficulties; absent or delayed speech development; seizures; deeply set eye”; <sup>i</sup>: “Global developmental delay, Microcephaly, Seizures”; <sup>j</sup>: “Microcephaly, ID, seizures, spasticity, global developmental delay, motor delay, hypertonia, abnormal facial shape, cortical dysplasia, short stature, cortical gyral simplification”; dn : de novo ; ACMG/AMP Criteria : PS2 : De novo confirmed ; PS3 : in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product; PM2 :Absent from controls ; in Exome Sequencing Project, 1000 Genomes or ExAC ; PP3 : in silico analysis support a deleterious effect on the gene product ; PP4 :Patient’s phenotype is highly specific for gene ; BS1: Allele frequency is greater than expected for disorder ; BS3 :Well-established in vitro or in vivo functional studies shows no damaging effect on protein function or splicing ; BP4 :in silico analysis suggest no impact on gene or gene product.

## **Table S8. Distal frameshift variants (last exon) identified in individuals with NDD**

Three distal frameshift variants are reported in Clinvar and in literature<sup>16</sup>. Their clinical interpretation remains ambiguous, especially for c.1726C>T p.Gln576\* and c.2040C>A p.Tyr680\*, for which inheritance is unknown and clinical manifestations do not really overlap those of *DYRK1A* syndrome. <sup>a</sup>: “Severe psychomotor delay, no walk aquisition and no language, severe amblyopia, self-injurious behavior, ASD, dysmorphic features (frontal bossing, hypertelorism, nystagmus, epicanthal folds, a flat nasal bridge, bilateral low-set ears, down-slanting palpebral fissures, a short philtrum, a high arched palate, downturned mouth and micrognathia). Relative macrocephaly (OFC: 52 cm, +0.6 SD; weight: 14.6 kg, -2.2 SD; height: 103.5 cm, -3.1 SD)” <sup>16</sup>; <sup>b</sup>: “Infantile spasms, west syndrome, ASD, aggression, hyperactivity” (personal communication Aida Telegrafi, Genedx); However, the most distal variant ever reported in *DYRK1A*, c.2213\_2218delinsAGAG p.Thr738fs, occurred *de novo* in an individual with clinical features consistent with *DYRK1A* syndrome. <sup>c</sup>: “Autism, microcephaly, seizures, history of failure to thrive and IUGR, mitochondrial complex IV deficiency noted on muscle biopsy” (personal communication Aida Telegrafi, Genedx).

## References

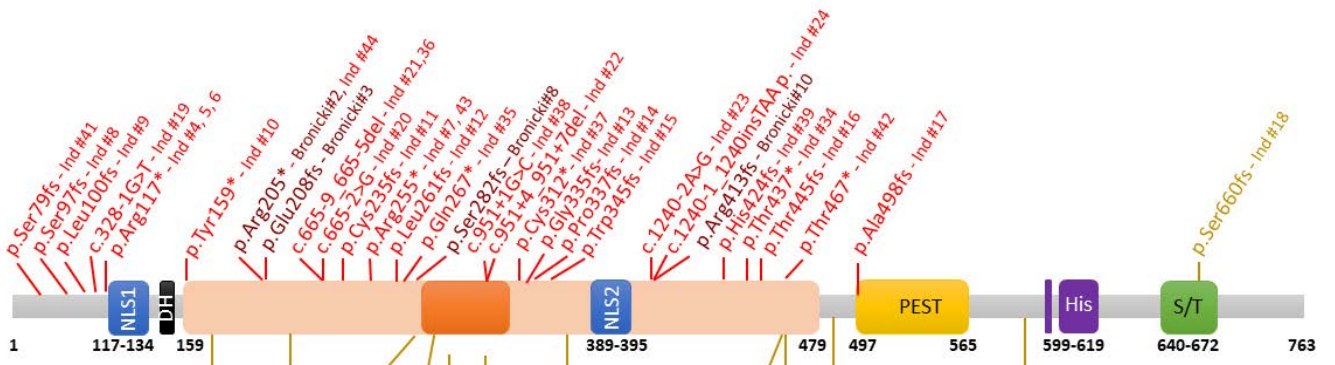
1. Widowati EW, Ernst S, Hausmann R, Müller-Newen G, Becker W. Functional characterization of DYRK1A missense variants associated with a syndromic form of intellectual deficiency and autism. *Biol Open*. 2018;7(4). doi:10.1242/bio.032862
2. Quartier A, Courraud J, Thi Ha T, et al. Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment. *Hum Mutat*. 2019;40(11):2021-2032. doi:10.1002/humu.23836
3. Mattioli F, Isidor B, Abdul-Rahman O, et al. Clinical and functional characterization of recurrent missense variants implicated in THOC6-related intellectual disability. *Hum Mol Genet*. 2019;28(6):952-960. doi:10.1093/hmg/ddy391
4. Lee K-S, Choi M, Kwon D-W, et al. A novel de novo heterozygous DYRK1A mutation causes complete loss of DYRK1A function and developmental delay. *Sci Rep*. 2020;10(1):9849. doi:10.1038/s41598-020-66750-y
5. Bronicki LM, Redin C, Drunat S, et al. Ten new cases further delineate the syndromic intellectual disability phenotype caused by mutations in DYRK1A. *Eur J Hum Genet*. 2015;23(11):1482-1487. doi:10.1038/ejhg.2015.29
6. Blackburn ATM, Bekheirnia N, Uma VC, et al. DYRK1A-related intellectual disability: a syndrome associated with congenital anomalies of the kidney and urinary tract. *Genet Med*. 2019;21(12):2755-2764. doi:10.1038/s41436-019-0576-0
7. Alvarez M, Estivill X, de la Luna S. DYRK1A accumulates in splicing speckles through a novel targeting signal and induces speckle disassembly. *J Cell Sci*. 2003;116(Pt 15):3099-3107. doi:10.1242/jcs.00618
8. Birling M-C, Dierich A, Jacquot S, Héroult Y, Pavlovic G. Highly-efficient, fluorescent, locus directed cre and FlpO deleter mice on a pure C57BL/6N genetic background. *Genesis*. 2012;50(6):482-489. doi:10.1002/dvg.20826
9. Nguyen TL, Duchon A, Manousopoulou A, et al. Correction of cognitive deficits in mouse models of Down syndrome by a pharmacological inhibitor of DYRK1A. *Dis Model Mech*. 2018;11(9). doi:10.1242/dmm.035634
10. Ung DC, Iacono G, Méziane H, et al. Ptchd1 deficiency induces excitatory synaptic and cognitive dysfunctions in mouse. *Mol Psychiatry*. 2018;23(5):1356-1367. doi:10.1038/mp.2017.39
11. Marechal D, Lopes Pereira P, Duchon A, Héroult Y. Dosage of the Abcg1-U2af1 region modifies locomotor and cognitive deficits observed in the Tc1 mouse model of Down syndrome. *PLoS ONE*. 2015;10(2):e0115302. doi:10.1371/journal.pone.0115302
12. Arbogast T, Iacono G, Chevalier C, et al. Mouse models of 17q21.31 microdeletion and microduplication syndromes highlight the importance of Kans11 for cognition. *PLoS Genet*. 2017;13(7):e1006886. doi:10.1371/journal.pgen.1006886
13. Dubos A, Méziane H, Iacono G, et al. A new mouse model of ARX dup24 recapitulates the patients' behavioral and fine motor alterations. *Hum Mol Genet*. 2018;27(12):2138-2153. doi:10.1093/hmg/ddy122
14. Karp NA, Meehan TF, Morgan H, et al. Applying the ARRIVE Guidelines to an In Vivo Database. *PLoS Biol*. 2015;13(5):e1002151. doi:10.1371/journal.pbio.1002151

15. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* 2010;8(6):e1000412. doi:10.1371/journal.pbio.1000412
16. van Bon BWM, Hoischen A, Hehir-Kwa J, et al. Intragenic deletion in *DYRK1A* leads to mental retardation and primary microcephaly. *Clin Genet.* 2011;79(3):296-299. doi:10.1111/j.1399-0004.2010.01544.x
17. van Bon BWM, Coe BP, Bernier R, et al. Disruptive de novo mutations of *DYRK1A* lead to a syndromic form of autism and ID. *Mol Psychiatry.* 2016;21(1):126-132. doi:10.1038/mp.2015.5
18. O'Roak BJ, Vives L, Fu W, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science.* 2012;338(6114):1619-1622. doi:10.1126/science.1227764
19. Courcet J-B, Faivre L, Malzac P, et al. The *DYRK1A* gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. *J Med Genet.* 2012;49(12):731-736. doi:10.1136/jmedgenet-2012-101251
20. Okamoto N, Miya F, Tsunoda T, et al. Targeted next-generation sequencing in the diagnosis of neurodevelopmental disorders. *Clin Genet.* 2015;88(3):288-292. doi:10.1111/cge.12492
21. Iglesias A, Anyane-Yeboah K, Wynn J, et al. The usefulness of whole-exome sequencing in routine clinical practice. *Genet Med.* 2014;16(12):922-931. doi:10.1038/gim.2014.58
22. Ruaud L, Mignot C, Guët A, et al. *DYRK1A* mutations in two unrelated patients. *Eur J Med Genet.* 2015;58(3):168-174. doi:10.1016/j.ejmg.2014.12.014
23. Ji J, Lee H, Argiropoulos B, et al. *DYRK1A* haploinsufficiency causes a new recognizable syndrome with microcephaly, intellectual disability, speech impairment, and distinct facies. *Eur J Hum Genet.* 2015;23(11):1473-1481. doi:10.1038/ejhg.2015.71
24. Rump P, Jazayeri O, van Dijk-Bos KK, et al. Whole-exome sequencing is a powerful approach for establishing the etiological diagnosis in patients with intellectual disability and microcephaly. *BMC Med Genomics.* 2016;9:7. doi:10.1186/s12920-016-0167-8
25. Luco SM, Pohl D, Sell E, Wagner JD, Dymant DA, Daoud H. Case report of novel *DYRK1A* mutations in 2 individuals with syndromic intellectual disability and a review of the literature. *BMC Med Genet.* 2016;17:15. doi:10.1186/s12881-016-0276-4
26. Murray CR, Abel SN, McClure MB, et al. Novel Causative Variants in *DYRK1A*, *KARS*, and *KAT6A* Associated with Intellectual Disability and Additional Phenotypic Features. *J Pediatr Genet.* 2017;6(2):77-83. doi:10.1055/s-0037-1598639
27. Evers JMG, Laskowski RA, Bertolli M, et al. Structural analysis of pathogenic mutations in the *DYRK1A* gene in patients with developmental disorders. *Hum Mol Genet.* 2017;26(3):519-526. doi:10.1093/hmg/ddw409
28. Lee K-S, Choi M, Kwon D-W, et al. A novel de novo heterozygous *DYRK1A* mutation causes complete loss of *DYRK1A* function and developmental delay. *Sci Rep.* 2020;10(1):9849. doi:10.1038/s41598-020-66750-y
29. Dang T, Duan WY, Yu B, et al. Autism-associated *Dyrk1a* truncation mutants impair neuronal dendritic and spine growth and interfere with postnatal cortical development. *Mol Psychiatry.* 2018;23(3):747-758. doi:10.1038/mp.2016.253
30. Qiao F, Shao B, Wang C, et al. A De Novo Mutation in *DYRK1A* Causes Syndromic Intellectual Disability: A Chinese Case Report. *Front Genet.* 2019;10:1194. doi:10.3389/fgene.2019.01194



31. Ernst J, Alabek ML, Eldib A, et al. Ocular findings of albinism in DYRK1A-related intellectual disability syndrome. *Ophthalmic Genet*. Published online August 24, 2020:1-6. doi:10.1080/13816810.2020.1814349
32. Tran KT, Le VS, Bui HTP, et al. Genetic landscape of autism spectrum disorder in Vietnamese children. *Sci Rep*. 2020;10(1):5034. doi:10.1038/s41598-020-61695-8
33. Møller RS, Kübart S, Hoeltzenbein M, et al. Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. *Am J Hum Genet*. 2008;82(5):1165-1170. doi:10.1016/j.ajhg.2008.03.001
34. Fujita H, Torii C, Kosaki R, et al. Microdeletion of the Down syndrome critical region at 21q22. *Am J Med Genet A*. 2010;152A(4):950-953. doi:10.1002/ajmg.a.33228
35. Oegema R, de Klein A, Verkerk AJ, et al. Distinctive Phenotypic Abnormalities Associated with Submicroscopic 21q22 Deletion Including DYRK1A. *Mol Syndromol*. 2010;1(3):113-120. doi:10.1159/000320113
36. Yamamoto T, Shimojima K, Nishizawa T, Matsuo M, Ito M, Imai K. Clinical manifestations of the deletion of Down syndrome critical region including DYRK1A and KCNJ6. *Am J Med Genet A*. 2011;155A(1):113-119. doi:10.1002/ajmg.a.33735
37. Valetto A, Orsini A, Bertini V, et al. Molecular cytogenetic characterization of an interstitial deletion of chromosome 21 (21q22.13q22.3) in a patient with dysmorphic features, intellectual disability and severe generalized epilepsy. *Eur J Med Genet*. 2012;55(5):362-366. doi:10.1016/j.ejmg.2012.03.011
38. Kim O-H, Cho H-J, Han E, et al. Zebrafish knockout of Down syndrome gene, DYRK1A, shows social impairments relevant to autism. *Mol Autism*. 2017;8:50. doi:10.1186/s13229-017-0168-2
39. Meissner LE, Macnamara EF, D'Souza P, et al. DYRK1A pathogenic variants in two patients with syndromic intellectual disability and a review of the literature. *Mol Genet Genomic Med*. Published online November 7, 2020:e1544. doi:10.1002/mgg3.1544
40. Matsumoto N, Ohashi H, Tsukahara M, Kim KC, Soeda E, Niikawa N. Possible narrowed assignment of the loci of monosomy 21-associated microcephaly and intrauterine growth retardation to a 1.2-Mb segment at 21q22.2. *Am J Hum Genet*. 1997;60(4):997-999.

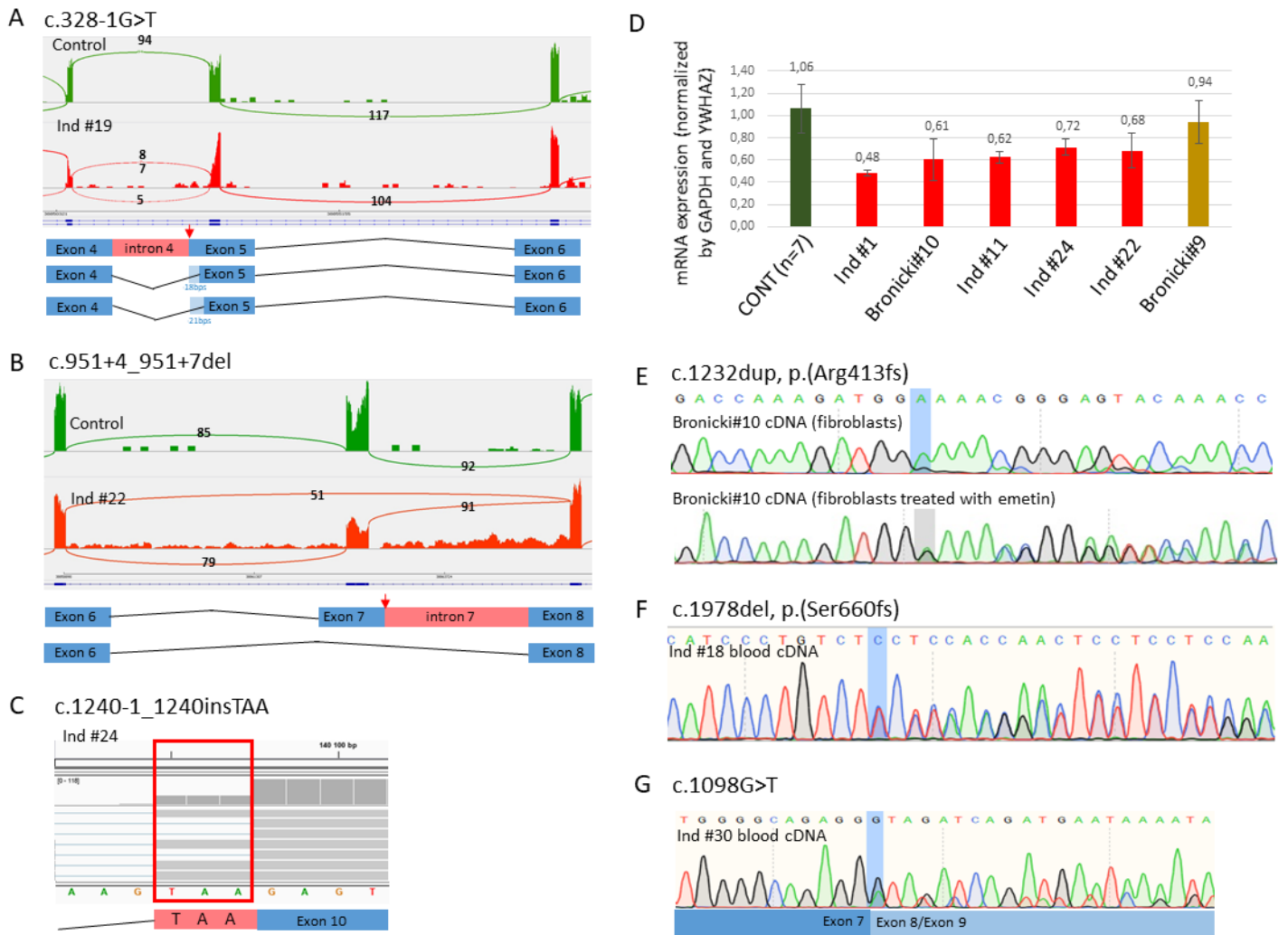
A



B



Figure S1



**Figure S2**



**Figure S3**

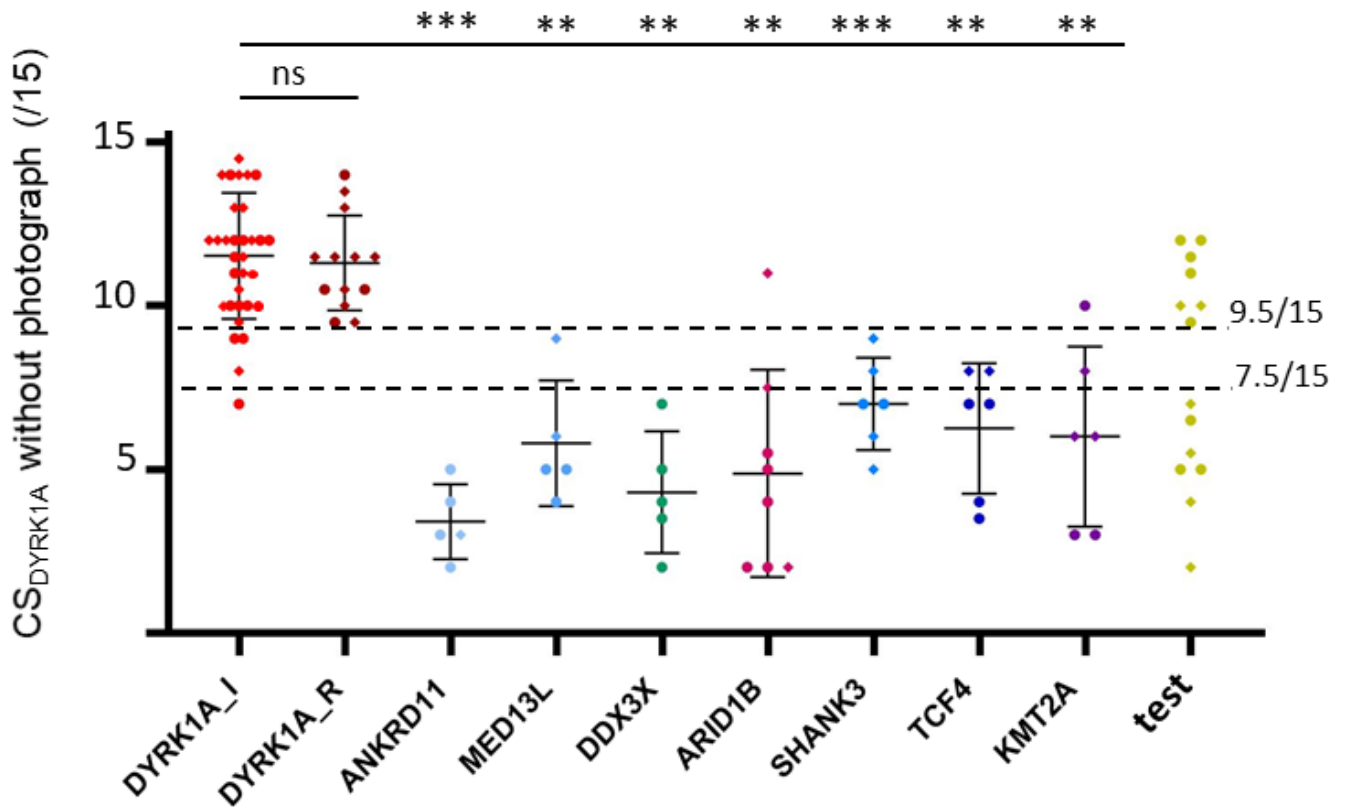
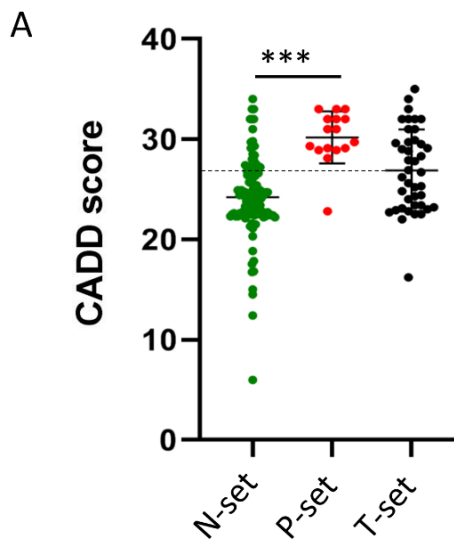


Figure S4



B

	CADD score			Conservation		
	$\geq 20$	$\geq 25$	$\geq 30$	V=100%	V=100% M $\geq 90\%$	V=100% M $\geq 90\%$ O $\geq 80\%$
<b>N-set</b> (n=115)	92,2% (106)	32,2% (37)	7% (8)	27,8% (32)	5% (6)	0,9% (1 <sup>a</sup> )
<b>P-set</b> (n=16)	100% (16)	93,75% (15)	50% (8)	93,75% (15)	87,5% (14)	81,25% (13 <sup>b</sup> )
<b>T-set</b> (n=44)	97,8% (43)	61,4% (27)	22,7% (10)	70,5% (31)	27,3% (12)	22,7% (10 <sup>c</sup> )

Figure S5

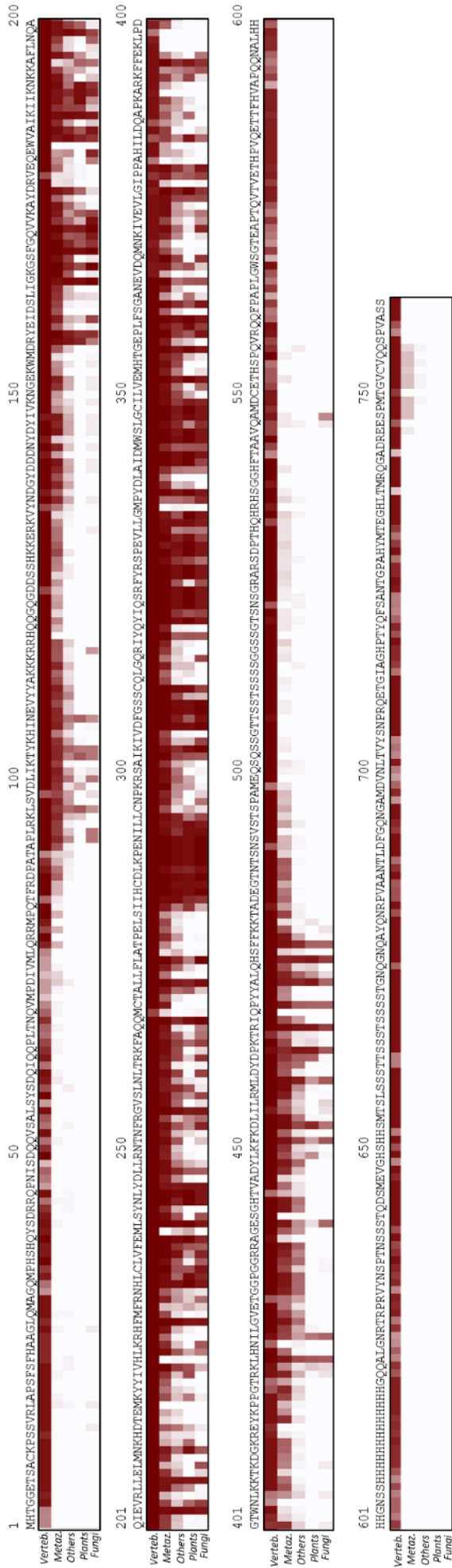
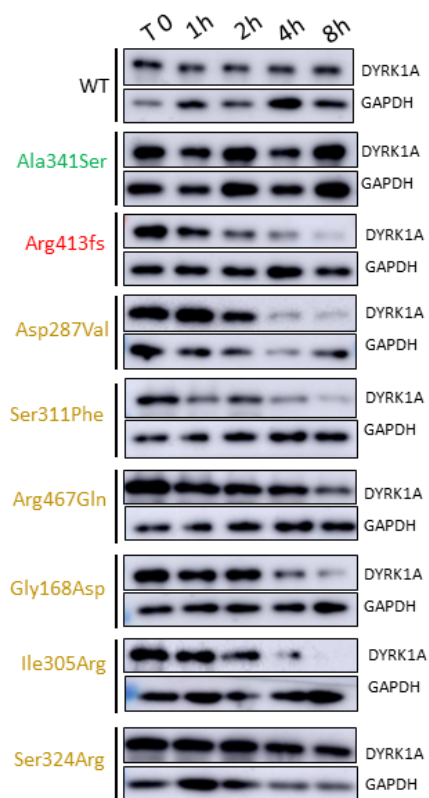
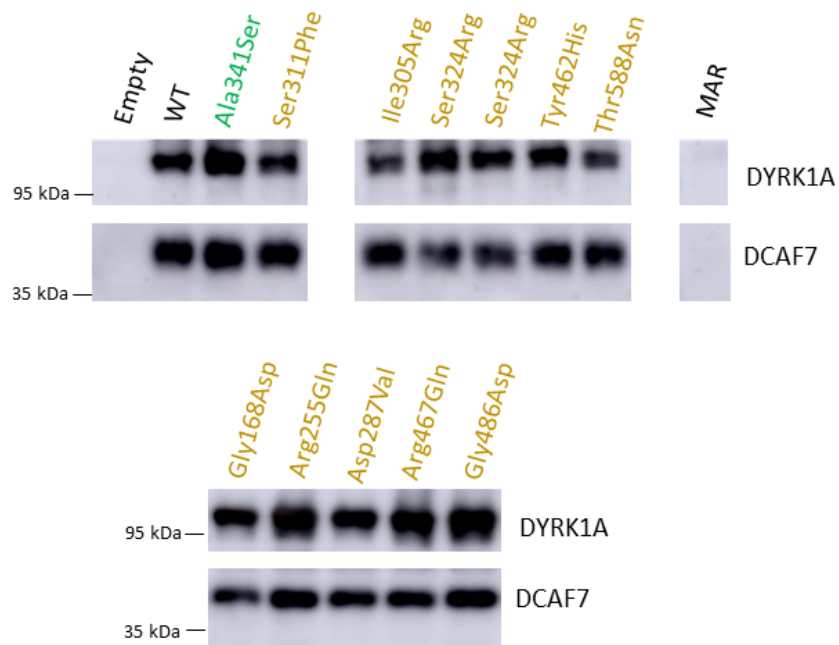


Figure S6



**A****B****Figure S7**

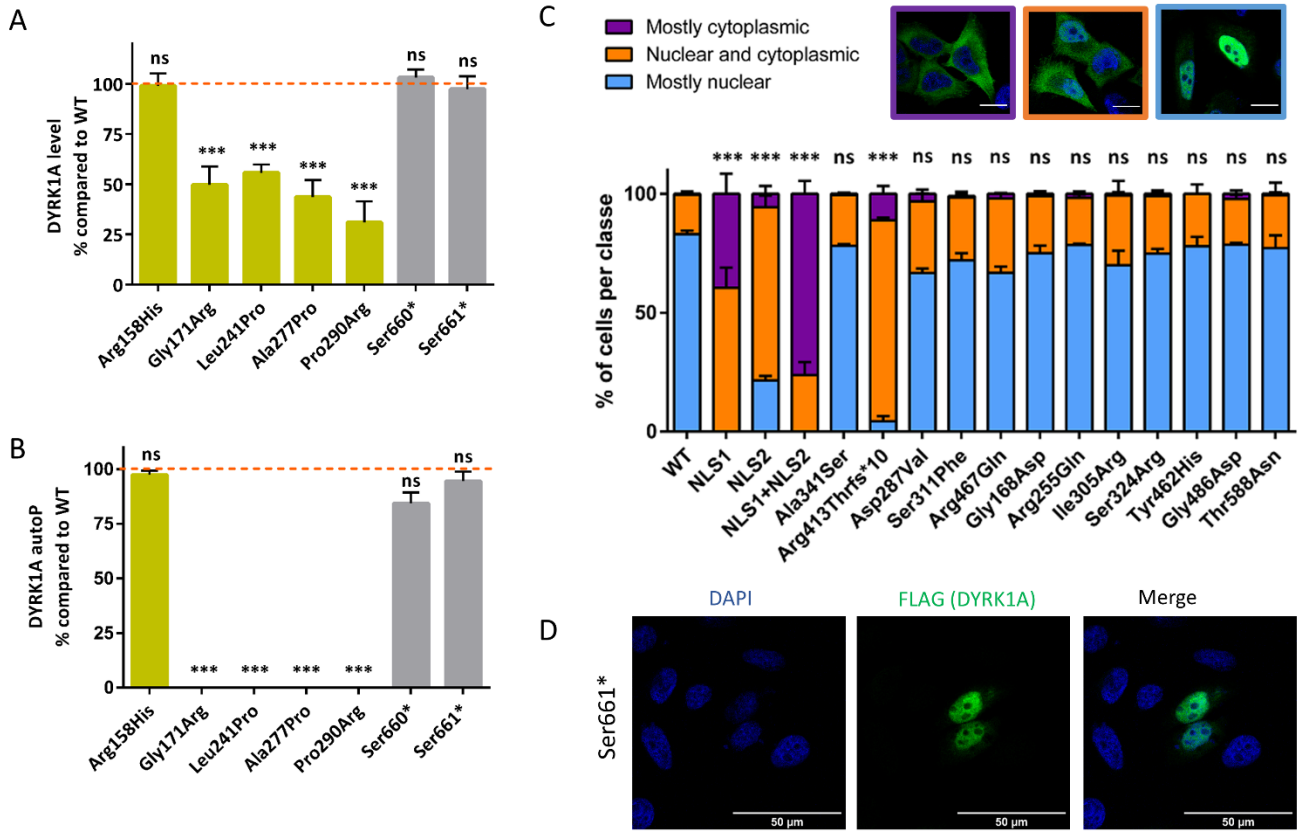


Figure S8



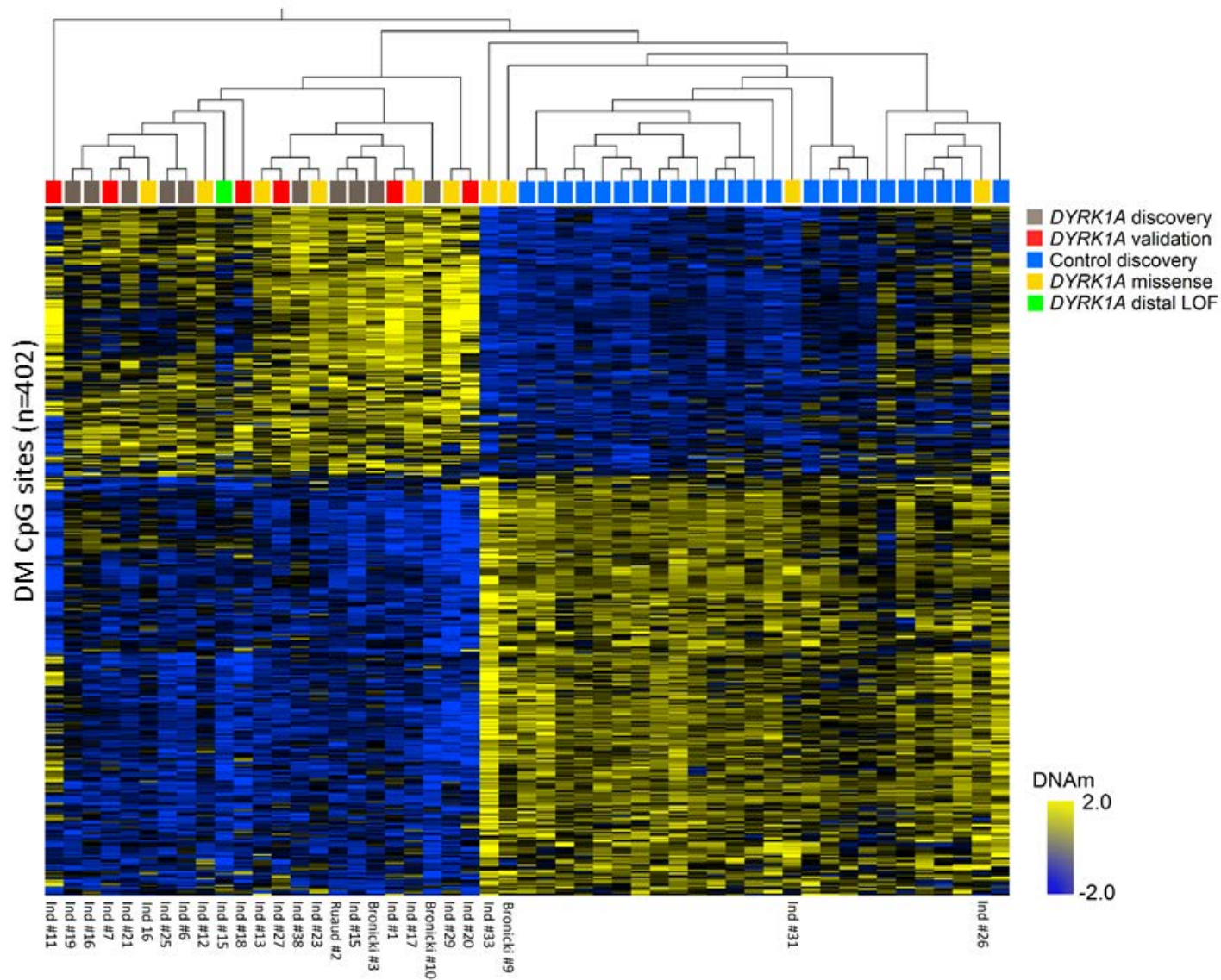


Figure S9

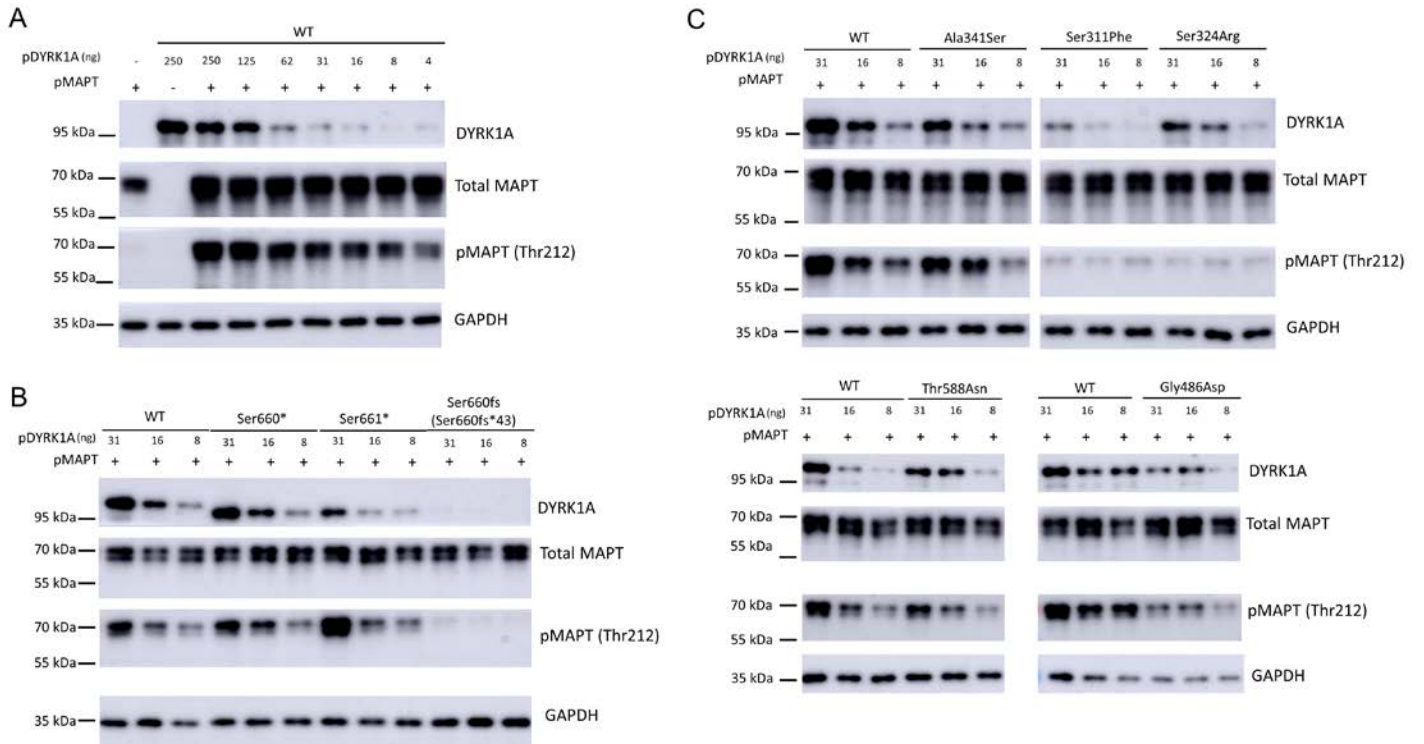


Figure S10

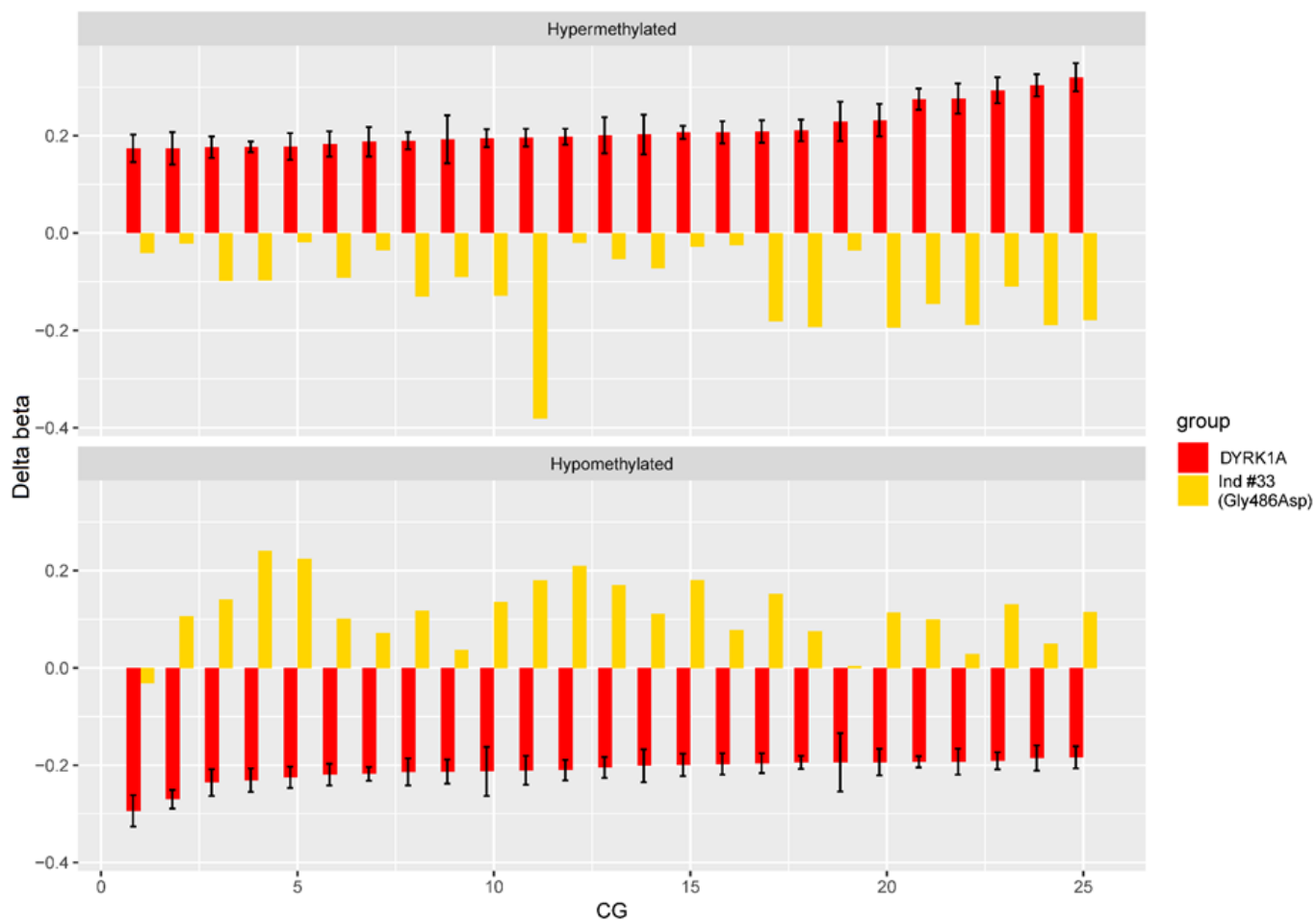


Figure S11

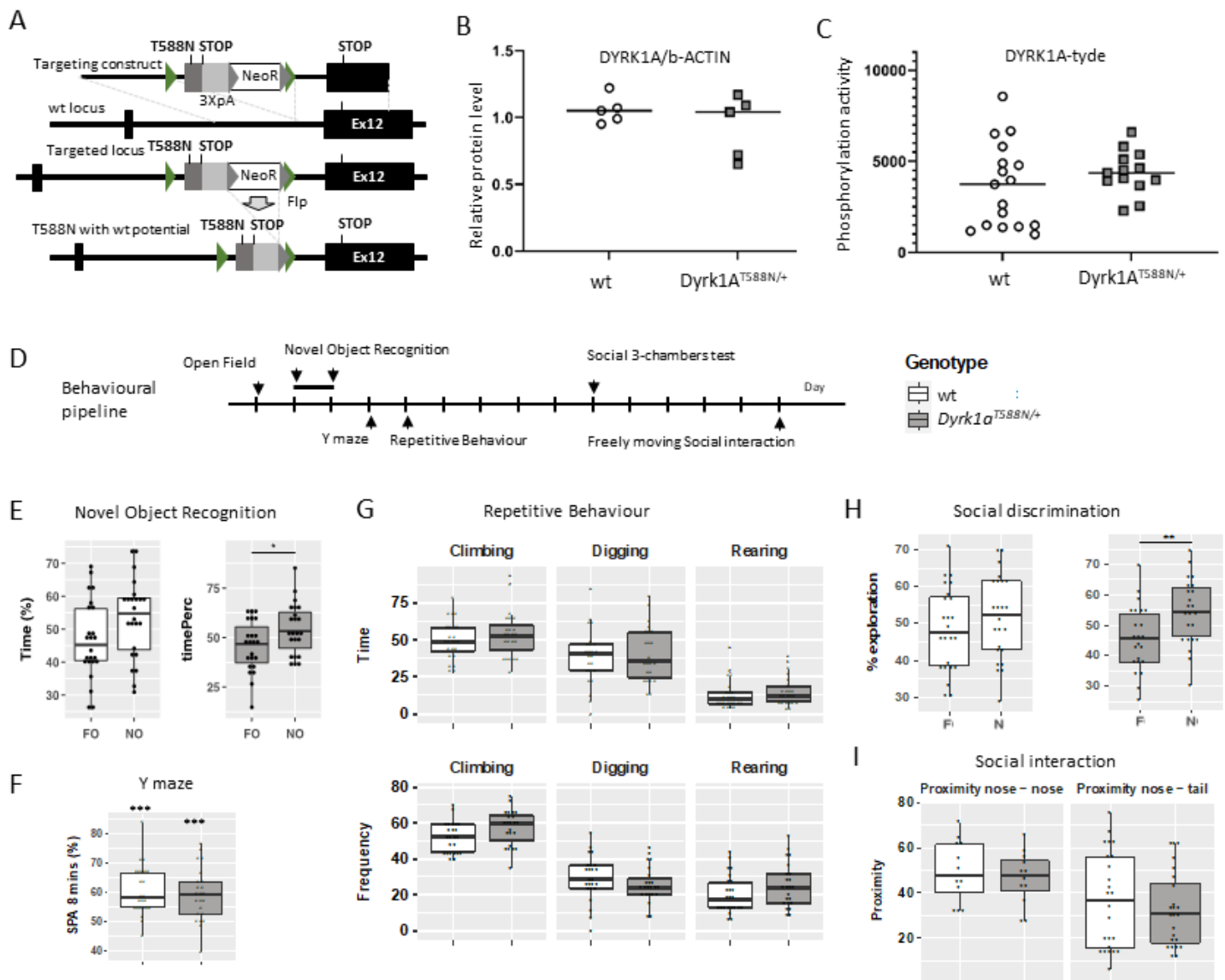


Figure S12

Table S1

		Ind #1	Ind #2	Ind #3	Ind #4	Ind #5	Ind #6	Ind #7	Ind #8	Ind #9	Ind #10	Ind #11
		M	F	M	M	F	M	M	M	M	F	M
<b>DYRK1A variant</b>	nomenclature (NM_001396.4)	g:38481804_4019 g.38722881_394 t(9;21)(p12;q22) c.349C>T										
	nomenclature (NP_001387.2)	-										
	variant type	deletion	deletion	translocation?	nonsense	nonsense	nonsense	nonsense	frameshift	frameshift	nonsense	frameshift
	inheritance	de novo	de novo	de novo	NA	mosaic father	de novo	NA	de novo	de novo	de novo	de novo
<b>Pregnancy and birth</b>	IUGR	yes	no	no	NA	NA	no	no	yes	yes	yes	no
	other	NA	NA	NA	antenatal anom	NA	NA	NA	NA	NA	NA	NA
	delivery	42	term	term	35	term	term	term	term	term	term	term
	APGAR	10/10	10/10	NA	low	NA	9/10	9/10	10/10	10/10	NA	10/10
	BW	2710	3430	2400	2060	3060	3150	3780	2290	2220	2710	2780
	BS	47.5	48	45	NA	47,5	45	51	45	43	46	48
	BOFC	32	33	29.5	31.3	NA	30	35	32	31	31	33
<b>Development</b>	<b>speech</b>	absence	NA	few words	few words	few sentences	few words	absence	absence	absence	few sentences	few words
	<b>ID level (estimated)</b>	moderate	NA	mild	severe	severe	severe	severe	moderate	severe	severe	moderate/severe
	<b>History of seizures</b>	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes
	<b>comment</b>	tonico-clonic	febrile, typical at	febrile, tonico cl	absence, tonico	febrile	NA	NA	febrile, then tonico-c	NA	febrile	absence
	<b>hypotonia</b>	no	yes	no	yes	no	NA	no	yes	yes	no	yes
	<b>motor delay walk acquisition &gt;=1</b>	yes (19.5 mo)	NA	no (15 mo)	yes (24 mo)	yes (18 mo)	yes (>36 mo)	no (16 mo)	yes (30 mo)	yes (18 mo)	yes (18 mo)	no (17 mo)
	<b>motor manifestations</b>	hypereflexia	no	no	no	hypereflexia, Bab	hypereflexia	no	ataxic walk	ataxic walk	hypereflexia	ataxic walk, dysmetri
<b>stereotypies/autistic traits/anxiety</b>	anxiety	no	no	stereotypies	stereotypies, anx	anxiety	stereotypies (ASD di	stereotypies (ASD di	stereotypies	stereotypies, anxiety	stereotypies	
<b>other</b>	behavioral problem	no	no	no	behavioral proble	sleep disorder	behavioral problem	sleep disorder, water	water fascination, be	behavioral problem	no	
<b>General exam</b>	age	10	0.5	12	3	NA	13	18	11	11.5	17	9
	weight (kg)	22	7.85	26.8	14.4	NA	23	66	24.5	22.5	49.5	26,3
	weight (SD)	-3.2	NA	NA	N	NA	-2.5	NA	-3.5	<-2	-2	-1
	size (cm)	128	NA	141	99	NA	142	168	129.5	133	148	135,2
	size (SD)	-1	NA	NA	N	NA	-1.5	NA	-1.5	-1.8	-2	N
	OFC (cm)	47	38	47.4	43.5	NA	48	53	47.5	47.5	49.5	46.5
	<b>OFC (SD)</b>	<-4	<-4	<-4	<-4	NA	-4,5	-2	<-4	<-4	-3,5	-3,5
	<b>skin</b>	thin skin, dermati	NA	thin skin, dermati	no	dermatitis	thin skin, dermat	severe dermatitis	thin skin, dermatitis	dermatitis at birth	no	no
	member anomalies	no	no	no	no	NA	NA	no	yes (syndactyly foot)	no	yes	no
	cardiac anomalies	no	no	no	yes (VSD)	no	no	no	no	no	no	no
	skeletal anomalies	no	no	no	no	mild pectus excav	NA	no	no	no	yes	no
	urogenital anomalies	NA	no	NA	cryptorchidism	no	NA	hypospade	cryptorchidism	no	no	unilateral cryptorchid
	other anomalies	inguinal hernia, py	NA	inguinal hernia	NA	bilateral coxa valg	NA	NA	NA	NA	NA	NA
	vision problems	no	NA	hypermetropia	no	NA	NA	NA	hypermetropia, astig	no	myopia, astigmatism	hypermetropia, astig
	other opthtalmic problems	no	NA	no	no	NA	NA	papillary pallor	no	no	no	no
<b>feeding problems</b>	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	
<b>comments</b>	difficulties in infan	difficult, vomitin	NA	frequent occlusi	NA	important difficul	mastication problem	in infancy, better no	low appetite, highly	sucking difficulty	in infancy	
constipation	no	no	yes	yes	NA	yes	NA	no	yes	no	no	
GER	no	no	yes	NA	NA	NA	yes	no	no	yes	no	
<b>MRI</b>	<b>performed</b>	done	done	done	done	done	done	NA	done	done	done	done
	<b>comment</b>	EV, CA	EV, CCH, gyratio	CA, CCH	hypersignal T2, i	nothing	EV, gyratio	anor	NA	EV	EV, ECC	nothing

DYRK1A\_I

Ind #12	Ind #13	Ind #14	Ind #15	Ind #16	Ind #17	Ind #19	Ind #20	Ind #21	Ind #22	Ind #23	Ind #24	Ind #34
F	F	F	F	F	F	F	M	M	M	M	F	F
c.782del	c.1004del	c.1008dup	c.1033del	c.1333dup	c.1491delC	c.328-1G>T	c.665-2A>G	c.665-9_665-5delTT	c.951+4_951+7delGT	c.1240-2A>G	c.1240-1_1240insTA	c.1309C>T
p.Leu261fs	p.Gly335fs	p.Pro337fs	p.Trp345fs	p.Thr445fs	p.Ala498fs	p.?	p.?	p.?	p.?	p.?	p.Glu414*	p.Arg437*
frameshift	Frameshift	frameshift	frameshift	frameshift	frameshift	splice	splice	splice	splice	splice	nonsense	nonsense
de novo	absent in the mother	de novo	de novo	de novo	de novo	de novo	de novo	de novo	de novo	de novo	de novo	NA
no	yes	no	yes	yes	yes	no	no	no	yes	yes	yes	yes
cerebellar anomaly	NA	NA	CGH-array + CMV + MRI	NA	NA	NA	NA	NA	NA	NA	caesarian	single umbilical artery
term	38	term	term	37	36	term	term	39	NA	term	term	term
10/10	NA	NA	NA	10/10	9	10	NA	10/10	NA	NA	10/10	10/10
3030	2400	3550	1865	1970	1990	3450	3110	3085	2690	3160	2410	2450
46	42	52	42	43,5	43	49	52	50	47	48	45	47
32	30.5	NA	29.5	30	28.5	32	33	34	NA	33	31	31
absence	few words	few words	few words/few sentences	few words	few sentences	few words	absence	few sentences	few words	absence	few sentences	few sentences
severe	severe	moderate	moderate to severe	severe	severe	severe	moderate	moderate	severe	severe	moderate	mild
yes	yes	yes	no	yes	no	yes	yes	yes	yes	yes	yes	yes
NA	atonic	febrile seizures	NA	febrile, then tonic-clonic	NA	febrile	febrile	tonic-clonic	febrile	febrile	febrile seizures + other febrile	
no	no	no	yes	yes	yes	yes	yes	no	NA	no	NA	yes
yes (19 mo)	yes (>36mo)	yes (>36mo)	yes (24 mo)	yes (28 mo)	no (16 mo)	yes (36 mo)	yes (31 mo)	no (16.5 mo)	yes (18 mo)	yes (18 mo)	yes (22 mo)	yes (23 mo)
no	no	ataxic walk	hypereflexia	hypereflexia	NA	ataxic walk, hypereflexia	NA	ataxic walk, hypereflexia	no	no	hypereflexia	no
stereotypies	stereotypies	stereotypies	stereotypies	stereotypies, anxiety	stereotypies	stereotypies	stereotypies	stereotypies (ASD diagnosis)	stereotypies, anxiety	stereotypies (ASD diagnosis)	stereotypies, anxiety	stereotypies, anxiety
sleep disorder, behavioral	no	no	sleep disorder	behavioral problem	NA	behavioral problem, water fascination	behavioral problem	no	no	sleep disorder, anxiety	sleep disorder, behavioral	NA
7.5	27	18.5	2	12.5	3.5	11	1.5	20	NA	7	NA	13
16.5	67	47.3	7.6	30	10.9	28.1	11.6	41	NA	18.1	11	27
-2	NA	-2.7	NA	-2	-2.5	NA	1	-4	NA	-2	-3	-2
118	150	164	76	140	93.5	149	82	172.5	NA	114.5	NA	130
-1	NA	-0.88	NA	-2	-2	NA	0	-0.5	NA	-1.3	-2	-2
46	48.5	50.5	40.5	48	44.5	47.5	45	51	NA	45	43	47
<-4	-3,5	<-3	-4	<-4		-4 <-4	-3	-4	-4	<-4	-5	-5
no	NA	thin skin	no	thin skin	thin skin, dermatitis	dermatitis at birth	thin skin	thin skin, dermatitis	no	thin skin	thin skin	thin skin, dermatitis
no	no	short 5th metatarsals, 5th toe hypoplasia	yes (adductus thumb)	no	NA	no	no	yes. Bilateral postaxial polydactyly	no	no	NA	no
no	NA	no	no	yes (AtSD)	NA	no	NA	no	no	no	NA	no
no	scoliosis, equinus feet	abnormalities of the feet	no	no	NA	no	no	no	no	no	NA	no
NA	anal malposition	no	no	no	NA	NA	NA	no	cryptorchidism	no	NA	grade II vesicoureteral reflux
inguinal hernia, vaso hiatal hernia	hiatal hernia	Acrocyanosis – hand	NA	NA	NA	NA	NA	NA	inguinal hernia	NA	NA	velopalatine insufficiency
astigmatism	NA	myopia	NA	myopia	hypermetropia	no	NA	myopia, astigmatism	myopia	NA	NA	NA
no	NA	no	no	no	NA	papillary pallor	papillary pallor	hypoplasia optic nerve	no	no	no	hypoplasia optic nerve
yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	yes
difficult in infancy	gastrostomy	breast, additional	NA	important feeding difficulties	very difficult between first months	low appetite	NA	highly selective	very selective	NA	very selective	very selective
no	no	no	yes	yes	yes	yes	no	yes	no	yes	yes	yes
no	yes	NA	NA	no	NA	yes	NA	no	NA	no	NA	yes
done	done	done	done	done	done	done	done	done	not done	done	done	done
CCH, CeA (pre and p EV)		nothing	NA	CCH, atrophy CSE	delayed myelination	CCH, CA, EV	EV, CA	EV	NA	CA, CeA	nothing	enlarged pericerebellum



										Total (n=34)		Previously reported <sup>a</sup> (n=80)	
Ind#35	Ind#36	Ind #37	Ind #38	Ind#39	Ind #40	Ind #41	Ind #42	Ind #43	Ind #44	Total	%	Total	%
M	M	F	M	F	M	F	F	F	M	17 F/ 17 M			
c.799C>T	c.665-9_665-5delTT<c.936T>A	c.936T>A	951+1G>C	c.1270del	g.38302140_400414 c.235dupA	c.235dupA	c.1399C>T	c.763C>T	c.613C>T				
p.Gln267*	p.?	p.Cys312*	p.?	p.His424fs	-	p.Arg79fs	p.Arg467*	p.Arg255*	p.Arg205*				
nonsense	splice	nonsense	splice	frameshift	deletion	frameshift	nonsense	nonsense	nonsense				
de novo	de novo	de novo	de novo	de novo	de novo	de novo	de novo	NA	NA				
no	yes	yes	no	yes	yes	yes	yes	yes	yes	20/32	63%	22/32	69%
NA	NA	NA	NA	NA	NA	NA	NA	NA	kidney asymetry				
term	term	term	term	term	term	term	term	35	term				
NA	10/10	NA	10/10	9/10	10/10	10/10	9/10	NA	NA				
	2880 2580	2175	2810	2700	3395	2160	2530	1695	2590				
	50 44	43	49	47	46	43.6	43	43.5	44				
	34 33	30,2	33	31.5	NA	30	31.5	28.5	32				
few sentences	absence	few words	NA	few sentences	absence	NA	few words	few words	absence	30/30 (A/FW: 22/31)	100% (71%)	74/74	100%
moderate	severe	NA	NA	moderate/severe	severe	NA	severe	severe	severe	32/32 (M/S:28/30)	100% (93%)	73/73	100%
yes	yes	yes	NA	yes	yes	yes	yes	yes	no	29/33 (febrile: 18)	88%	56/76 (febrile: 74%)	
febrile	febrile + tonicoclonic febrile	febrile	NA	febrile, absence, myr febrile, tonico-clonic	febrile, tonico-clonic	NA	absence, tonico-clon	NA	NA				
yes	yes	yes	yes	no	NA	yes	no	no	NA	16/29	55%		
yes (19 mo)	yes (21 mo)	yes (25 mo)	yes (>26 mo)	yes (24 mo)	yes (28 mo)	yes (21 mo)	no (17 mo)	yes (20 mo)	yes (>36 mo)	27/33	82%	59/67	88%
no	NA	digitigrade walk	hypereflexia	no	pyramidal signs	pyramidal signs, hyp	hypertonia	NA	NA	18/28	64%	32/44	73%
stereotypies	stereotypies, anxiet	stereotypies	NA	anxiety	stereotypies, anxiet	NA	anxiety	no	no	28/32	88%	40/45	89%
sleep disorder	behavioral problem	water fascination, be	NA	NA	sleep disorder, water	sleep disorders	sleep disorders, water	absence of fear	no				
23	18	3	NA	13	8.5	5	15	6	4				
63	40,5	12	NA	31.6	19	16.7	36.6	15.8	13.1				
N	NA	-1	NA	-1.3	-2	NA	-2.5	-1.4	-1.7				
173	159,5	87	NA	139	120	102	150.5	113	95				
N	NA	-2	NA	-1.7	-1.5	NA	-2	N	-2				
52.5	52	44	NA	50	45.5	45	49.3	44	44.5				
-3	-3	-3,5	NA	-2	-6	<-3	<-3	<-4	<-4	32/32	100%	72/77	94%
no	no	thin skin	NA	NA	thin skin	NA	thin skin	no	NA	19/28	68%	10/?	
no	no	yes	NA	long fingers, small th	no	NA	no	no	no	7			
no	no	NA	NA	pulmonary valve dys	no	NA	no	no	no	3			
cyphosis	scoliosis	pectus	NA	no	no	NA	no	no	no	7			
testicular ectopia	no	ovarian hernia	NA	no	NA	NA	no	no	micropenis	10/23	43%	13/?	
NA	NA	NA	NA	no	nail dystrophy	NA	no	no	no				
myopia	hypermetropia	no	NA	hypermetropia	NA	NA	hypermetropia	no	hypermetropia, astig	15/21	71%		
no	no	NO	strabism	nystagmus	hypolasia optic nerve	NA	no	strabism	no				
no	yes	yes	yes	yes	yes	yes	yes	yes	yes	31/33	94%	50/57	88%
NA	smashed food	NA	NA	first months	poor eating	NA	NA	very selective	NA				
yes	no	no	NA	NA	no	NA	yes	no	yes			12/?	
no	no	yes	yes	yes	no	NA	yes	no	NA			11/?	
not done	done	done	NA	NA	done	done	done	done	done				
NA	periventricular white	NA	NA	NA	CCH, enlarged cister	enlarged pericerebr	nothing	CA	CCH	22/27	81%	39/52	75%

DYRK1A_R (Bronicki et al., 2014)				Individuals with variants to test											
Bronicki_#2	Bronicki_#3	Bronicki_#8	Bronicki_#10	Ind #18	Ind #25	Ind #26	Ruud #2	Ind #29	Ind #31	Ind #30	Ind #28	Ind #32	Ind #27	Ind #33	Bronicki_#9
M	M	M	F	F	F	F	M	M	M	F	F	F	M	M	F
c.613C>T	c.621_624 delir	c.844dupA	c.1232dup	c.1978del	c.503G>A	c.764G>A	c.932C>T	c.972T>A	c.1384T>C	c.1098G>T	c.914T>G	c.1400G>A	c.860A>T	c.1457G>A	c.1763C>A
p.Arg205*	p.Glu208fs	p.(Ser282Lysfs	p.Arg413fs	p.Ser660fs	p.Gly168Asp	p.Arg255Gln	p.Ser311Phe	p.Ser324Arg	p.Tyr462His	p.Ile318_Glu366	p.Ile305Arg	p.Arg467Gln	p.Asp287Val	p.Gly486Asp	p.Thr588Asn
nonsense	frameshift	frameshift	frameshift	frameshift	missense	missense	missense	missense	missense	inframe del	missense	missense	missense	missense	missense
de novo	de novo	de novo	de novo	de novo	de novo	NA	de novo	de novo	de novo	de novo	de novo	de novo	de novo	de novo	de novo
NA	NA	no	yes	yes	NA	no	yes	yes	NA	yes	yes	NA	yes	NA	no
NA	NA	NA	NA	heart ventricular	NA	NA	artery	NA	NA	NA	NA	NA	NA	NA	NA
term	term	term	term	term	NA	term	term	term	term	term	term	NA	prematurity	NA	term
10/10	10/10	8/10/10	10/10	10/10	NA	10/10	NA	NA	7/10	9/10	NA	NA	NA	NA	8/3
2720	3280	3290	2215	2660	NA	3490	3170	2575	3440	2570	2550	NA	2200	NA	3000
47	49	47	46	46	NA	51,5	49	46	50,5	48	47	NA	NA	NA	48
32	33	34	32	31	NA	36	33.5	33	37	31	31.5	NA	NA	NA	32
few words	few words	absence	few words	few sentences	NA	few words	few words	few sentences	normal	absence	few words	NA	few words	normal	absence
moderate	moderate	moderate/sev	moderate/sev	moderate	NA	NA	moderate	mild	mild	severe	severe	NA	severe	no ID	severe
yes	absence?	yes	yes	no	NA	no	yes	yes	yes	yes	yes	no	yes	no	yes
NA	NA	febrile, tonico-	febrile, myoclon	NA	NA	NA	febrile	atonic	febrile	febrile, tonic-clo	febrile	NA	febrile	NA	absence, tonic
NA	NA	yes	NA	no	NA	no	NA	yes	no	yes	NA	NA	no	NA	yes
yes	yes (23mo)	no (16 mo)	yes (20 mo)	yes (18 mo)	NA	yes (18 mo)	yes	yes (22 mo)	no (13 mo)	yes (36 mo)	yes	NA	yes (20 mo)	no	yes (>36 mo)
NA	hypertonia	hypereflexia	NA	no	NA	no	NA	no	normal	no	ataxic walk, hy	NA	NA	no	ataxia
anxiety	stereotypies	ASD diagnosis	flapping	anxiety	NA	stereotypies	stereotypies	ASD, anxiety	anxiety	stereotypies	stereotypies	stereotypies	stereotypies	ASD	stereotypies
NA	sleep disorders	water fascinati	NA	NA	NA	behavioral prot	NA	sleep disorder,	aggressivity	sleep disorder,	water fascination,	behavioral prc	NA	NA	sleep disorder, water fascination
15	3	NA	NA	12	NA	3	NA	5	22,5	16	NA	NA	6	10	38
43.8	13.5	NA	NA	39	NA	15,7	NA	11,5	89	44	NA	NA	21	39.4	75
NA	-1	NA	-2	N	NA	N	NA	+1.2	NA	-2.3	-1.5	NA	NA	NA	+3.5
167	94	NA	NA	154	NA	NA	NA	101,5	181	162	NA	NA	117	156	170
NA	-1	NA	-0,5	+1.7	NA	NA	NA	-1,5	NA	-1	NA	NA	NA	NA	1
43,8	46	NA	NA	50	NA	NA	NA	52	60,5	47	NA	NA	46.5	58.5	NA
-3	-4	<-3	-3.5	-2.5	NA	2	<-4	+1	NA	<-3	-4	NA	<-3	NA	-3
thin skin	NA	no	thin skin	no	NA	no	NA	no	no	no	NA	dermatitis	NA	+2	thin skin
no	NA	no	NA	no	NA	no	no	clinodactyly 5tl	no	long fingers, can	NA		flat feet	NA	no
no	NA	VSD	NA	no	NA	no	no	no	no	NA	NA	NA	NA	NA	NA
pectus excavati	NA	scoliosis	NA	NA	NA	no	no	no	scoliosis	NA	NA	NA	NA	NA	NA
no	NA	NA	no	no	NA	no	no	no	no	NA	NA	NA	micropenis	NA	no
no	NA	NA	NA	NA	NA	no	no	no	no	NA	NA	NA	NA	NA	NA
hypermetropia	hypermetropia	hypermyopia	NA	hypermetropia	NA	no	NA	NA	hypermyopia	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	no	NA	NA	no	NA	no	NA	NA	NA	NA	NA	NA
yes	yes	yes	yes	no	NA	no	no	yes	no	yes	NA	yes	NA	NA	yes
NA	NA	gastrostomy	NA	NA	NA	NA	no	NA	NA	difficulty during	no	NA	NA	NA	mild
NA	yes	yes	NA	no	NA	no	NA	yes	no	yes	NA	NA	yes	NA	yes
NA	NA	yes	NA	no	NA	no	NA	yes	no	no	NA	NA	NA	NA	no
done	done	done	done	done	NA	done	done	done	done	done	done	done	NA	NA	not done
EV, plagioceph	EV	EV	EV, CCA	thin optic nerv	NA	nothing	EV	nothing	EV	CA, EV	CCH	normal	NA	NA	NA



Table S2

Cohort	Name	Ind	Sex	Mutation nomenclature	Type	Inheritance	Intellectual disability	Impaired language	Microcephaly	Feeding difficulty	History of seizures	Autistic traits / Anxiety	Brain anomaly on MRI	Motor delay	Ataxic walk/hypertonia	Atopic or transglutidase s	/15	Dysmorphology	total /20	calculated on clinical data coming from		
DYRK1A_I	this report	Ind #1	M	g:38481804_40190458del	-	deletion	de novo	1	2	2	2	1	1	1	0.5	0.5	12	4	16	this report		
		Ind #2	F	g:38722881_39426450del	-	deletion	de novo	NA	NA	2	2	2	0	1	NA	0	NA	7	NA	NA	this report	
		Ind #40	M	g:38302140_40041414del	-	deletion	de novo	2	2	2	2	2	1	1	1	0.5	0.5	14	NA	NA	this report	
		Ind #3	M	t(9;21)(p12;q22)	-	translocation?	de novo	0	2	2	2	2	0	1	0	0	0.5	9.5	3.5	13	this report	
		Ind #4	M	c.349C>T	p.Arg117*	nonsense	NA	2	2	2	2	1	1	1	1	0	0	12	NA	NA	this report	
		Ind #5	F	c.349C>T	p.Arg117*	nonsense	mosaic father	2	1	NA	2	2	1	0	1	0.5	0.5	10	NA	NA	this report	
		Ind #6	M	c.349C>T	p.Arg117*	nonsense	de novo	2	2	2	2	1	1	1	1	0.5	0.5	13	NA	NA	this report	
		Ind #10	F	c.477del	p.Tyr159*	nonsense	de novo	2	1	2	2	2	1	0	1	0.5	0	11.5	3.5	15	this report	
		Ind #44	M	c.613C>T	p.Arg205*	nonsense	NA	2	2	2	2	0	0	1	1	NA	NA	10	NA	NA	this report	
		Ind #7	M	c.763C>T	p.Arg255*	nonsense	NA	2	2	1	2	1	2	NA	0	0	0.5	10.5	4	14.5	this report	
		Ind #43	F	c.763C>T	p.Arg255*	nonsense	NA	2	2	2	2	1	0	1	1	NA	0	11	NA	NA	this report	
		Ind #35	M	c.799C>T	p.Gln267*	nonsense	de novo	1	1	2	0	2	1	NA	1	0	0	8	NA	NA	this report	
		Ind #37	F	c.936T>A	p.Cys312*	nonsense	de novo	1	2	2	2	2	1	NA	1	0.5	0.5	12	3	15	this report	
		Ind #34	F	c.1309C>T	p.Arg437*	nonsense	NA	0	1	2	2	2	1	0.5	1	0	0.5	10	NA	NA	this report	
		Ind #42	F	c.1372C>T	p.Arg467*	nonsense	de novo	2	2	2	1	1	1	0	0	0.5	0.5	10	4	14	this report	
		Ind #41	F	c.235dupA	p.Arg79fs	frameshift	de novo	NA	NA	2	NA	1	NA	0.5	1	0.5	NA	NA	NA	NA	NA	this report
		Ind #8	M	c.290_291 delCT	p.Ser97fs	frameshift	de novo	1	2	2	2	2	2	1	1	0.5	0.5	14	3.5	17.5	this report	
		Ind #9	M	c.297_301delCTTGA	p.Leu100fs	frameshift	de novo	2	2	2	2	0	1	1	1	0.5	0.5	12	4.5	16.5	this report	
		Ind #11	M	c.702_703delCT	p.Cys235fs	frameshift	de novo	2	2	2	2	1	1	1	0	0.5	NA	11.5	4	15.5	this report	
		Ind #12	F	c.782del	p.Leu261fs	frameshift	de novo	2	2	2	2	1	1	1	1	0	NA	12	4	16	this report	
		Ind #13	F	c.1004del	p.Gly335fs	frameshift	NA	2	2	2	2	1	1	1	1	0	NA	12	2.5	14.5	this report	
		Ind #14	F	c.1008dup	p.Pro337fs	frameshift	de novo	1	2	2	2	2	1	0	1	0.5	0.5	12	2.5	14.5	this report	
		Ind #15	F	c.1033del	p.Trp345fs	frameshift	de novo	2	1	2	2	0	1	0.5	1	0.5	0	10	3	13	this report	
		Ind #39	F	c.1270del	p.His424fs	frameshift	de novo	2	1	1	1	2	1	NA	1	0	0	9	NA	NA	this report	
Ind #16	F	c.1333dup	p.Thr445fs	frameshift	de novo	2	2	2	2	2	1	1	1	0.5	0.5	14	3.5	17.5	this report			
Ind #17	F	c.1491del	p.Ala498fs	frameshift	de novo	2	1	2	2	0	1	0.5	0	NA	0.5	9	NA	NA	this report			
Ind #19	F	c.328-1G>T	p.?	splice	de novo	2	2	2	2	2	1	1	1	0.5	0.5	14	3.5	17.5	this report			
Ind #20	M	c.665-2A>G	p.?	splice	de novo	1	2	2	2	2	1	1	1	0.5	0.5	13	4	17	this report			
Ind #21	M	c.665-9_665-5delTTCTC	p.?	splice	de novo	1	1	2	2	1	2	1	0	0.5	0.5	11	4.5	15.5	this report			
Ind #36	M	c.665-9_665-5delTTCTC	p.?	splice	de novo	2	2	2	2	2	2	1	1	NA	NA	14	2	16	this report			
Ind #38	M	c.951+1G>C	p.?	splice	de novo	NA	NA	NA	2	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	this report		
Ind #22	M	c.951+4_951+7delGTAA	p.?	splice	de novo	2	2	2	2	2	1	0	1	NA	0	12	4.5	16.5	this report			
Ind #23	M	c.1240-2A>G	p.?	splice	de novo	2	2	2	2	2	2	1	1	0	0.5	14.5	3	17.5	this report			
Ind #24	F	c.1240-1_1240insTAA	p.?	splice	de novo	1	1	2	2	2	2	1	0	1	0.5	0.5	11	3.5	14.5	this report		
DYRK1A_R	Bronicki et al. 2014	Bronicki_#2	M	c.613C>T	p.Arg205*	nonsense	de novo	1	2	2	2	0	1	0	1	NA	0.5	9.5	5	14.5	this report	
		Bronicki_#3	M	c.621_624 delinsGAA	p.Glu208fs	frameshift	de novo	1	2	2	2	1	1	1	0.5	NA	11.5	3.5	15	this report		
		Bronicki_#8	M	c.844dupA	p.Ser282fs	frameshift	de novo	2	2	2	2	2	2	1	0	0.5	0	13.5	3	16.5	this report	
		Bronicki_#10	F	c.1232dup	p.Arg413fs	frameshift	de novo	2	2	2	2	2	1	1	1	0.5	0.5	14	4.5	18.5	this report	
		Ruud et al., 2015	Ruud_#1	M	c.613C>T	p.Arg205*	nonsense	de novo	1	1	1	2	2	1	1	0.5	NA	11.5	4.5	16	literature	
		Van Bon et al. 2016	VanBon_SSC138F	F	c.1098+1G>A	p.?	splice	de novo	2	2	1	2	0	2	0	0.5	NA	9.5	4.5	14	literature	
		VanBon_SSC120C	M	c.143_144 del	p.Ile48fs	frameshift	de novo	1	2	2	2	2	2	NA	NA	0.5	NA	11.5	4	15.5	literature	
		VanBon_SSC135F	M	c.1491del	p.Ala498fs	frameshift	de novo	1	1	2	2	2	2	0	NA	NA	NA	10	3.5	13.5	literature	
		VanBon_UMCN1F	F	c.799C>T	p.Gln267*	nonsense	NA	2	2	2	2	2	NA	NA	NA	0.5	NA	10.5	3	13.5	literature	
		VanBon_UMCN2M	M	c.1240-2 A>G	p.?	splice	de novo	1	2	2	2	0	2	0	1	0.5	NA	10.5	3.5	14	literature	
VanBon_Troina1M	M	c.516+2 T>C	p.?	splice	de novo	2	2	2	2	2	1	1	NA	0.5	0.5	13	5	18	literature			
VanBon_GF2852M	M	c.367C>T	p.Gln123*	nonsense	de novo	1	2	2	2	2	2	0	0	0.5	NA	11.5	4.5	16	literature			
VanBon_Leuven3F	F	c.665-9_665-5delTTCTC	p.?	splice	de novo	1	2	2	2	0	2	0	1	0.5	NA	10.5	4	14.5	literature			
test	this report	Ind #18	F	c.1978del	p.Ser660fs	frameshift	de novo	1	0.5	1	0	0	1	0.5	1	0	0	5	2.5	7.5	this report	
		Ind #25	F	c.503G>A	p.Gly168Asp	missense	de novo	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	this report
		Ind #26	F	c.764G>A	p.Arg255Gln	missense	NA	1	2	0	0	0	1	0	1	0	0	5	NA	NA	this report	
		Ind #27	M	c.860A>T	p.Asp287Val	missense	de novo	2	2	2	NA	2	1	NA	1	NA	NA	10	NA	NA	this report	
		Ind #28	F	c.914T>G	p.Ile305Arg	missense	de novo	2	2	2	0	2	1	1	1	0.5	NA	11.5	4.5	16	this report	
		Ind #29	M	c.972T>A	p.Ser324Arg	missense	de novo	0	1	0	2	1	2	0	1	0	0	7	3.5	10.5	this report	
		Ind #30	F	c.1098G>T	p.Ile318_Glu366del	inframe	de novo	2	2	2	1	2	1	1	1	0	0	12	4	16	this report	
		Ind #31	M	c.1384T>C	p.Tyr462His	missense	de novo	0	0	0*	0	2	1	1	0	0	0	4	2.5	6.5	this report	
		Ind #32	F	c.1400G>A	p.Arg467Gln	missense	de novo	1	2	2	0	0	1	NA	0	NA	0.5	6.5	4	10.5	this report	
		Ind #33	M	c.1457G>A	p.Gly486Asp	missense	de novo	0	0	0*	NA	0	2	NA	0	NA	NA	2	NA	NA	this report	
Bronicki et al. 2014	Bronicki_#9	F	c.1763C>A	p.Thr588Asn	missense	de novo	2	2	2	1	1	1	NA	1	0.5	0.5	11	4.5	15.5	this report		
Ruud et al., 2015	Ruud_#2	M	c.932C>T	p.Ser311Phe	missense	de novo	1	2	2	0	2	1	1	1	NA	NA	10	2.5	12.5	this report		
Ji et al., 2015	Ji_#P7	F	c.563A>T	p.Lys188Ile	missense	de novo	1	2	1	2	0	1	1	1	0.5	NA	9.5	3.5	13	literature		
Ji_#P8	F	c.734T>G	p.Leu245Arg	missense	de novo	2	2	2	2	1	1	0.5	1	0.5	NA	12	NA	NA	literature			
Ji_#P9	M	c.883C>T	p.Leu295Phe	missense	de novo	0	2	2	0	0	1	NA	NA	0.5	NA	5.5	NA	NA	literature			
other monogenic forms	ANKRD11	ANKRD11_1	F	pathogenic variant		de novo	0	0	0	1	0	1	0	0	0	NA	2	1.5	3.5	this report		
		ANKRD11_2	F	pathogenic variant		inherited moth	1	1	0	2	0	0	0	1	0	0	5	1	6	this report		
		ANKRD11_3	M	pathogenic variant		de novo	1	1	0	0	NA	1	0	0	0	0	3	1	4	this report		
		ANKRD11_4	F	pathogenic variant		de novo	1	0	0	2	0	1	0	0	0	0	4	1	5	this report		
		ANKRD11_5	F	pathogenic variant		de novo	1	0	0	2	0	0	0	0	0	0	3	1	4	this report		
		MED13L	MED13L_1	M	pathogenic variant		de novo	1	1	0	0	2	1	0	1	0	NA	6	1.5	7.5	this report	
			MED13L_2	M	pathogenic variant		de novo	1	2	1	2	0	1	1	1	0	0	9	0.5	9.5	this report	
			MED13L_3	F	pathogenic variant		de novo	2	2	0	0	0	0	0	1	0	0	5	3	8	this report	
			MED13L_4	F	pathogenic variant		de novo	1	1	1	0	0	1	0	1	0	0	5	2	7	this report	
			MED13L_5	F	pathogenic variant		de novo	1	1	0	0	0	1	0	1	0	0	4				

Table S3

training set	source	gnomad allele		chr	pos	id	ref	alt	c.	p.	kinase domain	Freq Allele Reference				Freq Allele Variant					
		CADD	RefV									RefM	RefO	RefP	RefF	SubV	SubM	SubO	SubP	SubF	
N-set	gnomad>=2	2	21	38 845 018	rs7802	G	A	c.43G>A	p.Val15Ile	no	26.6	94	0	0	-1	0	0	0	0	-1	0
N-set	gnomad>=2	4	21	38 845 018	rs7802	G	T	c.43G>T	p.Val15Phe	no	28.5	94	0	0	-1	0	0	0	0	-1	0
T-test	clinvar_VUS	0	21	38 845 022		G	A	c.47G>A	p.Arg16Gln	no	23.2	94	0	0	-1	0	0	0	7	-1	0
N-set	gnomad>=2	2	21	38 845 030	rs7765	C	T	c.55C>T	p.Pro19Ser	no	27.5	94	0	0	-1	0	0	0	0	-1	0
T-test	clinvar_VUS	0	21	38 845 078		C	T	c.103C>T	p.Pro35Ser	no	22.5	78	-1	-1	-1	-1	21	-1	-1	-1	-1
N-set	gnomad>=2	4	21	38 845 097	rs7541G	A	C	c.122G>A	p.Ser41Asn	no	26.8	95	4	0	-1	-1	4	0	0	-1	-1
N-set	gnomad>=2	2	21	38 845 097	rs7541G	A	C	c.122G>C	p.Ser41Thr	no	26.4	95	4	0	-1	-1	0	0	0	-1	-1
N-set	gnomad>=2	8	21	38 845 102	rs3675C	T	C	c.127C>T	p.Arg43Cys	no	24	82	0	3	-1	-1	17	0	0	-1	-1
T-test	clinvar_VUS	1	21	38 845 103		G	A	c.128G>A	p.Arg43His	no	25.6	82	0	3	-1	-1	0	0	0	-1	-1
N-set	gnomad>=2	3	21	38 845 105	rs7585C	T	C	c.130C>T	p.Arg44Cys	no	26.5	56	4	3	-1	-1	0	0	0	-1	-1
N-set	gnomad>=2	2	21	38 845 106	rs7802G	A	C	c.131G>A	p.Arg44His	no	22.5	56	4	3	-1	-1	43	0	0	-1	-1
T-test	clinvar_VUS	0	21	38 845 114		A	G	c.139A>G	p.Asn47Asp	no	22.9	52	-1	-1	-1	-1	0	-1	-1	-1	-1
N-set	gnomad>=2	10	21	38 845 144	rs1045T	C	C	c.169T>C	p.Ser57Pro	no	5.976	39	0	0	-1	-1	52	0	1	-1	-1
N-set	gnomad>=2	2	21	38 850 498	rs8883G	A	A	c.223G>A	p.Val75Ile	no	22.3	39	4	0	-1	-1	0	0	0	-1	-1
N-set	gnomad>=2	3	21	38 850 514	rs1442G	A	C	c.239G>A	p.Arg80Gln	no	24.5	73	29	0	-1	-1	0	0	1	-1	-1
T-test	clinvar_VUS	1	21	38 850 514		G	T	c.239G>T	p.Arg80Leu	no	24.3	73	29	0	-1	-1	0	0	0	-1	-1
T-test	clinvar_VUS	0	21	38 850 529		T	G	c.254T>G	p.Phe85Cys	no	29.9	100	29	0	-1	-1	0	0	0	-1	-1
N-set	gnomad>=2	5	21	38 850 543	rs7674A	G	G	c.268A>G	p.Thr90Ala	no	21.1	47	19	15	-1	0	4	5	6	-1	0
N-set	gnomad>=2	2	21	38 850 552	rs1388C	G	C	c.277C>G	p.Leu93Val	no	23.4	100	76	24	0	53	0	8	18	58	0
T-test	Dang et al., 2018	0	21	38852967		C	T	c.355C>T	p.His119Tyr	no	22.8	100	-1	-1	-1	-1	0	-1	-1	-1	-1
T-test	Dang et al., 2018	0	21	38852988		G	T	c.376G>T	p.Asp126Tyr	no	31	100	41	0	-1	0	0	0	0	-1	0
T-test	clinvar_VUS	0	21	38 853 002		G	C	c.390G>C	p.Lys130Asn	no	25.3	100	71	9	-1	0	0	0	6	-1	7
T-test	Dang et al., 2018	0	21	38853010		G	A	c.398G>A	p.Arg133Gln	no	25.2	100	55	0	-1	7	0	0	0	-1	0
N-set	gnomad>=2	4	21	38 853 019	rs7501A	G	G	c.407A>G	p.Tyr136Cys	no	25.6	73	52	11	-1	0	0	0	0	-1	0
T-test	clinvar_VUS	0	21	38 853 043		A	G	c.431A>G	p.Asn144Ser	no	25.2	100	86	23	-1	7	0	0	0	-1	0
N-set	gnomad>=2	2	21	38 853 085	rs7566G	A	C	c.473G>A	p.Arg158His	no	29.8	100	96	92	86	52	0	0	0	0	0
N-set	gnomad>=2	3	21	38 853 112	rs1456A	G	C	c.500A>G	p.Lys167Arg	yes	29.1	100	100	64	0	9	0	0	4	0	0
P-set	clinvar_likpatho, this report	0	21	38 853 115		G	A	c.503G>A	p.Gly168Asp	yes	31	100	100	86	69	100	0	0	0	0	0
T-test	clinvar_VUS	0	21	38 853 123		G	A	c.511G>A	p.Gly171Arg	yes	32	100	100	86	66	80	0	0	0	0	0
P-set	clinvar_likpatho	0	21	38 858 785		A	G	c.533A>G	p.Asp178Gly	yes	32	100	93	86	60	15	0	0	0	0	5
N-set	gnomad>=2	3	21	38 858 795	rs7545G	T	C	c.543G>T	p.Glu181Asp	yes	12.42	100	57	8	3	0	0	11	3	3	0
P-set	clinvar_likpatho	0	21	38 858 815		A	T	c.563A>T	p.Lys188Ile	yes	31	100	96	97	95	92	0	0	0	0	0
T-test	NA	0	21	38 858 815		A	G	c.563A>G	p.Lys188Arg	yes	27.9	100	96	97	95	92	0	0	0	0	0
T-test	Dang et al., 2018	0	21	38 858 835		G	A	c.583G>A	p.Ala195Thr	yes	23.4	100	14	26	61	56	0	3	0	0	0
N-set	gnomad>=2	2	21	38 858 866	rs7577G	A	C	c.614G>A	p.Arg205Gln	yes	25.6	100	30	8	0	4	0	4	3	0	0
T-test	DDD, 2015	0	21	38 858 872		T	C	c.620T>C	p.Leu207Pro	yes	29.5	100	93	93	92	92	0	0	0	0	0
T-test	clinvar_VUS	0	21	38 862 480		A	G	c.668A>G	p.His223Arg	yes	22	100	0	3	0	0	0	35	59	93	63
N-set	gnomad>=2	2	21	38 862 488	rs1462C	T	C	c.676C>T	p.Arg226Cys	yes	32	100	59	5	0	0	0	0	3	0	0
N-set	gnomad>=2	3	21	38 862 489	rs7587G	A	C	c.677G>A	p.Arg226His	yes	25.7	100	59	5	0	0	0	0	9	0	5
T-test	clinvar_VUS	0	21	38 862 534		T	C	c.722T>C	p.Leu241Pro	yes	29.7	100	100	97	100	92	0	0	0	0	0
T-test	Ji, 2015	0	21	38 862 546		T	G	c.734T>G	p.Leu245Arg	yes	29	100	100	97	100	92	0	0	0	0	0
T-test	this report	0	21	38 862 576		G	A	c.764G>A	p.Arg255Gln	yes	24	100	75	23	46	57	0	0	5	0	5
T-test	Dang et al., 2018; DDD, 2015	0	21	38 862 589		G	T	c.777G>T	p.Leu259Phe	yes	24.8	100	89	84	67	9	0	3	0	0	0
N-set	gnomad>=2	3	21	38 862 619	rs7672G	A	C	c.807G>A	p.Met269Ile	yes	22.4	73	16	0	2	5	0	5	43	58	34
N-set	gnomad>=2	2	21	38 862 626	rs1385G	T	C	c.814G>T	p.Ala272Ser	yes	24.7	100	90	57	65	22	0	5	1	0	22
P-set	clinvar_likpatho	0	21	38 862 641		G	C	c.829G>C	p.Ala277Pro	yes	28.9	100	12	27	0	3	0	0	0	0	0
N-set	gnomad>=2	14	21	38 862 657	rs7588G	A	C	c.845G>A	p.Ser282Asn	yes	22.6	100	42	16	4	0	0	36	17	7	0
N-set	gnomad>=2	2	21	38 862 659	rs7803A	G	C	c.847A>G	p.Ile283Val	yes	22.4	100	89	53	79	31	0	10	32	6	5
T-test	Zhang et al., 2015	0	21	38 862 671		G	T	c.859G>T	p.Asp287Tyr	yes	32	100	100	97	95	100	0	0	0	0	0
T-test	Widowati	0	21	38 862 671		G	A	c.859G>A	p.Asp287Asn	yes	32	100	100	97	95	100	0	0	0	0	0
P-set	clinvar_patho, this report	0	21	38 862 672		A	T	c.860A>T	p.Asp287Val	yes	28.9	100	100	97	95	100	0	0	0	0	0
T-test	clinvar_VUS	0	21	38 862 681		C	G	c.869C>G	p.Pro290Arg	yes	28.3	100	100	97	95	100	0	0	0	0	0
P-set	clinvar_patho	0	21	38 862 695	rs797C	T	C	c.883C>T	p.Leu295Phe	yes	29.1	100	96	87	51	72	0	3	0	0	0
N-set	gnomad>=2	2	21	38 862 719	rs7778A	G	C	c.907A>G	p.Ile303Val	yes	24	100	100	74	67	65	0	0	12	28	19
T-test	this report	0	21	38 862 726		T	G	c.914T>G	p.Ile305Arg	yes	26.4	100	80	25	26	73	0	0	0	0	0
P-set	clinvar_likpatho	0	21	38 862 734		T	G	c.922T>G	p.Phe308Val	yes	29.1	100	100	89	65	94	0	0	0	0	0
P-set	clinvar_likpatho	0	21	38 862 735		T	C	c.923T>C	p.Phe308Ser	yes	32	100	100	89	65	94	0	0	0	0	0
P-set	clinvar_patho	0	21	38 862 744		C	T	c.932C>T	p.Ser311Phe	yes	33	100	100	71	30	27	0	0	0	0	0
N-set	gnomad>=2	2	21	38 862 751	rs7708G	T	C	c.939G>T	p.Gln313His	yes	14.51	100	71	9	0	0	0	11	12	0	4
P-set	clinvar_likpatho = this report	0	21	38 865 339		T	A	c.972T>A	p.Ser324Arg	yes	28.1	100	100	95	100	92	0	0	0	0	0
P-set	clinvar_likpatho	0	21	38 865 341		G	A	c.974G>A	p.Arg325His	yes	33	100	100	95	100	92	0	0	0	0	0
P-set	clinvar_likpatho	0	21	38 865 347		A	G	c.980A>G	p.Tyr327Cys	yes	29.3	100	100	95	100	92	0	0	0	0	0
T-test	clinvar_VUS, Stessman et al., 2017	0	21	38 865 349		C	T	c.982C>T	p.Arg328Trp	yes	34	100	100	95	100	92	0	0	0	0	0
N-set	gnomad>=2	89	21	38 865 388	rs2005G	T	C	c.1021G>T	p.Ala341Ser	yes	24.9	100	82	53	49	48	0	0	16	12	38
N-set	gnomad>=2	4	21	38 865 398	rs7466T	C	C	c.1031T>C	p.Met344Thr	yes	26.9	100	84	64	66	68	0	0	2	0	0
P-set	clinvar_likpatho	0	21	38 865 403		T	C	c.1036T>C	p.Ser346Pro	yes	29.7	100	100	92	100	92	0	0	0	0	0
P-set	clinvar_likpatho	0	21	38 865 404		C	T	c.1037C>T	p.Ser346Phe	yes	32	100	100	92	100	92	0	0	0	0	0
T-test	Trujillano et al., 2017	0	21	38 865 407		T	G	c.1040T>G	p.Leu347Arg	yes	31	100	96	84	47	83	0	0	0	0	0
N-set	gnomad>=2; clinvar_benign	573	21	38 865 433	rs1458A	G	C	c.1066A>G	p.Thr356Ala	yes	26.3	100	88	77	27	8	0	0	0	0	0
N-set	gnomad>=2	2	21	38 865 461	rs7563A	G	C	c.1094A>G	p.Asn365Ser	yes	22.5	100	89	22	0	0	0	3	25	84	71
N-set	gnomad>=2	2	21	38 865 461	rs7563A	T	C	c.1094A>T	p.Asn365Ile	yes	27.3	100</									



N-set	gnomad>=2	32	21	38 884 218	rs1441G	A	c.1676G>A	p.Arg559His	no	22.7	91	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 227	rs7457T	A	c.1685T>A	p.Phe562Tyr	no	16.75	53	-1	-1	-1	-1	40	-1	-1	-1	-1	
N-set	gnomad>=2	4	21	38 884 253	rs7797A	G	c.1711A>G	p.Thr571Ala	no	18.84	44	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	10	21	38 884 254	rs7466C	G	c.1712C>G	p.Thr571Ser	no	15.02	44	-1	-1	-1	-1	13	-1	-1	-1	-1	
N-set	gnomad>=2	4	21	38 884 262	rs7682C	G	c.1720C>G	p.Pro574Ala	no	24.7	80	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	4	21	38 884 271	rs7713G	C	c.1729G>C	p.Val577Leu	no	24.6	89	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	8	21	38 884 274	rs774CA	G	c.1732A>G	p.Thr578Ala	no	21.6	89	-1	-1	-1	-1	0	-1	-1	-1	-1	
T-test	clinvar_VUS	0	21	38 884 281	A	G	c.1739A>G	p.Glu580Gly	no	26.7	89	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 286	rs1293C	T	c.1744C>T	p.His582Tyr	no	24.8	89	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 287	rs7595A	C	c.1745A>C	p.His582Pro	no	24.9	89	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	3	21	38 884 289	rs7678C	A	c.1747C>A	p.Pro583Thr	no	27.2	89	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	3	21	38 884 301	rs3681A	G	c.1759A>G	p.Thr587Ala	no	26.5	89	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	44	21	38 884 304	rs1495A	C	c.1762A>C	p.Thr588Pro	no	23.4	85	-1	-1	-1	-1	0	-1	-1	-1	-1	
P-set	clinvar_likpatho	0	21	38 884 305	C	A	c.1763C>A	p.Thr588Asn	no	22.8	85	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	8	21	38 884 331	rs2008G	A	c.1789G>A	p.Ala597Thr	no	22.1	84	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	3	21	38 884 347	rs7583A	T	c.1805A>T	p.His602Leu	no	22.8	51	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 352	rs1468A	G	c.1810A>G	p.Asn604Asp	no	22.7	61	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 359	rs3675C	A	c.1817C>A	p.Ser606Tyr	no	24.2	72	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	4	21	38 884 388	rs7605C	G	c.1846C>G	p.His616Asp	no	26.6	88	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 398	rs754CA	G	c.1856A>G	p.His619Arg	no	23.2	88	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	3	21	38 884 412	rs7504T	G	c.1870T>G	p.Leu624Val	no	22.3	62	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	8	21	38 884 421	rs7583C	T	c.1879C>T	p.Arg627Trp	no	32	80	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	5	21	38 884 422	rs7614G	A	c.1880G>A	p.Arg627Gln	no	22.5	80	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	9	21	38 884 440	rs1476A	T	c.1898A>T	p.Tyr633Phe	no	24.7	85	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	12	21	38 884 443	rs370CA	G	c.1901A>G	p.Asn634Ser	no	22.6	77	-1	-1	-1	-1	8	-1	-1	-1	-1	
N-set	gnomad>=2	6	21	38 884 452	rs1421C	T	c.1910C>T	p.Thr637Met	no	23.4	81	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	6	21	38 884 455	rs7755A	G	c.1913A>G	p.Asn638Ser	no	21.5	72	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	4	21	38 884 484	rs2014G	A	c.1942G>A	p.Val648Ile	no	26.8	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 505	rs7654A	G	c.1963A>G	p.Met655Val	no	22.5	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	4	21	38 884 527	rs7515C	T	c.1985C>T	p.Thr662Met	no	24.9	51	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 556	rs7487A	G	c.2014A>G	p.Thr672Ala	no	24.4	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	9	21	38 884 567	rs7567A	C	c.2025A>C	p.Gln675His	no	22.3	60	-1	-1	-1	-1	34	-1	-1	-1	-1	
N-set	gnomad>=2	5	21	38 884 572	rs5438A	G	c.2030A>G	p.Asn677Ser	no	23.2	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2; clinvar_benign	388	21	38 884 577	rs5572G	C	c.2035G>C	p.Ala679Pro	no	25.2	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2; clinvar_benign	8	21	38 884 589	rs2015C	T	c.2047C>T	p.Arg683Cys	no	24.3	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	57	21	38 884 640	rs5405G	A	c.2098G>A	p.Val700Ile	no	22.5	38	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	23	21	38 884 652	rs2005G	A	c.2110G>A	p.Val704Ile	no	24.5	55	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	5	21	38 884 653	rs7637T	C	c.2111T>C	p.Val704Ala	no	22.4	55	-1	-1	-1	-1	38	-1	-1	-1	-1	
N-set	gnomad>=2	9	21	38 884 662	rs1513A	G	c.2120A>G	p.Asn707Ser	no	24.1	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 667	rs7816C	T	c.2125C>T	p.Arg709Cys	no	33	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	4	21	38 884 668	rs3747G	A	c.2126G>A	p.Arg709His	no	23.7	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	6	21	38 884 682	rs5546A	G	c.2140A>G	p.Ile714Val	no	22.2	55	-1	-1	-1	-1	8	-1	-1	-1	-1	
N-set	gnomad>=2	3	21	38 884 697	rs7744A	G	c.2155A>G	p.Thr719Ala	no	17.54	68	-1	-1	-1	-1	21	-1	-1	-1	-1	
N-set	gnomad>=2	23	21	38 884 713	rs1434C	G	c.2171C>G	p.Ala724Gly	no	24.6	51	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	2	21	38 884 715	rs1401A	C	c.2173A>C	p.Asn725His	no	26.5	90	-1	-1	-1	-1	0	-1	-1	-1	-1	
T-test	clinvar_VUS	1	21	38 884 719	C	G	c.2177C>G	p.Thr726Arg	no	24.4	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	12	21	38 884 722	rs7605G	T	c.2180G>T	p.Gly727Val	no	24.6	85	-1	-1	-1	-1	8	-1	-1	-1	-1	
N-set	gnomad>=2; clinvar_benign	10	21	38 884 757	rs2015A	G	c.2215A>G	p.Met739Val	no	21.5	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	23	21	38 884 757	rs2015A	T	c.2215A>T	p.Met739Leu	no	21.3	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	3	21	38 884 767	rs7498G	A	c.2225G>A	p.Gly742Glu	no	26.8	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
T-test	clinvar_VUS	0	21	38 884 785	C	T	c.2243C>T	p.Ser748Phe	no	28.8	94	25	0	-1	-1	0	0	0	0	-1	-1
N-set	clinvar_benign	0	21	38 884 788	rs7532C	T	c.2246C>T	p.Pro749Leu	no	29.3	94	25	0	-1	-1	0	0	0	0	-1	-1
N-set	gnomad>=2	22	21	38 884 793	rs1475A	T	c.2251A>T	p.Thr751Ser	no	22.7	77	0	0	-1	-1	0	0	0	0	-1	-1
N-set	gnomad>=2	2	21	38 884 830	rs7711C	G	c.2288C>G	p.Ser763Trp	no	32	94	-1	-1	-1	-1	0	-1	-1	-1	-1	
N-set	gnomad>=2	4	21	38 884 830	rs7711C	T	c.2288C>T	p.Ser763Leu	no	29.7	94	-1	-1	-1	-1	0	-1	-1	-1	-1	

Table S4

CoG	limma p adj	limma p	delta	beta	SVM	chr	pos	strand	Islands Name	Relation to	UCSC RefGene Name	UCSC RefGene Accession	UCSC RefGene Group	GOF site	
ce27246462	1.23e-05	0.0153	0.1386			chr1	2106365			OpenSea	PRKXZ,PRKXZ,PRKXZ	NM_001035881;NM_001035882;NM_001035883	Body;Body;Body	Yes	
ce23484809	7.28e-05	0.0335	0.1190	Included		chr1	2106400			OpenSea	PRKXZ,PRKXZ,PRKXZ	NM_001035881;NM_001035882;NM_001035883	Body;Body;Body	Yes	
ce02067531	1.29e-05	0.0156	0.1314	Included		chr1	16714027			OpenSea	SZRD1;SZRD1;SZRD1;SZRD1;SZRD1;SZRD1	NR_073503;NR_073502;NR_073501	Body;Body;Body;Body;Body;Body	Yes	
ce14076729	7.70e-07	0.0042	-0.1156	Included		chr1	38126268			OpenSea					
ce24375409	5.24e-05	0.0226	-0.1499	Included		chr1	32342020			OpenSea	EPHA10	NM_001094939	Body		
ce07865091	8.57e-07	0.0044	-0.1255	Included		chr1	43814306		chr1:43814305-43815277	OpenSea		NM_005373	Body		
ce11839415	5.04e-07	0.0034	-0.1249			chr1	43814764		chr1:43814305-43815277	OpenSea	MPL	NM_005373	Body	Yes	
ce08705657	3.82e-04	0.0050	-0.1474	Included		chr1	43814983		chr1:43814305-43815277	OpenSea	MPL	NM_005373	Body	Yes	
ce13939721	1.11e-06	0.0054	-0.1763			chr1	43815035		chr1:43814305-43815277	OpenSea	MPL	NM_005373	Body	Yes	
ce06496648	1.36e-05	0.0160	-0.1020	Included		chr1	47488961		chr1:47488226-47489633	N Shore	CYP4X1	NM_178033	TSS1500		
ce08817545	8.48e-07	0.0044	0.1316	Included		chr1	48559803			OpenSea					
ce09131818	7.16e-06	0.0119	0.1694	Included		chr1	54232631			OpenSea	NDC1;NDC1;NDC1	NM_001168551;NM_018087;NR_033	3'UTR;3'UTR;Body	Yes	
ce24803390	1.15e-07	0.0020	0.2879			chr1	56099793			OpenSea					
ce21039345	8.04e-06	0.0126	0.2211	Included		chr1	56107046			OpenSea					
ce45474764	1.63e-05	0.0172	-0.1193	Included		chr1	56184657			OpenSea					
ce24496745	0.0001	0.0400	-0.1191	Included		chr1	57285156			OpenSea	Clorf168;Clorf168	NM_001004303;NM_001004303	1stExon;5'UTR	Yes	
ce24495888	8.73e-06	0.0129	-0.1222			chr1	57285399			OpenSea	Clorf168	NM_001004303	TSS200		
ce05374238	1.94e-06	0.0068	-0.1554	Included		chr1	57285466			OpenSea	Clorf168	NM_001004303	TSS200		
ce09467404	2.41e-06	0.0074	-0.1105	Included		chr1	57424327			OpenSea	C8B;C8B;C8B	NM_01278544;NM_001278543;NM_01278544	Body;Body;Body	Yes	
ce17461600	0.0001	0.0390	-0.1076	Included		chr1	57983368			OpenSea	DAB1	NM_021080	5'UTR		
ce00026909	2.93e-07	0.0025	-0.1254	Included		chr1	58089001			OpenSea	DAB1	NM_021080	5'UTR		
ce20292791	1.70e-05	0.0177	-0.1124	Included		chr1	58099357			OpenSea	DAB1	NM_021080	5'UTR		
ce18163452	4.94e-06	0.0103	-0.1020	Included		chr1	63466707			OpenSea					
ce00320094	4.08e-05	0.0261	-0.1025	Included		chr1	68695183		chr1:68696639-68697628	N Shore	GPR177;GPR177	NM_001002292;NM_024911	Body;Body	Yes	
ce02553170	5.46e-05	0.0207	0.1696	Included		chr1	84556762			OpenSea	PRKACB;PRKACB	NM_020578;NM_002271	Body;Body		
ce26256793	0.0001	0.0424	-0.1291			chr1	103574469			OpenSea					
ce03520644	1.51e-05	0.0166	-0.1486			chr1	103574523			OpenSea					
ce06381331	6.15e-05	0.0151	-0.1665	Included		chr1	103574597			OpenSea					
ce19688885	6.23e-06	0.0118	-0.1918			chr1	103574619			OpenSea					
ce10691490	2.57e-07	0.0025	-0.1242	Included		chr1	103716210			OpenSea					
ce00583780	7.75e-05	0.0347	-0.1067	Included		chr1	105031120			OpenSea					
ce11452579	1.14e-07	0.0022	-0.1307	Included		chr1	105148900			OpenSea					
ce26004771	1.42e-05	0.0163	0.1047	Included		chr1	178456093		chr1:178455799-178456286	Island					
ce23898184	2.98e-06	0.0082	0.1183	Included		chr1	192921219			OpenSea					
ce10382206	1.03e-06	0.0079	-0.1101	Included		chr1	209379448			OpenSea					
ce18086868	5.86e-06	0.0109	-0.1011	Included		chr1	220463627			OpenSea					
ce04095097	3.06e-05	0.0229	-0.1100	Included		chr10	51551512			OpenSea					
ce06379227	4.73e-06	0.0097	-0.1334	Included		chr10	52179123			OpenSea					
ce07541020	0.0001	0.0399	0.1082	Included		chr10	44068714		chr10:44069359-44070179	N Shore	ZNF339;ZNF329;ZNF329	NM_001099284;NM_001099282;NM_001099281	5'UTR;5'UTR;5'UTR	Yes	
ce18768136	5.04e-05	0.0289	0.1198	Included		chr10	52096068			OpenSea					
ce12848457	0.0002	0.0481	-0.1152	Included		chr10	52566320			OpenSea					
ce00321393	7.00321393	0.0331	-0.1294	Included		chr10	54277038			OpenSea					
ce24059708	1.60e-05	0.0172	-0.1677	Included		chr10	57170612			OpenSea					
ce02281493	5.18e-05	0.0292	-0.1209	Included		chr10	67088005			OpenSea					
ce09355528	8.13e-06	0.0047	-0.1086	Included		chr10	71128300			OpenSea					
ce21324555	2.44e-05	0.0158	-0.1158	Included		chr10	72648679		chr10:72647738-72648317	S Shore					
ce10592592	1.47e-11	5.674E-06	-0.1228			chr10	74034663		chr10:74033167-74034668	Island					
ce20897136	2.09e-06	0.0120	-0.1187	Included		chr10	74034667		chr10:74033167-74034668	Island					
ce04856311	7.09e-06	0.0119	-0.1243	Included		chr10	74035570		chr10:74033167-74034668	S Shore					
ce14812036	0.0001	0.0434	0.1196	Included		chr10	82895774			OpenSea					
ce19686855	9.28e-06	0.0133	0.1149	Included		chr10	84402299			OpenSea					
ce12185615	8.12e-05	0.0062	0.1065	Included		chr10	86708533			OpenSea					
ce13463988	6.99e-05	0.0330	-0.1082	Included		chr10	87676204			OpenSea					
ce01102638	2.28e-07	0.0024	0.1829	Included		chr10	87679282			OpenSea					
ce09198129	5.76e-05	0.0079	-0.1020	Included		chr10	100134713			OpenSea					
ce19316579	4.26e-05	0.0265	-0.1361	Included		chr10	100843963			OpenSea					
ce02860671	7.88e-06	0.0125	0.1321	Included		chr10	103469861			OpenSea					
ce09796228	2.02e-06	0.0062	0.1121	Included		chr10	10737081			OpenSea					
ce14939821	0.0002	0.0471	0.1177	Included		chr10	128947230			OpenSea					
ce19966745	0.0001	0.0445	-0.1106	Included		chr10	134965215			OpenSea					
ce04023364	1.03e-05	0.0139	-0.1089	Included		chr11	850592			Island					
ce11286550	8.12e-05	0.0062	0.1065	Included		chr11	86748923			OpenSea					
ce09535011	3.09e-05	0.0230	0.1121	Included		chr11	18601039			OpenSea					
ce03466399	5.98e-05	0.0312	0.1233	Included		chr11	40422876			OpenSea					
ce25298685	2.73e-06	0.0047	-0.2075	Included		chr11	42517544			OpenSea					
ce15119262	0.0002	0.0465	0.1047	Included		chr11	70827457			OpenSea					
ce10832886	2.07e-05	0.0190	0.1294	Included		chr11	76587072			OpenSea					
ce09407810	2.02e-05	0.0067	0.1069	Included		chr11	80089753			OpenSea					
ce07236114	5.99e-07	0.0037	-0.1843			chr11	84254533			OpenSea					
ce19981409	8.32e-05	0.0357	-0.1147	Included		chr11	89225042		chr11:89224416-89224718	S Shore					
ce14384230	1.33e-05	0.0158	-0.1705	Included		chr11	90479256			OpenSea					
ce02719051	8.32e-05	0.0310	0.1154	Included		chr11	122775061			OpenSea					
ce07085632	3.85e-06	0.0094	-0.1058	Included		chr11	125974404			OpenSea					
ce12787465	3.13e-09	0.0002	-0.1045	Included		chr12	6876717		chr12:6875481-6876923	Island					
ce16116033	3.76e-05	0.0184	-0.234	Included		chr12	6876922		chr12:6875481-6876923	Island					
ce18732361	5.29e-10	4.5532E-05	-0.2418	Included		chr12	6877080		chr12:6875481-6876923	S Shore					
ce12140312	2.57e-05	0.0212	0.1297	Included		chr12	68153796			OpenSea					
ce12500766	1.23e-07	0.0062	0.1032	Included		chr12	72486848			OpenSea					
ce15026637	1.63e-05	0.0172	0.1022	Included		chr12	43295760			OpenSea					
ce12531275	7.23e-07	0.0041	-0.1791			chr12	58132093		chr12:58132478-58132734	N Shore	AGAP2;AGAP2	NM_001122772;NM_014770	TSS200;Body	Yes	
ce09196858	5.76e-06	0.0038	-0.1554	Included		chr12	58132105		chr12:58132478-58132734	N Shore	AGAP2;AGAP2	NM_001122772;NM_014770	TSS200;Body	Yes	
ce17921464	7.95e-06	0.0125	-0.1475			chr12	58132114		chr12:58132478-58132734	S Shore	AGAP2;AGAP2	NM_001122772;NM_014770	TSS200;Body	Yes	
ce08425810	4.02e-06	0.0096	-0.2773			chr12	58132558		chr12:58132478-58132734	Island	AGAP2;AG				

ce08401872	5.30E-07	0.0035	-0.1135	Included	chr2	218487226 -	OpenSea	DIRC3		NR 026597	Body	
ce02529599	4.60E-05	0.0276	-0.1299		chr2	218843435 +	N Shore					
ce07516871	3.07E-05	0.0236	-0.1361		chr2	218843460 +	Island					
ce05783384	7.81E-05	0.0348	-0.1326	Included	chr2	218843735 -	Island					
ce18951332	5.19E-05	0.0292	-0.1204	Included	chr2	220777552 -	OpenSea					
ce15595388	3.14E-05	0.0343	-0.1013	Included	chr2	223945451 +	OpenSea					
ce07554758	0.00004799	0.0393	-0.1657	Included	chr2	223904541 +	OpenSea					
ce25614265	6.49E-05	0.0322	-0.2045		chr2	223924453 -	OpenSea					
ce12333487	0.0278	0.0446	-0.1037	Included	chr20	6145814 +	OpenSea					
ce02046768	0.00014188	0.0441	-0.1235	Included	chr20	6145814 +	OpenSea					
ce24933489	9.72E-06	0.0137	-0.1101	Included	chr20	11550004	OpenSea					
ce19628072	2.21E-05	0.0196	-0.1084		chr20	22446131 +	OpenSea					
ce20659672	1.79E-05	0.0180	-0.1386		chr20	22588572 +	OpenSea					
ce11092487	6.14E-05	0.0316	-0.1115		chr20	33585188 -	OpenSea					
ce02209770	1.36E-05	0.0160	-0.1331		chr20	35062903 -	N Shore					
ce14179325	1.58E-05	0.0170	-0.1017		chr20	44872915 -	N Shelf					
ce09204484	4.30E-05	0.0266	-0.1221		chr20	50475464 -	OpenSea					
ce09184271	4.69E-05	0.0278	-0.1029		chr21	14917769 +	OpenSea					
ce05713399	6.75E-05	0.0326	-0.1212		chr21	17675215 -	OpenSea					
ce00285225	0.00014548	0.0441	-0.1194		chr21	24757350 -	OpenSea					
ce05156137	7.42E-05	0.0341	-0.1437		chr21	35898975 -	OpenSea					
ce13528349	0.00010278	0.0390	-0.1161		chr21	35899082 -	OpenSea					
ce12020137	0.00014874	0.0446	-0.1422		chr21	39089071 -	OpenSea					
ce02725398	1.94E-06	0.0068	-0.1245	Included	chr3	9988144 -	OpenSea					
ce24048493	1.84E-07	0.0022	-0.1036	Included	chr3	28013305 -	OpenSea					
ce08043536	1.67E-05	0.0175	-0.1322	Included	chr3	28075339 -	OpenSea					
ce12982775	1.22E-05	0.0152	-0.1049	Included	chr3	36668218 -	OpenSea					
ce25959131	5.85E-05	0.0309	-0.1130	Included	chr3	50854546 -	OpenSea					
ce08731951	5.08E-05	0.0237	-0.1642		chr3	50898023 -	OpenSea					
ce23759448	2.51E-04	0.0136	-0.1336		chr3	57432353 -	OpenSea					
ce09473396	6.78E-06	0.0116	-0.1099	Included	chr3	65131979 -	OpenSea					
ce06789445	1.19E-06	0.0053	-0.1349	Included	chr3	69942688 +	OpenSea					
ce00276101	5.84E-06	0.0147	-0.1571	Included	chr3	73789657 +	OpenSea					
ce21090457	5.28E-05	0.0295	-0.1750	Included	chr3	77573709 +	OpenSea					
ce15823315	1.88E-07	0.0022	-0.1906	Included	chr3	85139023 +	OpenSea					
ce11774500	9.17E-07	0.0011	-0.2220	Included	chr3	92242496 +	OpenSea					
ce19229209	2.63E-06	0.0078	-0.1478	Included	chr3	99440401 +	OpenSea					
ce15788038	2.53E-07	0.0025	-0.1265	Included	chr3	10601093 +	OpenSea					
ce12182020	4.19E-05	0.0271	-0.1471	Included	chr3	106287942 +	OpenSea					
ce15941063	5.19E-05	0.0292	-0.1428	Included	chr3	10623545 +	OpenSea					
ce1338761	1.92E-05	0.0185	-0.1073	Included	chr3	109056802 -	OpenSea					
ce08755525	2.15E-05	0.0194	-0.1083	Included	chr3	110344467 -	OpenSea					
ce002197479	0.00013749	0.0446	-0.1123	Included	chr3	110798738 +	OpenSea					
ce25779782	5.41E-05	0.0297	-0.1128	Included	chr3	124840638 -	OpenSea					
ce26047482	1.28E-05	0.0156	-0.1225		chr3	124840721 +	OpenSea					
ce13084900	0.00010776	0.0473	-0.1164	Included	chr3	134466358 -	OpenSea					
ce02782222	4.88E-07	0.0034	-0.2956	Included	chr3	138892919 -	OpenSea					
ce15901796	0.00012874	0.0424	-0.1029	Included	chr3	146521668 +	OpenSea					
ce01349384	2.00E-05	0.0265	-0.1281	Included	chr3	178326190 +	OpenSea					
ce05892994	2.00E-05	0.0188	-0.1029	Included	chr3	180522985 +	OpenSea					
ce20060398	3.42E-05	0.0239	-0.1637		chr3	181473901 +	OpenSea					
ce27657132	2.77E-05	0.0221	-0.1266	Included	chr3	181889731 +	OpenSea					
ce12465189	3.70E-07	0.0012	-0.1150	Included	chr3	182015336 -	OpenSea					
ce07228359	3.30E-07	0.0027	-0.1022	Included	chr3	190419123 +	OpenSea					
ce17430979	6.28E-05	0.0319	-0.1184	Included	chr4	6034828 -	OpenSea					
ce20793532	0.00013584	0.0499	-0.1111	Included	chr4	63445461 -	OpenSea					
ce07545526	0.00012565	0.0422	-0.1175	Included	chr4	13478708 -	OpenSea					
ce12850605	1.90E-05	0.0185	-0.1092	Included	chr4	16277758 +	OpenSea					
ce22055989	5.87E-06	0.0036	-0.1256	Included	chr4	16286075 +	OpenSea					
ce18825766	0.00018285	0.0494	-0.1003	Included	chr4	27362758 -	OpenSea					
ce21675464	1.50E-05	0.0166	-0.1483	Included	chr4	33717878 +	OpenSea					
ce26839512	0.00010771	0.0401	-0.1056	Included	chr4	52518489 -	S Shore					
ce18389260	0.00013183	0.0460	-0.1231	Included	chr4	58024966 +	OpenSea					
ce12695778	2.90E-06	0.0082	-0.1793	Included	chr4	60884636 -	OpenSea					
ce17185888	1.41E-05	0.0162	-0.1124	Included	chr4	66657583 +	OpenSea					
ce24510434	0.00014022	0.0407	-0.1077	Included	chr4	82646412 -	N Shore					
ce24060890	1.50E-05	0.0184	-0.1142	Included	chr4	93229269 -	S Shelf					
ce18516067	0.00013646	0.0434	-0.2082	Included	chr4	98291885 +	OpenSea					
ce10179984	0.00012	0.0452	-0.1108	Included	chr4	11920881 -	OpenSea					
ce16931969	5.41E-06	0.0106	-0.1495	Included	chr4	119385495 +	OpenSea					
ce18378811	6.66E-05	0.0324	-0.1154	Included	chr4	114900050 +	N Shore					
ce01397050	6.34E-06	0.0045	-0.1451		chr4	116809079 +	OpenSea					
ce11755558	2.65E-06	0.0078	-0.1390	Included	chr4	122103800 -	OpenSea					
ce22819413	2.37E-06	0.0074	-0.1379	Included	chr4	122670640 -	OpenSea					
ce11290351	1.03E-05	0.0189	-0.1134	Included	chr4	122919312 -	OpenSea					
ce01543488	5.02E-06	0.0029	-0.1748	Included	chr4	129496305 -	OpenSea					
ce04524417	6.44E-05	0.0321	-0.1372	Included	chr4	129605243 +	OpenSea					
ce17743555	1.90E-05	0.0185	-0.1346	Included	chr4	130647434 +	OpenSea					
ce06099975	5.87E-06	0.0047	-0.1608	Included	chr4	134018369 -	OpenSea					
ce03147713	2.59E-05	0.0213	-0.1037	Included	chr4	138950261 -	OpenSea					
ce03639671	3.40E-05	0.0239	-0.1104	Included	chr4	145430689 +	OpenSea					
ce23357404	0.00016089	0.0464	-0.1116	Included	chr4	145453234 -	OpenSea					
ce06505720	1.49E-05	0.0166	-0.2055		chr4	145512579 -	OpenSea					
ce07318204	1.36E-05	0.0160	-0.2273		chr4	145566441 -	Island					
ce02524475	1.93E-05	0.0185	-0.1430		chr4	145566595 +	Island					
ce11749822	5.37E-06	0.0156	-0.1271	Included	chr4	145566663 -	Island					
ce03408024	4.00E-06	0.0096	-0.1055	Included	chr4	147075042 -	OpenSea					
ce25989755	8.33E-06	0.0128	-0.1536	Included	chr4	151134483 +	OpenSea					
ce15490177	1.54E-05	0.0169	-0.1218	Included	chr4	151440382 -	OpenSea					
ce03065175	2.16E-06	0.0072	-0.1658		chr4	155661758 -	Island					
ce00113509	3.68E-05	0.0250	-0.1889		chr4	155661849 -	Island					
ce16510306	2.16E-06	0.0055	-0.1116	Included	chr4	155912121 -	OpenSea					
ce11707239	1.83E-05	0.0183	-0.1155	Included	chr4	159526798 -	OpenSea					
ce05493047	9.02E-05	0.0369	-0.1931	Included	chr4	164810672 +	OpenSea					
ce24308312	3.24E-05	0.0234	-0.1537	Included	chr4	164911219 -	OpenSea					
ce17376782	7.27E-06	0.0142	-0.1912	Included	chr4	170989358 +	OpenSea					
ce02300027	8.00E-07	0.0043	-0.1426	Included	chr4	177711356 +	N Shore					
ce24597533	1.27E-06	0.0054	-0.2207		chr4	177714618 -	N Shore					
ce20791945	7.03E-09	0.0003	-0.1834	Included	chr4	177713180 -	S Shore					
ce16039660	4.16E-06	0.0097	-0.1013	Included	chr4	184686701 +	OpenSea					



Table S5

Sample_Group	Cohort #	DYRK1A variant	SVM	SVM score
test	Ind #25	p.Gly168Asp	positive	0,7430
test	Ind #26	p.Arg255Gln	negative	0,0709
test	Ind #27	p.Asp287Val	positive	0,8478
test	Ind #28	p.Ile305Arg	positive	0,9347
test	Ruud_#2	p.Ser311Phe	positive	0,8190
test	Ind #29	p.Ser324Arg	positive	0,9613
test	Ind #31	p.Tyr462His	negative	0,0242
test	Ind #32	p.Arg467Gln	positive	0,7874
test	Ind #33	p.Gly486Asp	GoF?	0,0008
test	Bronicki_#9	p.Thr588Asn	negative	0,0233
test	Ind #18	p.Ser660fs	positive	0,9275
test_rep	Ind #18_repl	p.Ser660fs_repl	positive	0,9247
KMT2A	-	-	negative	0,0849
KMT2A	-	-	negative	0,0822
KMT2A	-	-	negative	0,0590
KMT2A	-	-	negative	0,0365
KMT2A	-	-	negative	0,0216
KMT2A	-	-	negative	0,0162
KMT2A	-	-	negative	0,0137
KMT2A	-	-	negative	0,0080
ARID1B	-	-	negative	0,0254
ARID1B	-	-	negative	0,0166
ARID1B	-	-	negative	0,0104
ARID1B	-	-	negative	0,0092
Control	-	-	negative	0,1394
Control	-	-	negative	0,1052
Control	-	-	negative	0,0920
Control	-	-	negative	0,0892
Control	-	-	negative	0,0827
Control	-	-	negative	0,0720
Control	-	-	negative	0,0682
Control	-	-	negative	0,0657
Control	-	-	negative	0,0583
Control	-	-	negative	0,0563
Control	-	-	negative	0,0509
Control	-	-	negative	0,0504
Control	-	-	negative	0,0501
Control	-	-	negative	0,0495
Control	-	-	negative	0,0483
Control	-	-	negative	0,0439
Control	-	-	negative	0,0403
Control	-	-	negative	0,0391
Control	-	-	negative	0,0372
Control	-	-	negative	0,0353
Control	-	-	negative	0,0351
Control	-	-	negative	0,0349
Control	-	-	negative	0,0345
Control	-	-	negative	0,0343
Control	-	-	negative	0,0309
Control	-	-	negative	0,0292
Control	-	-	negative	0,0289
Control	-	-	negative	0,0286
Control	-	-	negative	0,0286
Control	-	-	negative	0,0277
Control	-	-	negative	0,0270
Control	-	-	negative	0,0260
Control	-	-	negative	0,0258
Control	-	-	negative	0,0254
Control	-	-	negative	0,0251
Control	-	-	negative	0,0240
Control	-	-	negative	0,0223

Control	-	-	negative	0,0218
Control	-	-	negative	0,0215
Control	-	-	negative	0,0214
Control	-	-	negative	0,0212
Control	-	-	negative	0,0210
Control	-	-	negative	0,0210
Control	-	-	negative	0,0204
Control	-	-	negative	0,0196
Control	-	-	negative	0,0196
Control	-	-	negative	0,0190
Control	-	-	negative	0,0190
Control	-	-	negative	0,0186
Control	-	-	negative	0,0184
Control	-	-	negative	0,0183
Control	-	-	negative	0,0177
Control	-	-	negative	0,0168
Control	-	-	negative	0,0167
Control	-	-	negative	0,0162
Control	-	-	negative	0,0162
Control	-	-	negative	0,0159
Control	-	-	negative	0,0158
Control	-	-	negative	0,0152
Control	-	-	negative	0,0148
Control	-	-	negative	0,0147
Control	-	-	negative	0,0142
Control	-	-	negative	0,0138
Control	-	-	negative	0,0137
Control	-	-	negative	0,0137
Control	-	-	negative	0,0125
Control	-	-	negative	0,0125
Control	-	-	negative	0,0120
Control	-	-	negative	0,0120
Control	-	-	negative	0,0119
Control	-	-	negative	0,0115
Control	-	-	negative	0,0114
Control	-	-	negative	0,0112
Control	-	-	negative	0,0112
Control	-	-	negative	0,0110
Control	-	-	negative	0,0110
Control	-	-	negative	0,0103
Control	-	-	negative	0,0103
Control	-	-	negative	0,0099
Control	-	-	negative	0,0099
Control	-	-	negative	0,0096
Control	-	-	negative	0,0092
Control	-	-	negative	0,0089
Control	-	-	negative	0,0089
Control	-	-	negative	0,0088
Control	-	-	negative	0,0083
Control	-	-	negative	0,0081
Control	-	-	negative	0,0073
Control	-	-	negative	0,0072
Control	-	-	negative	0,0069
Control	-	-	negative	0,0067
Control	-	-	negative	0,0061
Control	-	-	negative	0,0056
Control	-	-	negative	0,0049



Table S6

Variant			In silico		In vitro studies				Individuals				Final classification	
Variant	Nb patients	Nb gnomad	CADD	Conserved <sup>1</sup>	Expression	AutoP	Localization	Results previously reported	Report	Clinical features (CS <sub>DYRK1A</sub> )	Inheritance	DNAm profil	ACMG/AMP criteria	Final classification
p.Ser660fs	1	0		NA	NA	NA	Aggr.	NA	this report Ind #18	poorly (7.5/20)	dn	positive		Lik Patho
p.Ala341Ser	0	89	24,9	midly	100%	100%	ns	NA	-	-	-	-	BS1, BS2, BS3, BP4	Benign
p.Gly168Asp	2	0	31	highly	<50%	0%	ns	NA	this report Ind #25	NA	dn	positive	PS2, PS3, PM2, PP3	Patho
									GeneDx (clinvar SCV00057310)	NA	NA	NA		
p.Arg255Gln	1	0	24	midly	100%	100%	ns	NA	this report Ind #26	poorly (5/15)	?	negative	BS3, BP4	Lik Benign
p.Asp287Val	3	0	28,9	highly	<25%	0%	ns	drastic decrease of autoP and kinase activity (1)	DDD (Decipher 258963)	highly <sup>a</sup>	dn	NA	PS2, PS3, PM2, PP3, PP4	Patho
									this report Ind #27 (Geneva Hospital, SCV000598121)	interm. (10/15)	dn	positive		
									Tubingen (clinvar SCV00144673)	intermediate <sup>i</sup>	NA	NA		
p.Ile305Arg	1	0	26,9	midly	<25%	0%	ns	NA	this report Ind #28	highly (16/20)	dn	positive	PS2, PS3, PM2, PP4	Patho
p.Ser311Phe	2	0	33	moderate	<50%	0%	ns	drastic decrease of autoP and kinase activity (1,2)	Ruaud_#2 (clinvar SCV000586742)	interm. (12,5/20)	dn	positive	PS2, PS3, PM2, PP3	Patho
									GeneDx (clinvar SCV000520979)	NA	NA	NA		
p.Ser324Arg	1	0	28,1	highly	80%	<25%	ns	NA	this report Ind #29 (SCV000902439)	interm. (10,5/20)	dn	positive	PS2, PS3, PM2, PP3	Patho
p.Glu366Asp*	1	0	35*	midly	100%	100%	ns	NA	this report Ind #30	highly (16/20)	dn	NA	PS2, PS3, PM2, PP4, PP3	Patho *
p.Tyr462His	1	0	29,6	midly	100%	100%	ns	NA	this report Ind #31	poorly (6,5/20)	dn	negative	PS2, PM2, BS3,	Lik Benign
p.Arg467Gln	4	0	33	highly	<50%	0%	ns	drastic decrease of autoP and kinase activity (1,2)	DDD (Decipher 270174)	poorly <sup>b</sup>	dn	NA	PS2, PS3, PM1, PM2, PP3	Patho
									Posey et al., 2016 (clinvar SCV000245477)	interm. <sup>c</sup>	dn	NA		
									GeneDx (clinvar SCV000534701)	NA	NA	NA		
									this report Ind #32	interm. (10.5/20)	dn	positive		
p.Gly486Asp	2	0	29,1	midly	100%	100%	ns	no decrease of kinase acti	Dang et al., 2018	NA	NA	NA	PS2, PM2, BS3	VUS
									this report Ind #33 (Geneva Hospital, SCV000747759)	poorly (2/15) <sup>d</sup>	dn	GoF?		
p.Thr588Asn	1**	0***	22,8	midly	100%	100%	ns	no effect on autoP and P	Bronicki_#9 (clinvar SCV000196065 = SCV000965705)**	highly (15.5/20)	dn	negative	PS2, PM2, PP4, BS3, BP4	Lik Benign
p.Arg158His	1	2	29,8	highly	100%	100%	NA	NA	DDD (Deciper _LikBenign)	NA <sup>e</sup>	dn	NA	PS2, PP3, BS2, BS3	Benign
p.Ala277Pro	2	0	28,9	midly	50%	0%	NA	drastic decrease of autoP	DDD (Deciper _Patho)	interm. <sup>f</sup>	dn	NA	PS2, PS3, PM2	Patho
									GeneDx (SCV000571206)	NA	NA	NA		
p.Gly171Arg	1	0	32	highly	50%	0%	NA	NA	Invitae (clinvar SCV000952112.2_LikelyPatho)	interm. <sup>g</sup>	dn	NA	PS2, PS3, PM2, PP3	Patho
p.Leu241Pro	1	0	29,7	highly	~50%	0%	NA	NA	EGL diagnostic (clinvar SCV000703832.2_VUS)	NA	NA	NA	PS3, PM2, PP3	Lik Patho
p.Pro290Arg	1	0	28,3	highly	<50%	0%	NA	NA	Ambry (clinvar SCV000741841.1_VUS)	interm. <sup>h</sup>	NA	NA	PS3, PM2, PP3	Lik Patho

Table S7

Variant			In silico		In vitro studies	Individuals			Final classification	
Variant	Nb patients	Nb gnomad	CADD	Highly Conserved <sup>1</sup>	Results previously reported	Report	Clinical features (CS <sub>DYRK1A</sub> )	Inheritance	ACMG/AMP criteria	Final classification
His119Tyr	1	0	22,8	no	no effect on neurite outgrowth (Dang et al.)	Dang et al., 2018	intermediate <sup>a</sup>	NA	PM2, BS3	Lik benign
Lys188Ile	1	0	31	yes	drastic decrease of autoP and kinase activity (Arranz)	Ji et al., 2015 (clinvar SCV000206791_LikPatho)	highly (13/20)	dn	PS2,PS3, PM2, PP3, PP4	Patho
Ala195Thr	1	0	23,4	no	Increase kinase activity (Arranz et al.)?, no effect on neurite outgrowth (Dang et al.)	Dang et al., 2018	poorly <sup>b</sup>	NA	PM2, BS3, BP4	Lik benign
Leu207Pro	1	0	29,5	yes	drastic decrease of autoP and kinase activity (Arranz)	DDD (Decipher 259211_Patho)	intermediate <sup>c</sup>	dn	PS2,PS3, PM2, PP3	Patho
His223Arg	1	0	22,0	no	Increase kinase activity (Arranz) et al.?	GeneDx (clinvar SCV000620751_VUS)	NA	NA	PM2, BS3, BP4	Lik benign
Leu245Arg	2	0	29	yes	drastic decrease of autoP and kinase activity (Widowati et al.,Arranz et al.)	Ji et al., 2015 (clinvar SCV000206792_LikPatho)	highly (12/15)	dn	PS2,PS3, PM2, PP4, PP3	Patho
						Baylor college (clinvar SCV000807304_VUS)	intermediate <sup>d</sup>	dn		
Leu259Phe	1	0	24,8	no	Increase kinase activity (Arranz et al.) ?, no effect on neurite outgrowth (Dang et al.)	Dang et al., 2018	poorly <sup>e</sup>	NA	PM2, BS3, BP4	Lik benign
Asp287Tyr	1	0	32	yes	drastic decrease of autoP and kinase activity (Arranz et al.)	Zhang et al., 2015	intermediate <sup>f</sup>	dn	PS2,PS3, PM2, PP3	VUS
Leu295Phe	2	0	29,1	yes	slight decrease kinase activity (Widowati et al.,Arranz et al.)	Ji et al., 2015 = SCV000206793 (UCLA_LikPatho)	poorly (5.5/15)	dn	PS2,PS3, PM2, PP3	VUS
						GeneDx Clinvar SCV000321572_Patho)	NA	NA		
Phe308Val	1	0	29,1	yes	drastic decrease of autoP and kinase activity (Widowati et al., Arranz et al.)	Chicago Hospital (clinvar SCV000247240_LikPatho)	NA	NA	PS3, PM2, PM2, PP3	Lik patho
Gln313His	0	2	14,5	no	slight decrease kinase activity (Widowati et al.,Arranz et al.)	-	-	-	-	-
Arg325His	1	0	33	yes	drastic decrease of autoP and kinase activity (Arranz et al.)	GeneDx (clinvar SCV000574147_LikPatho)	NA	dn	PS2,PS3, PM2, PP3	Lik patho
Tyr327Cys	1	0	29,3	yes	drastic decrease of autoP and kinase activity (Arranz et al.)	BC Hospital (clinvar SCV000599256_LikPatho)	NA	dn	PS2,PS3, PM2, PP3	Lik patho
Arg328Trp	1	0	34	yes	drastic decrease of autoP and kinase activity (Arranz et al.)	GeneDx (clinvar SCV000492371_VUS)	NA	NA	PS3, PM2, PP3	Lik patho
Ser346Phe	1	0	32	yes	drastic decrease of autoP and kinase activity (Arranz et al.)	LMM Cambridge (Clinvar SCV000712522_LikPatho)	intermediate <sup>g</sup>	dn	PS2,PS3, PM2, PP3	Patho
Ser346Pro	2	0	29,7	yes	drastic decrease of autoP and kinase activity (Widowati et al.,Arranz et al.)	Bronicki et al., 2015 (Clinvar SCV000196060_LikPatho)	highly <sup>h</sup>	dn	PS2,PS3, PM2, PP3, PP4	Patho
						DDD (decipher 260956_Patho)	intermediate <sup>i</sup>	dn		
Leu347Arg	1	0	31	yes	drastic decrease of autoP and kinase activity (Arranz et al.)	Trujillano et al., 2017	highly <sup>j</sup>	dn	PS2,PS3, PM2, PP3, PP4	Patho
Arg438His	1	1	33	no	sklight decrease in kinase activity (Arranz et al.)?	Wang et al., 2016	NA	dn	PS2, BS1, BS3	Lik benign
Arg458Met	1	0	32	no	Increase kinase activity (Arranz et al.)?	Dang et al., 2018	NA	NA	BS3, PM2	VUS
Arg528Trp	1	4	24	no	Increase kinase activity (Arranz et al.)?	Chicago Hospital (clinvar SCV000594478_VUS)	NA	NA	BS1, BS3, BP4	Benign

Name	Identifier	Classification	Inheritance	Clinical features
c.1726C>T p.Gln576*	Okamoto et al. 2014	considered as disease causing	NA	poorly <sup>a</sup>
c.2040C>A p.Tyr680*	ClinVar (Genedx, SCV000779549.2)	Likely pathogenic	NA	poorly to intermediate <sup>b</sup>
c.2213_2218delinsAGAG p.Thr738fs	ClinVar (Genedx, SCV000570988.4)	Likely pathogenic	<i>de novo</i>	intermediate <sup>c</sup>

**Table S8**

# Discussion de l'article 1

Le syndrome DYRK1A est une des formes monogéniques de DI les plus fréquentes et plus d'une centaine de faux-sens sont à ce jour recensés comme de signification inconnue (VSI) dans la base de données ClinVar. Nous avons proposé à travers ce travail une approche permettant d'interpréter les VSI identifiés dans le gène *DYRK1A*. Pour se faire, nous avons développé un score clinique spécifique du syndrome, plusieurs tests fonctionnels *in vitro* permettant d'étudier les conséquences des variants sur les protéines mutantes au niveau moléculaire sur le taux de protéines, leur localisation, leur capacité à s'autophosphoryler et à phosphoryler la protéine TAU, et défini un profil de méthylation spécifique des patients.

## **Le score clinique DYRK1A**

L'ensemble des patients avec variants pathogènes dans *DYRK1A* présentaient un score supérieur ou égal à 13, et c'est donc le seuil qui a été défini pour considérer un patient comme présentant un phénotype compatible avec le syndrome. Parmi les 41 patients atteints d'autres syndromes de DI également évalués, aucun d'entre eux ne présente une clinique évocatrice du syndrome mais deux individus (un patient avec mutation dans le gène *ARID1B* et un dans le gène *KMT2A*) présentent des scores à 12,5 et sont donc proches du seuil évocateur. Cette observation s'explique par le fait que certains symptômes du syndrome *DYRK1A* comme la DI, les troubles du langage, la microcéphalie ou encore l'épilepsie, sont relativement communs au sein des TND et partagés par plusieurs syndromes. Par exemple, des variants entraînant une perte de fonction de la protéine DYRK1A ont été récemment identifiés chez plusieurs patients initialement orientés vers un diagnostic de syndrome de Cockayne (Boujelbene, Courraud et al., manuscrit en préparation). De ce fait, il serait intéressant de conduire de nouvelles évaluations du score clinique DYRK1A sur des patients atteints d'autres syndromes de DI afin de confirmer sa spécificité. Il est également important de noter que si les variants reclassés comme bénins après notre étude ont été pour la plupart identifiés chez des individus ayant un score clinique faible, l'un d'entre eux (p.Thr588Asn) est présent chez un individu avec un score clinique élevé (15,5). De plus, certains individus avec variants classés pathogènes ne présentent pas de clinique très « évocatrice de DYRK1A ». En effet, trois d'entre eux (p.Ser311Phe ; p.Ser324Arg et p.Arg467Gln) présentent un score intermédiaire et un présente même un score inférieur à 10 (p.Ser660fs\*43). Nos résultats soulignent donc l'utilité du score clinique mais également ses limites. Si les seuils définis peuvent être affinés en augmentant le nombre

d'individus évalués, le score clinique reste un argument parmi d'autres pour interpréter les variants et doit être combiné à d'autres approches.

### **Limite des tests fonctionnels utilisés dans cette étude**

Les tests fonctionnels ont été réalisés en surexprimant des protéines mutantes dans des cellules en culture. Bien que la surexpression d'une protéine par le biais d'une construction plasmidique ne reflète pas les niveaux physiologiques de cette dernière, l'ensemble des tests réalisés repose sur la comparaison avec le comportement de la protéine sauvage. Les analyses mises au point durant ce travail nous permettent de mettre en évidence les variations entraînant une perte de la fonction kinase de la protéine. Cependant, la mise en évidence de potentiels variants ayant un effet gain de fonction (GoF) ou affectant d'autres propriétés indépendantes de la fonction kinase comme l'interactome, reste difficile. Concernant les GoF, le test de phosphorylation réalisé ne reflète en effet que la capacité de la protéine DYRK1A mutante à phosphoryler MAPT et ne renseigne pas sur l'état de phosphorylation des nombreuses autres cibles de DYRK1A. Ainsi, il serait pertinent de réaliser un test d'activité kinase globale pour les protéines suspectées GoF, en utilisant par exemple un peptide synthétique comportant une séquence phosphorylée par DYRK1A tel que le DYRKtide (Himpel et al., 2000) ou le Woodtide (Nguyen et al., 2018). On notera que les séquences consensus ciblées par les protéines kinases restent assez variables (Hornbeck et al., 2015) et la caractérisation globale de l'état de phosphorylation des cibles de DYRK1A doit passer par la réalisation d'un phosphoprotéome.

Le caractère gain de fonction d'un variant n'est pas restreint à l'effet sur l'activité enzymatique de DYRK1A. En effet, certains variants peuvent conduire à une augmentation d'affinité entre la protéine et un ou plusieurs de ses interacteurs. Par exemple, deux variants du gène *RARB* ont été décrits comme augmentant la liaison entre l'acide rétinolique et le récepteur RARB, conduisant à une augmentation de son activité transcriptionnelle (Srouf et al., 2016). Dans cette étude, les auteurs prédisent les modifications conformationnelles entraînées par ces variants et valident leurs effets par quantification d'expression des gènes cibles de RARB. Dans notre étude, nous avons évalué les conséquences des différents variants sur la liaison de DYRK1A avec DCAF7, cette interaction étant rapportée comme stimulant la transcription de myogénèse (Yu et al., 2019), mais n'avons identifié aucun effet. Mais, de la même manière, l'étude de l'interaction de DYRK1A avec DCAF7 ne reflète pas les conséquences du variant sur le reste des nombreuses interactions de DYRK1A. Ces données ne permettent également pas de mettre en évidence de potentielles nouvelles interactions qui seraient absentes dans la condition sauvage.

## **La signature de méthylation du syndrome DYRK1A**

Plusieurs dizaines de syndromes ont été décrits comme présentant une signature de méthylation mais il n'existe à l'heure actuelle pas de relation constante entre la fonction du gène affecté et la présence d'une telle signature dans le sang (Aref-Eshghi et al., 2020). Malgré les implications de *DYRK1A* dans la régulation épigénétique (Lepagnol et al., 2009, Jang et al., 2014 ; Li et al., 2018), l'identification d'un profil de méthylation spécifique des patients n'était pas assurée. Nous avons cependant pu mettre en évidence cette signature robuste dans le sang des patients mais n'avons observé aucune corrélation entre le phénotype des patients DYRK1A, les types de variation et les profils de méthylation obtenus. De précédentes études ont pu mettre en évidence des variants gain de fonction (GoF) par étude du méthylome comme un variant GoF dans le gène *EZH2* (Choufani et al., 2020) et nous avons également observé une signature opposée à celles obtenues pour les individus porteurs de mutations perte de fonction dans *DYRK1A* pour le variant p.Gly486Asp discuté par la suite. Au vu de la sensibilité de cette signature mais également de sa facilité de mise en place par une analyse d'ADN sanguin qui peut être faite sur le même prélèvement qui a servi à l'analyse génétique, nous proposons pour les futurs VSI identifiés dans *DYRK1A* d'analyser le profil de méthylation des patients afin de reclasser ces variants. Cependant, cette découverte ne permet pas de savoir si ce profil de méthylation est régulé par un ou des cibles/interacteurs de DYRK1A ou si cette signature est une conséquence cellulaire globale de l'inactivation de la kinase. Ainsi, il est possible que cette méthode ne permette pas de mettre en évidence l'effet de variants n'impactant pas l'activité kinase ou entraînant au contraire une activité enzymatique augmentée.

## **Trois Variants de Signification Inconnue aux profils atypiques**

### *Un potentiel gain de fonction : Gly486Asp*

Bien que le patient porteur du VSI p.Gly486Asp présente un profil de méthylation opposé à celui des patients DYRK1A, nous avons décidé de ne pas reclasser ce variant et de le laisser comme de signification inconnue. En effet, ce variant fut rapporté par une autre équipe comme présentant une autophosphorylation ainsi qu'une activité kinase supérieure à celle de la protéine sauvage (Arranz et al., 2019) mais nous n'avons pas observé dans notre test d'augmentation de la phosphorylation de MAPT (TAU). De plus, l'équipe qui a rapporté initialement cette augmentation d'activité kinase l'a également rapportée pour trois autres variants dont un présent quatre fois dans la base de données gnomAD (Arg258Trp) ce qui

soulève des questions sur un réel effet du variant et donc la sensibilité des tests utilisés. Dans ce cas particulier, une analyse de l'activité kinase globale est nécessaire à la mise en évidence d'une augmentation de l'activité enzymatique. Il est également possible que ce variant modifie l'interaction de la protéine avec les partenaires de *DYRK1A*, entraînant des pertes ou gains d'interaction. Dans ce cas, il sera nécessaire de comparer l'interactome de la protéine sauvage avec celle porteuse du variant. Le rôle et l'activité de *DYRK1A* peuvent varier selon les tissus et le stade de développement. Ainsi, il est possible que ce variant n'impacte l'activité de la protéine que dans certains tissus ou dans une fenêtre temporelle particulière. Au vu de la macrocéphalie présente chez le patient et de la robustesse du phénotype de microcéphalie dans l'ensemble des modèles animaux *DYRK1A* étudiés, il serait intéressant de générer un modèle porteur de la variation et d'évaluer la taille du cerveau. Le modèle murin reste le plus complet car il permet d'étudier les conséquences du variant au niveau du développement du cerveau ainsi qu'au niveau comportemental et reste relativement proche de l'Homme. Cependant, des modèles plus faciles à mettre en œuvre tels que la drosophile ou le poisson zèbre pourraient permettre de valider rapidement l'effet de ce variant. Enfin, la découverte d'un second individu porteur de cette variation, quel que soit son phénotype, serait un argument fort pour l'interprétation de ce VSI. A ce jour, un autre patient porteur de ce variant fut identifié (Dang et al., 2018) mais malgré plusieurs relances nous ne disposons pas à l'heure actuelle d'informations cliniques ou d'accès à l'ADN du patient. Cependant, dans le cas d'un potentiel GoF, il est possible que ce variant ait pu être écarté lors d'analyses génétiques en raison de son association avec un phénotype très éloigné des patients *DYRK1A*. Enfin, dans l'hypothèse d'un GoF enzymatique, nous avons également étudié le profil de méthylation de deux patients porteurs de duplications comprenant le gène *DYRK1A* et avons pu observer qu'ils présentaient un profil semblable aux individus contrôles. Cependant, nous ne disposons que de l'ADN extrait de sang de ces patients et n'avons pu confirmer l'augmentation d'expression de *DYRK1A* au niveau ARNm ou protéique. Cependant, nos collaborateurs ont également pu comparer les profils de méthylation de patients atteints du SD par rapport aux contrôles et aux patients *DYRK1A*. Ces résultats préliminaires semblent indiquer des profils opposés entre patients SD et *DYRK1A*. De manière très intéressante, le variant Gly486Asp semble clusteriser avec les patients SD mais des analyses supplémentaires restent nécessaires pour valider ces résultats. Pour terminer, il reste possible que le patient porteur de ce variant soit atteint d'une autre forme génétique de TND possédant également une signature de méthylation propre. De ce fait, ce variant ne clusteriserait ni avec les patients *DYRK1A* ni avec les contrôles.

### *Le décalage du cadre de lecture distal : Ser660fs\*43*

La protéine s'est révélée insoluble et les analyses d'expression et d'autophosphorylation n'ont de ce fait pas pu être considérées. Etant donné l'observation des agrégats dans un contexte de surexpression, nous souhaitons investiguer les conséquences de ce variant dans un contexte physiologique en utilisant une culture de fibroblastes primaires issus du patient, qui sera générée prochainement. En effet, les protéines mutantes pourraient posséder une activité partielle en contexte physiologique, expliquant la plus faible sévérité des traits cliniques du patient. De plus, je n'ai pas détecté d'effet délétère associé aux variants Ser660\* et Ser661\*, questionnant les conséquences d'une perte de l'extrémité C-terminale de DYRK1A. Ainsi, il paraît clair que l'interprétation des variants tronquant doit être réalisée avec précaution, en particulier pour ceux dans le dernier exon et échappant au système NMD. A l'heure actuelle, trois frameshifts distaux sont recensés dans la base de données ClinVar et il serait intéressant de tester la signature épigénétique de ces individus. De ce fait, il serait également intéressant d'investiguer l'interactome des protéines mutantes porteuses de variations non-sens mais également celui de la protéine comportant le variant p.Ser660fs\*43 afin d'améliorer la caractérisation de variants identifiés dans cette région.

### *Une potentielle phénocopie : Thr588Asn*

Le variant p.Thr588Asn fut initialement identifié *de novo* par séquençage d'exome en trio et rapporté comme pathogénique en raison de la clinique évocatrice du patient (Bronicki et al., 2015). Le score clinique réalisé lors de cette étude (15,5) révèle effectivement un phénotype évocateur du syndrome DYRK1A. Cependant, les tests fonctionnels ainsi que le modèle murin Knock-in n'ont montré aucune différence significative entre la protéine mutante et la protéine sauvage. Deux autres variants sont recensés sur la même position : Thr588Ala et Thr588Pro respectivement 2 et 44 fois dans gnomAD et le profil de méthylation de ce patient s'est révélé semblable à la cohorte d'individus contrôles. C'est pour ces raisons que nous avons choisi de reclasser ce variant en classe 2 (probablement bénin) et le cas de ce patient est considéré comme une phénocopie. Une analyse plus détaillée de la clinique de la patiente nous a permis de nous rendre compte qu'elle présentait également des caractéristiques non classiques du syndrome DYRK1A, comme une obésité (alors que les patients avec syndrome DYRK1A sont plutôt maigres). La potentielle cause génétique de DI reste de ce fait inconnue et aucun autre variant candidat ne fut retenu lors de l'analyse de l'exome en trio. C'est pourquoi un séquençage de génome en trio sera prochainement réalisé. La mise en évidence d'une variation entraînant une perte de fonction de *DYRK1A* non détectée lors de l'analyse d'exome (variant dans une région



régulatrice ou intronique profond) est aujourd'hui considérée comme improbable au vu du profil de méthylation.

## **Conclusion**

Nous avons lors de ce travail développé une combinaison d'outils permettant d'interpréter de manière efficace les VSI identifiés dans le gène *DYRK1A*. Les résultats obtenus indiquent que l'interprétation des variants doit être réalisée avec prudence, en particulier pour les décalages de cadre de lecture distaux, et en considérant l'ensemble des éléments cliniques et moléculaires disponibles. Nous avons pu développer une signature de méthylation spécifique de patients *DYRK1A* mais également la valider au niveau moléculaire. Ces tests fonctionnels nous ont permis de reclasser 12 VSI sur 13 initialement testés et de prouver le caractère bénin d'un variant précédemment classé comme probablement pathogène. En ce qui concerne le VSI restant, des études supplémentaires restent nécessaires afin de confirmer le potentiel caractère GoF du variant. En ce qui concerne les futurs VSI identifiés dans *DYRK1A*, nous recommandons de procéder à une analyse du profil de méthylation si le sang du patient est accessible. Dans le cas contraire, nous conseillons de tester l'effet du variant sur la capacité d'autophosphorylation de la protéine, qui reste le test le plus sensible.

**Article 2 : Role of *DYRK1A* in neural  
progenitor cells**

---



## Résumé

Le syndrome DYRK1A est une des formes de DI les plus fréquente et représente près de 0,5% des patients DI. *DYRK1A* code pour une kinase à double spécificité Tyrosine et Serrine/Thréonine au implications multiples et dont les rôles durant le neurodéveloppement restent à l'heure actuelle, mal compris. L'objectif de ce travail était de caractériser les effets d'une diminution d'expression de *DYRK1A* dans un modèle de précurseurs neuronaux humain. Nous avons dans un premier temps caractérisé l'interactome de DYRK1A dans le modèle hNSCs par spectrométrie de masse. Ainsi, nous avons pu mettre en évidence 38 protéines interagissant potentiellement avec DYRK1A dans ces cellules. Parmi ces interacteurs, nous avons retrouvé plusieurs éléments composant le complexe APC, une machinerie à activité ubiquitine ligase régulant notamment le cycle cellulaire. Dans un second temps, nous avons étudié les conséquences d'une inactivation de DYRK1A sur la régulation de l'expression des gènes par analyse transcriptomique. Nous avons souhaité par la suite valider ces variations dans un second modèle de hNSCs et rapporté 18 gènes dont l'expression était significativement modulée suite à l'inactivation de *DYRK1A* dans les deux lignées de hNSCs. En raison de leur rôle dans le neurodéveloppement, nous avons choisi de tester les niveaux protéiques de deux gènes dont l'expression était régulée par *DYRK1A* : *NMNAT2* et *PTBP2*. Etonnamment, nous n'avons pas observé de variation du niveau protéique de NMNAT2 malgré une diminution significative de son niveau d'ARNm. De manière très intéressante, nous avons pu mettre en évidence une diminution d'expression protéique de PTBP2 alors que son ARNm était retrouvé comme sur exprimé lors de l'inactivation de *DYRK1A*. Par la suite, nous avons associé l'inactivation partielle de *DYRK1A* à a surexpression d'une isoforme du gène *CDKN1A* (P21). Le gène *CDKN1A* code pour une protéine aux rôles variés mais notamment très impliquée dans la régulation du cycle cellulaire et nous avons pu souligner une diminution d'expression de la protéine CDKN1A après inactivation de *DYRK1A*. CDKN1A est un acteur clef du processus de sénescence mais nous n'avons observé aucune différence du nombre de cellules entrant en sénescence entre les conditions contrôles et les hNSCs inactivées pour *DYRK1A*. Parallèlement, au vu de l'implication de DYRK1A et de ses interacteurs dans la régulation du cycle cellulaire, nous avons investigué les conséquences de son inactivation sur la prolifération des lignées hNSCs. De ce fait, nous rapportons associée à l'inactivation de *DYRK1A*, une diminution de la capacité de prolifération des cellules ainsi qu'une entrée significativement plus élevée de ces hNSCs dans le processus d'apoptose.

Pour conclure, ce travail nous a permis de mieux comprendre le rôle de *DYRK1A* dans un modèle de hNSCs, pertinent dans l'étude des troubles du neurodéveloppement. La caractérisation des partenaires mais également des gènes dont l'expression est régulée de manière directe ou indirecte par *DYRK1A* pourra permettre de mieux comprendre son rôle dans le développement du cerveau et les mécanismes physiopathologiques associés au syndrome.

**Ce travail est présenté sous forme d'article mais ne correspond pas à la version définitive. En effet, plusieurs résultats restent à valider et différentes pistes sont actuellement investiguées.**

## **DYRK1A roles in human neural progenitors**

Jeremie Courraud<sup>1,2,3,4</sup>, Angélique Quartier<sup>1,2,3,4</sup>, Nathalie Drouot<sup>1,2,3,4</sup>, Irene Zapata-Bodalo<sup>1,2,3,4</sup>, Johan Gilet<sup>1,2,3,4</sup>, Bill Keyes<sup>1,2,3,4</sup>, Alexandra Benchoua<sup>5</sup>, Jean-Louis Mandel<sup>1,2,3,4</sup>, Amélie Piton<sup>1,2,3,4</sup>

<sup>1</sup> Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France

<sup>2</sup> Centre National de la Recherche Scientifique, UMR7104, Illkirch, France

<sup>3</sup> Institut National de la Santé et de la Recherche Médicale, U964, Illkirch, France

<sup>4</sup> Université de Strasbourg, 67 400 Illkirch, France

<sup>5</sup> I-Stem, Evry, France

<sup>6</sup> Genetic diagnosis laboratory, Strasbourg University Hospital, 67 090, Strasbourg, France

<sup>7</sup> Institut Universitaire de France

### **Abstract**

Mutations in *DYRK1A* lead to one of the most frequent monogenic form of intellectual disability (ID). *DYRK1A* code for a dual specificity tyrosine and serine/threonine kinase which has been involved in various cellular process and is a key regulator of nervous system development. For the first time, we characterized *DYRK1A*'s interactome in a model of human neural stem cells (hNSC) and identified 38 potential interactors for which 14 have never been described. Among them, five are part of the TPR lobe of the Anaphase Promoting Complex (APC) and one is another ubiquitine ligase, RNF114 (alias ZNF313) known to target p21/CDKN1A. We also performed transcriptomic studies in order to describe genes and pathways regulated by *DYRK1A* in hNSC. Most of the genes found to be significantly deregulated after *DYRK1A* KD are down-regulated, with an enrichment in genes encoding components of extra cellular matrix (ECM)(ex: *COL6A3*, *THBS2*, *LUM*, etc). Other genes down-regulated included proteins involved in neural development and in regulation of cell cycle (ex: *CCND2*). Moreover, we identified an increased in the expression of a non-canonical isoform of *CDKN1A*, a multifaceted gene involved among other in cell cycle and senescence regulation. However, we showed that *DYRK1A* KD leads to a decrease of CDKN1A/p21 protein level. *DYRK1A* KD does not induce senescence but decreases cell proliferation and increases apoptosis. In parallel, we found that *DYRK1A* KD in hNSC leads to a decrease of the growth-associated cascade ERK1/2, and we could speculate that it participates to these observed phenomenons. In conclusion, our study provides information about the roles of *DYRK1A* in human neural stem cells, a relevant model to study neurodevelopmental disorders such as *DYRK1A* syndrome.

**Keywords :** *DYRK1A*, hNSC, interactome, gene expression, CDKN1A/P21, cell proliferation

## Introduction

Intellectual disability (ID) and autism spectrum disorder (ASD) are two groups of neurodevelopmental disorders (NDD) presenting an important genetic contribution and overlapping strongly both at the clinical and genetic levels. Single genetic events account for an important part of ID cases and a non-negligible part of ASD cases. Around a thousand of genes have been involved in these monogenic forms of ID or ASD. For most of them, little is known about the pathophysiological mechanisms which can lead to ID and ASD. Some are more frequently mutated than others, such as *DYRK1A* (*dual specificity tyrosine phosphorylation regulated kinase 1A*) (Gonzalez-Mantilla et al. 2016). The first disruptions of *DYRK1A* were identified in individuals with intrauterine growth retardation (IUGR), primary microcephaly and epilepsy (Møller et al. 2008). The clinical spectrum associated to heterozygote mutations in *DYRK1A* was further refined with publication of additional patients and comprises: ID, feeding difficulties, poor or absent language, microcephaly, autistic traits, epilepsy, and a specific facial dysmorphism (Bronicki et al. 2015; van Bon et al. 2016) (Courraud et al. 2021). Mutations were also identified in cohorts of ASD individuals (O’Roak et al. 2012; Iossifov et al. 2014), but all have ID. Interestingly, increased dosage of *DYRK1A*, located on chromosome 21, is thought to participate in the cognitive manifestations of Down syndrome (Altafaj et al. 2001), suggesting that a correct balance of *DYRK1A* dosage is essential for brain development and cognitive function.

The *DYRK1A* gene codes a dual tyrosine-serine / threonine (Tyr-Ser/Thr) kinase protein which belongs to the family of DYRKs. Its main isoform comprises 753 amino acids (Becker 2011). *DYRK1A* has the general structure of kinases, with a DH (DYRK Homology-box) domain, two nuclear localization signal sequence (NLS), a central catalytic domain (including Tyrosine 321, involved in the activation of *DYRK1A* by autophosphorylation), a leucine zipper (bZIP) domain whose presence suggests that *DYRK1A* is a protein that forms dimers or multimers with other nuclear proteins such as transcription factors, some Ser/Thr repeats, involved in the interaction with its target proteins, a poly-histidine sequence for its addressing to the nuclear speckles compartments and a PEST domain necessary for its degradation and the fine regulation of its cell concentration. *DYRK1A* is ubiquitously expressed during embryonic development and in adult. Its location is both cytoplasmic and nuclear and depends on cell type and stage of development (Hämmerle et al. 2008). By the number and the diversity of its protein targets, *DYRK1A* regulate numerous cellular functions (review (Tejedor and Hämmerle 2011; Duchon and Herault 2016)). *DYRK1A* regulates cytoskeleton-associated proteins such as TAU,

MAP1B,  $\beta$ -tubulin and others, and have been shown to play a role in regulation of dendritic morphogenesis in rodents and drosophila (Ryoo et al. 2007; Ori-McKenney et al. 2016). DYRK1A modulates synaptic plasticity by regulating NMDA receptors expression at the membrane surface. DYRK1A interacts with proteins involved in endocytosis such as Dynamin1a or Amphiphysin1 and interacts with the light chain of Clathrin and Endophilin1. DYRK1A phosphorylates several proteins involved in cell cycle regulation, such as cyclin D1 or p27<sup>Kip1</sup>, and therefore is involved in regulation of neural progenitor proliferation (Hämmerle et al. 2008). In addition to its cytoplasmic targets, DYRK1A also regulates numerous nuclear proteins: it interacts with and phosphorylates transcription factors and proteins from Histone H3 (Jang et al. 2014), acetyl transferases CBP and p300 (Li et al. 2018), binds to the chromatin remodeling SWI/SNIF complex (Lepagnol-Bestel et al. 2009) as well as RNA polymerase type II (Di Vona et al. 2015), regulate in one hand positively CREB1 (Yang 2001) and GLI1 (Ehe 2017) activity, NFAT nuclear translocation (liu 2017) and, in other hand, negatively REST complex stability (Lu 2011). We recently demonstrated that loss-of-function variants in *DYRK1A* induce specific changes in DNA methylation in blood, confirming the nuclear role of *DYRK1A* in chromatin remodeling (Courraud et al. 2021). Other nuclear substrates of DYRK1A have reported, including splicing factors (Shi et al. 2008).

Although numerous studies have unraveled some of the functions of DYRK1A, most of the works about its role in brain have been performed in mouse (Fotaki et al. 2002; Fotaki et al. 2004), drosophila (11,11,13) or more recently, in xenopus models (Blackburn et al. 2019; Willsey et al. 2020). Therefore, little is known about its role in human neural precursors, and about what are, among its cellular functions, those that are critical for human brain development and functioning. In order to characterize the role of DYRK1A in human neuronal progenitors we performed proteomic and transcriptomic studies to identify its interactome as well as the genes deregulated after its Knock-Down (KD). We also analyzed the cellular consequences of DYRK1A *KD* and identify that it leads to a decrease of cell proliferation coupled with an increase of apoptosis.

## **Material and methods**

### **Western Blot**

Cells were lysed in RIPA buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.25% sodium deoxycholate, 1% NP-40) supplemented with protease inhibitor cocktail and phosphatase



inhibitor cocktail. 5 to 50µg of protein lysate were separated on 10% SDS-PAGE and transferred to polyvinylidene fluoride membrane. Membranes were blocked in 5% nonfat dry milk diluted in tris buffered saline with tween 20 (50mM Tris, 150mM NaCl, 0.05% Tween 20) and probed using the antibodies overnight at 4°C. GAPDH was used as loading control. Incubation with appropriate secondary HRP-labelled antibody (less than 1h) was followed by detection with Immobilon western chemiluminescent HRP substrate (Merck Millipore, Darmstadt, Germany). List of antibodies used is available in table S1.

### **Cell culture, transfection and proliferation assay**

Cell Culture and Transfection Human neuronal stem cells (hNSCs) were derived from human embryonic stem cell line (SA001) and from reprogrammed fibroblasts (GM01869). They were obtained from I-Stem (Cellartis, work supervised by the French Bioethics Agency, and Coriell Institute for Medical Research) as described previously (Boissart et al. 2012). hNSCs were seeded on poly-ornithine- and laminin-coated dishes and maintained in N2B27 medium (DMEM/F12 and Neurobasal medium [1:1] supplemented with N2, B27, 2-mercaptoethanol [all from Invitrogen]), BDNF (20 ng/mL), FGF-2 (10 ng/mL) (both from PeproTech), and EGF (R&D Systems; 10 ng/mL). Culture and quality controls of hNSCs were performed as described in Quartier et al. (Quartier et al. 2018). hNSCs were transfected using INTERFERin reverse transfection protocol (Polyplus-transfection) with scramble siRNA, DYRK1A siRNA (pool of siRNA, at 120 nM final concentration), or transfecting agent only. Cells were stopped 48 h after transfection for RNA and protein extractions. Controls and patient's derived fibroblasts were cultured in DMEM (Gibco) supplemented with glucose 1g/L, 10% FCS and gentamycin. All cell lines were grown at 37°C in 5% CO<sub>2</sub>. For proliferation assay, reverse transfection of SA001 and GM01869 hNSCs (20,000 cells/cm<sup>2</sup>) was performed in 96-well plates using InterferIN reagent and 20nM siRNA according to manufacturer's recommendation. Each day a plate was fixed and stained with DAPI and number of nucleus in each condition counted using CellInsight automated microscope and HCS Studio software (Thermo Fisher Scientific).

### **Immunoprecipitation coupled to mass spectrometry (IP-MS)**

Proteins were extracted from untreated hNSC-1 and hNSC-2 cells and immunoprecipitated with respectively DYRK1A antibodies directed against N-terminal and C-terminal of the protein (Cohesion Biosciences #CPA1357 (Immunogen sequence not provided) and Abnova H00001859-M01 (aa 674-763)). As negative controls we used antibody free beads and MAR

and RAM antibodies (Mouse Against Rabbit (211-002-171, Jackson ImmunoResearch) and Rabbit against Mouse (315-005-044, Jackson ImmunoResearch) as described (Mattioli et al. 2019). Immunoprecipitation were validated by WB as previously done (Mattioli et al. 2019) before the mass spectrometry analyses (Proteomic platform, IGBMC). Briefly, samples were treated with LysC/Trypsine for liquid digestion and injected in Orbitrap ELITE / C18 Accucore 50cm (20 $\mu$ L 0.1%TFA / 1 $\mu$ L) for 2h runs in triplicate. Data were processed with Proteome Discoverer 2.2 software using Homosapiens\_190716\_reviewed. fasta and contaminants\_190528.fasta databases. Threshold were put at 1% FDR with a minimum of 2 peptides per protein. To consider a protein as a candidate interactor, we applied the thresholds to keep only protein with 1) Peptide-Spectrum Matching (PSM) sum of control conditions (Beads, MAR, GAR) equal to 0 and 2) a sum of PSM for the three replicates of the test conditions (DYRK1A C-term and DYRK1A N-term) greater or equal to 5. Enrichment analysis were realized using DAVID (Database for Annotation, Visualization and Integrated Discovery) or the GO (Gene Ontology Resource).

### **RNA sequencing and RT-qPCR**

RNA were extracted from two series (biological replicates) of SA001-derived hNSCs treated by siRNA against DYRK1A, scramble siRNA, or transfecting agent only, as described in Mattioli et al 2019. RNA isolation, libraries preparation, sequencing and statistical analysis were realized as described in Mattioli et al. 2018 (Mattioli et al. 2020)(Gene Expression Omnibus accession number: GSEXXXX). All 48 DEGs were tested in a third serie of siRNA treated SA001 hNSCs but also in three series (biological replicates) of siRNA treated GMO1869 hNSCs after RNA isolation, Reverse transcription reaction (Invitrogen SuperScript IV) on 200-500ng RNA and qPCR performed on cDNA on LightCycler 480 II (Roche) using the QuantiTect SYBR Green PCR Master Mix (Qiagen) or RT-qPCR multiplex using 48\*48 array Fluidigm Biomark technology (GenomEast platform).

### **SA- $\beta$ galactosidase assay and cells irradiation**

Cells were fixed 4 days after siRNA transfections with 0.5% glutaraldehyde solution for 15mn at room temperature and washed twice with PBS/MgCl<sub>2</sub> pH 5.95. Fixed cells were incubated with freshly prepared X-gal staining solution (1 mg/ml X-gal), 5 mM Potassium ferrocyanide, and 5 mM Potassium ferricyanide in PBS/MgCl<sub>2</sub> pH 6 at 37°C for 20 to 48 hours. Cells were

washed 3 times with H<sub>2</sub>O and pictured in bright field (20X magnification). Counting was done manually using Cell Counter plugin in FIJI software and 1400 to 3200 cells were counted for each condition. As a positive control of SA- $\beta$  galactosidase labelling, a batch of untreated cells was also irradiated at 8Gy using “CellRad System Precision” fixed and labeled as previously described.

## RESULTS

### Characterization of DYRK1A interactome in human NSCs

Currently, 409 proteins are described as interacting with DYRK1A and 44 of them have been found in at least 2 different studies (Skurat and Dietrich 2004; Varjosalo et al. 2013; Huttlin et al. 2015; Huttlin et al. 2017; Menon et al. 2019; Roewenstrunk et al. 2019; Guard et al. 2019). To identify protein partners of DYRK1A in human neural stem cells, proteins from hNSC-1 were immunoprecipitated with DYRK1A antibodies recognizing either the C-terminal (Cter) or the N-terminal (Nter) parts of the protein. Mass spectrometry analysis of immunoprecipitated proteins revealed, in addition to DYRK1A, 38 potential interactors with 14 found in both conditions and respectively 14 and 11 found specifically in Cter and Nter conditions (Figure 1A). Among them, 14 were previously described as DYRK1A partners in literature (BioGRID), thereby confirming the robustness of our findings and showing that these interactions are conserved in neural cells (**Table 1**). We report also 23 potential additional DYRK1A interactors. Among the partners identified, an enrichment in proteins involved in Anaphase Promoting Complex (APC) (APC, ubiquitin ligase activity, APC catabolic process) were found (**Table S2**): 5 members of the APC, two of them (CDC16 = ANAPC6 and ANAPC7) have never been described as DYRK1A interactors and three of them (CDC27=ANAPC3, ANAPC4, CDC23=ANAPC8) were already identified in a previous study (Guard et al. 2019). Interestingly, four of these core subunits (ANAPC3-6-7 and 8) compose the tetratricopeptide repeat (TPR) lobe APC subcomplex, playing a role in the global scaffolding of the complex and substrate recognition (Sivakumar and Gorbsky 2015). This would suggest that DYRK1A may interact with APC through TPR lobe and given the kinase activity of DYRK1A, could possibly phosphorylate the subunit and therefore play a role in the global activity of the complex. We also found that DYRK1A interact with RNF114, alias ZNF313, a zinc-finger E3 ligase involved in cell cycle progression and senescence repression that has never been linked to human pathologies but cancer. ZNF313 expression is increased in G1 to S transition and it has been shown to ubiquitinate p21/CDKN1A causing its degradation. Phosphorylation analysis on IP-

MS data identified phosphorylation events on DYRK1A (Tyr321) and in three DYRK1A partners with undescribed roles: FAM117B (Ser106), GLCC11 (Ser30 and Ser303) and FAM53C (Ser324) but they do not fall in consensus DYRK1A target sequence (R-P-x(1,3)S/T-P).

### **DYRK1A KD in hNSC induces changes in gene expression but not on splicing events**

DYRK1A has been previously described as a both direct and indirect transcriptional regulator. Indeed, DYRK1A has been shown to modulate remodeling complex and of epigenetic marks (Lepagnol Bestel et al. 2009, Li et al. 2018) but also interact with various TFs such as NFAT (Liu 2017) and Gli1 (Ehe 2017) and with RNA Pol II (Yu et al. 2019). In order to identify the effect of DYRK1A loss on the regulation of gene expression and alternative splicing, we performed sequencing of mRNA extracted from hNSC-1 line after *DYRK1A* KD (48h treatment with *DYRK1A* siRNA which led to a reduction of 50% in average of its mRNA, **Figure S1**). Transcriptomic analysis revealed 48 significantly deregulated genes (DEG) after *DYRK1A* KD compared to a treatment with lipofectant only (INTERFERin), most of them being downregulated (n=45), with the second most significant being *DYRK1A* (log2 fold change = -0.66, adjusted p-value = 2.26E-05) (**Figure 2A, Table S3**). In comparison, no gene was found significantly DE in hNSCs treated with nonspecific siRNA (scramble). Enrichment analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) demonstrated that the set of genes was enriched for only one Gene Ontology (GO) term, related to extra cellular matrix (ECM) components (fold enrichment: 18, adjusted p-value = 3.9E-4): COL6A1, COL6A3, LUM, THBS2, etc. Other GO terms related to peripheral nervous system development, positive regulation of gene expression, axon guidance and neuronal cell body are enriched but not significantly after Benjamini correction. We then checked by multiplex RT-qPCR the expression of all 48 DEG in a new serie of hNSC-1 and three series of hNSC-2 treated by *DYRK1A* siRNA and were able to confirm a significant change in genes expression in both cell lines for 18 of them (**Figure 2B, Table 2**). We searched for motif enrichment in the promoter region (-1000/+100 bp) of the DEG using AME tool (Analysis of Motif Enrichment, version 5.1.1) but did not find any significant enrichment. Because of their role in neural development, we tested by Western Blot the effect of DYRK1A KD on protein levels of NMNAT2 and PTBP2 but we could not observe any change in NMNAT2 protein level even after 72h and 96h. ore surprisingly, the analysis of PTBP2 revealed a significant decrease of the protein level after DYRK1A KD in both hNSC lines in contradiction with the decrease observed

at the mRNA level (**Figure S2**). In parallel we also performed sequencing of mRNA extracted from primary cultures of fibroblasts obtained from patients carrying pathogenic variants in *DYRK1A* (n=7) compared to controls individuals (n=6) and found changes in expression of 88 protein-coding genes (**Table S4, Figure S3**). As for hNSC, most of them (70/88) were down-regulated. No significant enrichment of GO terms could be identified using DAVID.

### **DYRK1A inactivation leads to particular CDKN1A isoform expression**

Splicing analysis using LeafCutter failed to identify change in alternative splicing but revealed changes in the expression of some exons of the *CDKN1A* gene (**Table S5**). The *CDKN1A* isoform whose expression was the most significantly increased after *DYRK1A* KD in hNSC was NM\_001291549.1 (*CDKN1A\_1291549*) (**Figure 3A**). We confirmed by qPCR that *DYRK1A* KD increases the mRNA level of *CDKN1A\_1291549* isoform (**Figure 3B**) with no effect on the shorter isoform NM\_00389.5 (*CDKN1A\_389*). *CDKN1A* encodes the p21 protein, involved in cell cycle progression, proliferation, survival and senescence (Dutto et al. 2015), but it is not currently known if the different isoforms exert different functions. Interestingly, at the protein level, we showed that *DYRK1A* KD leads to a significant decrease of p21/*CDKN1A* protein (**Figure S4**) suggesting that this increase in mRNA level could be a compensatory mechanism of this decrease in protein level.

### **DYRK1A inactivation reduces hNSCs proliferation**

Considering the previously described role of *DYRK1A* in regulation of cell proliferation, the fact that it interacts with APC members in hNSC and that its KD leads to changes in expression of genes associated to cell-cycle regulation (*CCND2*, *CDKN1A*, etc), we decided to investigate the effects of *DYRK1A* KD on hNSC proliferation. We observed a significant decrease in cell proliferation 72 hours after *DYRK1A* KD in hNSC-1 when compared to transfection agent only (INTERFERin) or Scramble siRNA (**Figure 4A**). Similar results were obtained for hNSC-2 at 96h (**Figure 4A**). Because *DYRK1A* has also been linked to apoptosis process for example by phosphorylating Caspase-9 (Seifert et al. 2008; Laguna et al. 2013) or SIRT1 factor (Guo et al. 2010), we hypothesized that the decreased of cell number after inactivation could be partially explained by an increase in apoptosis. We analyzed by Western Blot the amount of cleaved Caspase-3 (cCASP3) protein in the supernatant of hNSCs, 48 hours after treatment by *DYRK1A* siRNA and noticed an increase of cCASP3 compared to cells treated with lipofectant alone

(INF) or scramble siRNA (**Figure 4B**). Therefore, a significant portion of the cells entry into apoptosis process following *DYRK1A* KD which can participate to the proliferation decrease. Because of the implication of p21/CDKN1A in cellular senescence, we also wanted to test if *DYRK1A* KD could induce senescence in hNSCs. We treated hNSCs with transfection agent only (INTERFERin), *Scramble* or *DYRK1A* siRNA and evaluate the number of cells entering into senescence using SA- $\beta$  galactosidase assay 4 days after transfection. Counting revealed that few cells turned positive after SA- $\beta$  galactosidase labelling and that neither *Scramble* or *DYRK1A* siRNA treatments seem to affect the number of senescent cells (**Figure 4C, Figure S5, Table S6**).

### **DYRK1A KD in hNSC leads to a decrease of ERK activation**

Recently, a study conducted in *DYRK1A* conditional heterozygous mice model identified several alterations in growth signaling cascades (Levy et al. 2021). Thus, we decided to evaluate by Western Blot the phosphorylation levels of major regulator proteins such as mTOR, AKT, ERK1/2 and PKC (**Figure S4**) in both hNSC lines. No significant variation in phosphorylation levels of mTOR, PKC and AKT proteins was observed, however a significant decrease in ERK phosphorylation was noticed 48h after treatment with si*DYRK1A* (**Figure S4**). Interestingly, such a decrease in level of phosphorylated ERK1/2 was described by Levy et al in homozygous P0 cKO (but not in heterozygous P0, P7 or adult mice).

## **DISCUSSION**

In this study we identified genes and exon differentially expressed in hNSC transiently KD for *DYRK1A* as well as in primary fibroblasts from patients carrying *DYRK1A* LoF variants. No overlap was identified between the genes identified in these cell types. This can be explained by tissue-specific effects but also by the fact the difference between cells with an innate dysfunction of *DYRK1A* (patients' cells) which can lead to feedback/compensation mechanisms and hNSC which are facing a transient and sudden knock-down of *DYRK1A* expression. We compared our results with previous transcriptomic studies conducted in murine B and T cells (Thompson et al. 2015), zebrafish brain (Cho et al. 2019), xenopus brain (Willsey et al. 2020) and several mis-regulated genes identified following ChipSeq analysis in HeLa cells (Di Vona et al. 2015). Surprisingly, even considering the high conservation of *DYRK1A*

across evolution, we found no overlap of DEG genes with these studies. However, we were able to confirm the deregulation of 18 genes in a second line of hNSC with another genetic background, indicating the robustness of the results obtained in the hNSC model.

In hNSC, *DYRK1A* KD leads to a deregulation in expression of gene encoding proteins linked to the extra cellular matrix (ECM)(ex: *COL6A3*, *LUM*, *THBS2*, etc) and proteins involved in cell cycle regulation (ex: *CCND2*, *CDKN1A*), in neural development such as axon and synapse formation (ex: *NMNAT2*, *SEMA3C*, *PTBP2*). Among these DEG, an enrichment in genes known or suspected to be involved in NDD (n=17, SySID database p-value=4.1E-4) have been observed. Among the HPO terms related to *DYRK1A* syndrome, half of them (35/68) are also described associated to mutations of some of the DEG (**Table S7**). Besides ID, HPO terms found the most frequently associated to *DYRK1A* KD DEG are « Global developmental delay », « Seizure », « Absent speech » and « Scoliosis » but also « Protruding ear », « Strabismus », « Deeply set eye », « Motor delay » and « Hypoplasia of the corpus callosum ».

Interestingly, despite a strong upregulation of *PTBP2* mRNA following *DYRK1A* KD, we showed that the expression of the PTBP2 protein was on the contrary significantly decreased. Similar results were obtained for *CDKN1A*, as the mRNA level of a long isoform is increased after *DYRK1A* KD while the amount of P21/*CDKN1A* protein is strongly decreased. We could therefore hypothesize that *DYRK1A* KD might lead to a decrease/degradation of the proteins which will activate compensatory mechanisms leading to an activation of the transcription of the corresponding mRNA.

PTBP proteins (PTBP1 and PTBP2) are RBPs controlling alternative splicing, polyadenylation, mRNA stability and translation. They are mostly expressed in neurons and hematopoietic cells and PTBP1 and 2 are known to regulate each other transcripts stabilities (Hu et al. 2018). Both proteins have redundant role and therefore can compensate in case of decreased protein levels. PTBP2 regulate neural development by regulation of alternative splicing in brain tissues (Hu et al. 2018), contributing to fate determinism and synaptogenesis. PTBP2 coordinates axogenesis in mice (Zhang et al. 2019) and is expressed in differentiating neurons (Mireya Reynage 2018). PTBP2 has also been shown to modulate mutually exclusion ratio of exons 9a and 9b of *Dnm1* mRNA, a protein associated with early infantile encephalopathy also described as a phosphorylated target of *DYRK1A* (Huang et al. 2004). Regarding these results, it would be interesting to evaluate the consequences of PTBP2 inactivation on proliferation and differentiation of hNSC.

P21/CDKN1A has been first described as a mitogen factor and inhibiting G1 to S transition and is a major marker of senescence entry (Stein et al. 1999). However, we did not observe any effect of DYRK1A KD on senescence, but time induction of senescence might be variable depending of stimulus and cell type and can occurs for example after a dozen of days in murine NSCs (Dong et al. 2014). Moreover, P21 has been described to play a role in regulation of cell cycle, of apoptosis (Gartel and Tyner 2002), DNA repair (Mauro et al. 2012) or cellular differentiation (Ghanem and Steinman 2006).

In this study we highlighted a decreased of proliferation associated to DYRK1A KD in two hNSC lines. Most of the studies previously published report a decreased of proliferation associated to DYRK1A KD, in drosophila (Tejedor and Hämmerle 2011) or in mice model (Guedj et al. 2012; Laguna et al. 2013; Barallobre et al. 2014). A role of DYRK1A in neuronal precursor proliferation has been described since its discovery in *Drosophila mnb* model (K.F Fishbach et M. Heisenber 1984) where the authors noticed reduced number of cells and brain size in the mutant flies. Up today, DYRK1A has been described as a positive and negative cellular proliferation regulator in both context of over or under expression and in various models. Decreased of DYRK1A expression level leads generally to decrease cell proliferation in various models. For example, heterozygous KO *Dyrk1a +/-* drosophila, mice, zebrafish or xenopus showed reduced brain size (Tejedor et al. 1995; Fotaki et al. 2002; Kim et al. 2017; Willsey et al. 2020). Effect of DYRK1A inhibition on proliferation human neural stem cells has not been investigated before our study, however, Bellmaine et al. underlined that DYRK1A inhibition blocks neural specification of human ESC (Bellmaine et al. 2017). We also report a significant increase of cleaved CASP3 protein suggesting that the proportion of cells entering into apoptosis is increased when DYRK1A is inactivated. This result is consistent with the previously described role of DYRK1A on inhibiting of apoptosis process by phosphorylation of Caspase-9 in U2OS and mouse retina cells (Laguna et al. 2008; Seifert et al. 2008). The decrease of p21/CDKN1A might contribute to the increase of apoptosis observed after *DYRK1A* KD.

Finally, we found a significant decrease of ERK1/2 phosphorylation in hNSC after *DYRK1A* KD. A decreased phosphorylation of ERK1/2 was previously reported in cortex of homozygous cKO mice at P0 (Levy et al. 2021). ERK1/2 is an actor or the MAPK Ras-Raf pathway and one of its roles is to control cell cycle progression through activation of the CDK4-6 complex.



Additional experiments are needed to understand how the decrease of ERK activation is linked to the changes we observed at the molecular and cellular level in hNSC after *DYRK1A* KD.

In parallel to study the consequences of *DYRK1A* KD in human neural stem cells, we also characterized *DYRK1A* interactome in these cells and identified 38 potential partners of *DYRK1A*, with an important overlap with those previously identified in other cell types (24/34). Among them, five (ANAPC3-4-6-7 and 8) are components of the Anaphase Promoting Complex (APC) and four of them compose the TPR lobe subunit (Sivakumar and Gorbsky 2015). APC is a large E3 ubiquitin ligase machinery involved in various process including of course cell cycle regulation but also DNA repair (Pal & Summers), cellular growth (Skaar and Pagano 2009), or neurons death (Maestre et al. 2008). Recently, another work also highlighted *DYRK1A* interaction with three APC factors (CDC27=ANAPC3, ANAPC4, CDC23=ANAPC8) (Guard et al. 2019) that we validated in this study. Despite no phosphorylation event was detected on these proteins on our IP-MS data, *DYRK1A* might regulate APC activity by phosphorylating APC subunits. We cannot exclude however another regulation independent of its kinase activity, as APC subunits are known to be modulated by proteins interacting with the TPR lobe interactor, such as CDH1 and CDC20 (Vodermaier et al. 2003), it might be possible that *DYRK1A* plays a role in APC regulation by modulating these interactions. However, preliminary Western Blot experiment to assess expression level of two APC<sup>CHD1</sup> targets, PLK1 (Skaar and Pagano 2009) and FMRP (Huang et al. 2015), failed to show any significant difference in protein levels caused by *DYRK1A* KD. Among the 14 novel *DYRK1A* interactors we identified, one is also a ubiquitin E3 ligase, RNF114/ZNF313. This protein plays a role in regulation of cell cycle and repression of senescence and has been shown to ubiquitinate p21/CDKN1A leading to its degradation (Han et al. 2013). Whether this interaction leads to the phosphorylation of RNF114/ZNF313 and participates to the decrease of P21/CDKN1A protein observed after *DYRK1A* KD remain to be investigated.

In conclusion, we identified in this study, as consequences of *DYRK1A* knock-down in neural stem cells, a change in expression of genes encoding components of the extracellular matrix, a decrease of p21/CDKN1A protein level, a decrease of ERK activation, a decrease of cell proliferation and an increase of apoptosis. The link between these different observations and the known and novel *DYRK1A* interactors we identified, and especially members of APC complex and RNF114/ZNF313, should be investigated further to better understand the physiopathology of the *DYRK1A* syndrome.

## **DATA AVAILABILITY**

Data have been submitted to Gene Expression Omnibus.

## **ACKNOWLEDGEMENTS**

The authors would like to thank the families for their participation and support. The authors also thank the Agence de Biomédecine, Fondation APLM and Fondation Jérôme Lejeune for financial support. We also thank people from GenomEast and from the Mass spectrometry and High-throughput screening platforms.

## **CONFLICTS OF INTEREST**

None

## **ETHICS DECLARATION**

Skin biopsies were collected as part of the Strasbourg Hospital Diagnostic Laboratory collection for research use approved by the local Ethic committee (Comité de protection des personnes, CPP IVB est). All patients or their legal representative enrolled in this study have signed informed consent for research use.

## **REFERENCES**

- Altafaj X, Dierssen M, Baamonde C, Martí E, Visa J, Guimerà J, Oset M, González JR, Flórez J, Fillat C, Estivill X (2001) Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (minibrain), a murine model of Down's syndrome. *Hum Mol Genet* 10:1915–1923
- Barallobre MJ, Perier C, Bové J, Laguna A, Delabar JM, Vila M, Arbonés ML (2014) DYRK1A promotes dopaminergic neuron survival in the developing brain and in a mouse model of Parkinson's disease. *Cell Death Dis* 5:e1289. <https://doi.org/10.1038/cddis.2014.253>
- Becker W (2011) Recent insights into the function of DYRK1A. *FEBS J* 278:222. <https://doi.org/10.1111/j.1742-4658.2010.07953.x>
- Bellmaine SF, Ovchinnikov DA, Manallack DT, Cuddy CE, Elefanty AG, Stanley EG, Wolvetang EJ, Williams SJ, Pera M (2017) Inhibition of DYRK1A disrupts neural lineage specification in human pluripotent stem cells. *Elife* 6:e24502. <https://doi.org/10.7554/eLife.24502>

- Blackburn ATM, Bekheirnia N, Uma VC, Corkins ME, Xu Y, Rosenfeld JA, Bainbridge MN, Yang Y, Liu P, Madan-Khetarpal S, Delgado MR, Hudgins L, Krantz I, Rodriguez-Buritica D, Wheeler PG, Al-Gazali L, Mohamed Saeed Mohamed Al Shamsi A, Gomez-Ospina N, Chao H-T, Mirzaa GM, Scheuerle AE, Kukulich MK, Scaglia F, Eng C, Willsey HR, Braun MC, Lamb DJ, Miller RK, Bekheirnia MR (2019) DYRK1A-related intellectual disability: a syndrome associated with congenital anomalies of the kidney and urinary tract. *Genet Med* 21:2755–2764. <https://doi.org/10.1038/s41436-019-0576-0>
- Boissart C, Nissan X, Giraud-Triboulet K, Peschanski M, Benchoua A (2012) miR-125 potentiates early neural specification of human embryonic stem cells. *Development* 139:1247–1257. <https://doi.org/10.1242/dev.073627>
- Bronicki LM, Redin C, Drunat S, Piton A, Lyons M, Passemard S, Baumann C, Faivre L, Thevenon J, Rivière J-B, Isidor B, Gan G, Francannet C, Willems M, Gunel M, Jones JR, Gleeson JG, Mandel J-L, Stevenson RE, Friez MJ, Aylsworth AS (2015) Ten new cases further delineate the syndromic intellectual disability phenotype caused by mutations in DYRK1A. *Eur J Hum Genet* 23:1482–1487. <https://doi.org/10.1038/ejhg.2015.29>
- Courraud J, Chater-Diehl E, Durand B, Vincent M, del Mar Muniz Moreno M, Boujelbene I, Drouot N, Genschik L, Schaefer E, Nizon M, Gerard B, Abramowicz M, Cogné B, Bronicki L, Burglen L, Barth M, Charles P, Colin E, Coubes C, David A, Delobel B, Demurger F, Passemard S, Denommé A-S, Faivre L, Feger C, Fradin M, Francannet C, Genevieve D, Goldenberg A, Guerrot A-M, Isidor B, Johannesen KM, Keren B, Kibæk M, Kuentz P, Mathieu-Dramard M, Demeer B, Metreau J, Møller RS, Moutton S, Pasquier L, Pilekær Sørensen K, Perrin L, Renaud M, Saugier P, Svane J, Thevenon J, Them FTM, Tronhjem CE, Vitobello A, Layet V, Birling M-C, Drunat S, Bayat A, Dubourg C, Chehadeh SE, Fagerberg C, Mignot C, Guipponi M, Bienvenu T, Herault Y, Thompson J, Willems M, Mandel J-L, Weksberg R, Piton A (2021) Integrative approach to interpret DYRK1A variants, leading to a frequent neurodevelopmental disorder. *medRxiv* 2021.01.20.21250155. <https://doi.org/10.1101/2021.01.20.21250155>
- Di Vona C, Bezdán D, Islam ABMMK, Salichs E, López-Bigas N, Ossowski S, de la Luna S (2015) Chromatin-wide profiling of DYRK1A reveals a role as a gene-specific RNA polymerase II CTD kinase. *Mol Cell* 57:506–520. <https://doi.org/10.1016/j.molcel.2014.12.026>
- Dong C-M, Wang X-L, Wang G-M, Zhang W-J, Zhu L, Gao S, Yang D-J, Qin Y, Liang Q-J, Chen Y-L, Deng H-T, Ning K, Liang A-B, Gao Z-L, Xu J (2014) A stress-induced cellular aging model with postnatal neural stem cells. *Cell Death Dis* 5:e1116. <https://doi.org/10.1038/cddis.2014.82>
- Duchon A, Herault Y (2016) DYRK1A, a Dosage-Sensitive Gene Involved in Neurodevelopmental Disorders, Is a Target for Drug Development in Down Syndrome. *Front Behav Neurosci* 10:104. <https://doi.org/10.3389/fnbeh.2016.00104>
- Dutto I, Tillhon M, Cazzalini O, Stivala LA, Prosperi E (2015) Biology of the cell cycle inhibitor p21(CDKN1A): molecular mechanisms and relevance in chemical toxicology. *Arch Toxicol* 89:155–178. <https://doi.org/10.1007/s00204-014-1430-4>
- Eifler K, Cuijpers SAG, Willemstein E, Raaijmakers JA, El Atmioui D, Ovaas H, Medema RH, Vertegaal ACO (2018) SUMO targets the APC/C to regulate transition from metaphase to anaphase. *Nat Commun* 9:1119. <https://doi.org/10.1038/s41467-018-03486-4>
- Fischbach, K F, et M Heisenberg. « NEUROGENETICS AND BEHAVIOUR IN INSECTS ». *The company of Biologist limited* 112,65-93.

- Fotaki V, Dierssen M, Alcántara S, Martínez S, Martí E, Casas C, Visa J, Soriano E, Estivill X, Arbonés ML (2002) Dyrk1A haploinsufficiency affects viability and causes developmental delay and abnormal brain morphology in mice. *Mol Cell Biol* 22:6636–6647. <https://doi.org/10.1128/mcb.22.18.6636-6647.2002>
- Fotaki V, Martínez De Lagrán M, Estivill X, Arbonés M, Dierssen M (2004) Haploinsufficiency of Dyrk1A in mice leads to specific alterations in the development and regulation of motor activity. *Behav Neurosci* 118:815–821. <https://doi.org/10.1037/0735-7044.118.4.815>
- Gartel AL, Tyner AL (2002) The role of the cyclin-dependent kinase inhibitor p21 in apoptosis. *Mol Cancer Ther* 1:639–649
- Ghanem L, Steinman RA (2006) p21Waf1 inhibits granulocytic differentiation of 32Dcl3 cells. *Leuk Res* 30:1285–1292. <https://doi.org/10.1016/j.leukres.2005.12.023>
- Gonzalez-Mantilla AJ, Moreno-De-Luca A, Ledbetter DH, Martin CL (2016) A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *JAMA Psychiatry* 73:275–283. <https://doi.org/10.1001/jamapsychiatry.2015.2692>
- Guard SE, Poss ZC, Ebmeier CC, Pagratis M, Simpson H, Taatjes DJ, Old WM (2019) The nuclear interactome of DYRK1A reveals a functional role in DNA damage repair. *Sci Rep* 9:6539. <https://doi.org/10.1038/s41598-019-42990-5>
- Guedj F, Pereira PL, Najas S, Barallobre M-J, Chabert C, Souchet B, Sebric C, Verney C, Herault Y, Arbones M, Delabar JM (2012) DYRK1A: a master regulatory protein controlling brain growth. *Neurobiol Dis* 46:190–203. <https://doi.org/10.1016/j.nbd.2012.01.007>
- Guo X, Williams JG, Schug TT, Li X (2010) DYRK1A and DYRK3 promote cell survival through phosphorylation and activation of SIRT1. *J Biol Chem* 285:13223–13232. <https://doi.org/10.1074/jbc.M110.102574>
- Hämmerle B, Elizalde C, Tejedor FJ (2008) The spatio-temporal and subcellular expression of the candidate Down syndrome gene Mnb/Dyrk1A in the developing mouse brain suggests distinct sequential roles in neuronal development. *Eur J Neurosci* 27:1061–1074. <https://doi.org/10.1111/j.1460-9568.2008.06092.x>
- Han J, Kim Y-L, Lee K-W, Her N-G, Ha T-K, Yoon S, Jeong S-I, Lee J-H, Kang M-J, Lee M-G, Ryu B-K, Baik J-H, Chi S-G (2013) ZNF313 is a novel cell cycle activator with an E3 ligase activity inhibiting cellular senescence by destabilizing p21(WAF1.). *Cell Death Differ* 20:1055–1067. <https://doi.org/10.1038/cdd.2013.33>
- Hu J, Qian H, Xue Y, Fu X-D (2018) PTB/nPTB: master regulators of neuronal fate in mammals. *Biophys Rep* 4:204–214. <https://doi.org/10.1007/s41048-018-0066-y>
- Huang J, Ikeuchi Y, Malumbres M, Bonni A (2015) A Cdh1-APC/FMRP Ubiquitin Signaling Link Drives mGluR-Dependent Synaptic Plasticity in the Mammalian Brain. *Neuron* 86:726–739. <https://doi.org/10.1016/j.neuron.2015.03.049>
- Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, Szpyt J, Tam S, Zarraga G, Pontano-Vaites L, Swarup S, White AE, Schweppe DK, Rad R, Erickson BK, Obar RA, Guruharsha KG, Li K, Artavanis-Tsakonas S, Gygi SP, Harper JW (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature* 545:505–509. <https://doi.org/10.1038/nature22366>

- Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, Wühr M, Chick J, Zhai B, Kolippakkam D, Mintseris J, Obar RA, Harris T, Artavanis-Tsakonas S, Sowa ME, De Camilli P, Paulo JA, Harper JW, Gygi SP (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162:425–440. <https://doi.org/10.1016/j.cell.2015.06.043>
- Iossifov I, O’Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, Smith JD, Paepfer B, Nickerson DA, Dea J, Dong S, Gonzalez LE, Mandell JD, Mane SM, Murtha MT, Sullivan CA, Walker MF, Waqar Z, Wei L, Willsey AJ, Yamrom B, Lee Y, Grabowska E, Dalkic E, Wang Z, Marks S, Andrews P, Leotta A, Kendall J, Hakker I, Rosenbaum J, Ma B, Rodgers L, Troge J, Narzisi G, Yoon S, Schatz MC, Ye K, McCombie WR, Shendure J, Eichler EE, State MW, Wigler M (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515:216–221. <https://doi.org/10.1038/nature13908>
- Kim O-H, Cho H-J, Han E, Hong TI, Ariyasiri K, Choi J-H, Hwang K-S, Jeong Y-M, Yang S-Y, Yu K, Park D-S, Oh H-W, Davis EE, Schwartz CE, Lee J-S, Kim H-G, Kim C-H (2017) Zebrafish knockout of Down syndrome gene, DYRK1A, shows social impairments relevant to autism. *Mol Autism* 8:50. <https://doi.org/10.1186/s13229-017-0168-2>
- Laguna A, Aranda S, Barallobre MJ, Barhoum R, Fernández E, Fotaki V, Delabar JM, de la Luna S, de la Villa P, Arbonés ML (2008) The protein kinase DYRK1A regulates caspase-9-mediated apoptosis during retina development. *Dev Cell* 15:841–853. <https://doi.org/10.1016/j.devcel.2008.10.014>
- Laguna A, Barallobre M-J, Marchena M-Á, Mateus C, Ramírez E, Martínez-Cue C, Delabar JM, Castelo-Branco M, de la Villa P, Arbonés ML (2013) Triplication of DYRK1A causes retinal structural and functional alterations in Down syndrome. *Hum Mol Genet* 22:2775–2784. <https://doi.org/10.1093/hmg/ddt125>
- Lepagnol-Bestel A-M, Zvara A, Maussion G, Quignon F, Ngimbous B, Ramoz N, Imbeaud S, Loe-Mie Y, Benihoud K, Agier N, Salin PA, Cardona A, Khung-Savatovsky S, Kallunki P, Delabar J-M, Puskas LG, Delacroix H, Aggerbeck L, Delezoide A-L, Delattre O, Gorwood P, Moalic J-M, Simonneau M (2009) DYRK1A interacts with the REST/NRSF-SWI/SNF chromatin remodelling complex to deregulate gene clusters involved in the neuronal phenotypic traits of Down syndrome. *Hum Mol Genet* 18:1405–1414. <https://doi.org/10.1093/hmg/ddp047>
- Levy JA, LaFlamme CW, Tsaprilis G, Crynen G, Page DT (2021) Dyrk1a Mutations Cause Undergrowth of Cortical Pyramidal Neurons via Dysregulated Growth Factor Signaling. *Biol Psychiatry*. <https://doi.org/10.1016/j.biopsych.2021.01.012>
- Maestre C, Delgado-Esteban M, Gomez-Sanchez JC, Bolaños JP, Almeida A (2008) Cdk5 phosphorylates Cdh1 and modulates cyclin B1 stability in excitotoxicity. *EMBO J* 27:2736–2745. <https://doi.org/10.1038/emboj.2008.195>
- Mattioli F, Hayot G, Drouot N, Isidor B, Courraud J, Hinckelmann M-V, Mau-Them FT, Sellier C, Goldman A, Telegrafi A, Boughton A, Gamble C, Moutton S, Quartier A, Jean N, Van Ness P, Grotto S, Nambot S, Douglas G, Si YC, Chelly J, Shad Z, Kaplan E, Dineen R, Golzio C, Charlet-Berguerand N, Mandel J-L, Piton A (2020) De Novo Frameshift Variants in the Neuronal Splicing Factor NOVA2 Result in a Common C-Terminal Extension and Cause a Severe Form of Neurodevelopmental Disorder. *Am J Hum Genet* 106:438–452. <https://doi.org/10.1016/j.ajhg.2020.02.013>

- Mauro M, Rego MA, Boisvert RA, Esashi F, Cavallo F, Jasin M, Howlett NG (2012) p21 promotes error-free replication-coupled DNA double-strand break repair. *Nucleic Acids Res* 40:8348–8360. <https://doi.org/10.1093/nar/gks612>
- Menon VR, Ananthapadmanabhan V, Swanson S, Saini S, Sesay F, Yakovlev V, Florens L, DeCaprio JA, Washburn MP, Dozmorov M, Litovchick L (2019) DYRK1A regulates the recruitment of 53BP1 to the sites of DNA damage in part through interaction with RNF169. *Cell Cycle* 18:531–551. <https://doi.org/10.1080/15384101.2019.1577525>
- Møller RS, Kübart S, Hoeltzenbein M, Heye B, Vogel I, Hansen CP, Menzel C, Ullmann R, Tommerup N, Ropers H-H, Tümer Z, Kalscheuer VM (2008) Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. *Am J Hum Genet* 82:1165–1170. <https://doi.org/10.1016/j.ajhg.2008.03.001>
- Ori-McKenney KM, McKenney RJ, Huang HH, Li T, Meltzer S, Jan LY, Vale RD, Wiita AP, Jan YN (2016) Phosphorylation of  $\beta$ -Tubulin by the Down Syndrome Kinase, Minibrain/DYRK1a, Regulates Microtubule Dynamics and Dendrite Morphogenesis. *Neuron* 90:551–563. <https://doi.org/10.1016/j.neuron.2016.03.027>
- O’Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K, Munson J, Hiatt JB, Turner EH, Levy R, O’Day DR, Krumm N, Coe BP, Martin BK, Borenstein E, Nickerson DA, Mefford HC, Doherty D, Akey JM, Bernier R, Eichler EE, Shendure J (2012) Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338:1619–1622. <https://doi.org/10.1126/science.1227764>
- Pal & Summers The Roles of Cullin RING Ligases and the Anaphase Promoting Complex/Cyclosome in the Regulation of DNA Double Strand Break Repair
- Quartier A, Chatrousse L, Redin C, Keime C, Haumesser N, Maglott-Roth A, Brino L, Le Gras S, Benchoua A, Mandel J-L, Piton A (2018) Genes and Pathways Regulated by Androgens in Human Neural Cells, Potential Candidates for the Male Excess in Autism Spectrum Disorder. *Biol Psychiatry* 84:239–252. <https://doi.org/10.1016/j.biopsych.2018.01.002>
- Roewenstrunk J, Di Vona C, Chen J, Borrás E, Dong C, Arató K, Sabidó E, Huen MSY, de la Luna S (2019) A comprehensive proteomics-based interaction screen that links DYRK1A to RNF169 and to the DNA damage response. *Sci Rep* 9:6014. <https://doi.org/10.1038/s41598-019-42445-x>
- Ryoo S-R, Jeong HK, Radnaabazar C, Yoo J-J, Cho H-J, Lee H-W, Kim I-S, Cheon Y-H, Ahn YS, Chung S-H, Song W-J (2007) DYRK1A-mediated hyperphosphorylation of Tau. A functional link between Down syndrome and Alzheimer disease. *J Biol Chem* 282:34850–34857. <https://doi.org/10.1074/jbc.M707358200>
- Seifert A, Allan LA, Clarke PR (2008) DYRK1A phosphorylates caspase 9 at an inhibitory site and is potently inhibited in human cells by harmine. *FEBS J* 275:6268–6280. <https://doi.org/10.1111/j.1742-4658.2008.06751.x>
- Shi J, Zhang T, Zhou C, Chohan MO, Gu X, Wegiel J, Zhou J, Hwang Y-W, Iqbal K, Grundke-Iqbal I, Gong C-X, Liu F (2008) Increased dosage of Dyrk1A alters alternative splicing factor (ASF)-regulated alternative splicing of tau in Down syndrome. *J Biol Chem* 283:28660–28669. <https://doi.org/10.1074/jbc.M802645200>

- Sivakumar S, Gorbsky GJ (2015) Spatiotemporal regulation of the anaphase-promoting complex in mitosis. *Nat Rev Mol Cell Biol* 16:82–94. <https://doi.org/10.1038/nrm3934>
- Skaar JR, Pagano M (2009) Control of cell growth by the SCF and APC/C ubiquitin ligases. *Curr Opin Cell Biol* 21:816–824. <https://doi.org/10.1016/j.ceb.2009.08.004>
- Skurat AV, Dietrich AD (2004) Phosphorylation of Ser640 in muscle glycogen synthase by DYRK family protein kinases. *J Biol Chem* 279:2490–2498. <https://doi.org/10.1074/jbc.M301769200>
- Stein GH, Drullinger LF, Soulard A, Dulić V (1999) Differential roles for cyclin-dependent kinase inhibitors p21 and p16 in the mechanisms of senescence and differentiation in human fibroblasts. *Mol Cell Biol* 19:2109–2117. <https://doi.org/10.1128/MCB.19.3.2109>
- Tejedor F, Zhu XR, Kaltenbach E, Ackermann A, Baumann A, Canal I, Heisenberg M, Fischbach KF, Pongs O (1995) minibrain: a new protein kinase family involved in postembryonic neurogenesis in *Drosophila*. *Neuron* 14:287–301. [https://doi.org/10.1016/0896-6273\(95\)90286-4](https://doi.org/10.1016/0896-6273(95)90286-4)
- Tejedor FJ, Hämmerle B (2011) MNB/DYRK1A as a multiple regulator of neuronal development. *FEBS J* 278:223–235. <https://doi.org/10.1111/j.1742-4658.2010.07954.x>
- van Bon BWM, Coe BP, Bernier R, Green C, Gerds J, Witherspoon K, Kleefstra T, Willemsen MH, Kumar R, Bosco P, Fichera M, Li D, Amaral D, Cristofoli F, Peeters H, Haan E, Romano C, Mefford HC, Scheffer I, Gecz J, de Vries BBA, Eichler EE (2016) Disruptive de novo mutations of DYRK1A lead to a syndromic form of autism and ID. *Mol Psychiatry* 21:126–132. <https://doi.org/10.1038/mp.2015.5>
- Varjosalo M, Keskitalo S, Van Drogen A, Nurkkala H, Vichalkovski A, Aebersold R, Gstaiger M (2013) The protein interaction landscape of the human CMGC kinase group. *Cell Rep* 3:1306–1320. <https://doi.org/10.1016/j.celrep.2013.03.027>
- Vodermaier HC, Gieffers C, Maurer-Stroh S, Eisenhaber F, Peters J-M (2003) TPR subunits of the anaphase-promoting complex mediate binding to the activator protein CDH1. *Curr Biol* 13:1459–1468. [https://doi.org/10.1016/s0960-9822\(03\)00581-5](https://doi.org/10.1016/s0960-9822(03)00581-5)
- Willsey HR, Xu Y, Everitt A, Dea J, Exner CRT, Willsey AJ, State MW, Harland RM (2020) The neurodevelopmental disorder risk gene DYRK1A is required for ciliogenesis and control of brain size in *Xenopus* embryos. *Development* 147. <https://doi.org/10.1242/dev.189290>

### **Figure 1. Representation of DYRK1A protein interactome in hNSC**

(A) Schematic representation of interactors identified using immunoprecipitation coupled to mass spectrometry (IP-MS) in hNSC-1 cell line depending of DYRK1A antibody used: antibody directed against C-terminal part of DYRK1A (Cterm, in green) and antibody direct against N-terminal part of the protein (Nterm, in purple). (B) Protein interactome of DYRK1A: boxes are filled with a yellow to dark red gradient corresponding to a gradient of Peptide-Spectrum Matching (PSM) values (from 5 to 94 ; proteins found using both Cterm and Nterm antibodies are circled in blue ; proteins found using only Nterm antibody are circled in purple ; protein found using only Cterm antibody are circled in green ; proteins previously described as

interactors of DYRK1A are represented in circles; proteins discovered in this study as interacting with DYRK1A are represented in rectangular boxes.

**Figure 2. Transcriptomic changes induced by DYRK1A knock-down (KD) by siRNA in hNSC**

(A) Volcano plot showing RNA sequencing data (INTERFERin VS siDYRK1A treated hNSC-1). Genes not significantly deregulated are shown in black; genes with an adjusted p-value are between 0.5 and 0.1 are shown in orange while genes significantly deregulated (DEG, adjusted p-value <0.05) are shown in red. (B) RT-qPCR analysis of 28 DEGs (normalized on *GAPDH* and *YWHAZ*) after knock-down (KD) of *DYRK1A* in a third series of hNSC-1 and in 3 series of hNSC-2. Multiple comparisons tests were performed using Student T test t: ns: not significant; \*: p-value <0.01; \*\*: p-value <0.01 \*\*\*: p-value <0.001; Errors bars represent SEM.

**Figure 3. DYRK1A inactivation leads to an increase of CDKN1A isoform**

(A) IGV visualization of RNAseq data from hNSCs treated by *Scramble* or *DYRK1A* siRNA. The red frame indicate the event detected as significantly increased in *DYRK1A* siRNA compared to INTERFERin ; qPCR primers (1, 2 and 3) used for RT-qPCR validation are represented by black arrows. (B) RT-qPCR analysis of *CDKN1A* isoform NM\_000389.5 (389) and NM\_001291549.1 isoform (1291549) (normalized on *GAPDH* and *YWHAZ*) in both hNSCs lines (hNSC-1 and hNSC-2) after transient knock-down (KD) of *DYRK1A* using respectively primer couples 2-3 and 1-3 (n=3). Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; \*\*\*: p-value <0.001; Errors bars represent SEM (standard error of the mean).

**Figure 4. Effect of DYRK1A knock-down on hNSCs proliferation**

(A) Proliferation assay performed on both hNSC lines (hNSC-1 and hNSC-2) treated with lipofectant alone (INTERFERin) or transfected with *Scramble* siRNA (siScramble); *DYRK1A* siRNA (siDYRK1A) and *PLK1* siRNA (siPLK1), a control leading to cell apoptosis. At each time point (day 0, 1, 2, 3, and 4), cells were counted and data normalized by INTERFERin treatment (n=3 per line). Student's t test comparison was done comparing to INTERFERin: \*: p-value <0.05; Errors bars represent SEM. (B) Level of cleaved Caspase-3 in total extract protein from supernatant of both hNSCs lines after knock-down (KD) of *DYRK1A* normalized; level of total Caspase-3 in total extract protein from cells associated to supernatant (n=2 per line). (C) Evaluation of *DYRK1A* KD effect on senescence induction by quantification of number SA- $\beta$  galactosidase positive cells in both hNSC lines (hNSC-1 and hNSC-2) treated



with lipofectant alone (INTERFERin) or transfected with *Scramble* siRNA (siScramble); *DYRK1A* siRNA (siDYRK1A). Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; Errors bars represent SEM (standard error of the mean).

### **Table 1. List of DYRK1A interactors identified by Mass-Spectrometry in hNSCs**

This list includes DYRK1A's protein interactors identified through mass spectrometry analysis. Associated to these proteins, the number of peptide iteration (PSM) found using C-terminal and N-terminal DYRK1A antibodies, information related to their previously described interaction with DYRK1A (1: Guard et al. 2019 (HeLa); 2:Varjosalo et al., 2013 (HEK293); 3:Huttlin et al., 2015 (HEK293); 4: Huttlin et al. 2017 (HEK293); 5: Menon et al., 2019 (U-2 OS); 6: Roewenstrunk et al., 2019 (HeLa)), their functions and associated disease.

### **Table 2. List of genes DE in DYRK1A KD hNSCs**

This list includes names of the genes investigated at transcriptional level in this study. Associated to these genes, the information related to RNA sequencing: rank classed by adjusted p-values in RNAseq analysis, log2FC, p-value and adjusted p-value but also data related to RT-qPCR analysis in hNSCs-1 and hNSCs-2 lines with their associated p-value calculated using Student test. The "Gtex brain" column indicate approximative expression levels of the genes in the total brain: 0: not expressed ; +: low expressed ; ++ : well expressed ; +++ ; strongly expressed. The gene included in the list of genes involved in ID, ASD or EPI; ID<sup>1</sup>: gene well established as involved in ID (SysID primary list); ID<sup>2</sup>: candidate gene for ID (SysID candidate genes), ASD<sup>1</sup>: gene strongly associated to ASD (TADA102 gene list); ASD2: gene involved or associated to ASD with high confidence (from SFARI Gene database, categories: Syndromic, 1, 2); EPI: gene involved in epilepsy (genes included in the panel used by the French Epigene diagnosis network): Exp. Corr. BS: Correlation coefficient between the expression of the gene and that of *DYRK1A* in different brain tissues over time (between 8 weeks of pregnancy and 40 years, BrainSpan data, 254 brain samples). Finally, related phenotypes of KO mice models (jaxlab phenotype database) and associated pLI (gnomAD) are also indicated.

Figure 1

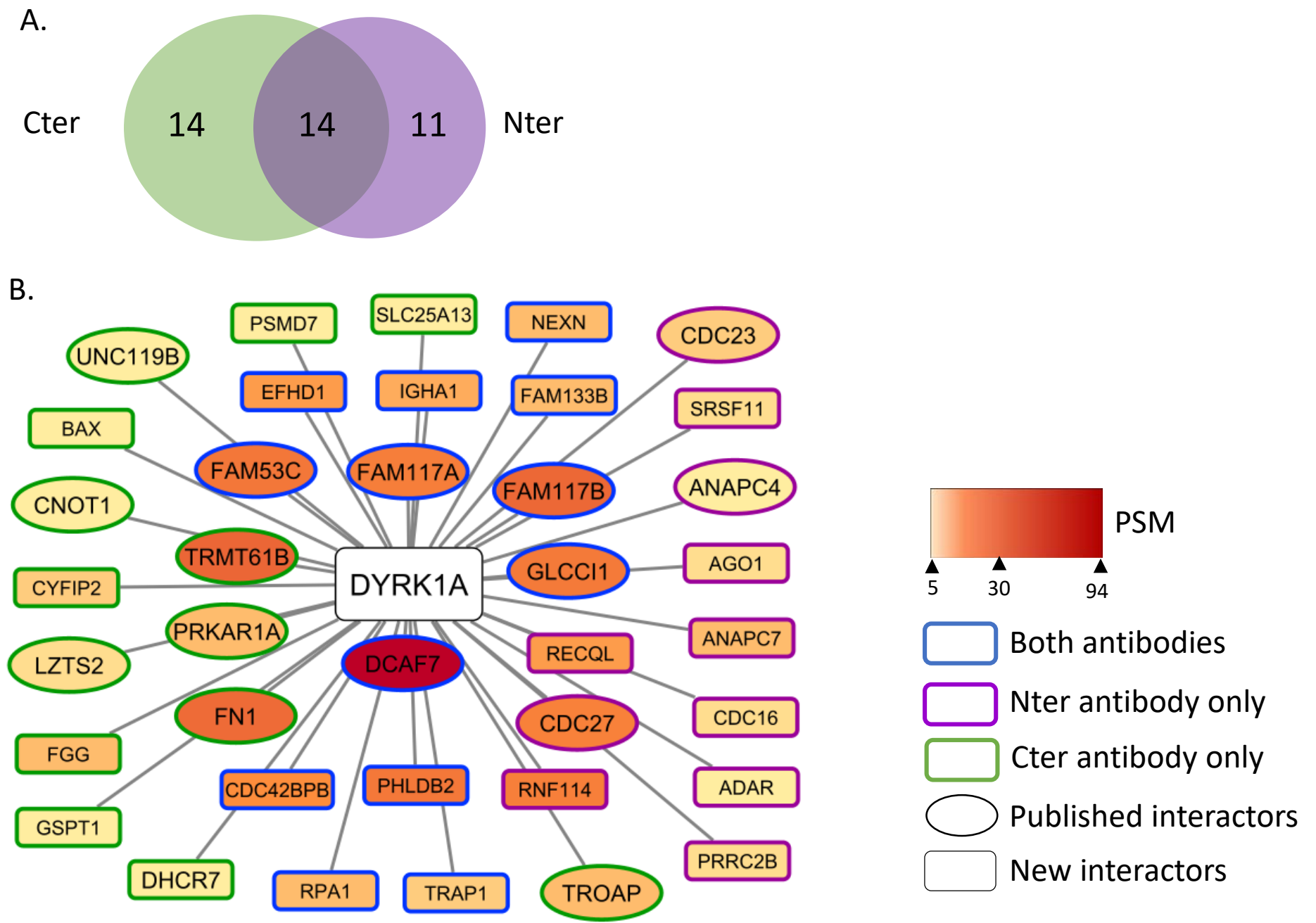


Figure 2

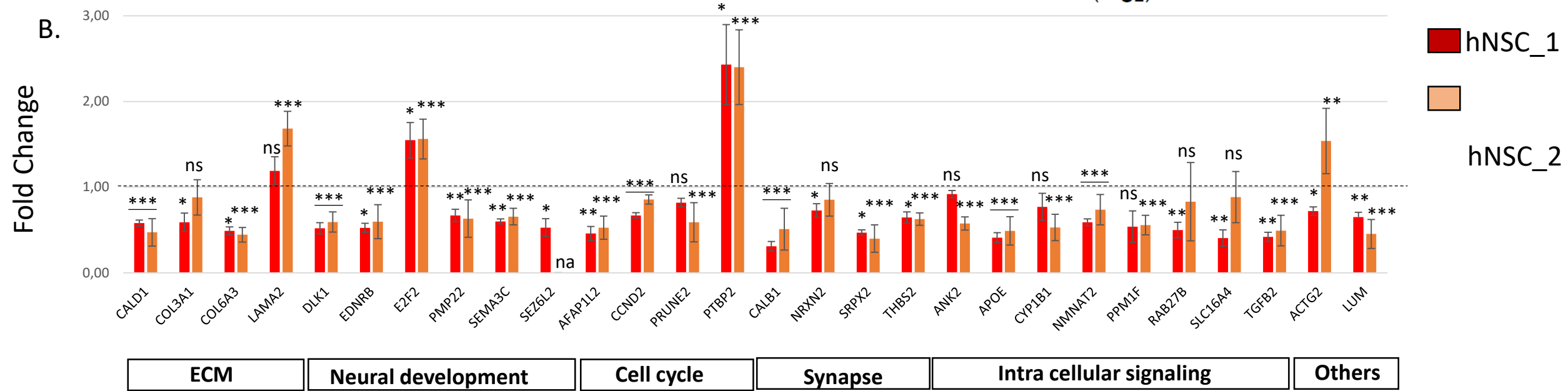
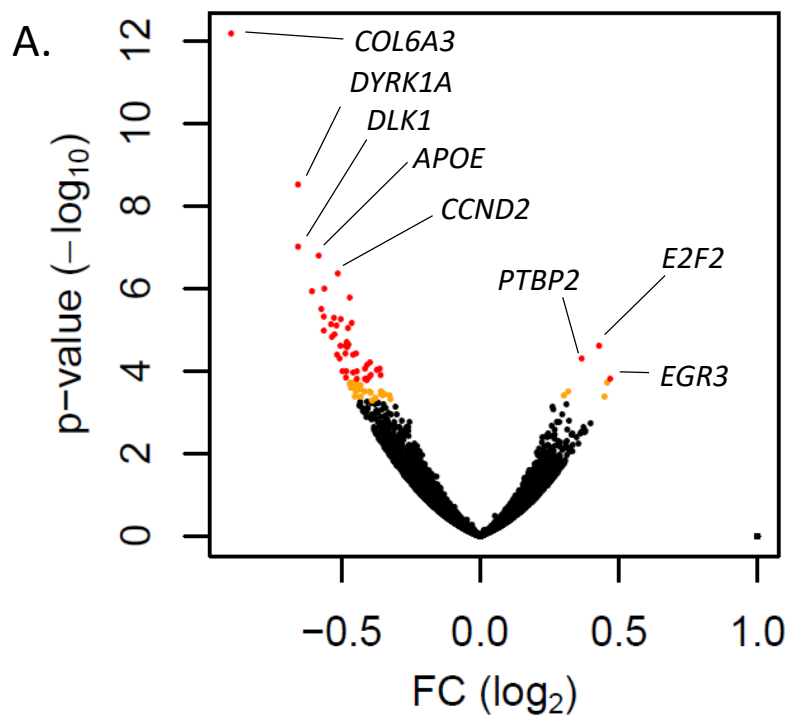


Figure 3

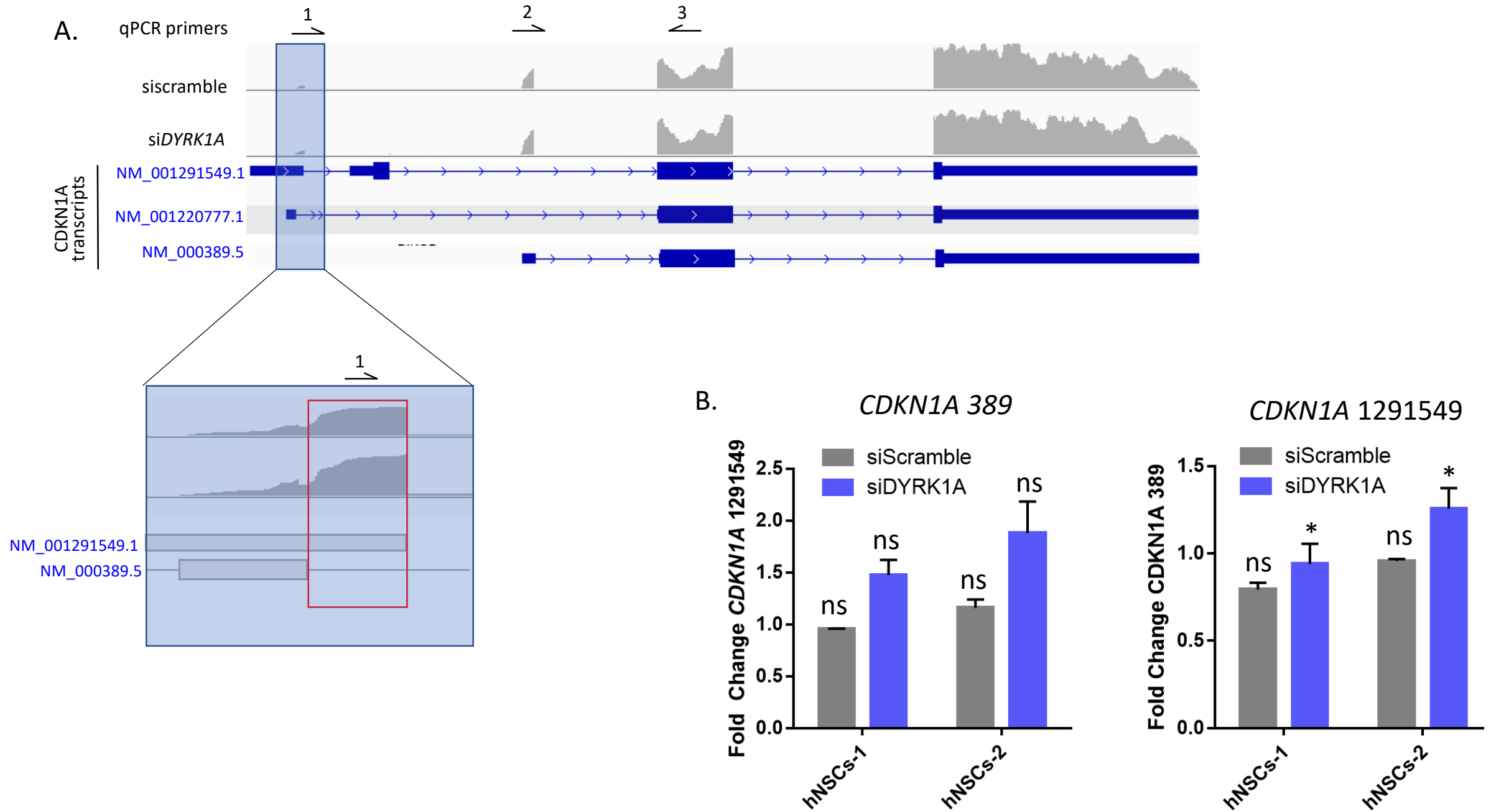
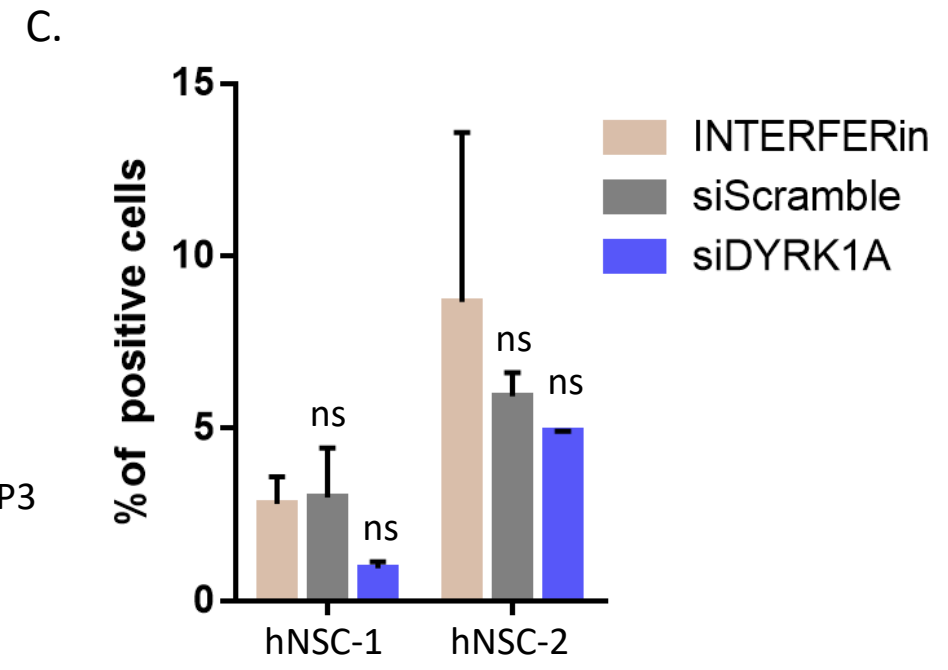
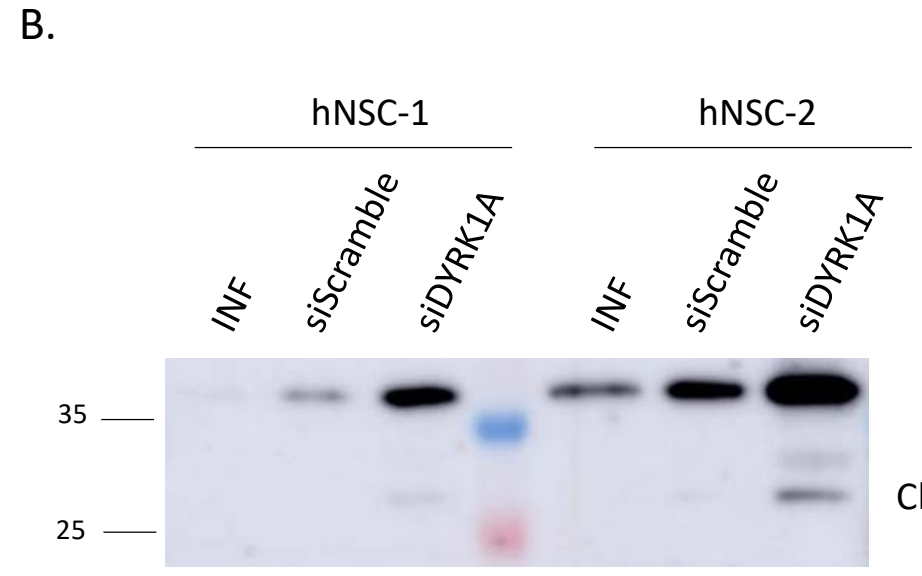
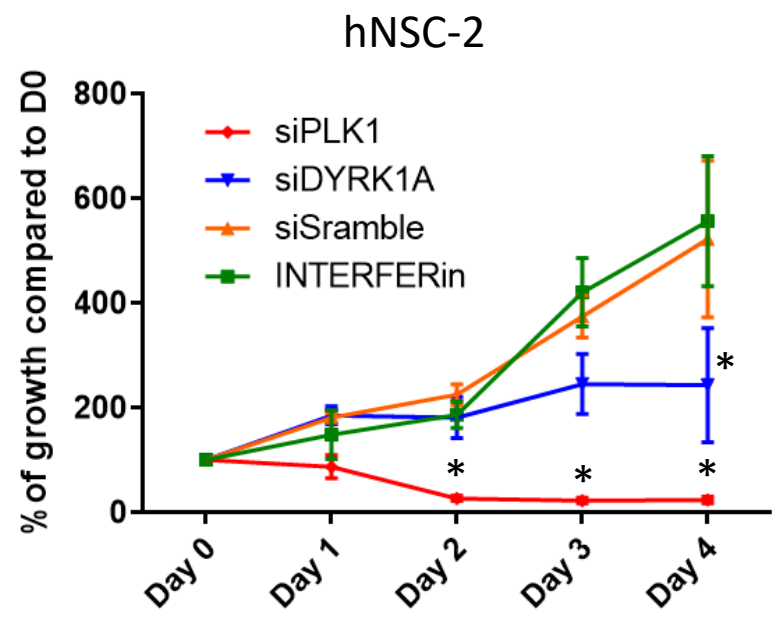
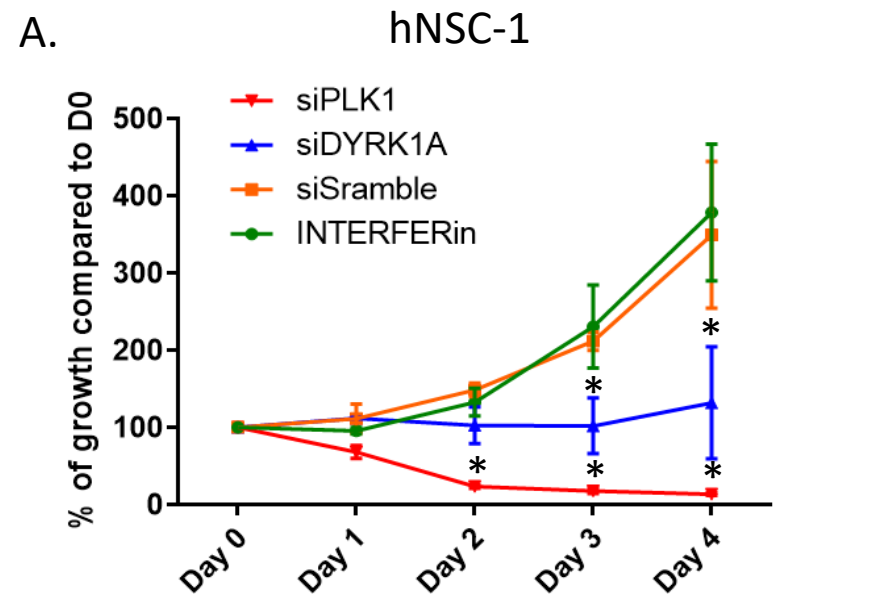


Figure 4



**Table 1**

Names	PSM C ter	PSM Nter	Phosphorylation event	Antibody	Interaction with DYRK1A	Disease OMIM/SyID	pLI
DYRK1A	97	80	Y321	both	NA	DYRK1A syndrome / MRD7	1
DCAF7	56	38		both	Skurat 2004 >10 studies	-	0,93
TRMT61B	34	0		C ter	1	-	0
FAM117B	21	12	S106	both	1, 2, 4, 5	-	0,01
FN1	31	0		C ter	1	Glomerulopathy with Fibronectin Deposits 2 and Spondylometaphyseal Dysplasia	0
FAM53C	13	11	S324	both	1, 2, 3, 4, 5, 6	-	0,07
PHLDB2	8	15		both	-	-	0
GLCCI1	16	6	S30, S303	both	1, 2, 5, 6	-	0,06
FAM117A	15	5		both	2, 5	-	0
RNF114	0	19		N ter	-	possible association with psoriasis	0,05
CDC27	0	18		N ter	1	-	1
CDC42BPB	3	8		both	-	Autism candidate genes, Current primary ID genes	1
EFHD1	5	5		both	-	-	0,01
RECQL	0	10		N ter	-	Bloom Syndrome	0
IGHA1	2	7		both	-	-	-
ANAPC7	0	8		N ter	-	-	0,85
FAM133B	5	3		both	-	-	0
FGG	8	0		C ter	-	Dysfibrinogenemia, afibrinogenemia	0
NEXN	3	5		both	-	Dilated Cardiomyopathy 1CC and Hypertrophic Cardiomyopathy 20	1
PRKAR1A	8	0		C ter	2, 5	Papillary thyroid carcinoma and leukemia	0,09
RPA1	2	6		both	-	Cancer	0
TROAP	7	1		C ter	1, 2, 3, 4, 5	-	0
CDC23	0	7		N ter	1	NDD with ataxia, hypotonia and mirocephaly	0,99
CYFIP2	6	1		C ter	-	Epileptic encephalopathy-65	1
TRAP1	5	2		both	-	ID candidate gene	0
AGO1	0	6		N ter	-	ID gene	1
CDC16	0	6		N ter	-	-	0,61
LZTS2	6	0		C ter	2, 5	-	0,05
PRRC2B	0	6		N ter	-	-	1
SRSF11	0	6		N ter	-	ID candidate gene	1
ADAR	0	5		N ter	-	Aicardi-Goutieres syndrome 6 and Dyschromatosis Symmetrica Hereditaria 1	0,61
ANAPC4	0	5		N ter	1	-	0
BAX	5	0		C ter	-	Cancer	0,32
CNOT1	5	0		C ter	5	Holoprosencephaly 12 and Vissers-Bodmer Syndrome	1
DHCR7	5	0		C ter	-	Smith-Lemli-Opitz syndrome	0
GSPT1	5	0		C ter	-	-	1
PSMD7	5	0		C ter	-	-	0,26
SLC25A13	5	0		C ter	-	Neonatal-Onset Type II Citrullinemia	0
UNC119B	5	0		C ter	5	Late-onset cone-rod dystrophy	0,78

Table 2

Rank RNAseq	Name	description	hNSCs-1 RNASeq		hNSCs-1 qPCR		hNSCs-2 qPCR		Gtex brain	gene list	Exp. Corr. BS	NDD in animal models	pLI
			log2FC	p-value adj	log2FC	p-value	log2FC	p-value					
1	COL6A3	collagen, type VI, alpha 3	-0,90	9,80E-09	-1,03	1,3E-02	-1,17	2,6E-01	0	ID <sup>1</sup>		0,01 -	0
2	DYRK1A	dual-specificity Tyrosine regulated kinase 1A	-0,66	2,26E-05	-1,47	5,1E-05	-2,05	1,8E-09	++	NA		1,00 NA	NA
3	DLK1	delta-like 1 homolog (Drosophila)	-0,66	4,84E-04	-0,94	2,2E-05	-0,75	2,2E-05	0	-		0,11 regulate post natal neurogenesis	0,03
4	APOE	apolipoprotein E	-0,58	5,82E-04	-1,29	7,0E-04	-1,03	1,0E-05	+	-		-0,68 Decreased brain cholesterol level ; abnormal astrocyte physiology	0
5	CCND2	cyclin D2	-0,51	1,34E-03	-0,58	2,1E-04	-0,22	7,9E-05	+	ID <sup>1</sup>		0,51 Abnormal cerebellar tissues and morphology	0,99
6	LAMA4	laminin, alpha 4	-0,56	2,46E-03	-0,76	2,1E-02	-0,95	7,4E-08	0	-		0,09 Abnormal neuromuscular jonction	0
7	PLP1	proteolipid protein 1	-0,61	2,46E-03	-1,03	1,1E-02	-0,44	3,9E-03	Brain only	ID <sup>1</sup>		-0,62 Abnormal proliferation in subventricular zone, demyelination	0,93
8	NMNAT2	nicotinamide nucleotide adenyltransferase 2	-0,47	3,09E-03	-0,76	1,0E-03	-0,92	5,9E-07	+++	ID <sup>2</sup>		-0,10 Nerve/axon defect and truncated axons	0,99
9	AFAP1L2	actin filament associated protein 1-like 2	-0,57	5,10E-03	-1,12	2,7E-03	-0,94	6,6E-04	+	-		-0,55 -	0
10	LAMA2	laminin, alpha 2	-0,56	6,80E-03	0,25	2,5E-01	0,75	1,7E-05	0/+	ID <sup>1</sup>		-0,13 Abnormal : axon, myelination, reflex, nerve conduction; ataxia	0
12	EDNRB	endothelin receptor type B	-0,50	6,80E-03	-0,93	1,8E-02	-0,74	6,8E-04	++	-		-0,47 Abnormal nervous system, development of enteric neuroblastes	0,01
11	SEZ6L2	seizure related 6 homolog (mouse)-like 2	-0,53	6,80E-03	-0,92	1,2E-02	NA	NA	+++	-		-0,43 Motor discoordination	0
14	CALB1	calbindin 1, 28kDa	-0,54	7,69E-03	-1,69	1,1E-04	-0,97	4,8E-04	++	-		-0,15 ataxia, abnormal : reflex/ electrophysiology	0,24
13	PMP22	peripheral myelin protein 22	-0,46	7,69E-03	-0,58	1,5E-02	-0,66	1,1E-03	0/+	-		-0,01 Abnormal axons, myelination	0,91
15	SCN3A	sodium channel, voltage-gated, type III, alpha subunit	-0,52	7,69E-03	-0,43	3,6E-02	-0,11	3,5E-01	++	ID <sup>1</sup> , EPI		0,62 Seizures	1
16	MYORG	KIAA1161	-0,48	8,50E-03	0,32	3,9E-01	0,80	9,4E-04	++	-		-0,19 -	0
17	RGS6	regulator of G-protein signaling 6	-0,56	9,53E-03	-1,36	4,9E-04	-1,12	3,3E-06	++	ID <sup>1</sup>		-0,24 -	0,44
18	RAB27B	RAB27B, member RAS oncogene family	-0,53	1,11E-02	-1,00	7,4E-03	-0,27	4,5E-01	++	-		0,31 Abnormal sleep	0,04
19	SCG2	secretogranin II	-0,53	1,20E-02	-0,29	4,7E-02	0,69	6,0E-05	++	ID <sup>2</sup>		0,72 -	0
20	CALD1	caldesmon 1	-0,48	1,53E-02	-0,79	5,5E-04	-1,08	4,7E-07	0	-		-0,41 -	1
23	TGFB2	transforming growth factor, beta 2	-0,49	1,55E-02	-1,25	1,1E-02	-1,02	9,9E-05	+	-		-0,45 -	0
21	E2F2	E2F transcription factor 2	0,43	1,55E-02	0,63	3,9E-02	0,64	1,2E-04	0	-		0,34 -	0,26
24	CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1	-0,48	1,55E-02	-0,38	2,1E-01	-0,92	5,3E-04	0/+	-		-0,11 -	0
22	PRUNE2	prune homolog 2 (Drosophila)	-0,50	1,55E-02	-0,29	5,7E-02	-0,76	6,9E-04	+	-		-0,08 Normal nervous sytem, neurological and behavioral phenotype	0
25	CUBN	cubilin (intrinsic factor-cobalamin receptor)	-0,47	1,55E-02	-0,20	4,2E-01	-0,10	5,3E-01	0/+	-		0,19 -	0
27	SEMA3C	semaphorin 3C	-0,48	2,11E-02	-0,74	7,9E-03	-0,61	2,4E-07	0/+	-		-0,15 abnormal neural crest migration	0
26	COL3A1	collagen, type III, alpha 1	-0,45	2,11E-02	-0,76	2,8E-02	-0,18	2,1E-01	0	-		0,00 -	1
29	SRPX2	sushi-repeat containing protein, X-linked 2	-0,52	2,13E-02	-1,09	6,0E-03	-1,32	6,5E-07	0/+	ID <sup>2</sup>		-0,06 abnormal response to social novelty, decreased CNS synapse formation	0,05
28	UNC13A	unc-13 homolog A (C. elegans)	-0,46	2,13E-02	-0,86	1,1E-02	-0,84	4,0E-05	++	ID <sup>2</sup> , ASD <sup>2</sup>		-0,17 abnormal synaptic : transmission, vesicles number, recycling, morphology	1
30	PTBP2	polypyrimidine tract binding protein 2	0,37	2,46E-02	1,28	1,8E-02	1,26	6,5E-07	++	-		0,80 Abnormal neurone differentiation, brain structures, precursor number	1
31	CDKN2B	cyclin-dependent kinase inhibitor 2B (p15)	-0,51	2,46E-02	0,11	7,2E-01	1,10	4,8E-04	0/+	-		-0,54 -	0,01
32	CLCN4	chloride channel, voltage-sensitive 4	-0,40	2,80E-02	0,15	4,4E-01	0,57	4,6E-05	++	ID <sup>1</sup> , ASD <sup>2</sup> , EPI		0,50 -	1
33	NACAD	NAC alpha domain containing	-0,41	3,22E-02	-0,18	2,1E-01	0,48	8,9E-04	++	-		-0,27 -	0,16
35	PPM1F	protein phosphatase, Mg2+/Mn2+ dependent, 1F	-0,36	3,66E-02	-0,89	9,9E-02	-0,84	3,1E-04	0/+	-		-0,22 Hyperactivity	0
34	ANK2	ankyrin 2, neuronal	-0,42	3,66E-02	-0,12	1,7E-01	-0,79	3,3E-04	+++	ID <sup>2</sup> , ASD <sup>1,2</sup>		0,37 Abnormal : axon morphology, locomotor behavior ; optic nerve	1
36	BMF	Bcl2 modifying factor	-0,37	3,87E-02	0,08	7,4E-01	0,81	1,4E-03	0/+	-		0,38 -	0,01
37	ACTG2	actin, gamma 2, smooth muscle, enteric	-0,44	4,01E-02	-0,47	4,5E-02	0,62	7,9E-03	0	-		-0,19 -	0,07
39	CHL1	cell adhesion molecule L1-like	-0,48	4,01E-02	-0,47	7,5E-02	-0,49	9,0E-02	+	ID <sup>2</sup>		0,69 Abnormal: cognition, neurones, cortex	0
38	SLC16A4	solute carrier family 16, member 4	-0,50	4,01E-02	-1,30	9,0E-03	-0,18	3,5E-01	+	-		-0,26 -	0
40	F2RL2	coagulation factor II (thrombin) receptor-like 2	-0,46	4,05E-02	0,19	4,0E-01	0,70	3,8E-05	0/+	-		0,04 -	0
41	THBS2	thrombospondin 2	-0,39	4,50E-02	-0,63	3,1E-02	-0,67	1,3E-06	0/+	-		0,34 -	0,56
42	FAM214B	family with sequence similarity 214, member B	-0,36	4,56E-02	NA	NA	NA	NA	++	-		0,72 -	0,8
47	LUM	lumican	-0,41	4,85E-02	-0,62	3,6E-03	-1,14	7,1E-06	0	-		-0,03 -	0,08
43	EGR3	early growth response 3	0,47	4,85E-02	0,39	1,0E-02	0,73	6,3E-05	+	-		-0,21 abnormal: electrophysiology, neurones	0,95
46	DCX	doublecortin	-0,40	4,85E-02	-0,45	8,9E-02	-0,28	3,1E-02	++ (brain only)	ID <sup>1</sup>		0,75 Abnormal : Brain structure, behavior	0,3
44	NRXN2	neurexin 2	-0,42	4,85E-02	-0,46	2,7E-02	-0,23	1,4E-01	++	ID <sup>2</sup> , ASD <sup>2</sup>		-0,45 Abnormal: novel object, social investigation, normal nervous system	1
48	FAT4	FAT atypical cadherin 4	-0,44	4,85E-02	-0,34	1,5E-01	-0,19	2,8E-01	0/+	ID <sup>1</sup>		-0,49 Aborman neural tube morphology	1
45	NALCN	sodium leak channel, non-selective	-0,48	4,85E-02	NA	NA	NA	NA	+++	ID <sup>1</sup>		0,57 Abnormal: brain wave, nervous system electrophysiology	0
49	TNFRSF19	tumor necrosis factor receptor, member 19	-0,41	4,98E-02	0,07	8,0E-01	0,29	1,5E-02	0/+	-		-0,17 abnormal neuron physiology and blood-brain barrier function	0,03
11730	CDKN1A (NV cyclin dependent kinase inhibitor 1A		-0,02	9,98E-01	-0,09	0,67	0,33	0,01	0/+	-		0,00 Abnormal neural precursor proliferation	0
NA	CDKN1A NM cyclin dependent kinase inhibitor 1A		NA	NA	0,56	0,03	0,91	0,00	0/+	NA	NA	NA	NA

## Supplementaries

### **Figure S1. *DYRK1A* KD using siRNA transfection leads to decrease of *DYRK1A* mRNA and protein expression.**

(A) RT-qPCR analysis of *DYRK1A* mRNA level (normalized on GAPDH and YWHAZ) performed on both hNSC lines (hNSC-1 and hNSC-2) transfected with *Scramble* siRNA (siScramble) or *DYRK1A* siRNA (siDYRK1A) (n=3 per line) (B) Western Blot analysis on total protein extract from hNSC-1 treated with lipofectant alone (INF) or transfected with *Scramble* or *DYRK1A* siRNA. Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; \*\*\*: adjusted p-value<0.001; Errors bars represent SEM (standard error of the mean).

### **Figure S2. Effect of *DYRK1A* KD on NMNAT2 and PTBP2 expression levels.**

(A) Western Blot analysis on total protein extract hNSC-1 line treated with lipofectant alone (INF) or transfected with *Scramble* or *DYRK1A* siRNA for 48h (n=3); NMNAT2 expression level quantification using GAPDH as a calibrator. (B) Western Blot analysis on total protein extract of hNSC (hNSC-1 and 2) treated with lipofectant alone (INF) or transfected with *Scramble* or *DYRK1A* siRNA for 48h (n=3); PTBP2 expression level quantification using GAPDH as a calibrator. Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; \*\*: adjusted p-value<0.01; \*: adjusted p-value<0.05; Errors bars represent SEM (standard error of the mean).

### **Figure S3. Transcriptomic analysis in fibroblasts from individuals carrying *DYRK1A* variants compared to control individuals**

(A) Principal Component Analysis of fibroblast transcriptomic analysis of control individuals (n=6, in pink) and individuals carrying *DYRK1A* variant (n=7, in blue). (B) Volcano plot showing RNA sequencing data. P-value (-log10) is expressed for each gene in function of log2fold change (log2FC). Genes not significantly regulated are shown in black; while genes significantly regulated (adjusted p-value <0.05) are shown in red.

### **Figure S4. Effect of *DYRK1A* KD on p21/CDKN1A protein level and mTOR, PKC and ERK phosphorylation levels.**



(A) Western Blot analysis on total protein extract from both hNSC (hNSC-1 and 2) treated with lipofectant alone (INF) or transfected with *Scramble* or *DYRK1A* siRNA for 48h. (B) Phosphorylation ratio of mTOR (n=2), PKC (n=3) and ERK (n=4) after Western Blot quantification using GAPDH as a calibrator and normalizing on INF; p21/CDKN1A protein level (n=3 per line) using GAPDH as a calibrator and normalized on INF. Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; \*\*: adjusted p-value<0.01; \*: adjusted p-value<0.05; Errors bars represent SEM (standard error of the mean).

**Figure S5. *DYRK1A* KD does not affect senescence in hNSC.**

Representative pictures of both hNSCs lines (hNSCs-1 and 2) after 8Gy irradiation treatment or treated by with lipofectant alone (INF) or transfected with *Scramble* or *DYRK1A* siRNA for 48h and labeled by SA- $\beta$  galactosidase assay. Cells were fixed 4 days after irradiation or transfection. Scales bars (in black) represent 50 $\mu$ m.

**Table S1. References of the different antibodies used during this study**

**Table S2. Enriched GO terms associated to *DYRK1A* interactors**

**Table S3. List of DEG identified after *DYRK1A* KD hNSC**

This list includes raw information of the DE genes identified through RNA sequencing in KD hNSCs-1. Associated to each gene its normalized number of reads, log<sub>2</sub> Fold Change, p-values and adjusted p-values of siDYRK1A condition compared to INF.

**Table S4. List of DEG identified in fibroblasts of individuals carrying pathogenic variants in *DYRK1A* compared to control individuals**

This list includes raw information of the DE genes identified through RNA sequencing control individuals (in pink) and patients carrying *DYRK1A* LoF variants (in blue), patients present in previous study (Courraud et al. 2021) are labelled "Crd" and their number in this paper. Associated to each gene its normalized number of reads, log<sub>2</sub> Fold Change and adjusted p-values of controls VS *DYRK1A* patients' conditions.

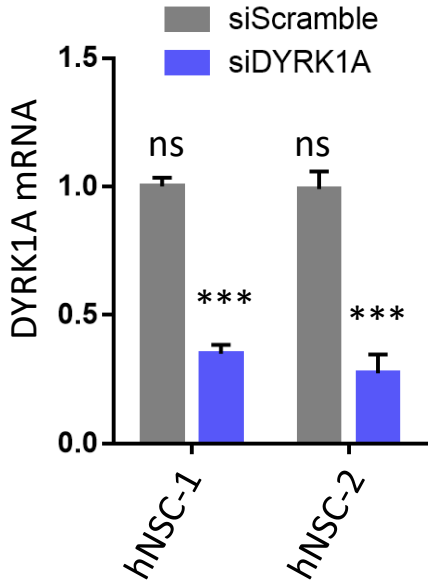
**Table S5. List of significant changes identified in splicing/isoform after *DYRK1A* KD in hNSC by LeafCutter program**

**Table S6. Number of cells counted following SA- $\beta$ -GAL staining in both hNSC-1 and hNSC-2 KD 4 days *DYRK1A* KD**

**Table S7. List of HPO terms associated to DE genes identified by transcriptomic study in KD hNSCs-1**

Figure S1

A.



B.

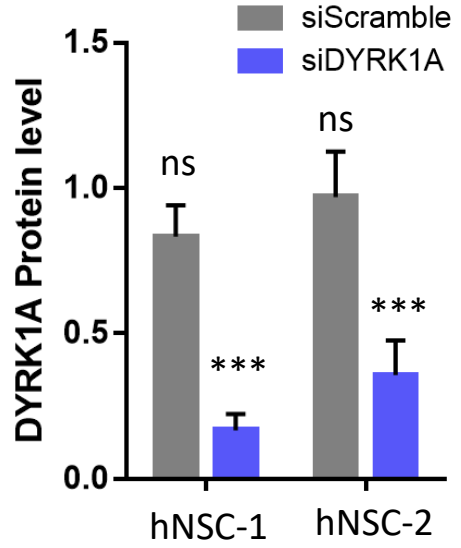
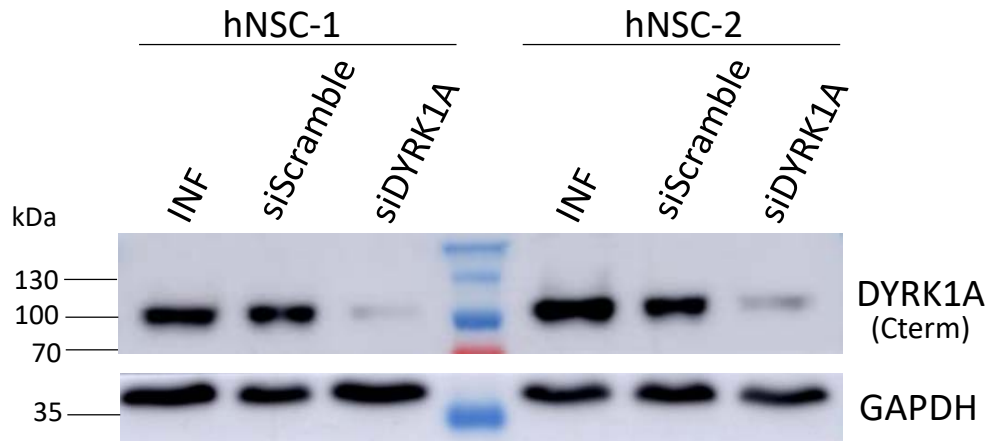
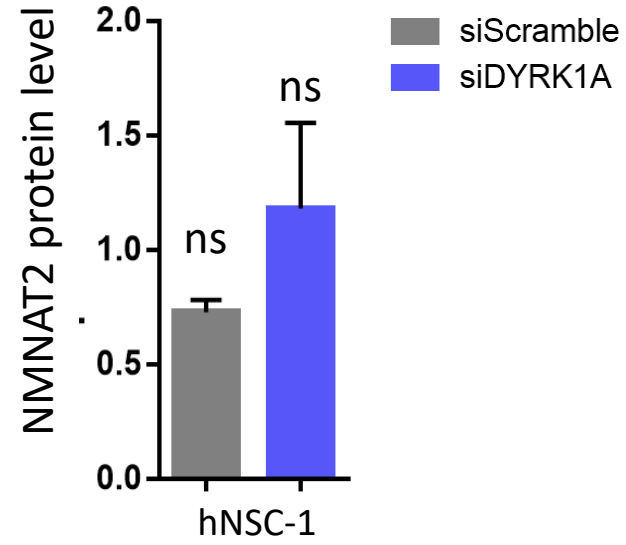
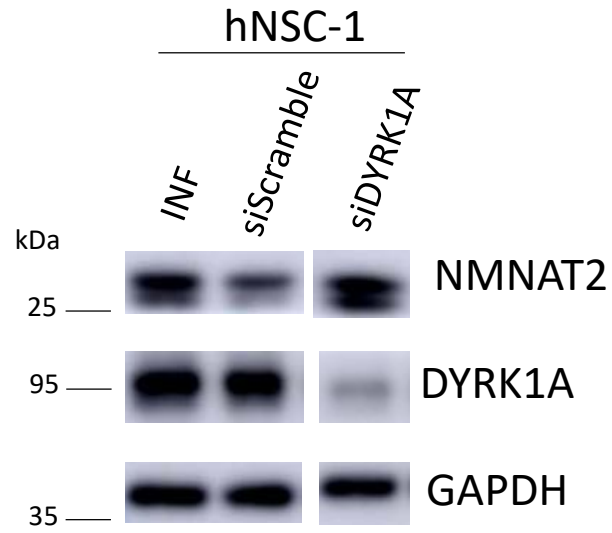


Figure S2

A.



B.

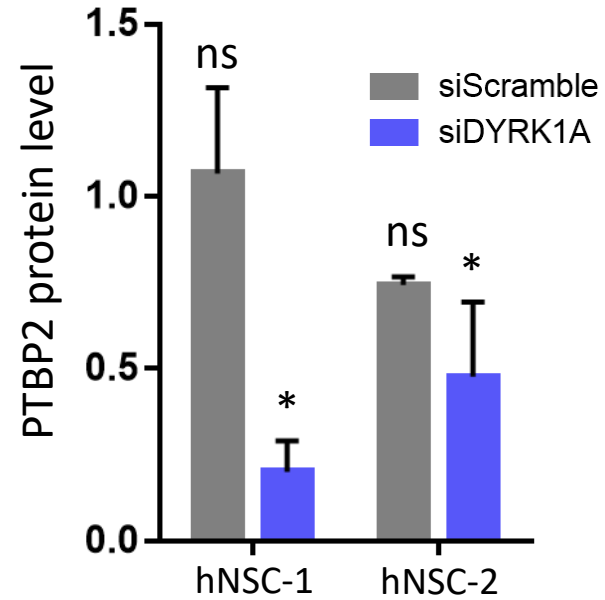
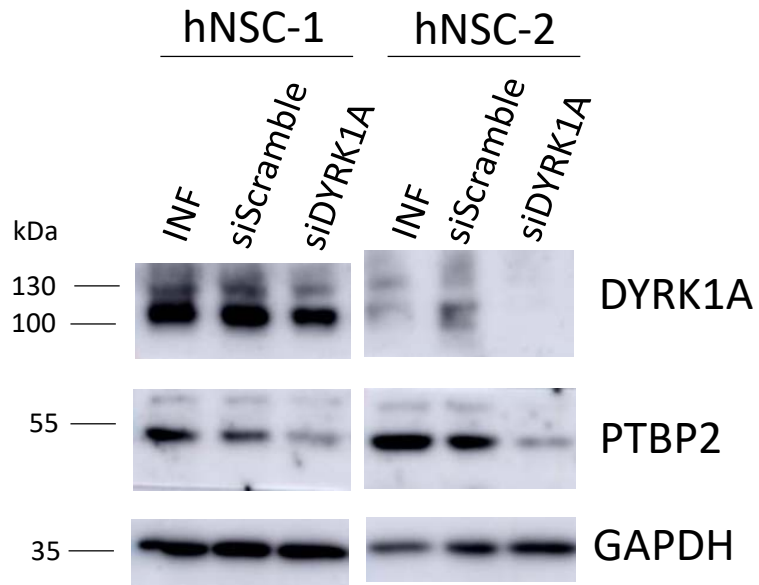
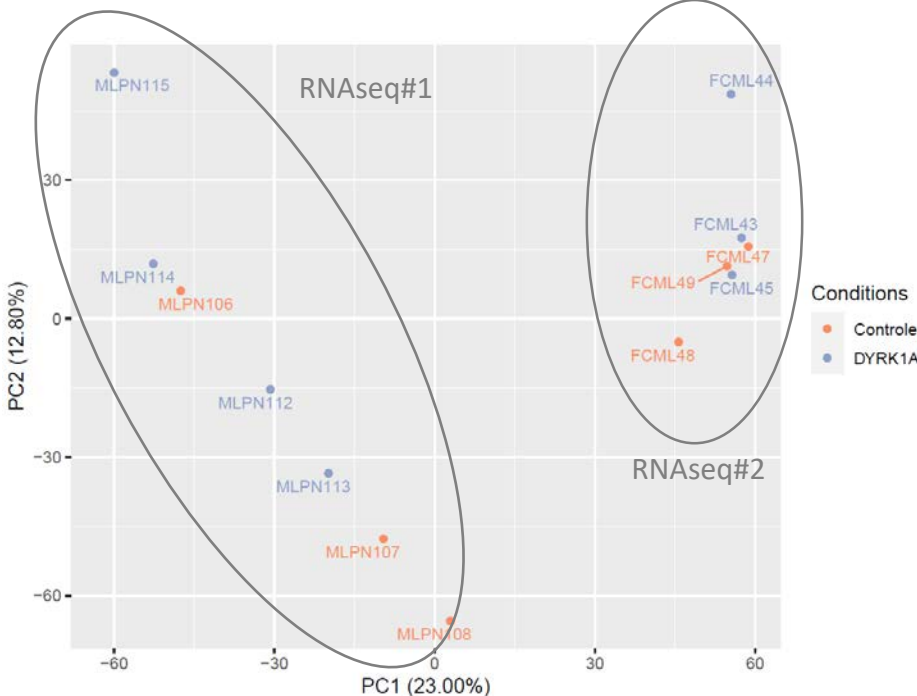


Figure S3

A.



B.

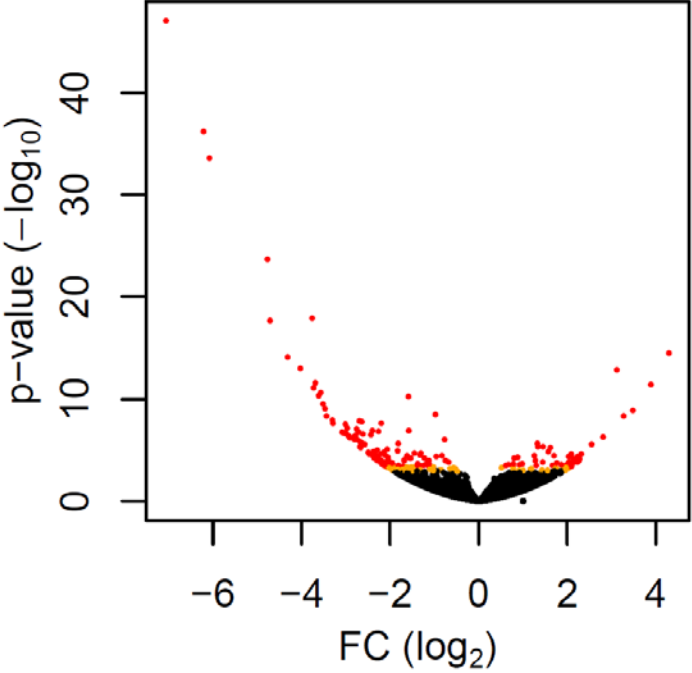
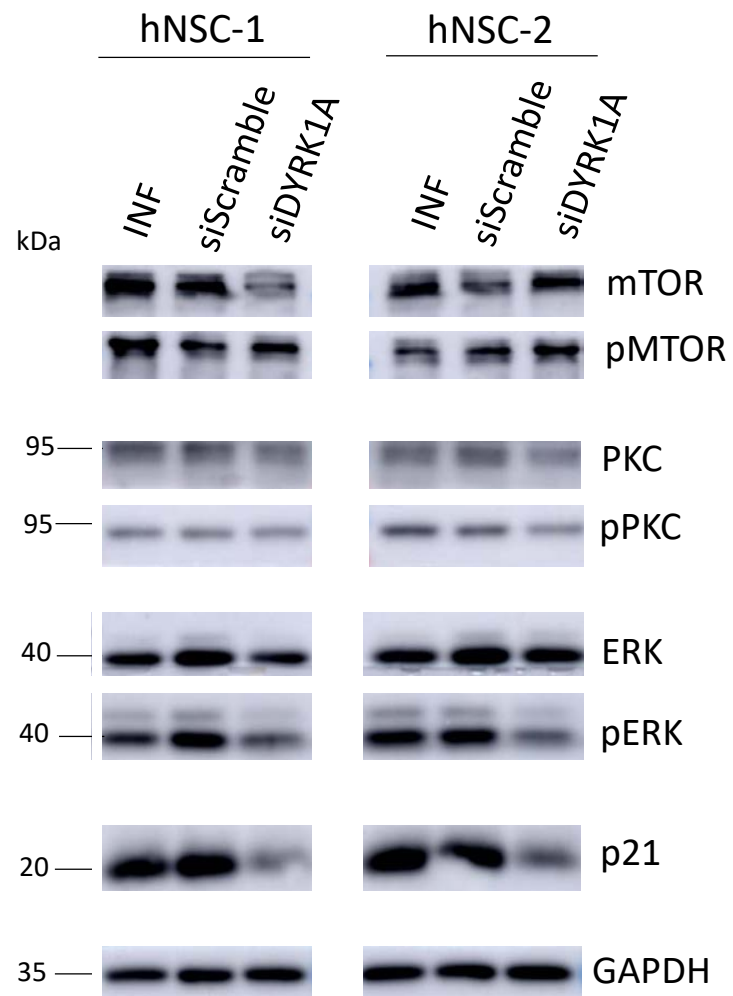


Figure S4

A.



B.

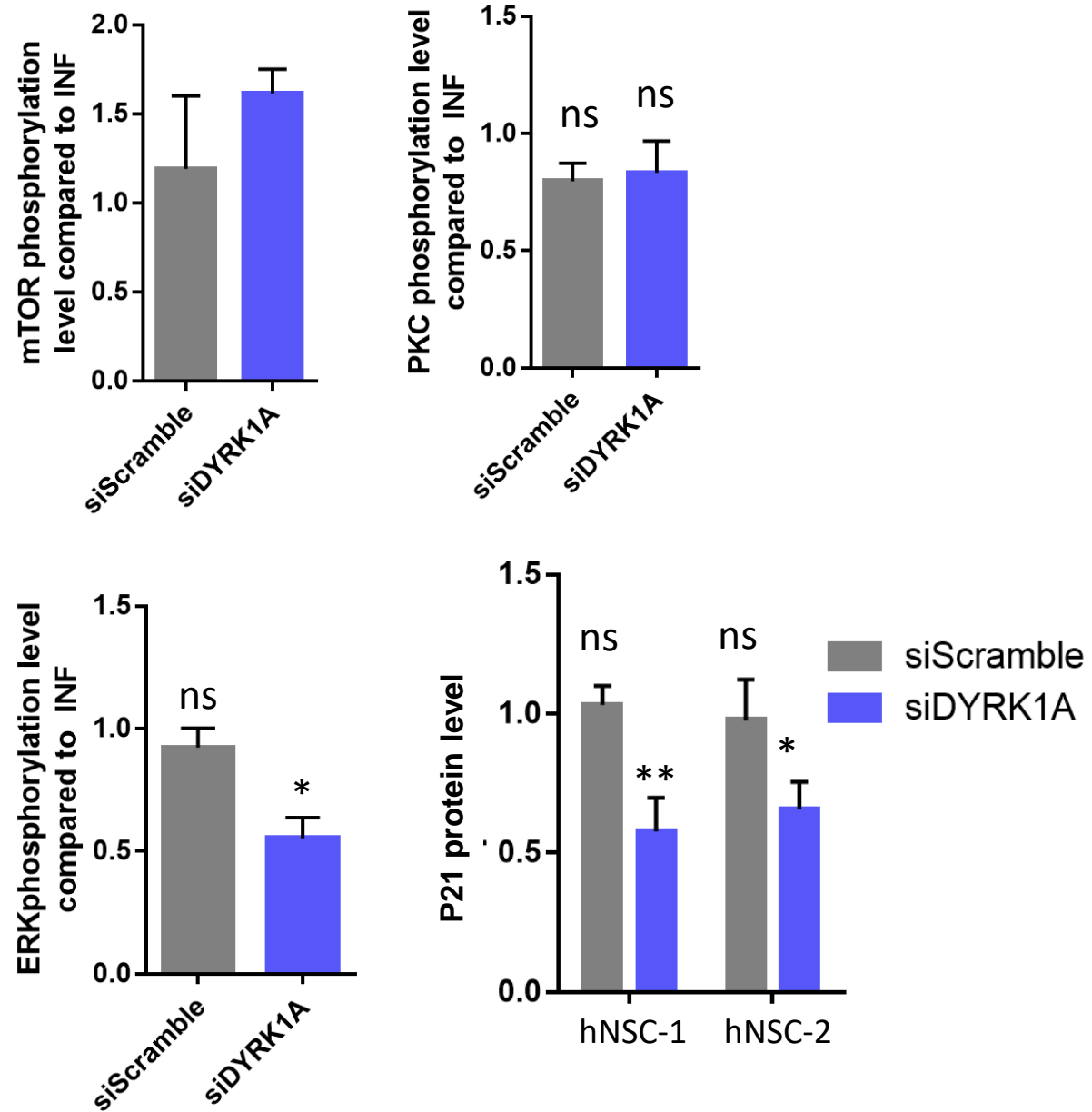
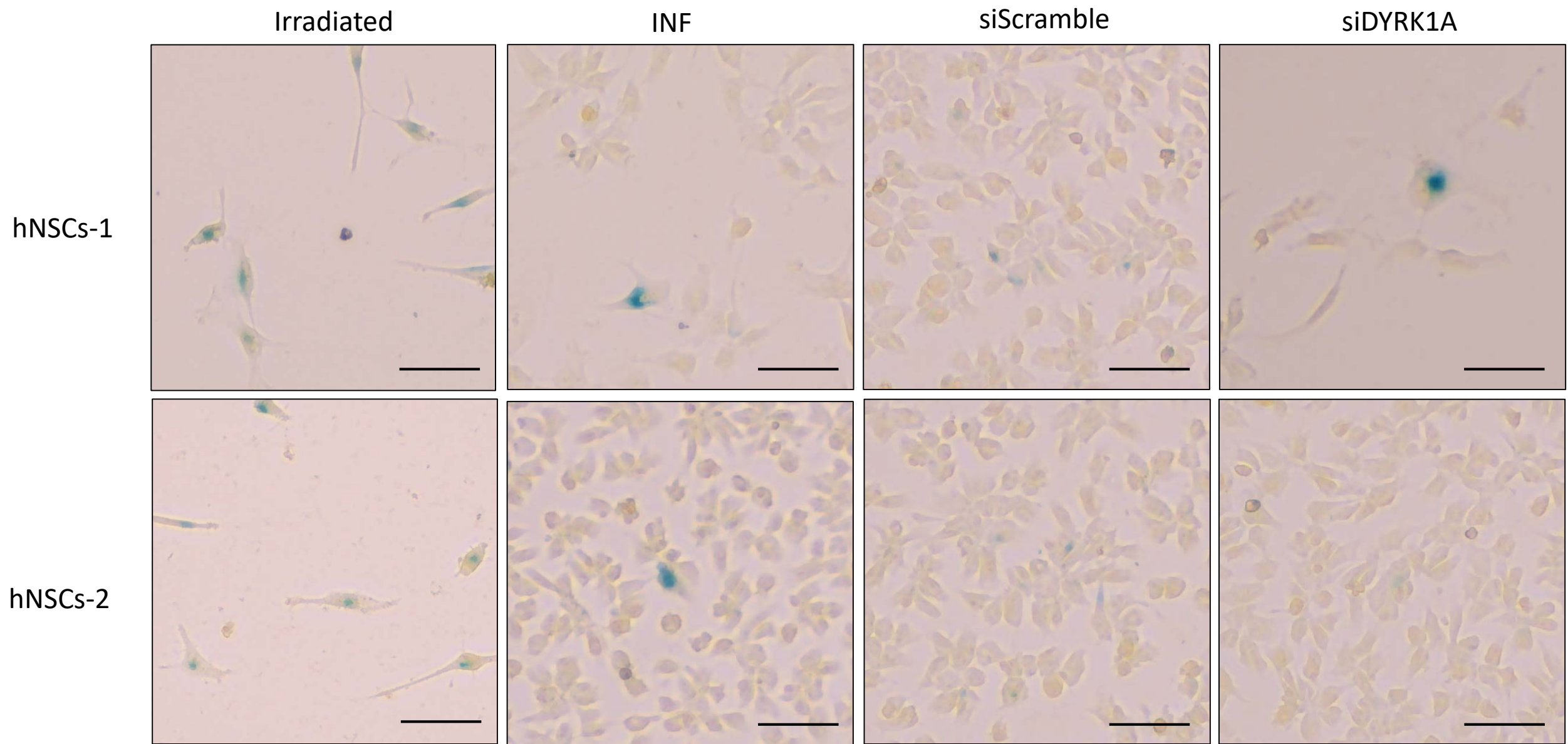


Figure S5



**Table S1**

<b>Proteins</b>	<b>Reference</b>
GAPDH	G9545 Sigma Aldrich
DYRK1A Nterm	CPA1357 Cohesion Biosciences
DYRK1A Cterm	H00001859-M01 Abnova
Cleaved Caspase 3	AB3623 Merck
pan mTOR	2972 Cell signaling technology
phospho mTOR	2974 Cell signaling technology
pan PKC	P5704 clone MC5 Sigma
phospho PKC $\alpha$	sc-208 Santa Cruz
pan ERK1/2	4696S Cell signaling technology
phospho ERK1/2	9101 Cell signaling technology
P21	2947S Cell signaling technology
NMNAT2	STJ115774 St John's Laboratory
PTBP2	#4535 Cell signaling Technology

**Table S2**

<b>GO term Category</b>	<b>GO term nomenclature</b>	<b>nber of genes</b>	<b>enrichment</b>	<b>p-value</b>	<b>adjusted p-value</b>
CC	anaphase-promoting complex	5	12,8	9,80E-08	9,20E-06
BP	regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	5	12,8	1,20E-07	3,50E-05
BP	protein K11-linked ubiquitination	5	12,8	2,40E-07	3,50E-05
BP	negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	6	15,4	3,00E-07	3,50E-05
BP	positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	6	15,4	4,20E-07	3,60E-05
BP	anaphase-promoting complex-dependent catabolic process	6	15,4	5,10E-07	3,60E-05
BP	proteasome-mediated ubiquitin-dependent protein catabolic process	6	15,4	5,10E-05	3,10E-03
BP	protein ubiquitination involved in ubiquitin-dependent protein catabolic process	5	12,8	2,50E-04	1,30E-02
BP	cell division	6	15,4	6,50E-04	2,90E-02



Table S3

## Reads normalized and divided by gene length in kb

INF#1	INF#2	siSrc#1	siSrc#2	siDYRK1A#1	siDYRK1A#2	Gene name	log2(FC)	Adjusted p-value INF VS siDYRK1A
27	1389	22	1155	12	413	COL6A3	-0,90	9,80E-09
4657	5159	4640	4846	3202	2167	DYRK1A	-0,66	2,26E-05
239	741	195	717	110	362	DLK1	-0,66	4,84E-04
426	2455	414	2654	252	1401	APOE	-0,58	5,82E-04
7502	62842	8232	60053	5509	35472	CCND2	-0,51	1,34E-03
4942	1531	4764	1518	2967	848	LAMA4	-0,56	2,46E-03
5590	1551	4860	1545	2958	686	PLP1	-0,61	2,46E-03
1265	1310	1179	1227	910	819	NMNAT2	-0,47	3,09E-03
26	345	28	332	16	169	AFAP1L2	-0,57	5,10E-03
1393	1366	1329	1250	951	782	EDNRB	-0,50	6,80E-03
315	1016	318	1004	211	545	SEZ6L2	-0,53	6,80E-03
459	7742	398	6596	349	3192	LAMA2	-0,56	6,80E-03
2053	8115	1835	7861	1526	4884	PMP22	-0,46	7,69E-03
686	414	602	382	411	252	SCN3A	-0,52	7,69E-03
276	1134	265	1033	154	650	CALB1	-0,54	7,69E-03
310	639	301	581	233	362	KIAA1161	-0,48	8,50E-03
54	172	34	180	16	89	RGS6	-0,56	9,53E-03
1000	1737	915	1486	670	880	RAB27B	-0,53	1,11E-02
1089	1511	1071	1414	672	738	SCG2	-0,53	1,20E-02
43963	71637	48303	66226	36220	35316	CALD1	-0,48	1,53E-02
1396	762	1282	752	1837	1208	E2F2	0,43	1,55E-02
1482	2908	1668	2463	1190	1521	CUBN	-0,47	1,55E-02
1277	6338	1242	5772	940	3476	CYP1B1	-0,48	1,55E-02
690	961	689	815	478	551	TGFB2	-0,49	1,55E-02
841	1713	861	1408	639	825	PRUNE2	-0,50	1,55E-02
536	10244	505	9843	353	6872	COL3A1	-0,45	2,11E-02
3718	37279	3661	32713	2214	23454	SEMA3C	-0,48	2,11E-02
110	259	84	234	77	158	UNC13A	-0,46	2,13E-02
171	235	159	204	95	119	SRPX2	-0,52	2,13E-02
3433	2267	3134	2226	4302	3275	PTBP2	0,37	2,46E-02
392	1043	441	884	273	474	CDKN2B	-0,51	2,46E-02
547	1224	506	1121	399	856	CLCN4	-0,40	2,80E-02
906	661	1070	722	732	399	NACAD	-0,41	3,22E-02
6251	10283	6578	10241	5213	6800	PPM1F	-0,36	3,66E-02
4917	6871	4766	5949	3747	4346	ANK2	-0,42	3,66E-02
3196	2102	3335	2120	2497	1444	BMF	-0,37	3,87E-02
694	1062	728	940	559	566	ACTG2	-0,44	4,01E-02
614	595	504	509	344	358	CHL1	-0,48	4,01E-02
92	258	75	264	58	111	SLC16A4	-0,50	4,01E-02
47	544	52	576	33	315	F2RL2	-0,46	4,05E-02
1760	1196	1889	1169	1275	831	THBS2	-0,39	4,50E-02
828	800	826	801	666	547	FAM214B	-0,36	4,56E-02
46	23	70	22	100	63	EGR3	0,47	4,85E-02
699	132	628	117	489	93	DCX	-0,40	4,85E-02
947	1372	897	1384	670	917	LUM	-0,41	4,85E-02
569	680	564	712	366	491	NRXN2	-0,42	4,85E-02
348	789	292	625	282	409	FAT4	-0,44	4,85E-02
23	147	16	141	11	67	NALCN	-0,48	4,85E-02
687	1083	754	1011	545	652	TNFRSF19	-0,41	4,98E-02

Table S4

Reads normalized and divided by gene length in kb

Control individuals						Individuals with pathogenic variants in <i>DYRK1A</i>							Gene name	log2(FC)	P-value	Adjusted p-value
MLPN106	MLPN107	MLPN108	FCML47	FCML48	FCML49	MLPN112_Crd#24	MLPN113_Crd#11	MLPN114	MLPN115	FCML43_Crd#1	FCML44	FCML45_Crd#22				
1	1	0	0	0	1	81	13	42	17	3	21	8	SHISA2	3,41	1,1E-13	2,3E-09
119	2186	4857	153	1601	2104	74	23	42	113	57	28	57	OLR1	-3,41	2,3E-13	2,4E-09
61	198	299	12	567	66	2	6	10	3	1	10	2	AARD	-3,26	1,6E-11	1,1E-07
617	692	349	860	561	741	1573	1148	1564	1643	1642	1764	1446	ADAMTS6	1,25	7,3E-11	3,9E-07
276	14	48	13	5	13	1	6	2	1	0	0	1	ZNF385B	-3,11	1,3E-10	5,4E-07
0	0	0	0	0	0	3	1	5	3	8	4	1	PCDHAC2	2,96	4,0E-10	1,4E-06
53	24	64	37	50	31	5	2	16	5	5	14	6	NOX5	-2,12	1,4E-08	4,0E-05
20	63	260	34	26	96	5	3	20	7	3	0	2	ENPP4	-2,71	1,5E-08	4,0E-05
6	135	1194	16	121	151	26	9	22	15	1	0	2	TMEM52B	-2,79	1,9E-08	4,4E-05
2	0	1	11	2	5	22	14	2	4	175	48	110	CCNJL	2,56	2,6E-08	5,6E-05
21	44	79	78	80	28	261	44	527	3133	899	225	1075	SCG2	2,53	1,1E-07	2,1E-04
441	374	688	1026	1608	1247	158	157	251	250	262	372	672	HOXA11	-1,33	1,4E-07	2,1E-04
314	124	281	158	255	388	70	86	41	18	25	70	11	VSTM2L	-2,05	1,4E-07	2,1E-04
234	626	1752	208	11582	3938	46	188	14	14	359	145	303	SLC40A1	-2,55	1,4E-07	2,1E-04
21	22	64	16	52	3	1	1	4	2	5	1	2	FTCDNL1	-2,44	2,3E-07	3,2E-04
21	72	115	126	65	797	16	3	4	5	27	19	56	CYP39A1	-2,29	5,2E-07	6,5E-04
3	1	4	11	2	13	15	12	6	15	83	40	32	NUDT10	1,96	6,0E-07	7,1E-04
825	460	1953	293	163	168	37	179	27	0	36	28	6	PDE1A	-2,37	9,0E-07	1,0E-03
1228	3341	2198	1168	2882	2168	617	1001	485	1026	666	951	630	NR2F1	-1,37	1,2E-06	1,2E-03
125	103	187	127	444	277	21	44	38	50	70	16	98	TMEM229B	-1,72	2,1E-06	2,0E-03
2	12	10	15	40	31	2	0	6	1	2	2	2	BHMT	-2,17	2,6E-06	2,4E-03
113	424	12	0	1127	289	2	2	0	11	5	10	1	PITX1	-2,34	3,5E-06	2,9E-03
2	2	30	0	1	1	0	0	1	0	0	0	0	OTOA	-2,34	3,5E-06	2,9E-03
87	133	286	179	461	181	42	18	67	38	15	79	86	KLHL3	-1,75	4,6E-06	3,5E-03
103	78	106	156	378	208	44	42	21	27	134	114	92	RRAD	-1,21	4,9E-06	3,6E-03
28911	38491	19661	9721	68522	123942	3362	10208	1732	1400	12307	18423	14932	MASP1	-1,93	5,2E-06	3,6E-03
137	22	24	1	38	18	2	11	7	7	1	3	2	MRAP2	-2,17	5,9E-06	4,0E-03
12	16	42	52	35	986	2	4	7	3	19	36	30	MAP1LC3C	-2,12	6,9E-06	4,6E-03
9	13	84	35	24	29	9	5	3	10	6	3	7	NOSTRIN	-1,88	7,3E-06	4,7E-03
10	13	52	8	23	13	2	8	0	3	1	0	2	MUC19	-2,11	9,1E-06	5,6E-03
90	30	19	7	25	52	11	12	4	8	5	0	2	METTL7B	-2,00	1,1E-05	6,6E-03
226	188	933	218	1070	2091	42	210	91	26	97	233	113	MGARP	-1,92	1,3E-05	7,0E-03
612	76	412	141	30	48	1	0	9	1	3	18	0	CDH10	-2,20	1,3E-05	7,0E-03
90	122	658	157	271	202	94	61	58	59	77	73	21	ALS2CL	-1,65	1,4E-05	7,4E-03
1	1	1	28	2	2	36	7	14	29	79	59	35	PLEKHG4B	2,05	1,5E-05	7,9E-03
5	72	999	62	826	1110	22	8	4	4	163	51	18	INMT	-2,16	1,7E-05	8,4E-03
432	359	327	396	309	370	575	496	555	447	508	567	586	ING1	0,54	1,8E-05	8,6E-03
45	38	782	1245	26	74	18	6	18	7	96	4	10	DUOX1	-2,15	1,8E-05	8,6E-03
174	185	196	81	299	158	62	73	55	43	117	91	45	RAVER2	-1,27	1,9E-05	9,0E-03
352	1027	280	114	1161	791	214	195	202	35	106	56	61	ANKRD29	-1,83	2,1E-05	9,7E-03
42	41	24	18	14	15	74	47	54	56	25	36	35	HIC2	0,82	2,2E-05	1,0E-02
1308	1576	1884	1001	2130	1717	738	782	959	123	450	288	243	ATL1	-1,50	2,5E-05	1,1E-02
18	25	27	54	10	10	2	2	0	0	3	10	0	ENPP5	-2,06	2,6E-05	1,1E-02
3	2	6	5	8	11	1	0	2	0	0	0	0	APOBEC3H	-2,07	3,4E-05	1,4E-02
530	118	13263	1522	105	323	28	43	334	236	22	198	55	SCIN	-2,08	3,5E-05	1,4E-02
1360	3680	8173	736	9163	2737	1307	1466	697	411	890	209	928	FRY	-1,80	3,6E-05	1,4E-02
180	310	540	282	623	357	101	153	191	68	201	126	54	EPB41	-1,37	4,6E-05	1,8E-02
281	483	1092	110	9888	1051	27	75	86	243	15	11	8	SYPL2	-2,05	4,7E-05	1,8E-02
0	0	0	0	4	3	2	3	16	72	4	5	69	CACNG8	2,04	4,9E-05	1,8E-02
0	0	1	1	0	1	5	3	2	1	26	0	4	PLCH2	2,03	5,0E-05	1,8E-02
7	12	4	9	3	5	216	2	20	11	1110	231	20	BACH2	2,04	5,2E-05	1,9E-02
23	76	19	27	130	96	4	4	2	2	33	30	55	CCDC170	-1,74	5,4E-05	1,9E-02
1	3	1	3	1	0	4	5	5	40	24	3	19	PAK3	1,98	5,8E-05	2,1E-02
122	715	1554	1393	1898	24828	24	30	37	359	32	841	638	RGCC	-2,00	5,9E-05	2,1E-02
4	14	61	47	72	9	1	9	0	4	8	1	6	PIFO	-1,92	6,3E-05	2,2E-02
177	182	431	594	156	138	68	43	141	3	11	36	43	PLAC8	-1,83	6,6E-05	2,2E-02
221	223	291	285	356	332	141	163	129	147	250	239	242	C5orf63	-0,57	7,2E-05	2,3E-02
72	178	30	50	73	403	7	5	17	5	3	28	81	SLC14A1	-1,88	7,2E-05	2,3E-02
15	6	27	6	16	21	5	0	0	0	4	0	1	ENAM	-1,96	7,4E-05	2,4E-02
312	264	860	278	371	314	169	188	248	202	133	111	210	ARSG	-1,08	8,1E-05	2,5E-02
71	96	238	202	124	69	14	19	26	4	63	78	64	CCL28	-1,61	8,1E-05	2,5E-02
7834	29676	52456	14856	35653	31311	11421	8577	8880	9079	9602	13805	11626	GABARAPL1	-1,29	8,2E-05	2,5E-02
8	4	3	4	3	0	14	16	23	31	9	5	15	SLC05A1	1,65	8,4E-05	2,6E-02
39	28	29	61	36	27	0	4	0	2	18	26	4	FAM84A	-1,83	8,7E-05	2,6E-02
76	78	97	62	81	119	26	24	51	30	61	48	30	CCDC13	-1,06	9,7E-05	2,8E-02
1	0	0	5	0	4	1	1	2	78	12	310	258	CYP24A1	1,95	1,0E-04	2,9E-02
7	4	35	26	88	18	5	7	2	2	10	7	4	CNNM1	-1,71	1,1E-04	3,0E-02
9297	10393	13094	11208	15435	11356	7572	7945	7595	9298	7392	7417	8692	ANXA4	-0,54	1,1E-04	3,2E-02
158	42	25	36	47	83	2659	96	1054	1439	279	183	22	DUSP2	1,83	1,3E-04	3,5E-02
1	3	20	57	4	3	2	2	2	0	2	3	1	DUOX2	-1,85	1,3E-04	3,6E-02
367	2039	974	1060	3866	934	23	226	47	14	797	75	792	DAPK1	-1,82	1,4E-04	3,8E-02
36	16	40	35	2	20	16	2	5	7	1	4	1	GNG4	-1,74	1,6E-04	4,1E-02
15566	15086	12895	11849	20938	20372	4588	6007	4707	11562	5543	7502	12658	CSF1	-1,00	1,7E-04	4,2E-02
7753	8633	11097	16718	13598	13314	7278	6842	6335	6980	8558	10943	10171	BLCAP	-0,48	1,7E-04	4,3E-02
50	3	74	18	226	19	2	0	0	5	21	17	14	FYB	-1,85	1,8E-04	4,4E-02
269	430	1587	1656	1175	1112	511	266	281	111	309	464	171	MAP6	-1,41	1,8E-04	4,4E-02
1	2	3	2	5	6	0	0	1	0	1	0	0	SEC14L6	-1,86	1,9E-04	4,4E-02
108	90	151	62	157	70	29	67	61	59	29	35	57	TNFSF13	-1,08	1,9E-04	4,4E-02
590	219	1288	497	644	2279	144	116	129	43	179	702	550	CDA	-1,56	1,9E-04	4,4E-02
5614	539	6024	5695	2745	482	115	136	182	540	1158	2018	403	SERPINB7	-1,77	2,0E-04	4,4E-02
435	208	219	367	493	558	988	341	1074	2189	1106	3281	312	IL11	1,49	2,0E-04	4,5E-02
78	101	33	70	50	38	148	109	346	261	170	84	81	GAREM	1,20	2,0E-04	4,5E-02
360	378	465	983	1000	1492	308	309	252	236	400	673	686	FBXO6	-0,73	2,0E-04	4,5E-02
353	91	154	104	170	205	2	39	19	14	65	68	123	IGJ	-1,61	2,1E-04	4,6E-02
67	133	1107	169	1692	1191	11	207	110	74	11	113	19	TMEM130	-1,82	2,2E-04	4,6E-02
65	104	5														

**Table S5**

genes	intron	Ajusted p-val logef	DYRK1A	other	deltapsi	description
CDKN1A	chr6:36644347:36645602:clu_14171	0,03	-0,23	0,008	0,01	0,00 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36644347:36645565:clu_14171	0,03	-0,13	0,007	0,01	0,00 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36645990:36651874:clu_14171	0,03	0,44	0,001	0,00	0,00 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36644301:36651874:clu_14171	0,03	-0,14	0,025	0,02	-0,01 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36644347:36645679:clu_14171	0,03	-0,23	0,004	0,00	0,00 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36644347:36645467:clu_14171	0,03	-0,21	0,002	0,00	0,00 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36644347:36651874:clu_14171	0,03	-0,15	0,121	0,09	-0,03 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36644347:36645599:clu_14171	0,03	0,30	0,001	0,00	0,00 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36646716:36651874:clu_14171	0,03	0,36	0,005	0,01	0,00 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36645710:36651874:clu_14171	0,03	-0,25	0,025	0,02	-0,01 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1A	chr6:36646575:36651874:clu_14171	0,03	0,23	0,802	0,85	0,05 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
IFT140	chr16:1657298:1659402:clu_6597	0,03	12,35	0,000	0,20	0,20 intraflagellar transport 140 homolog (Chlamydomonas)
IFT140	chr16:1657298:1660665:clu_6597	0,03	-12,35	1,000	0,80	-0,20 intraflagellar transport 140 homolog (Chlamydomonas)

**Table S6**

<b>Cell line</b>	<b>Treatment</b>	<b>Total cells counted</b>	<b>total positive cell counted</b>	<b><u>Positive %</u></b>
hNSC-1_#1	Interferin	1722	39	2,25
	siScramble	1429	28	1,98
	siDYRK1A	1456	16	1,08
hNSC-1_#2	Interferin	1837	62	3,36
	siScramble	1881	75	4,01
	siDYRK1A	1438	12	0,81
hNSC-2_#1	Interferin	2822	343	12,14
	siScramble	2778	178	6,42
	siDYRK1A	2838	140	4,92
hNSC-2_#2	Interferin	3254	169	5,19
	siScramble	3223	176	5,45
	siDYRK1A	3414	168	4,91

Table S7

HPO_TERM_ID	HPO_TERM_NAME	DYRK1A	COL6A3	DLK1	APOE	CCND2	LAMA4	PLP1	NMNAT2	AFAP1L2	LAMA2	SEZ6L2	EDNRB	PMP22	CALB1	SCN3A	KIAA1161/MYORG	RG56
HP:0001249	Intellectual disability	Y	0 NA	0	0	0	0	1 NA	NA	NA	1 NA	1	0 NA	1	0 NA	1	0 NA	0 NA
HP:0001263	Global developmental delay	Y	0 NA	0	0	0	0	1 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001250	Seizure	Y	0 NA	0	0	0	0	1 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001344	Absent speech	Y	0 NA	0	0	0	0	1 NA	NA	NA	0 NA	0	0 NA	0	0 NA	1	0 NA	0 NA
HP:0002650	Scoliosis	Y	1 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000411	Protruding ear	Y	1 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000486	Strabismus	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000490	Deeply set eye	Y	0 NA	0	0	0	0	1 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001270	Motor delay	Y	1 NA	0	0	0	0	0 NA	NA	NA	1 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002079	Hypoplasia of the corpus callosum	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000750	Delayed speech and language development	Y	0 NA	0	0	0	0	1 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0007957	Corneal opacity	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	1 NA
HP:0000054	Micropenis	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000047	Hypospadias	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001511	Intrauterine growth retardation	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001999	Abnormal facial shape	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000426	Prominent nasal bridge	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	1	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002119	Ventriculomegaly	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0007018	Attention deficit hyperactivity disorder	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000767	Pectus excavatum	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001643	Patent ductus arteriosus	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001659	Aortic regurgitation	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001629	Ventricular septal defect	Y	0 NA	0	1	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0006466	Ankle flexion contracture	Y	1 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002020	Gastroesophageal reflux	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000540	Hypermetropia	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000126	Hydronephrosis	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001508	Failure to thrive	Y	1 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001518	Small for gestational age	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000341	Narrow forehead	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002120	Cerebral cortical atrophy	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002363	Abnormal brainstem morphology	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000708	Behavioral abnormality	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001561	Polyhydramnios	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002808	Kyphosis	Y	1 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001650	Aortic valve stenosis	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0012785	Flexion contracture of finger	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002013	Vomiting	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002247	Duodenal atresia	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0011968	Feeding difficulties	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0011757	Posterior pituitary hypoplasia	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000646	Amblyopia	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000483	Astigmatism	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000504	Abnormality of vision	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000577	Exotropia	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000541	Retinal detachment	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000543	Optic disc pallor	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000545	Myopia	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000028	Cryptorchidism	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000122	Unilateral renal agenesis	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000125	Pelvic kidney	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000107	Renal cyst	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0003561	Birth length less than 3rd percentile	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002719	Recurrent infections	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001182	Tapered finger	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001773	Short foot	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001831	Short toe	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001288	Gait disturbance	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0410263	Brain imaging abnormality	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0002269	Abnormality of neuronal migration	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000739	Anxiety	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000733	Stereotypy	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000729	Autistic behavior	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0000252	Microcephaly	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0011171	Simple febrile seizure	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0001562	Oligohydramnios	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0003319	Abnormality of the cervical spine	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA
HP:0009121	Abnormal axial skeleton morphology	Y	0 NA	0	0	0	0	0 NA	NA	NA	0 NA	0	0 NA	0	0 NA	0	0 NA	0 NA

**Table S7**

HPO_TERM_ID	HPO_TERM_NAME	RAB27B	SCG2	CALD1	E2F2	PRUNE2	TGFB2	CYP1B1	CUBN	COL3A1	SEMA3C	UNC13A	SRPX2	PTBP2	CDKN2B	CLCN4	NACAD	ANK2
HP:0001249	Intellectual disability	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	1	NA	NA	0	NA	0
HP:0001263	Global developmental delay	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	1	NA	NA	1	NA	0
HP:0001250	Seizure	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0001344	Absent speech	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	1	NA	0
HP:0002650	Scoliosis	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0000411	Protruding ear	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0000486	Strabismus	NA	NA	NA	NA	NA	0	1	0	0	NA	NA	0	NA	NA	1	NA	0
HP:0000490	Deeply set eye	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0001270	Motor delay	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0002079	Hypoplasia of the corpus callosum	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	1	NA	0
HP:0000750	Delayed speech and language development	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0007957	Corneal opacity	NA	NA	NA	NA	NA	0	1	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000054	Micropenis	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000047	Hypospadias	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0001511	Intrauterine growth retardation	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	1	NA	NA	0	NA	0
HP:0001999	Abnormal facial shape	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000426	Prominent nasal bridge	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0002119	Ventriculomegaly	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0007018	Attention deficit hyperactivity disorder	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	1	NA	NA	1	NA	0
HP:0000767	Pectus excavatum	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	1	NA	NA	0	NA	0
HP:0001643	Patent ductus arteriosus	NA	NA	NA	NA	NA	1	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001659	Aortic regurgitation	NA	NA	NA	NA	NA	1	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001629	Ventricular septal defect	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0006466	Ankle flexion contracture	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0002020	Gastroesophageal reflux	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000540	Hypermetropia	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0000126	Hydronephrosis	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001508	Failure to thrive	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001518	Small for gestational age	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	1	NA	0
HP:0000341	Narrow forehead	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0002120	Cerebral cortical atrophy	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	1	NA	0
HP:0002363	Abnormal brainstem morphology	NA	NA	NA	NA	NA	0	0	0	1	NA	NA	0	NA	NA	0	NA	0
HP:0000708	Behavioral abnormality	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	1	NA	0
HP:0001561	Polyhydramnios	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0002808	Kyphosis	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001650	Aortic valve stenosis	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0012785	Flexion contracture of finger	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0002013	Vomiting	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0002247	Duodenal atresia	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0011968	Feeding difficulties	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0011757	Posterior pituitary hypoplasia	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000646	Amblyopia	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000483	Astigmatism	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000504	Abnormality of vision	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000577	Exotropia	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000541	Retinal detachment	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000543	Optic disc pallor	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000545	Myopia	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000028	Cryptorchidism	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000122	Unilateral renal agenesis	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000125	Pelvic kidney	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000107	Renal cyst	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0003561	Birth length less than 3rd percentile	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0002719	Recurrent infections	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001182	Tapered finger	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001773	Short foot	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001831	Short toe	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001288	Gait disturbance	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0410263	Brain imaging abnormality	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0002269	Abnormality of neuronal migration	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000739	Anxiety	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000733	Stereotypy	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000729	Autistic behavior	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0000252	Microcephaly	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0011171	Simple febrile seizure	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0001562	Oligohydramnios	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0003319	Abnormality of the cervical spine	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0
HP:0009121	Abnormal axial skeleton morphology	NA	NA	NA	NA	NA	0	0	0	0	NA	NA	0	NA	NA	0	NA	0

**Table S7**

HPO_TERM_ID	HPO_TERM_NAME	PPM1F	BMF	ACTG2	SLC16A4	CHL1	F2RL2	THBS2	FAM214B	EGR3	NRXN2	NALCN	DCX	LUM	FAT4	TNFRSF19	CDKN1A	
HP:0001249	Intellectual disability	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	1 NA	NA	NA	7
HP:0001263	Global developmental delay	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	1 NA	NA	NA	6
HP:0001250	Seizure	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	1 NA	NA	NA	4
HP:0001344	Absent speech	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	0 NA	NA	NA	4
HP:0002650	Scoliosis	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	1 NA	NA	NA	4
HP:0000411	Protruding ear	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	0 NA	NA	NA	3
HP:0000486	Strabismus	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	0 NA	NA	NA	3
HP:0000490	Deeply set eye	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	0 NA	NA	NA	3
HP:0001270	Motor delay	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	3
HP:0002079	Hypoplasia of the corpus callosum	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	1 NA	NA	NA	3
HP:0000750	Delayed speech and language development	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	0 NA	NA	NA	3
HP:0007957	Corneal opacity	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	2
HP:0000054	Micropenis	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	1 NA	0 NA	1 NA	NA	NA	2
HP:0000047	Hypospadias	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	1 NA	NA	NA	2
HP:0001511	Intrauterine growth retardation	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	0 NA	NA	NA	2
HP:0001999	Abnormal facial shape	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	1 NA	NA	NA	2
HP:0000426	Prominent nasal bridge	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	0 NA	NA	NA	2
HP:0002119	Ventriculomegaly	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	1 NA	NA	NA	2
HP:0007018	Attention deficit hyperactivity disorder	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	2
HP:0000767	Pectus excavatum	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	2
HP:0001643	Patent ductus arteriosus	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0001659	Aortic regurgitation	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0001629	Ventricular septal defect	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0006466	Ankle flexion contracture	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0002020	Gastroesophageal reflux	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	0 NA	NA	NA	
HP:0000540	Hypermetropia	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0000126	Hydronephrosis	NA	NA	1 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0001508	Failure to thrive	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0001518	Small for gestational age	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0000341	Narrow forehead	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	1 NA	NA	NA	
HP:0002120	Cerebral cortical atrophy	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0002363	Abnormal brainstem morphology	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0000708	Behavioral abnormality	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0001561	Polyhydramnios	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	1	0 NA	0 NA	0 NA	NA	NA	
HP:0002808	Kyphosis	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0001650	Aortic valve stenosis	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0012785	Flexion contracture of finger	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0002013	Vomiting	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	
HP:0002247	Duodenal atresia	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0011968	Feeding difficulties	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0011757	Posterior pituitary hypoplasia	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000646	Amblyopia	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000483	Astigmatism	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000504	Abnormality of vision	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000577	Exotropia	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000541	Retinal detachment	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000543	Optic disc pallor	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000545	Myopia	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000028	Cryptorchidism	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000122	Unilateral renal agenesis	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000125	Pelvic kidney	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000107	Renal cyst	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0003561	Birth length less than 3rd percentile	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0002719	Recurrent infections	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0001182	Tapered finger	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0001773	Short foot	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0001831	Short toe	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0001288	Gait disturbance	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0410263	Brain imaging abnormality	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0002269	Abnormality of neuronal migration	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000739	Anxiety	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000733	Stereotypy	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000729	Autistic behavior	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0000252	Microcephaly	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0011171	Simple febrile seizure	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0001562	Oligohydramnios	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0003319	Abnormality of the cervical spine	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0
HP:0009121	Abnormal axial skeleton morphology	NA	NA	0 NA	NA	NA	NA	NA	NA	NA	NA	0	0 NA	0 NA	0 NA	NA	NA	0

# Discussion de l'article 2

Le syndrome DYRK1A est une des formes de DI monogénique les plus fréquentes mais les mécanismes physiopathologiques qui lui sont associés restent mal décrits. Les rôles, interacteurs et cibles de DYRK1A sont nombreux. En effet DYRK1A est impliqué dans la régulation de l'expression des gènes (complexes de remodelage, épigénétique, transcription et épissage) mais régule également le cycle cellulaire, interagit avec plusieurs protéines du cytosquelette, module la plasticité synaptique et le métabolisme cellulaire (Duchon et al., 2016, Review ; Pathak et al., 2018, Review ; Arbones et al., 2019, Review) Nous avons lors de ce travail étudié l'interactome de DYRK1A dans un modèle de cellules souches neurales (hNSC) et étudié les conséquences de son inactivation sur la régulation de l'expression génique. Nous avons également pu mettre en évidence une baisse de prolifération des hNSC partiellement inactivées pour DYRK1A ainsi qu'une mort par apoptose d'une partie de ces cellules.

## **Caractérisation de l'interactome de DYRK1A dans le modèle hNSC**

A l'heure actuelle, 409 interacteurs de DYRK1A sont référencés sur la base de données BioGRID et plusieurs analyses protéomiques ont déjà pu être réalisées dans le but d'étudier l'interactome de cette kinase. Cependant, l'ensemble des études réalisées dans des modèles humains concernent des lignées cellulaires peu pertinentes pour l'étude du rôle de DYRK1A dans le neurodéveloppement (HEK293, HeLa, T98G ou lignées de glioblastomes). C'est pourquoi nous avons choisi d'étudier l'interactome de DYRK1A dans une lignée hNSCs. Parmi les interacteurs, 14 sont déjà décrits dans la littérature, confirmant la robustesse de notre expérience et validant leur interaction avec DYRK1A dans le modèle hNSC. De plus, cinq d'entre elles (ANAPC3-4-6-7 et 8) font partie de l'Anaphase Promoting Complex (APC), un complexe d'ubiquitination caractérisé initialement comme régulant l'entrée des cellules en anaphase mais sont également impliquées dans d'autres mécanismes tels que la réparation de l'ADN (Pal et Summers, 2018, Review), la croissance cellulaire (Skaar et al., 2009), la croissance axonale (Kannan et al., 2012) et dendritique (Yang et al., 2009) ou encore la mort des neurones (Maestre et al., 2008). Récemment, une équipe a également pu montrer l'interaction de DYRK1A avec trois des cinq protéines identifiées dans ce travail (ANAPC3-4 et 8) dans un modèle de HEK293 (Guard et al., 2019). De manière intéressante, quatre des cinq protéines que nous avons pu identifier composent le « lobe TPR » de l'APC, une sous unité impliquée dans le repliement et la stabilité du complexe et permettant sa liaison aux substrats



et à des sous unités régulatrices (Barford, 2011). Le lobe TPR permet notamment l'interaction de l'APC avec différents facteurs, comme les protéines CDC20 et CDH1 (Vodermaier et al., 2003). L'interaction de l'APC avec CDC20 ou CDH1 module son activité ainsi que la reconnaissance de ses substrats, lui conférant des rôles différents en fonction de ses partenaires. En effet, le complexe APC<sup>CDC20</sup> serait actif durant la phase M tandis que le complexe APC<sup>CDH1</sup> serait formé durant la phase G1 et orienterait le devenir cellulaire (Han et al., 2019), toujours par phosphorylation du TPR lobe. Ainsi, il est possible que DYRK1A puisse réguler l'association d'APC avec CDC20 ou CDH1, modulant de ce fait son activité. Même si l'analyse par spectrométrie de masse n'a pas détecté de phosphorylation des protéines de l'APC interagissant avec DYRK1A, il est également possible que DYRK1A module l'activité du complexe indépendamment de son activité kinase. Les cibles ubiquitinées par l'APC sont nombreuses (Huang et Bonni, 2016, Review) et des expériences préliminaires de Western Blot réalisées dans les hNSCs partiellement inactivées pour DYRK1A n'ont pas montré de variation d'expression pour deux des cibles de l'APC : PLK1 (Skaar et al., 2009) et FMRP (Huang et Bonni, 2016, Review). Il serait cependant intéressant d'investiguer le rôle régulateur potentiel de DYRK1A sur l'APC, en étudiant l'interaction de l'APC avec ces facteurs et la balance APC<sup>CDC20</sup>/APC<sup>CDH1</sup> dans le cas d'une inactivation de DYRK1A en hNSCs synchronisées. L'analyse de phosphorylation réalisée sur ces données a comme attendu mis en évidence la présence d'un phosphate sur le résidu Tyrosine 321 de DYRK1A. De plus, nous avons pu détecter trois interacteurs présentant une phosphorylation : FAM117B, FAM117C (GLCC1) et FAM53C (C5Orf6). Le rôle de ces protéines reste à l'heure actuelle inconnu et nous ne pouvons affirmer que ces trois facteurs sont phosphorylés par DYRK1A. En effet, les régions peptiques encadrant les acides aminés phosphorylés ne comportent pas la séquence consensus reconnue par DYRK1A (R-P-x(1,3)S/T-P). Enfin, nous avons également pu mettre en évidence l'interaction de DYRK1A avec ZNF313 (RNF114), une E3 ubiquitine ligase impliquée dans la régulation du cycle cellulaire et la répression de la sénescence, régulant entre autres la stabilité de la protéine P21. De manière intéressante, si une augmentation d'une isoforme de CDKN1A, qui code la protéine p21, a été mise en évidence après une inactivation de *DYRK1A*, une diminution du taux de protéine p21/CDKN1A a au contraire été observée. Au vu de la connexion entre DYRK1A et P21 mise en évidence durant ce travail, il sera intéressant de valider dans un premier temps l'interaction avec la protéine ZNF313 et d'essayer de comprendre si ZNF313 est impliquée dans la régulation du taux de protéine p21 par DYRK1A. Nous pourrions donc investiguer si la diminution du taux de protéines P21 causée par l'inactivation de DYRK1A pourrait être reversée par une inhibition de ZNF313.

## ***DYRK1A* et prolifération cellulaire dans les hNSCs**

Depuis la découverte de son homologue *minibrain* chez la Drosophile, l'impact de la sur ou sous expression de *DYRK1A* sur la prolifération cellulaire fut largement décrit, et ce dans différents modèles animaux et cellulaires. Il est aujourd'hui clair que l'action de *DYRK1A* sur le cycle cellulaire dépend de plusieurs facteurs, entre autres notamment du type cellulaire et de la fenêtre développementale. Nous avons donc souhaité décrire l'impact d'une inactivation partielle de *DYRK1A* sur la prolifération des hNSCs. Les résultats obtenus indiquent une diminution significative de la capacité de prolifération cellulaire 3 jours après inactivation de *DYRK1A*. Ces résultats sont en accord avec plusieurs études ayant rapporté une diminution de prolifération et une réduction du nombre de neurones chez la drosophile et dans le modèle murin (Tejedor et al., 1995 ; Barallobre et al., 2014). Ces observations sont également en accord avec le phénotype de microcéphalie observé chez les patients et l'ensemble des modèles animaux étudiés y compris le poisson zèbre (Kim et al., 2017) et le xénope (Willsey et al., 2020), défaut de développement dont l'origine est dans la plupart des syndromes due à des défauts de neurogénèse et/ou de division cellulaire (Fernandez et al., 2016). Cependant, la diminution de prolifération reste une observation globale et plusieurs processus peuvent en être à l'origine (mort cellulaire, blocage ou ralentissement du cycle, sénescence, différenciation, dommages ADN, etc ...). Au vu de l'implication de *DYRK1A* dans la régulation des processus apoptotiques (Seifert et al. 2008 ; Guo et al., 2010), nous avons succinctement évalué le phénotype d'apoptose dans les hNSCs inactivées pour *DYRK1A* en comptant le nombre de cellules positives au marquage Caspase-3 clivée et en quantifiant la quantité de protéine Caspase-3 clivée présente dans le milieu de culture. Nous avons ainsi pu mettre en évidence une mort par apoptose significativement plus importante des cellules inactivées pour *DYRK1A*. Cette augmentation de l'apoptose avait également été observée dans différents modèles comme par exemple *in vivo* chez la souris (Laguna et al., 2008, Barallobre et al., 2014) ou également sur cellules en culture (Guo et al., 2010). Il reste important de rappeler que le modèle hNSC ne reflète pas l'environnement 3D et physiologique présent lors du développement du cerveau. Ainsi, l'observation de ce phénotype pourrait par exemple être due à une délamination des cellules en culture, entraînant leur mort et il est possible que les lignées NSC humaines *in vivo* stoppent leur division ou entrent en quiescence. De ce fait, il serait intéressant d'étudier les effets d'une inactivation partielle de *DYRK1A* dans un modèle 3D tel que celui des cultures organoïdes, comportant les différents types cellulaires présents lors du développement cérébral. De plus, l'inactivation de *DYRK1A* par la méthode d'interférence ARN conduit à une

inactivation hétérogène du gène au sein de la culture. Comme décrit dans l'introduction, nous savons que *DYRK1A* est un gène « sensible au dosage » et que son knock-out à l'état homozygote est léthal durant l'embryogénèse. Ainsi, il est possible que certains phénotypes, notamment l'apoptose, soient associés à une diminution trop forte de *DYRK1A* dans certaines cellules. De plus, la diminution d'expression de *DYRK1A* est également décrite comme favorisant le processus d'entrée des NSCs en différenciation neuronale (Yabut et al., 2010 ; Chen et al., 2021), dérégulant la balance synapse excitatrice/inhibitrice (Arranz et al., 2019) et diminuant la complexité des réseaux de neurones (Dang et al., 2018). Nous avons donc lors d'expériences préliminaires investigué les conséquences de l'inactivation de *DYRK1A* sur l'état de différenciation de la lignée SA001 en mesurant l'expression de gènes spécifiques des précurseurs en division comme *SOX2* ainsi que de gènes spécifiques d'une différenciation précoce (*TBR1* et *TBR2*) ou tardive (*TUBB3*) des neurones corticaux. Nous n'avons cependant pas observé de différence significative entre les cellules non traitées ou partiellement inactivées pour *DYRK1A* jusqu'à 6 jours après transfection. Nous prévoyons donc d'investiguer les conséquences de l'inactivation de *DYRK1A* sur la différenciation grâce à un modèle de hNSCs stable portant une variation perte de fonction à l'état hétérozygote suite à une édition du génome. Cette lignée permettra d'évaluer les conséquences de la perte de fonction du gène dans des conditions d'inactivation homogènes mais également sur un temps de différenciation plus long. Enfin, plusieurs études rapportent le lien entre *DYRK1A* et des cascades de signalisation impliquées dans les troubles du neurodéveloppement et la prolifération cellulaire. En effet, la surexpression de *DYRK1A* fut reliée à une suractivation de la voie ERK/AKT dans des souris CBS (Cystathionine beta syntase) (Abekhoukh et al., 2013). On sait également que la voie PI3K/AKT/mTOR est aussi suractivée dans des extraits de cortex frontaux de patients DS (Perluigi 2014). Nous avons donc testé par western blot et dans les deux lignées de hNSC, l'état de phosphorylation d'acteurs (mTOR, PKC, AKT et ERK) de ces cascades de signalisation. Nous n'avons noté aucune différence concernant les niveaux de phosphorylation des protéines mTOR, PKC et AKT dans les hNSCs traitées par siRNA *DYRK1A*. En revanche, nous avons pu mettre en évidence une diminution de la phosphorylation des protéines ERK1/2 associée à une inactivation de *DYRK1A*. Les protéines ERK font parties de la famille des MAP kinases et de la voie Ras-Raf-MEK-ERK, elles sont généralement décrites comme des « mitogen factors » conduisant à la transcription de facteurs prolifératifs (Roskoski et al., 2012). Les rôles de ERK1/2 et des MAP kinases restent cependant très nombreux et de plus amples investigations sont nécessaires pour comprendre les conséquences de cette diminution de phosphorylation. On sait cependant que la phosphorylation de ERK1/2 peut conduire à l'activation du facteur S6K

(Anjum et Bleis, 2008, Review) et une étude récente a pu mettre en évidence une diminution de la phosphorylation de S6 et S6K dans des cortex de souris inactivés pour *DYRK1A* (Levy et al., 2021). C'est pourquoi nous souhaitons poursuivre nos investigations autour de la cascade de signalisation ERK1/2, en commençant par valider la diminution du taux de phosphorylation de S6 et S6K. Il est également intéressant de noter que les auteurs de cette étude rapportent un phénotype de microcéphalie sans diminution du nombre de neurones et mettent en évidence une répartition plus dense des cellules ainsi qu'une diminution de la taille de leur soma. L'ensemble de ces investigations permettra de mieux comprendre le rôle de *DYRK1A* dans le développement du cerveau chez l'Homme, grâce au modèle hNSC.

### **Conséquences de la régulation de P21 (CDKN1A) par *DYRK1A***

La protéine P21 (CDKN1A) est un acteur majeur du cycle cellulaire aux rôles multiples. P21 est largement décrite comme un inhibiteur des CDK2 et CDK4 régulant la transition G1-S et est également un des facteurs clef d'entrée des cellules en sénescence. A l'heure actuelle une étude fait état d'un lien potentiel entre *DYRK1A* et P21. Cet article rapporte que l'inhibition pharmacologique de *DYRK1A* peut conduire à une diminution de la prolifération cellulaire due à une augmentation de l'expression de la Cycline D1 et du facteur P21 (Chen et al., 2013). Lors de ce travail, nous avons dans un premier temps identifié l'augmentation d'expression d'une isoforme codée par le gène *CDKN1A* suite à l'inactivation de *DYRK1A* et ce, dans les deux lignées de hNSCs. Le rôle de cette isoforme ainsi que de la séquence nucléotidique additionnelle sont à l'heure actuelle inconnus.

Il serait donc intéressant de cloner cette isoforme et d'étudier les conséquences de sa surexpression en culture cellulaire. Au vu du rôle de *CDKN1A* dans la régulation du cycle cellulaire, nous avons souhaité investiguer son lien potentiel avec *DYRK1A* dans les hNSCs. Ainsi, nous avons également pu mettre en évidence une diminution du niveau protéique de CDKN1A\_S dans les deux lignées de hNSCs, 48 heures après inactivation de *DYRK1A*. Nos résultats semblent indiquer que l'inactivation de *DYRK1A* conduit à une diminution d'expression protéique de CDKN1A, contrairement aux résultats obtenus par Chen et al.. Cependant, Chen et al. rapporte une augmentation de CDKN1A après traitement des cellules par l'harmine (un inhibiteur de *DYRK1A*) pendant 4 heures. Outre la possibilité d'une inactivation d'autres kinases par l'harmine, ces résultats sont obtenus seulement quatre heures après inactivation de *DYRK1A*. Ainsi, il est possible que le niveau de P21 augmente suite à l'inhibition de *DYRK1A* afin d'exercer son rôle d'inhibiteur du cycle cellulaire et que la protéine voit son niveau diminuer après l'arrêt définitif de la capacité de la cellule à se diviser

ou son entrée en sénescence. En effet, P21 étant un marqueur établi de l'entrée en sénescence (Stein et al., 1999), nous avons évalué en collaboration avec l'équipe du Docteur Bill Keyes à l'IGBMC, le nombre de cellules entrant en sénescence 2, 4 et 6 jours après inactivation de *DYRK1A* dans les deux lignées hNSCs. Étonnamment, aucune différence significative ne fut observée dans les deux lignées de hNSCs. De plus, nous n'avons pas non plus observé de variations significatives au niveau de l'expression de l'ARNm (à 48 heures post transfection) codant pour d'autres protéines indiquant une entrée en sénescence (*CDKN2A*, *CDKN2B*, *IL6*, *IL1A*, *IL1B*, *IL8*, *CSF1*, *CCL2* ...). Il reste important de noter que la vitesse d'induction du processus de sénescence dépend du traitement appliqué mais également du type cellulaire et peut par exemple varier dans les NSCs murines de quelques dizaines d'heures à plus d'une dizaine de jours (Wang et al., 2014). Ainsi, il est possible que nous n'ayons pu détecter le lien entre *DYRK1A* et le processus de sénescence dans notre modèle. Afin de confirmer ces résultats, il serait pertinent de réitérer ce protocole sur une durée plus longue et dans un modèle de hNSCs exprimant une variation perte de fonction stable. Il reste cependant important de noter qu'une étude récente a souligné que l'activité  $\beta$ -GAL utilisée pour le marquage des cellules en sénescence pouvait également être augmentée dans les neurones sains (Mera-Rodriguez et al., 2021). Les auteurs rapportent également une expression de P21 augmentée dans les neurones post mitotiques précoces, cette expression étant à priori reliée à la sortie irréversible de ces neurones du cycle cellulaire.

De plus, l'entrée en sénescence n'est pas la seule activité de P21. Comme rappelé précédemment, P21 est impliquée dans la régulation de l'apoptose (Gartel et al., 2002), la réparation de l'ADN (Mauro et al., 2012), le cycle cellulaire (Romanov et Rudolph, 2016) ou la différenciation cellulaire (Ghanem et al., 2006). Ainsi, nous souhaitons également investiguer les variations d'expression de P21 lors de la différenciation des hNSCs en neurones en conditions contrôles et lors d'une inactivation de *DYRK1A*.

## Conclusion

Ce travail nous a permis de mieux comprendre le rôle de *DYRK1A* dans la régulation de l'expression des gènes dans un modèle de précurseurs neuronaux humains. Nous avons également pu investiguer les conséquences d'une inactivation de *DYRK1A* sur la prolifération de ces cellules et avons pu mettre en évidence une apoptose significativement plus élevée dans le cas d'une diminution d'expression de *DYRK1A*. Nous avons également souligné l'effet d'une perte de *DYRK1A* sur l'augmentation d'une isoforme du gène *CDKN1A* au niveau ARNm mais

également sa diminution au niveau protéique. Malgré une recherche active et une littérature très riche, les mécanismes moléculaires responsables de la physiopathologie du syndrome DYRK1A et plus généralement les rôles multiples de cette kinase, restent aujourd'hui mal compris. Ainsi, ces investigations du rôle de DYRK1A dans le modèle hNSC permettront de mieux comprendre le rôle de cette protéine dans ce modèle cellulaire particulier et de ce fait, dans le développement du cerveau. A terme, la meilleure compréhension des mécanismes physiopathologiques pourra permettre une meilleure prise en charge de ces patients.



**Article 3 : Molecular consequences of  
variants in *PQBP1*, involved in the  
X-linked Renpenning syndrome**

---





## Résumé

Le syndrome de Renpenning est une rare forme de DI syndromique liée à l’X causée par des variations entraînant une perte de fonction du gène *PQBPI*. Ce gène code pour la Polyglutamine Binding Protein 1, une protéine notamment impliquée dans la régulation de la transcription et de l’épissage alternatif. L’objectif de ce travail était d’une part, de mieux comprendre les rôles transcriptionnel et d’épissage alternatif de PQBP1 dans des lignées LCL et dans des lignées de hNSCs, mais également de mettre au point, basé sur ces connaissances, une signature transcriptomique spécifique d’altérations du gène *PQBPI* permettant d’évaluer le caractère bénin ou pathogène de n’importe quel VSI identifié dans *PQBPI*. Pour se faire nous avons réalisé une première analyse transcriptomique dans 5 lignées LCL porteuses de variants pathogènes dans le gène *PQBPI*. Dans un second temps, nous avons pu réaliser une deuxième analyse transcriptomique dans une lignée de hNSCs traités par siRNA dirigés contre *PQBPI*. Ces expériences m’ont entre autres permis de mettre en évidence une augmentation d’expression d’une isoforme particulière du gène *UPF3B* (*UPF3B\_S*), plus courte que les deux isoformes de référence (*UPF3B\_L*) dans les lignées LCL issues de patients PQBP1 mais également dans une première lignée hNSCs partiellement inactivée. J’ai par la suite pu valider cette augmentation spécifique d’*UPF3B\_S* dans une seconde série de LCL (comportant contrôles, variants pathogènes ou VSI dans le gène *PQBPI* et patients atteints du syndrome de l’X fragile) ainsi que dans différents modèles cellulaires (une seconde lignée de hNSCs, HeLa, HEK293, SHSY-5Y) traités par siRNA *PQBPI*. Ces résultats m’ont permis de conclure que cette augmentation était non seulement robuste car retrouvée dans tous les types cellulaires testés mais semblait également spécifique d’altération de *PQBPI* car absente chez les patients FXS. Par la suite, Camille Engel et Nathalie Drouot ont également pu montrer qu’*UPF3B\_S* était augmenté dans des extraits d’ARN issus de sang de patients Renpenning. Enfin, elles ont pu mettre en évidence par expérience de complémentation que l’augmentation d’*UPF3B\_S*, consécutive à l’inactivation de *PQBPI* pouvait être ramenée au niveau basal en cas de co-transfections avec un plasmide codant pour la protéine PQBP1 sauvage mais pas avec une protéine non fonctionnelle. Ainsi, l’augmentation d’expression de cette isoforme semble être un moyen robuste de tester l’impact de VSI identifiés dans *PQBPI*, par une simple prise de sang.

J’ai par la suite cherché à comprendre le rôle d’*UPF3B\_S* et les conséquences de l’augmentation de son expression. Le gène *UPF3B* code pour un acteur clef du système NMD. En effet, la protéine UPF3B a pour rôle la reconnaissance de l’ARNm porteur d’un codon STOP

prématuré en interagissant avec RBM8A (ou Y14), un composant de l'EJC. J'ai donc dans un premier temps voulu comparer les interactomes protéiques d'UPF3B\_S et d'UPF3B\_L par spectrométrie de masse. J'ai ainsi pu mettre en évidence que comme attendu, UPF3B\_L et UPF3B\_S étaient capables de se lier aux complexes EJC par l'intermédiaire de RBM8A. En revanche, il est apparu que contrairement à UPF3B\_L, la protéine UPF3B\_S n'est pas capable d'interagir avec le facteur UPF2 et donc, de recruter la machinerie du NMD. J'ai donc par la suite investigué l'effet potentiel d'*UPF3B\_S* sur la régulation du NMD mais n'ai pas observé de différence significative quant à la stabilité d'ARNm porteurs de codons STOP prématurés dans le cadre d'une surexpression d'*UPF3B\_S* ou d'une inactivation de *PQBPI*. Les analyses protéomiques ont également révélé qu'UPF3B\_L comportait un nombre d'interactants bien plus grand (124 protéines) qu'UPF3B\_S (31 protéines). Enfin, j'ai pu montrer que l'inactivation de *PQBPI* conduisait à la diminution de la capacité proliférative de deux lignées hNSCs.

Pour conclure, nous avons pu mettre en évidence l'augmentation d'expression de *UPF3B\_S* dans différents modèles cellulaires inactivés pour par siRNA *PQBPI* ou provenant de cellules de patients Renpenning. Cette augmentation semble spécifique d'une altération de la fonction de *PQBPI* et est un marqueur biologique potentiel permettant d'évaluer le caractère délétère de VSI dans *PQBPI*. Même si nous avons pu montrer l'absence d'impact d'*UPF3B\_S* sur l'efficacité du NMD, son rôle et les conséquences de sa surexpression restent inconnus. Enfin, j'ai pu montrer que l'inactivation de *PQBPI* dans les modèles de hNSCs conduisait à une diminution de leur prolifération.

**Ce travail est présenté sous forme d'article mais ne correspond pas à la version définitive. En effet, plusieurs résultats restent à valider et différentes pistes sont actuellement investiguées.**

## **Molecular consequences of variants in *PQBPI*, involved in the X-linked Renpenning syndrome**

Jérémie Courraud<sup>1,2,3,4</sup>, Camille Engel<sup>1,2,3,4\*</sup>, Nathalie Drouot<sup>1,2,3,4\*</sup>, Angélique Quartier<sup>1,2,3,4\*</sup>, Ursula Houessou<sup>1,2,3,4</sup>, Damien Plassard<sup>1,2,3,4</sup>, Arthur Sorlin<sup>5</sup>, Elise Boucher<sup>6</sup>, Lionel Van Maldergem<sup>6</sup>, Massimiliano Rossi<sup>7</sup>, Patrick Edery<sup>7</sup>, Brigitte Gilbert<sup>8</sup>, Vera Kalscheuer<sup>10</sup>, Jean-Louis Mandel<sup>1,2,3,4</sup>, Amélie Piton<sup>1,2,3,4,10,11</sup>

<sup>1</sup> Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France

<sup>2</sup> Centre National de la Recherche Scientifique, UMR7104, Illkirch, France

<sup>3</sup> Institut National de la Santé et de la Recherche Médicale, U964, Illkirch, France

<sup>4</sup> Université de Strasbourg, 67 400 Illkirch, France

<sup>5</sup> Université de Bourgogne, CHU F. Mitterrand, UMR1231, 21079 Dijon, France

<sup>6</sup> Centre de Génétique Humaine, Université de Franche-Comté, Besançon, France

<sup>8</sup> Centre de Recherche en Neurosciences, UMR 5292, Université Claude Bernard, Lyon, France

<sup>9</sup> LBMC, Univ Claude Bernard Lyon 1, UMR5239, Lyon, France

<sup>10</sup> Genetic diagnosis laboratory, Strasbourg University Hospital, 67 090, Strasbourg, France

<sup>11</sup> Institut Universitaire de France

\* Equal contribution

## Abstract

Mutations in the *PQBPI* gene (Polyglutamine-binding protein 1) are responsible for syndromic X-linked form of intellectual disability (ID), the Renpenning syndrome. *PQBPI* encodes a protein that plays a role in the regulation of gene expression, splicing and mRNA translation. To investigate the consequences of mutations in *PQBPI*, we performed transcriptomic studies in 1) patient's lymphoblastoid cell lines (LCL) carrying pathogenic variants in *PQBPI* and 2) in human neural stem cells (hNSC) knock-down (KD) for *PQBPI* with siRNA. This led to the identification of a hundred genes whose expression is deregulated. In particular, we identified an increase in the expression of a non-canonical isoform of another X-linked DI gene, *UPF3B*. *UPF3B* plays a crucial role during neurodevelopment by coding for an important actor of the Nonsense mRNA Mediated Decay (NMD) system involved in regulation of protein translation, however, the exact function of this isoform, UPF3B\_S, is currently unknown. In order to investigate the role of UPF3B\_S isoform, we compared the protein interactomes of UPF3B\_S and of the canonical isoform UPF3B\_L. We confirmed that, on the contrary to UPF3B\_L, UPF3B\_S does not interact with the UPF2/UPF1 complex while it still interacts with exon junction complexes (EJC). However, no notable decrease of NMD pathways was observed in patient's LCL or in cells KD for *PQBPI*. We identified several additional protein interactors specific to UPF3B\_S. Moreover, we used the increase of UPF3B\_S mRNA as a molecular marker to test the pathogenicity of variants of unknown significance identified in individuals with ID in *PQBPI*. We analyzed patient's LCL mRNA as well as blood mRNA samples and performed complementation studies in HeLa cells by overexpressing WT and mutant *PQBPI* cDNA. We showed that all these three approaches were efficient to test the effect of variants. In conclusion, our study provides information on how *PQBPI* loss can affect gene expression, and in particular the short isoform of UPF3B, which informs us about the pathological mechanisms involved in Renpenning syndrome but also allows to propose a functional test for variants of unknown significance identified in *PQBPI*.

**Keywords:** PQBP1, Renpenning syndrome, gene expression, UPF3B, variants of unknown significance

## INTRODUCTION

More than a dozen mutations, mostly truncated, have been described in *PQBP1* (Polyglutamine-binding protein 1), responsible for syndromic X-linked intellectual disabilities (Renpenning, Sutherland-Haan, Hamel, Porteous, and Golabi-Ito-Hall)(Kalscheuer et al. 2003; Lenski et al. 2004; Kleefstra et al. 2004; Cossée et al. 2006; [CSL STYLE ERROR: reference with no printed form.]) regrouped now under the unique term of Renpenning syndrome(Stevenson et al. 2005). Clinical manifestations associated with these syndromes include microcephaly, growth retardation, lean profile and specific facial features (Germanaud et al. 2011).

The *PQBP1* (*Polyglutamine Binding Protein 1*) gene encodes a protein involved in different cellular processes such as regulation of transcription, splicing, translation or even response to retroviral infection. PQBP1 was initially described as a protein interacting with polyglutamine tracts such as Huntingtin or Ataxin1, though its PRD (polar amino-acid-rich domain) domain (Okazawa et al. 2002). It also contains on its N-terminal side a WW domain which has a transcriptional activity and interacts with the splicing factor SIPP1/WP11. The C-terminal part of PQBP1 includes a domain (CTD) interacting with other splicing factors such as U5-15kDa/TXNL4A(Waragai et al. 2000). Moreover, a Nuclear Localization Signal (NLS) allows its addressing to the nucleus through the Kap $\beta$ 2 receptor (Lee et al. 2006; Liu et al. 2020). In the nucleus, PQBP1 is involved in regulation of transcription through its interaction with activated RNA polymerase II (Okazawa et al. 2002) and various transcription factors such as POU3F2/Brn2(Waragai et al. 1999). PQBP1 has been found to be located in nuclear speckles, suggesting a role in the regulation of splicing, consistent with its interactions with the different splicing factors cited above and with the general splicing alterations observed after *PQBP1* knock-down in a model of primary neurons (Wang et al. 2013). PQBP1 is nuclear-cytoplasmic shuttling protein playing also various roles in the cytoplasm. Indeed, it can be localized at stress granules (Kunde et al. 2011) and was shown to play a role in translation of messenger RNAs (Shen et al. 2021). PQBP1 binds to the elongation factor eEF2 via its WW domain and suppress its phosphorylation-mediated inactivation. Loss of *PQBP1* leads to an increase of the phosphorylation of eEF2 stopping translational elongation and resulting in a global decrease of protein synthesis, which affect protein synthesis-dependent form of synaptic plasticity in hippocampus (Shen et al. 2021). PQBP1 is involved in the regulation of neuronal ciliogenesis via its interaction with Dynamin2 (Ikeuchi et al. 2013). Finally PQBP1 can also play a role in response to retroviral infection, interacting with reverse-transcribed HIV-1 DNA, probably

though its CTD domain, and to the viral DNA sensor cGAS probably through its WW domain, and therefore contributing to the innate immune response (Yoh et al. 2015).

If constitutive knock-out of *PQBPI* in mice appears to be lethal ([www.mousephenotype.org/](http://www.mousephenotype.org/)), mice models where *PQBPI* was knock-down (~50%) display abnormal anxiety-related behavior, and a decrease in anxiety-related cognition (Ito et al. 2009). Nestin-Cre conditional knock-out (*PQBPI* cKO), which are microcephalic, also display abnormal anxiety-related behavior, as well as abnormal fear conditioning and motor dysfunction at the rotarod test (Ito et al. 2015). Recently, the same group demonstrated that cKO mice showed a short stature and a reduction in bone mass, and display impairment in bone formation and chondrocyte deficiency with reduced osteoblast and chondrocyte-related gene expression (Yang et al. 2020). The cKO *PQBPI* showed microcephaly, probably resulting from an elongated cell cycle of neural progenitors (Ito et al. 2015).

Most of the mutations described are small indels, located in the AG hexamer of exon 4 or downstream, resulting in truncated proteins lacking their C-terminal domain. Only one distal frameshift variant, p.Phe240fs, have been reported to date, leading to a protein 10 amino acids longer than the wild-type (Hu et al. 2016). Only few missense variants have been reported in *PQBPI*. For years, only a unique missense mutation (p.Tyr65Cys) was described in a family with intellectual disability (Lubs et al. 2006). This missense variant affects a very conserved amino acid position located in the WW domain and disrupts the interaction between *PQBPI* and the splicing factor WBP11 leading to a decrease in pre-mRNA splicing efficiency (Tapia et al. 2010). Two other missense variations (p.Arg243Trp and p.Pro244Leu), located in the CTD domain were identified in patients with ID (Redin et al. 2014; Hu et al. 2016). The Pro244Leu change was shown to disrupt *PQBPI* binding to the splicing factor U5-15kDa/TXNL4A (Liu et al. 2020). We report here two additional missense variants identified in individuals with NDD, p.Arg10Pro and p.Arg159Trp. As they are not located in the WW or CTD domain, their functional consequences and therefore pathogenicity remains unknown.

We described in this study the changes in gene expression induced by a loss of *PQBPI* in human cells, in lymphoblastoid cell lines (LCL) of patients carrying pathogenic variants in *PQBPI* and in human neural stem cells (hNSC) where *PQBPI* was knock-down by siRNA. We reported in particular the induction of a short isoform of *UPF3B*, another gene involved in XLID involved in nonsense mRNA decay (NMD) and regulation of translation. Finally, we showed that we can use this increase as a biomarker of Repenning syndrome to test the pathogenicity of variant of unknown significance in *PQBPI*.

## Material and methods

### *Subjects*

Clinical information was retrieved from clinicians for individuals carrying pathogenic variants identified by panel or exome sequencing. All the variants were named according to the isoform NM\_005710.2 which encodes a protein of 265 amino acids. The potential effect of missense variants were predicted using SIFT (<http://sift.bii.a-star.edu.sg/>), PolyPhen2 (<http://genetics.bwh.harvard.edu/pph2/>) and CADD (<https://cadd.gs.washington.edu/>). LCL were available for eleven individuals with variants in *PQBP1*, seven control individuals and four individuals with mutation in *FMR1*. Blood RNA samples (Paxgene) was available for three individuals with variants in *PQBP1* and four individuals with other genetic forms of ID (Table S1).

### *Plasmids and antibodies*

Human *PQBP1* cDNA NM\_005710.2 sequence (including UTR) was subcloned into PCS2+ plasmid under CMV promoter and optimized to escape to degradation by siRNA against *PQBP1* by introducing xx silent point mutations. Benign, pathogenic and VUS variants were introduced by site-directed mutagenesis. Human *UPF3B\_L* cDNA NM\_080632.2 and *UPF3B\_S* ENST00000276201 were subcloned into pcDNA3 plasmid under CMV promoter and tagged with HA motif (YPYDVPDYA) at the C-terminal side of the protein. For western blots and immunostaining, specific primary antibodies were used: anti-UPF3B (NSJ BIOREAGENTS), anti-*PQBP1* (NOVUSBIO), anti-APP (Invitrogen Invitrogen 51-2700), anti-GAPDH (G9545 Sigma Aldrich). The secondary antibodies used were HRP-labelled goat anti-mouse or anti-rabbit IgG and HRP (Jackson Immunoresearch, Baltimore, PA, USA).

### *Cell culture, transfection and proliferation assay*

Lymphoblastoid cell lines (LCL) were previously (Kalscheuer et al. 2003) or newly established by infection of blood lymphocytes using Epstein-Barr virus and were maintained in RPMI without HEPES and 10% fetal calf serum. Two lines of human neuronal stem cells (hNSC)



were used: hNSC-1 (SA001) and hNSC-2 (GM1869), previously described (Boissart et al. 2012; Quartier et al. 2018). They were seeded on poly-ornithine and laminin-coated (Sigma-Aldrich) dishes and maintained in DMEM/F12 and Neurobasal medium (1:1) supplemented with N2, B27, 2-mercaptoethanol (Invitrogen, Carlsbad, CA, USA), BDNF (20ng/mL), FGF-2 (10ng/mL) (PeproTech, Rocky Hill, NJ, USA) and EGF (R&D Systems; 10ng/mL) as described (Quartier et al. 2018). HEK293 and HeLa cells were maintained respectively in DMEM (1g glucose / liter), 10% fetal calf serum and gentamicin (40 $\mu$ g/mL) in CellBind® flasks (Corning, NY, USA) for HEK293T, DMEM (1g glucose / liter), 5% fetal calf serum and gentamicin (40 $\mu$ g/mL) in CellBind® flasks (Corning, NY, USA) for HeLa. Cells were transfected using InterferIN (Polyplus) or Lipofectamine 2000 transfection protocol (Thermo Fisher Scientific) for siRNA and plasmids. Cells were stopped at 24/48 and 96 hours after transfection for protein and RNA extractions. For proliferation assay, reverse transfection of SA001 and GM1869 hNSCs (20,000 cells/cm<sup>2</sup>) was performed in 96-well plates using InterferIN reagent and 20nM siRNA according to manufacturer's recommendation. Each day a plate was fixed and stained with DAPI and number of nucleus in each condition counted using CellInsight automated microscope and HCS Studio software (Thermo Fisher Scientific).

### ***mRNASeq***

For the LCL cells, RNA-seq libraries of template molecules were prepared as previously described (Quartier et al. 2018). For the hNSCs, RNA-seq libraries of template molecules suitable for high throughput sequencing were established from 300ng of total RNA using the KAPA RNA HyperPrep Kit after a first step of purification using poly-T oligo-attached magnetic beads (Roche). Libraries were sequenced on the Illumina HiSeq 4000 sequencer as paired-end 100 base reads. Reads were mapped onto the hg38 assembly of the human genome using Tophat 2.0.10 or 2.0.14 (Kim et al. 2013) and the bowtie version 2-2.1.0 aligner (Langmead and Salzberg 2012) and only uniquely mapped reads were kept. Gene expression was quantified using HTSeq-0.6.1 (Anders et al. 2015) and gene annotations from Ensembl release 75. Read counts have then been normalized across libraries with the median-of-ratios method proposed by Anders and Huber (Anders and Huber 2010). Comparisons of interest were performed using the method proposed by Love et al. (Love et al. 2014) and adjustment for multiple testing was performed with the Benjamini and Hochberg method (Benjamini, Y). Alternative splicing and isoform-specific differential expression were identified using DEXSeq (Anders et al. 2012).

### ***RT-qPCR***

Total RNA were extracted from cells using the RNeasy extraction kit and treated with RNase free DNase set during 20 min at 37°C (Qiagen, Valencia, CA, USA) or from blood using PaxGene blood RNA kit extraction (Preanalytix, Hombrechtikon, Switzerland). RNA levels and quality were quantified using a Nanodrop spectrophotometer and then with a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) for the RNA used for RNA-sequencing. For reverse transcription-PCR (RT-PCR), 500ng to 1µg of total RNA was reverse transcribed into cDNA using random hexamers and SuperScript IV reverse transcriptase according to manufacturer's recommendation. Real-time PCR quantification (qPCR) was performed on LightCycler 480 II (Roche, Basel, Switzerland) using the QuantiTect SYBR Green PCR Master Mix (Qiagen). All qPCR reactions were performed in triplicate. Primer sets are listed in **Table S2**. Reaction specificity was controlled by post-amplification melting curve analysis. The relative expression of gene-of-interest vs *GAPDH* and *YWHAZ* was calculated using the 2- $(\Delta\Delta C_t)$  method and a parametric Student's t-test was performed in order to compare control vs patient cells or untreated vs siRNA treated cells. Error bars represent standard error of the mean (SEM).

### ***Western Blot***

Cells were lysed in RIPA buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.25% sodium deoxycholate, 1% NP-40) supplemented with protease inhibitor cocktail and phosphatase inhibitor cocktail. 5 to 50µg of protein lysate were separated on 10% SDS-PAGE and transferred to polyvinylidene fluoride membrane. Membranes were blocked in 5% nonfat dry milk diluted in tris buffered saline with tween 20 (50mM Tris, 150mM NaCl, 0.05% Tween 20) and probed using the antibodies overnight at 4°C. GAPDH was used as loading control. Incubation with appropriate secondary HRP-labelled antibody (less than 1h) was followed by detection with Immobilon western chemiluminescent HRP substrate (Merck Millipore, Darmstadt, Germany).

### ***Immunoprecipitation and Mass Spectrometry analyses***

Protein were extracted from HEK293 cells overexpressing UPF3B\_L (ENST00000636792) or UPF3B\_S (ENST00000276201) and subjected to immunoprecipitation with Pierce Anti-HA

Magnetic Beads (ThermoFisher 88836) according to manufacturer protocol. Immunoprecipitations were validated by WB as previously done (Mattioli et al. 2019) before the mass spectrometry analyses (Proteomic platform, IGBMC). Briefly, samples were treated with LysC/Trypsine for liquid digestion and injected in Orbitrap ELITE / C18 Accucore 50cm (20 $\mu$ L 0.1% TFA / 1 $\mu$ L) for 2h runs in triplicate. Data were processed with Proteome Discoverer 2.2 software using Homosapiens\_190716\_reviewed.fasta and contaminants\_190528.fasta databases. To consider a protein as a candidate interactor, we applied the thresholds to keep only protein with 1) at least two unique peptides, 2) a positive value of Peptide-Spectrum Matching (PSM) in each replicate 3) a positive eXtracted Ion Chromatogram (XIC) in at least one replicate, 4) a sum of PSM for the three replicates of the control (empty) condition inferior to 5, 5) a sum of PSM for the three replicates of the test condition (UPF3B\_L or UPF3B\_L) greater or equal to 5 and 6) a ratio between the Normalized Spectral Abundance Factor (NSAF) of the test condition (UPF3B\_L or UPF3B\_L) and the NSAF of the control (empty) condition greater or equal to 2. Enrichment analysis were realized using DAVID (Database for Annotation, Visualization and Integrated Discovery) or the GO (Gene Ontology Resource).

## RESULTS

### Identification of genes differentially expressed in individuals with Renpenning syndrome

In order to identify the effect of *PQBPI* loss on the regulation of gene expression and alternative splicing, we performed sequencing of mRNA extracted from lymphoblastoid cell lines (LCL) obtained from male individuals carrying pathogenic or likely pathogenic variants in *PQBPI* (n=5) to that of unrelated control males (n=3) (**Figure 1A-B, Figure S1A**). RT-qPCR was used to confirm the results obtained by RNAseq on LCL mRNA from the same individuals and from a second serie of LCL including additional control individuals (n=3), individuals with *PQBPI* frameshift variants (n=2), individuals with *PQBPI* variant of unknown significance (VUS, n=3) and individual with a neurodevelopmental condition from another genetic origin, the fragile-X syndrome (FXS, n=4). Analysis of gene expression revealed 79 protein-coding genes differentially expressed (DE) in individuals with variants in *PQBPI*, 53 up-regulated in patients and 26 down regulated (**Figure S1B, Table S3**). We observed for instance a strong increase in APP expression in individuals with pathogenic variant in *PQBPI*. However, if this increase has been confirmed by RT-qPCR on the same samples at the mRNA level (**Figure S1C**) but also at the protein level (**Figure S1D**), this was also observed for FXS individuals and failed to be

validated in the analysis of the second set of LCL (**Figure S1E**), which detect a large variability in APP expression, regardless of the status of the individuals. Similar results were obtained for another gene found upregulated in the RNASeq analysis, *TSPAN5* (**Figure 1F-G**). Transcriptomic analysis performed at exon level using DEXseq identified 117 different expressed (DE) exons belonging to 97 protein-coding genes (**Table S4**). Among them, four exons corresponding to a short isoform (ENST00000276201, UPF3B\_S) of the *UPF3B* gene, another gene involved in XLID, were significantly more covered in patients (log2FC: 1.06-2.17, adjusted p-value: 1.4E-3 – 1.78E-5) (**Figure 2A-B, Figure S2A**). This time, no change was observed for individuals with FXS syndrome, and the increase was found significant in the second set of patients vs controls (**Figure 2C**). This increase was specific to the short isoform as no change was observed in RNAseq or RT-qPCR for the long UPF3B isoform (ENST00000636792, UPF3B\_L) (**Figure S2B**). To confirm that *PQBPI* loss leads to an upregulation of the short isoform (UPF3B\_S), we treated two human Neural Stem Cell lines (hNSC) which are self-renewal homogenous precursors of cortical neurons with a pool of siRNA directed against *PQBPI*. A knock-down (KD) of 40% of *PQBPI* expression was obtained (**Figure S3**). A significant increase of *UPF3B\_S* expression was identified in the KD condition while no change was observed when a nonspecific siRNA (scramble) was used (**Figure 2D**) with no effect on UPF3B\_L expression (**Figure S2C**), confirming what was observed in patients with pathogenic variants in *PQBPI*.

### **Identification of genes differentially expressed after *PQBPI* knock-down in human neural stem cells (hNSC)**

In order to identify additional genes and exons DE after a KD of *PQBPI* in hNSC, we performed RNAseq for the first hNSC line (hNSC\_1, in duplicates). Transcriptomic analysis revealed 59 genes significantly DE (12 upregulated and 47 down regulated) with *PQBPI* siRNA compared to cells treated with the transfection agent only (INTERFERin), while no gene was found significantly DE in the scramble siRNA condition (**Table S5, Figure 3**). Unsurprisingly, the most significantly DE gene was *PQBPI* (log2FC=-0.82, adjusted p-value =3.7E-8). To confirm these changes in gene expression, RT-qPCR was performed for the best candidate genes on a third serie of hNSC-1 (n=3 samples per condition) as well as on three series of another cell line, hNSC-2 (n=9 samples per condition in total) (**Figure 3B**). Among others, we identified that *PQBPI* KD leads to a decrease in expression of genes encoding proteins essential for brain development, including transcription factors (*FOXJ1, NPAS3*),

neurotrophic factor (*GFMB*) or autophagic proteins (*AMBRA1*). A decrease of proteins involved in neurite outgrowth and synaptic formation (*LRFN4*, *APC2* or *FIGNL2*) or synaptic function (*SYT2*, *NRXN2*, *SARM1*, *MAPK8IP2* or *RNF167*) was also observed. Two RNA-binding proteins involved in mRNA metabolism and particularly regulation of alternative splicing (*CELF2* and *RBPMS2*) were found to be decrease after *PQBP1* KD. Decrease in expression of genes encoding metabolic enzymes, and especially enzymes involved in lipid or steroid metabolism (*HSDL2*, *MIDIIP1*) have also been observed. As *PQBP1* is known to play a role of cell cycle regulator in neural progenitors (Ito et al. 2015), which is consistent with the microcephaly observed in individuals affected by Renpenning syndrome, we decided to test if *PQBP1* KD affect their proliferation of our hNSC model. We observed a significant decrease in hNSCs proliferation 96 hours after *PQBP1* KD in hNSC-1 when compared to transfection agent only (INTERFERin) or Scramble siRNA (**Figure 3C**). A similar tendency was obtained for hNSC-2 (p=0.059). Despite the described role of *PQBP1* in splicing regulation, difference in exon coverage were observed in few genes only using DEXSeq (n=3).

#### **The UPF3B\_S isoform, increased after *PQBP1* KD, does not bind to NMD actors UPF1/2**

The UPF3B\_L isoform encodes a protein involved in mRNA nonsense mediated decay (NMD), but the role of UPF3B\_S is currently not known. This isoform is particularly expressed in testis and in EBV-immortalized lymphocytes but is also detected at lower level in other tissue such as cerebellum (GTex database). In order to better understand the role of the protein encoded by UPF3B\_S, we generated plasmids containing human UPF3B\_L and UPF3B\_S sequences. We performed immunoprecipitation-coupled mass spectrometry on proteins extracted from HEK293 cells transfected with these plasmids in order to identify protein partners for both isoforms. We identified a total of 133 proteins: 23 proteins interacting with both isoforms, 101 proteins interacting preferentially or exclusively with UPF3B\_L and only 9 preferentially or exclusively with UPF3B\_S (**Table S6**). We found, among the interactors identified, an enrichment in proteins involved in mRNA metabolism (splicing, mRNA export from nucleus, NMD, protein translation, etc) but also in cell-cell adhesion, metabolism (glycogenesis) and immune system (immunoglobulin receptor binding, positive regulation of B cell activation, antigen binding, phagocytosis, antibacterial humoral response, etc) (**Table S6**). We retrieved six of the ten known UPF3B interactors annotated in the STRING database (**Figure 4**). As expected, we found that UPF1, UPF2 and ERF3A interact only with UPF3B\_L and not

UPF3B\_S, which lacks the interaction domain (**Figure 4**), while RBM8A (alias Y14) interact with both isoforms. Among the proteins interacting preferentially with UPF3B\_S, we identified proteins involved in antibacterial humoral response (LTF, SEMG1, LYZ) and in redox homeostasis (PRDX4, SDHA).

### ***PQBP1* inactivation does not affect NMD**

UPF3B\_L is a major actor of NMD system through its binding to UPF2 and to RBM8A (EJC). As UPF3B\_S does not bind UPF2 but is still able to bind RBM8A, we could hypothesize that UPF3B\_S could act as a competitive inhibitor of UPF3B\_L by interacting with EJC through RBM8A but failing to recruit UPF2 and the NMD machinery. To test that hypothesis, we first searched for enrichment of NMD targeted genes in DE genes in patients' LCL or in *PQBP1* KD hNSCs but we found no overlap. Furthermore, not significant overlap was detected when comparing the genes DE in patients' LCL and those previously reported DE in LCL obtained from individuals with pathogenic variants in *UPF3B*. To go further, we tested by RT-qPCR whether *PQBP1* inactivation affects *SIX3* mRNA expression, since *SIX3* mRNA was found upregulated in murine *UPF3B* KO neural cells (Jolly et al. 2013) as well as in UPF3B patient's LCL (Nguyen et al. 2012). We did not find any significant variation of *SIX3* mRNA expression in LCL from controls and individuals with variants in *PQBP1* as well as in *PQBP1* KD hNSCs (**Figure S4**). Finally, we knocked-down *PQBP1* in fibroblasts carrying a truncating variant in the *DYRK1A* gene (c.1232dup, p.Arg413fs) known to be degraded by the NMD system (Courraud et al. 2021) but observed no increase of the mutant transcript, confirming that *PQBP1* inactivation has no obvious effect on NMD (data not shown).

### **Increase in *UPF3B\_S* expression is useful for variant testing**

We identified two additional missense variants in *PQBP1* in individuals with ID: 29G>C, p.Arg10Pro and c.475C>T, p.Arg159Trp. Both variants are not located in a known functional domain and remains therefore of unknown significance. 21 additional missense variants of unknown significance are reported in the ClinVar database. We tested the effect of truncating and missense variants identified in *PQBP1* in patients with ID on the level or the stability of PQBP1 protein and found no effect (**Figure 1B**). Therefore, it would be important to develop a functional assay to test variants' effect. Increase of UPF3B\_S expression was the only event observed both in individuals with Renpenning syndrome and after transient KD of *PQBP1* in

the two lines of hNSC. We therefore decided to investigate if we could use this event to test the functional consequences of variants of unknown significance (VUS) in *PQBPI*. We have previously shown that this increase is detectable in LCL of individuals with pathogenic variants in *PQBPI* (**Figure 2B-C**). We observed a significant increase of *UPF3B\_S* in LCL obtained from the two brothers (V1, V2) carrying the distal frameshift variant p.Phe240fs previously identified by Hu et al (Hu et al. 2016), confirming the pathogenicity of this variant (**Figure 5A**). At the opposite, no significance increase was identified for LCL of the patient carrying the variant 29G>C, p.Arg10Pro (V4). In order to propose a functional assay which would not depend on having access to LCLs, we first confirmed that KD of *PQBPI* in HeLa cells leads to increase of *UPF3B\_S* expression (**Figure 5C-D**). These cells are easily transfectable and could be used to perform a complementation test, by co-transfecting a *PQBPI* siRNA with a plasmid containing the wild-type (WT) human *PQBPI* cDNA optimized not to be targeted by this siRNA (see Methods). We showed that this co-transfection of *PQBPI* siRNA with optimized WT *PQBPI* significantly decreases the expression of endogenous *PQBPI* (**Figure 5C**) and reverts *UPF3B\_S* expression to its basal level (**Figure 5D**). On the contrary, co-transfection of *PQBPI* siRNA with mutant Arg154fs\* *PQBPI* fails to revert *UPF3B\_S* increase, confirming the loss-of-function effect of this variant.

Therefore, increase of *UPF3B\_S* expression appears to be a relevant readout, and the complementation assay we performed was found efficient to evaluate the pathogenicity of variants in *PQBPI*. However, this assay requires different steps of bacterial cloning and cell transfection, which are time-consuming and not always possible to perform in a routine diagnostic practice. We therefore decided to test if we could measure increase of *UPF3B\_S* expression on mRNA extracted directly from patients' blood. As expected, no expression was detected in three control individuals and in two individuals with FXS, while *UPF3B\_S* was expressed in blood from three individuals with pathogenic variants in *PQBPI* (**Figure 5B**). However, no expression was detected for the individual harboring the p.Arg159Trp variant (V4), suggesting that this variant is not-disease causing.

## DISCUSSION

In this study, we identified genes and exons differentially expressed in individuals with pathogenic variants in *PQBPI* as well as those differentially expressed after a transient knock-down of *PQBPI* in human neural stem cells (NSC). We observed a high variability in gene

expression between the two series of LCL analyze, which prevented us to confirm the deregulation observed in the first serie of patients and controls. However, for the changes in gene expression observed in neural stem cells, we successfully replicated them in a second cell line, showing their robustness. Decrease in the expression of genes involved in regulation of neurite outgrowth and synapse formation or lipid metabolism are consistent with the roles of PQBP1 previously described by others. As an example, a role of PQBP1 in the regulation of dendrite length and branching has been described in mouse embryonic primary cortical neurons (Wang et al. 2013). PQBP1 was also implicated in the regulation of lipid metabolism, with a decrease of the lipid content in intestinal cells observed in nematode which could be involved in the low body mass index observed in patients (Takahashi et al. 2009). A significant decrease in the expression of an RNA binding protein involved, among other things, in the regulation of splicing, *CELF2*, was observed after *PQBP1* KD in human neural stem cells. Interestingly, heterozygote mutations in *CELF2* were recently reported to be responsible for a neurodevelopmental syndrome including ID and epilepsy (Itai et al. 2021).

Exon analysis revealed that some isoforms of different genes were deregulated (as *UPF3B\_S* for instance) but failed to identify important changes in alternative splicing events, despite the known role of PQBP1 in splicing regulation and the fact that the variants present in the patients analyzed are known to disrupt interactions with splicing factors (Tapia et al. 2010; Liu et al. 2020). Additional programs might be used to try to detect more changes in alternative splicing.

No overlap was identified between the genes and exons identified are significantly deregulated in patients' LCL and after transient PQBP1 KD in hNSC, except a deregulation of *KCNJ2* mRNA which was not replicated by RT-qPCR in hNSC-2. This can be explained by the tissue differences, but also by the difference between an innate dysfunction of PQBP1 (in patients' cells) which can lead to feedback/compensation regulations and a transient knock-down. However, it should be noted that the threshold used for RNAseq analysis is stringent and therefore only a limited number of genes were identified for each condition, but this does not mean that there is no common genes deregulated in both conditions, as illustrated by the example of *UPF3B\_S*. Indeed, if no significant difference in *UPF3B\_S* mRNA level was detected by RNA-Seq analysis in hNSC KD for *PQBP1*, qPCR analysis revealed however that *UPF3B\_S* mRNA was actually also significantly increased in the two hNSC lines.

Increase of *UPF3B\_S* mRNA level was also confirmed in other cell types, such as HeLa cells and therefore appears to be robustly related to the loss of *PQBP1*. We showed that this increase of *UPF3B\_S* mRNA level could be detected in blood RNA extracted from Renpenning patients



and are not detected in at least five other genetic forms of ID. This means that together with clinical and segregation data, pathogenicity of future VUS identified in *PQBP1* could be evaluated from a simple blood RNA extraction. In order to use this biomarker for routine diagnostic practice, we should confirm the results obtained in this study in a larger number of individuals with Renpenning syndrome (>10) and patients with other genetic forms of ID (>50).

There is no report about the biological function of this short isoform of *UPF3B*. *UPF3B\_S* mRNA appears to be low expressed, except in testis, according to GTex database. It is notable that genital manifestations, and especially microorchidia, were reported for individuals with Renpenning syndrome. A mild expression in EBV-immortalized lymphoblastoid cells is also reported in GTex, and might explain why we particularly detected this increase in LCL. If no expression of *UPF3B\_S* is reported in GTEx in adult brain tissue, we cannot exclude that it could be expressed in a transient way during brain development. UPF3B protein contains several sequences known to interact with RBM8A (alias Y14) in exon E'2 but not that which interacts with UPF2/UPF1, meaning that it could bind to EJC but not to NMD effectors. We could therefore speculate that *UPF3B\_S* could compete with *UPF3B\_L* for EJC binding and act as a negative dominant form of UPF3B. However, we did not observe NMD dysfunction after *PQBP1* KD in fibroblasts carrying a premature stop codon (PTC) in another gene involved in NDD and no overlap between genes DE in Renpenning patients' LCL or after *PQBP1* KD and genes known to be natural target of NMD. *In vitro*, UPF3B is involved in the regulation of translation termination by interacting with ERF3A (Neu-Yilik et al. 2017) in a context of PTC and might also play a role in the regulation of normal translation termination. *PQBP1* was shown to reduce translational elongation by preventing the phosphorylation of the elongation factor *EEF2* (Shen et al. 2021). Therefore, it would be interesting to investigate the effect of *PQBP1* KD and of *UPF3B\_S* overexpression on global protein translation in our cell models.

In order to better understand the role of *UPF3B\_S* isoform, we performed immunoprecipitation coupled to Mass spectrometry (IP-MS) to identify difference in protein partners between *UPF3B\_S* and *UPF3B\_L*. We confirmed that UPF3B isoforms interact with proteins involved in mRNA binding and regulation, and in particular regulation of translation (ribosomal proteins and translation initiation or elongation factors). In particular, both isoforms interact with the translation elongation factor *EEF2*, which is known to be also bound by *PQBP1*, which regulated its activity by inhibiting its phosphorylation (Shen et al. 2021). We confirmed that, contrary to *UPF3B\_L*, *UPF3B\_S* is not able to bind NMD effectors UPF2 and UPF1. We also

detect few proteins interacting with UPF3B\_S but not with the canonical isoform, some playing a role in immune defense against bacterial infection (DEFA1, LTF, LYZ and SEMG1). Interestingly, PQBP1 is known to play a role in response to infection as it can recognize viral DNA, interact with cGAS and regulate immune response (Yoh et al. 2015).

Several of the novel interactors of UPF3B\_S (PRDX4, GLUL) are implicated in regulation of cell cycle. In this study, we also showed that *PQBP1* KD induces a decrease in hNSC ability to proliferate. This result is consistent with the microcephaly observed in individuals with Renpenning syndrome and with what was previously reported by other in non-human neuronal progenitors (Ito et al. 2015). Ito et al. observed, in nestin-Cre conditional KO mice for *Pqbp1*, which present with microcephaly, a loss of proliferation of neural progenitors probably due to an elongated cell cycle. As *Upf3b* KD in mouse neural progenitor cells leads on the contrary to an increase of cell proliferation (Jolly et al. 2013), additional experiments would be needed to test the role of an increase of *UPF3B\_S* expression and how if it participates to the decrease of hNSC proliferation induced by *PQBP1* KD.

## **DATA AVAILABILITY**

Data have been submitted to Gene Expression Omnibus.

## **ACKNOWLEDGEMENTS**

The authors would like to thank the families for their participation and support. The authors also thank the Agence de Biomédecine, Fondation APLM and Fondation Jérôme Lejeune for financial support. We also thank the GenomEast sequencing platform for performing RNASeq and for their help in data analysis, as well as Bastien Morlet from the Mass spectrometry platform and Anne Maglott from the high-throughput screening platform.

## **CONFLICTS OF INTEREST**

None

## **ETHICS DECLARATION**

Individuals were referred by clinical geneticists for genetic testing as part of routine clinical care. All patients enrolled and/or their legal representative have signed informed consent for research use and authorization for publication. IRB approval was obtained from the Ethics Committee of the Strasbourg University Hospital (CCPPRB) as well as from local institutions.

## WEB RESSOURCES

The URLs for online tools and data presented herein are:

CADD: <https://cadd.gs.washington.edu/>

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

Clustal Omega: <https://www.ebi.ac.uk/Tools/msa/clustalo/>

Ensembl: <https://www.ensembl.org/>

GEO: <https://www.ncbi.nlm.nih.gov/geo/>

*GnomAD*: <http://gnomad.broadinstitute.org/>

GTEx : <https://gtexportal.org/home/>

Integrative Genomics Viewer (IGV): <http://www.broadinstitute.org/igv/>

OMIM: <http://www.omim.org/>

OrthoInspector database: <https://www.lbgi.fr/orthoinspectorv3/databases>

UCSC: <http://genome.ucsc.edu/>

## REFERENCES

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106. <https://doi.org/10.1186/gb-2010-11-10-r106>

Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169. <https://doi.org/10.1093/bioinformatics/btu638>

Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* 22:2008–2017. <https://doi.org/10.1101/gr.133744.111>

Benjamini, Y Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing

Boissart C, Nissan X, Giraud-Triboult K, Peschanski M, Benchoua A (2012) miR-125 potentiates early neural specification of human embryonic stem cells. *Development* 139:1247–1257. <https://doi.org/10.1242/dev.073627>

Cossée M, Demeer B, Blanchet P, Echenne B, Singh D, Hagens O, Antin M, Finck S, Vallee L, Dollfus H, Hegde S, Springell K, Thelma BK, Woods G, Kalscheuer V, Mandel J-L (2006) Exonic microdeletions in the X-linked PQBP1 gene in mentally retarded patients: a pathogenic mutation and in-frame deletions of uncertain effect. *Eur J Hum Genet* 14:418–425. <https://doi.org/10.1038/sj.ejhg.5201593>

Courraud J, Chater-Diehl E, Durand B, Vincent M, del Mar Muniz Moreno M, Boujelbene I, Drouot N, Genschik L, Schaefer E, Nizon M, Gerard B, Abramowicz M, Cogné B, Bronicki L, Burglen L, Barth M, Charles P, Colin E, Coubes C, David A, Delobel B, Demurger F, Passemard S, Denommé A-S, Faivre L, Feger C, Fradin M, Francannet C, Genevieve D, Goldenberg A, Guerrot A-M, Isidor B, Johannesen KM, Keren B, Kibæk M, Kuentz P, Mathieu-Dramard M, Demeer B, Metreau J, Møller RS, Moutton S, Pasquier L, Pilekær Sørensen K, Perrin L, Renaud M, Saugier P, Svane J, Thevenon J, Them FTM, Tronhjem CE, Vitobello A, Layet V, Birling M-C, Drunat S, Bayat A, Dubourg C, Chehadeh SE, Fagerberg C, Mignot C, Guipponi M, Bienvenu T, Herault Y, Thompson J, Willems M, Mandel J-L, Weksberg R, Piton A (2021) Integrative approach to interpret DYRK1A variants, leading to a frequent neurodevelopmental disorder. *medRxiv* 2021.01.20.21250155. <https://doi.org/10.1101/2021.01.20.21250155>

Germanaud D, Rossi M, Bussy G, Gérard D, Hertz-Pannier L, Blanchet P, Dollfus H, Giuliano F, Bennouna-Greene V, Sarda P, Sigaudy S, Curie A, Vincent MC, Touraine R, des Portes V (2011) The Renpenning syndrome spectrum: new clinical insights supported by 13 new PQBP1-mutated males. *Clin Genet* 79:225–235. <https://doi.org/10.1111/j.1399-0004.2010.01551.x>

He T-S, Dang L, Zhang J, Zhang J, Wang G, Wang E, Xia H, Zhou W, Wu S, Liu X (2021) The Hippo signaling component LATS2 enhances innate immunity to inhibit HIV-1 infection through PQBP1-cGAS pathway. *Cell Death Differ*. <https://doi.org/10.1038/s41418-021-00849-1>

Hu H, Haas SA, Chelly J, Van Esch H, Raynaud M, de Brouwer APM, Weinert S, Froyen G, Frints SGM, Laumonnier F, Zemojtel T, Love MI, Richard H, Emde A-K, Bienek M, Jensen C, Hambrock M, Fischer U, Langnick C, Feldkamp M, Wissink-Lindhout W, Lebrun N, Castelnau L, Rucci J, Montjean R, Dorseuil O, Billuart P, Stuhlmann T, Shaw M, Corbett MA, Gardner A, Willis-Owen S, Tan C, Friend KL, Belet S, van Roozendaal KEP, Jimenez-Pocquet M, Moizard M-P, Ronce N, Sun R, O’Keeffe S, Chenna R, van Bömmel A, Göke J, Hackett A, Field M, Christie L, Boyle J, Haan E, Nelson J, Turner G, Baynam G, Gillissen-Kaesbach G, Müller U, Steinberger D, Budny B, Badura-Stronka M, Latos-Bieleńska A, Ousager LB, Wieacker P, Rodríguez Criado G, Bondeson M-L, Annerén G, Dufke A, Cohen M, Van Maldergem L, Vincent-Delorme C, Echenne B, Simon-Bouy B, Kleefstra T, Willemsen M, Fryns J-P, Devriendt K, Ullmann R, Vingron M, Wrogemann K, Wienker TF, Tzschach A, van Bokhoven H, Gecz J, Jentsch TJ, Chen W, Ropers H-H, Kalscheuer VM (2016)

- X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Mol Psychiatry* 21:133–148. <https://doi.org/10.1038/mp.2014.193>
- Ikeuchi Y, de la Torre-Ubieta L, Matsuda T, Steen H, Okazawa H, Bonni A (2013) The XLID protein PQBP1 and the GTPase Dynamin 2 define a signaling link that orchestrates ciliary morphogenesis in postmitotic neurons. *Cell Rep* 4:879–889. <https://doi.org/10.1016/j.celrep.2013.07.042>
- Itai T, Hamanaka K, Sasaki K, Wagner M, Kotzaeridou U, Brösse I, Ries M, Kobayashi Y, Tohyama J, Kato M, Ong WP, Chew HB, Rethanavelu K, Ranza E, Blanc X, Uchiyama Y, Tsuchida N, Fujita A, Azuma Y, Koshimizu E, Mizuguchi T, Takata A, Miyake N, Takahashi H, Miyagi E, Tsurusaki Y, Doi H, Taguri M, Antonarakis SE, Nakashima M, Saitsu H, Miyatake S, Matsumoto N (2021) De novo variants in CELF2 that disrupt the nuclear localization signal cause developmental and epileptic encephalopathy. *Hum Mutat* 42:66–76. <https://doi.org/10.1002/humu.24130>
- Ito H, Shiwaku H, Yoshida C, Homma H, Luo H, Chen X, Fujita K, Musante L, Fischer U, Frints SGM, Romano C, Ikeuchi Y, Shimamura T, Imoto S, Miyano S, Muramatsu S - i, Kawauchi T, Hoshino M, Sudol M, Arumughan A, Wanker EE, Rich T, Schwartz C, Matsuzaki F, Bonni A, Kalscheuer VM, Okazawa H (2015) In utero gene therapy rescues microcephaly caused by Pqbp1-hypofunction in neural stem progenitor cells. *Mol Psychiatry* 20:459–471. <https://doi.org/10.1038/mp.2014.69>
- Ito H, Yoshimura N, Kurosawa M, Ishii S, Nukina N, Okazawa H (2009) Knock-down of PQBP1 impairs anxiety-related cognition in mouse. *Hum Mol Genet* 18:4239–4254. <https://doi.org/10.1093/hmg/ddp378>
- Jolly LA, Homan CC, Jacob R, Barry S, Gecz J (2013) The UPF3B gene, implicated in intellectual disability, autism, ADHD and childhood onset schizophrenia regulates neural progenitor cell behaviour and neuronal outgrowth. *Hum Mol Genet* 22:4673–4687. <https://doi.org/10.1093/hmg/ddt315>
- Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7:1511–1522. <https://doi.org/10.1038/nprot.2012.085>
- Kalscheuer VM, Freude K, Musante L, Jensen LR, Yntema HG, Gécz J, Sefiani A, Hoffmann K, Moser B, Haas S, Gurok U, Haesler S, Aranda B, Nshedjan A, Tzschach A, Hartmann N, Roloff T-C, Shoichet S, Hagens O, Tao J, Van Bokhoven H, Turner G, Chelly J, Moraine C, Fryns J-P, Nuber U, Hoeltzenbein M, Scharff C, Scherthan H, Lenzner S, Hamel BCJ, Schweiger S, Ropers H-H (2003) Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. *Nat Genet* 35:313–315. <https://doi.org/10.1038/ng1264>
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36. <https://doi.org/10.1186/gb-2013-14-4-r36>
- Kleefstra T, Franken CE, Arens YHJM, Ramakers GJA, Yntema HG, Sistermans EA, Hulsmans CFCH, Nillesen WN, van Bokhoven H, de Vries BBA, Hamel BCJ (2004) Genotype-phenotype studies in three families with mutations in the polyglutamine-

- binding protein 1 gene (PQBP1). *Clin Genet* 66:318–326.  
<https://doi.org/10.1111/j.1399-0004.2004.00308.x>
- Kunde SA, Musante L, Grimme A, Fischer U, Müller E, Wanker EE, Kalscheuer VM (2011) The X-chromosome-linked intellectual disability protein PQBP1 is a component of neuronal RNA granules and regulates the appearance of stress granules. *Hum Mol Genet* 20:4916–4931. <https://doi.org/10.1093/hmg/ddr430>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Lee BJ, Cansizoglu AE, Süel KE, Louis TH, Zhang Z, Chook YM (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* 126:543–558.  
<https://doi.org/10.1016/j.cell.2006.05.049>
- Lenski C, Abidi F, Meindl A, Gibson A, Platzer M, Frank Kooy R, Lubs HA, Stevenson RE, Ramser J, Schwartz CE (2004) Novel truncating mutations in the polyglutamine tract binding protein 1 gene (PQBP1) cause Renpenning syndrome and X-linked mental retardation in another family with microcephaly. *Am J Hum Genet* 74:777–780.  
<https://doi.org/10.1086/383205>
- Liu X, Dou L-X, Han J, Zhang ZC (2020) The Renpenning syndrome-associated protein PQBP1 facilitates the nuclear import of splicing factor TXNL4A through the karyopherin  $\beta 2$  receptor. *J Biol Chem* 295:4093–4100.  
<https://doi.org/10.1074/jbc.RA119.012214>
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>
- Lubs H, Abidi FE, Echeverri R, Holloway L, Meindl A, Stevenson RE, Schwartz CE (2006) Golabi-Ito-Hall syndrome results from a missense mutation in the WW domain of the PQBP1 gene. *J Med Genet* 43:e30. <https://doi.org/10.1136/jmg.2005.037556>
- Mattioli F, Isidor B, Abdul-Rahman O, Gunter A, Huang L, Kumar R, Beaulieu C, Gez J, Innes M, Mandel J-L, Piton A (2019) Clinical and functional characterization of recurrent missense variants implicated in THOC6-related intellectual disability. *Hum Mol Genet* 28:952–960. <https://doi.org/10.1093/hmg/ddy391>
- Nguyen LS, Jolly L, Shoubridge C, Chan WK, Huang L, Laumonier F, Raynaud M, Hackett A, Field M, Rodriguez J, Srivastava AK, Lee Y, Long R, Addington AM, Rapoport JL, Suren S, Hahn CN, Gamble J, Wilkinson MF, Corbett MA, Gez J (2012) Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability. *Mol Psychiatry* 17:1103–1115.  
<https://doi.org/10.1038/mp.2011.163>
- Okazawa H, Rich T, Chang A, Lin X, Waragai M, Kajikawa M, Enokido Y, Komuro A, Kato S, Shibata M, Hatanaka H, Mouradian MM, Sudol M, Kanazawa I (2002) Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death. *Neuron* 34:701–713. [https://doi.org/10.1016/s0896-6273\(02\)00697-9](https://doi.org/10.1016/s0896-6273(02)00697-9)

- Quartier A, Chatrousse L, Redin C, Keime C, Haumesser N, Maglott-Roth A, Brino L, Le Gras S, Benchoua A, Mandel J-L, Piton A (2018) Genes and Pathways Regulated by Androgens in Human Neural Cells, Potential Candidates for the Male Excess in Autism Spectrum Disorder. *Biol Psychiatry*. <https://doi.org/10.1016/j.biopsych.2018.01.002>
- Redin C, Gérard B, Lauer J, Herenger Y, Muller J, Quartier A, Masurel-Paulet A, Willems M, Lesca G, El-Chehadeh S, Le Gras S, Vicaire S, Philipps M, Dumas M, Geoffroy V, Feger C, Haumesser N, Alembik Y, Barth M, Bonneau D, Colin E, Dollfus H, Doray B, Delrue M-A, Drouin-Garraud V, Flori E, Fradin M, Francannet C, Goldenberg A, Lumbroso S, Mathieu-Dramard M, Martin-Coignard D, Lacombe D, Morin G, Polge A, Sukno S, Thauvin-Robinet C, Thevenon J, Doco-Fenzy M, Genevieve D, Sarda P, Edery P, Isidor B, Jost B, Olivier-Faivre L, Mandel J-L, Piton A (2014) Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *J Med Genet* 51:724–736. <https://doi.org/10.1136/jmedgenet-2014-102554>
- Shen Y, Zhang ZC, Cheng S, Liu A, Zuo J, Xia S, Liu X, Liu W, Jia Z, Xie W, Han J (2021) PQBP1 promotes translational elongation and regulates hippocampal mGluR-LTD by suppressing eEF2 phosphorylation. *Mol Cell* 81:1425-1438.e10. <https://doi.org/10.1016/j.molcel.2021.01.032>
- Stevenson RE, Bennett CW, Abidi F, Kleefstra T, Porteous M, Simensen RJ, Lubs HA, Hamel BCJ, Schwartz CE (2005) Renpenning syndrome comes into focus. *Am J Med Genet A* 134:415–421. <https://doi.org/10.1002/ajmg.a.30664>
- Takahashi K, Yoshina S, Masashi M, Ito W, Inoue T, Shiwaku H, Arai H, Mitani S, Okazawa H (2009) Nematode homologue of PQBP1, a mental retardation causative gene, is involved in lipid metabolism. *PLoS One* 4:e4104. <https://doi.org/10.1371/journal.pone.0004104>
- Tapia VE, Nicolaescu E, McDonald CB, Musi V, Oka T, Inayoshi Y, Satteson AC, Mazack V, Humbert J, Gaffney CJ, Beullens M, Schwartz CE, Landgraf C, Volkmer R, Pastore A, Farooq A, Bollen M, Sudol M (2010) Y65C missense mutation in the WW domain of the Golabi-Ito-Hall syndrome protein PQBP1 affects its binding activity and deregulates pre-mRNA splicing. *J Biol Chem* 285:19391–19401. <https://doi.org/10.1074/jbc.M109.084525>
- Wang Q, Moore MJ, Adelmant G, Marto JA, Silver PA (2013) PQBP1, a factor linked to intellectual disability, affects alternative splicing associated with neurite outgrowth. *Genes Dev* 27:615–626. <https://doi.org/10.1101/gad.212308.112>
- Waragai M, Junn E, Kajikawa M, Takeuchi S, Kanazawa I, Shibata M, Mouradian MM, Okazawa H (2000) PQBP-1/Npw38, a nuclear protein binding to the polyglutamine tract, interacts with U5-15kD/dim1p via the carboxyl-terminal domain. *Biochem Biophys Res Commun* 273:592–595. <https://doi.org/10.1006/bbrc.2000.2992>
- Waragai M, Lammers CH, Takeuchi S, Imafuku I, Udagawa Y, Kanazawa I, Kawabata M, Mouradian MM, Okazawa H (1999) PQBP-1, a novel polyglutamine tract-binding protein, inhibits transcription activation by Brn-2 and affects cell survival. *Hum Mol Genet* 8:977–987. <https://doi.org/10.1093/hmg/8.6.977>

Yang S-S, Ishida T, Fujita K, Nakai Y, Ono T, Okazawa H (2020) PQBP1, an intellectual disability causative gene, affects bone development and growth. *Biochem Biophys Res Commun* 523:894–899. <https://doi.org/10.1016/j.bbrc.2019.12.097>

Yoh SM, Schneider M, Seifried J, Soonthornvacharin S, Akleh RE, Olivieri KC, De Jesus PD, Ruan C, de Castro E, Ruiz PA, Germanaud D, des Portes V, García-Sastre A, König R, Chanda SK (2015) PQBP1 Is a Proximal Sensor of the cGAS-Dependent Innate Response to HIV-1. *Cell* 161:1293–1305. <https://doi.org/10.1016/j.cell.2015.04.050>

A two base pair deletion in the PQBP1 gene is associated with microphthalmia, microcephaly, and mental retardation - PubMed. <https://pubmed-ncbi-nlm-nih-gov.proxy.insermbiblio.inist.fr/17033686/>. Accessed 1 Jul 2021

## Figures

### Figure 1. Variants in *PQBP1* in individuals with ID

(A) Schématic representation of PQBP1 protein with its different domains (WW: WW domain, PRD: Proline-rich domain, NLS: nuclear localization signal, CTD :C-terminal domain) and the different variants analyzed in this study (on the top: missense variants, on the bottom : truncating variants ; in green : a missense variants found in two males from gnomAD populations and considered therefore as non-disease causing ; in red one variant reported as disease causing (Lubs 2006); in orange 2 variants considered as likely pathogenic ; in blue 3 variants of unknown significance). (B-C) Level of PQBP1 WT and mutant proteins extracted from control (in green), patients' carrying pathogenic variants in *PQBP1* (in red) and patients carrying VUS in *PQBP1* (in blue) LCL (B) or overexpressed in HeLA cells (C), (normalized with Beta tubuline).

### Figure 2. *PQBP1* inactivation leads to an increase of a short 3' isoform of *UPF3B*.

(A) Schematic representation of the different exons composing the canonical longest isoform of *PQBP1* (ENST00000636792, alias UPF3B\_L) as well as the shorter 3' isoform (ENST00000276201, alias UPF3B\_S). IGV visualization of RNAseq data from one control individual and one patient are represented. The stars indicate the exons detected as significantly increased in patients vs controls (DEXseq); qPCR primer couples used for RT-qPCR validation are represented by black arrows.(B) RT-qPCR analysis of UPF3B 3' short isoform (*UPF3B\_S*)



(normalized on GAPDH and YWHAZ) in LCL from control individuals (Cont, in green), individuals with pathogenic variants in *PQBP1* (Patho, in red) and in individuals affected from X-fragile syndrome (FXS, in yellow) ( $n \geq 2$  per cell line) (C) and in a second set of LCL from controls individuals (in green), 2 individuals with the same frameshift variants in *PQBP1* (in red) and individual affected from X-fragile syndrome (in yellow) ( $n=3$  per line). (D) RT-qPCR analysis of UPF3B 3' short (*UPF3B\_S*) isoform (normalized on GAPDH and YWHAZ) in both hNSCs cells lines (hNSCs-1 and hNSCs-2) after a transient knock-down (KD) of *PQBP1* ( $n=3$  per line). Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; \*\*\*: p-value <0.001; Errors bars represent SEM (standard error of the mean).

**Figure 3. Identification of Differentially Expressed Genes (DEG) in human Neural Stem Cells (hNSCs) knock-down (KD) for *PQBP1* and effect on the global proliferation.**

(A) Volcano plot showing RNA sequencing data (INTERFERin VS siPQBP1 treated hNSCs-1). Genes not significantly deregulated are shown in black; genes with an adjusted p-value are between 0.5 and 0.1 are shown in orange while genes significantly deregulated (adjusted p-value <0.05) are shown in red. (B) RT-qPCR analysis of 24 DEGs (normalized on GAPDH and YWHAZ) after knock-down (KD) of *PQBP1* in a third serie of hNSC-1 and in 3 series of hNSC-2. Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; \*: p-value <0.01; \*\*: p-value<0.01 \*\*\*: p-value <0.001; Errors bars represent SEM. (C) Proliferation assay performed on both hNSC lines (hNSC-1 and hNSC-2) treated with lipofectant alone (INTERFERin) or transfected with *Scramble* siRNA (siScramble); *PQBP1* siRNA (siPQBP1) and *PLK1* siRNA (siPLK1), a transfection control leading to apoptosis of the cells. At each time point (day 0, 1, 2, 3, and 4), cells were counted and data normalized by INTERFERin treatment ( $n=3$  per line). Student's t test comparison was done comparing to INTERFERin: \*: p-value<0.05; Errors bars represent SEM.

**Figure 4. Representation and protein interactomes of UPF3B protein isoforms.**

(A) Schematic representation of UPF3B long (UPF3B\_L) and UPF3B 3' short (UPF3B\_S) proteins with their domains known to interact with UPF2, GSTP1 and RBM8A proteins. (B) Protein interactome of UPF3B long (UPF3B\_L) and 3' short (UPF3B\_S) isoforms characterized by immunoprecipitation coupled to mass spectrometry (IP-MS) in HEK293 cells.

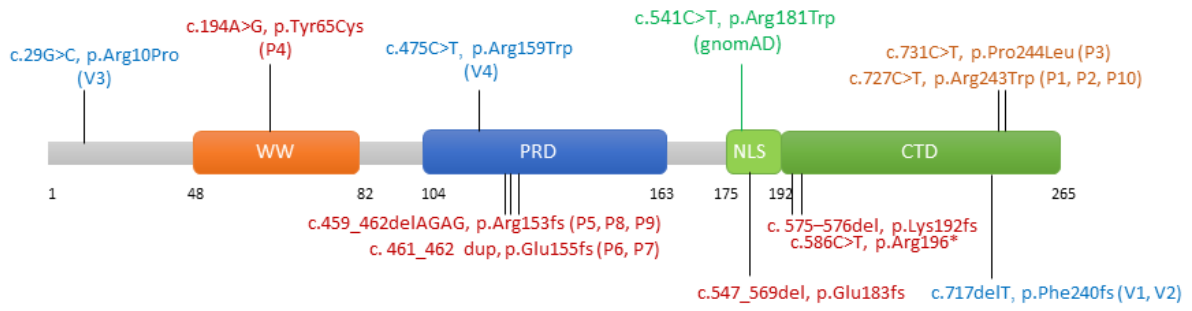
In orange: interactors specific to UPF3B\_L; in Blue: interactors specific to UPF3B\_S; in green: interactors of both UPF3B\_L and UPF3B\_S isoforms. Protein already described as interacting with UPF3B\_L are circled with a black line.

**Figure 5. UPF3B\_S is a robust marker of *PQBPI* loss of function and can be used for VUS testing**

(A) RT-qPCR analysis of UPF3B 3' short isoform (*UPF3B\_S*) (normalized on GAPDH and YWHAZ) in LCL from control individuals (in green), individuals with VUS in *PQBPI* (in blue) and in one individual affected from X-fragile syndrome (FXS, in yellow) (n=3 per cell line). control LCL from the first set were used a calibrator. (B) RT-qPCR analysis of UPF3B 3' short isoform (*UPF3B\_S*) (normalized on GAPDH and YWHAZ) in blood RNA samples (Paxgene) from controls individuals (in green), individuals with pathogenic variants in *PQBPI* (in red), individual with VUS in *PQBPI* (in blue) and individuals affect from X-Fragile syndrome (in yellow) (n=1 per Paxgene). (C) RT-qPCR analysis of endogenous *PQBPI* mRNA or (D) UPF3B 3' short isoform (*UPF3B\_S*) (normalized on GAPDH and YWHAZ) in HeLa cells transfected co-transfected by scramble or *PQBPI* siRNA and overexpressing WT or truncated (Arg153fs) *PQBPI* protein. Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; \*\*: p-value <0.01; \*\*\*: p-value <0.001; Errors bars represent SEM.

Figure 1

A



B

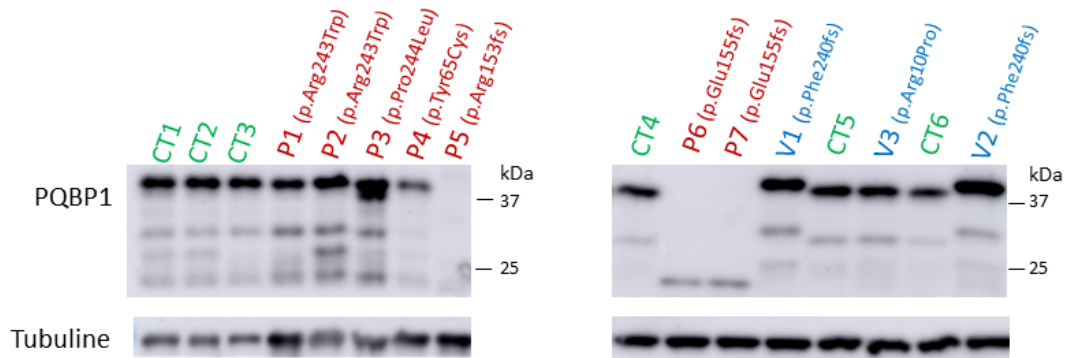


Figure 2

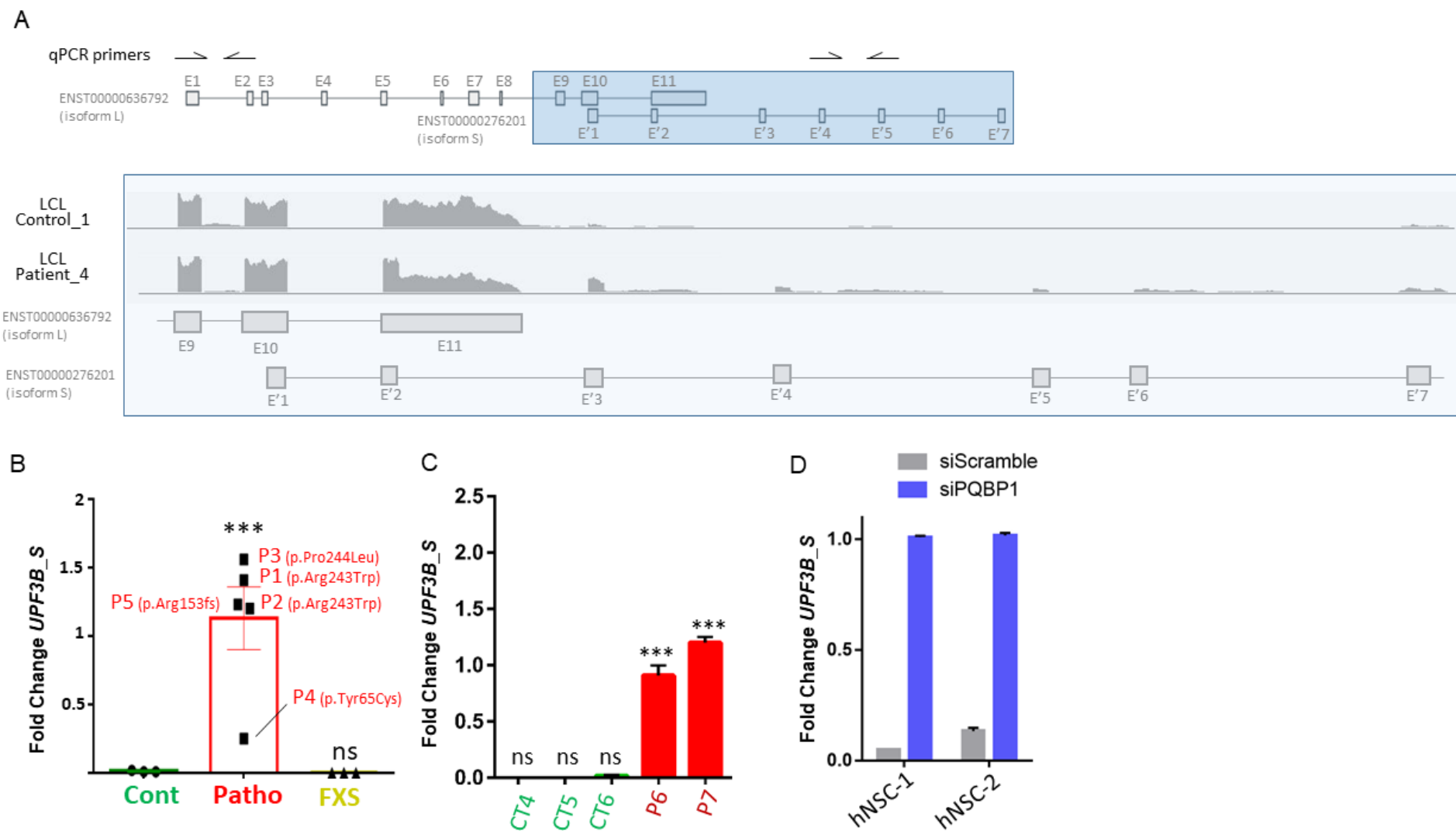


Fig 2

Figure 3

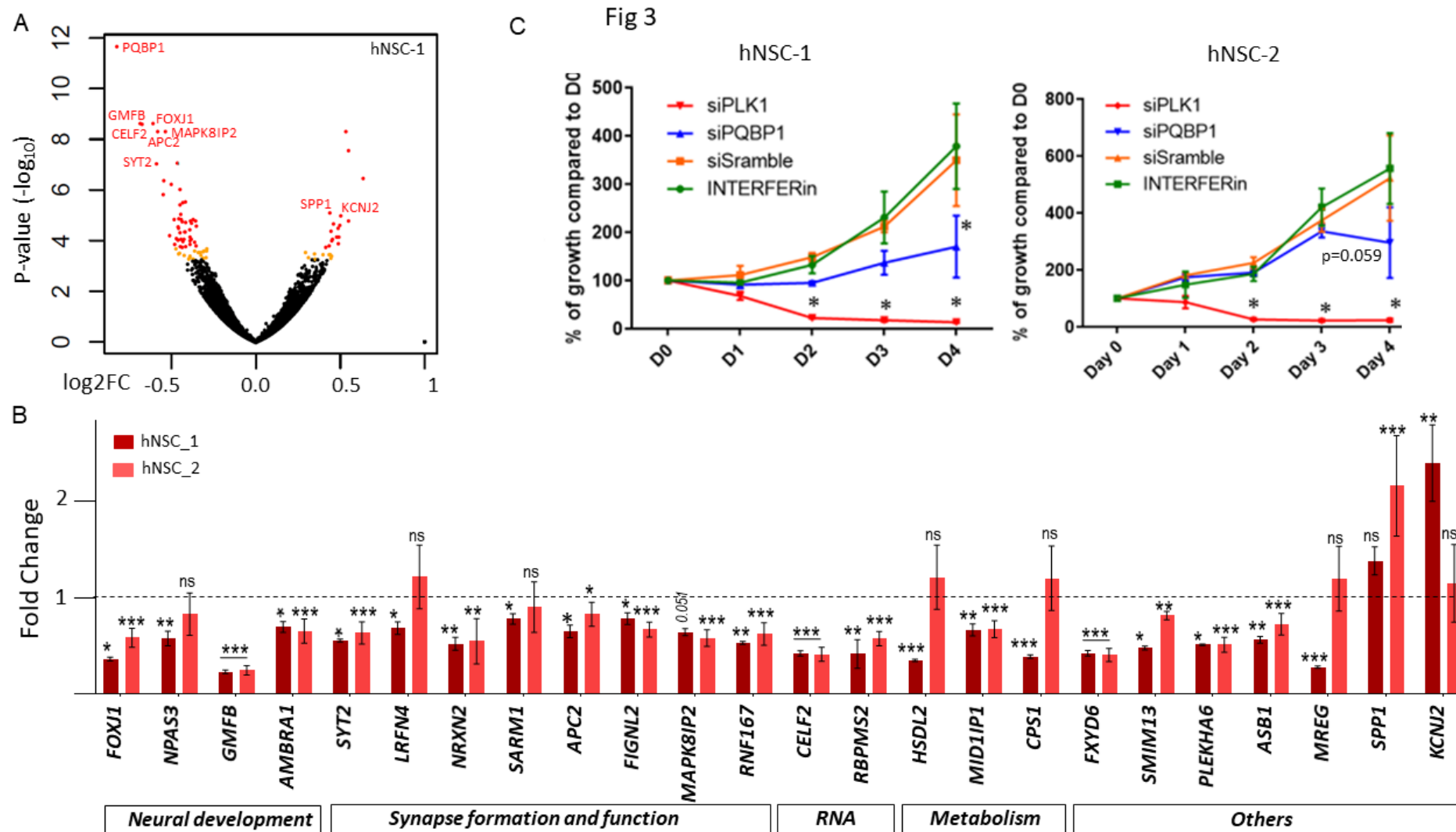
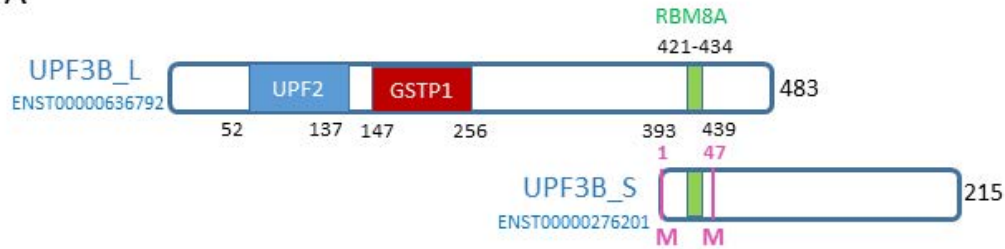


Figure 4

A



B

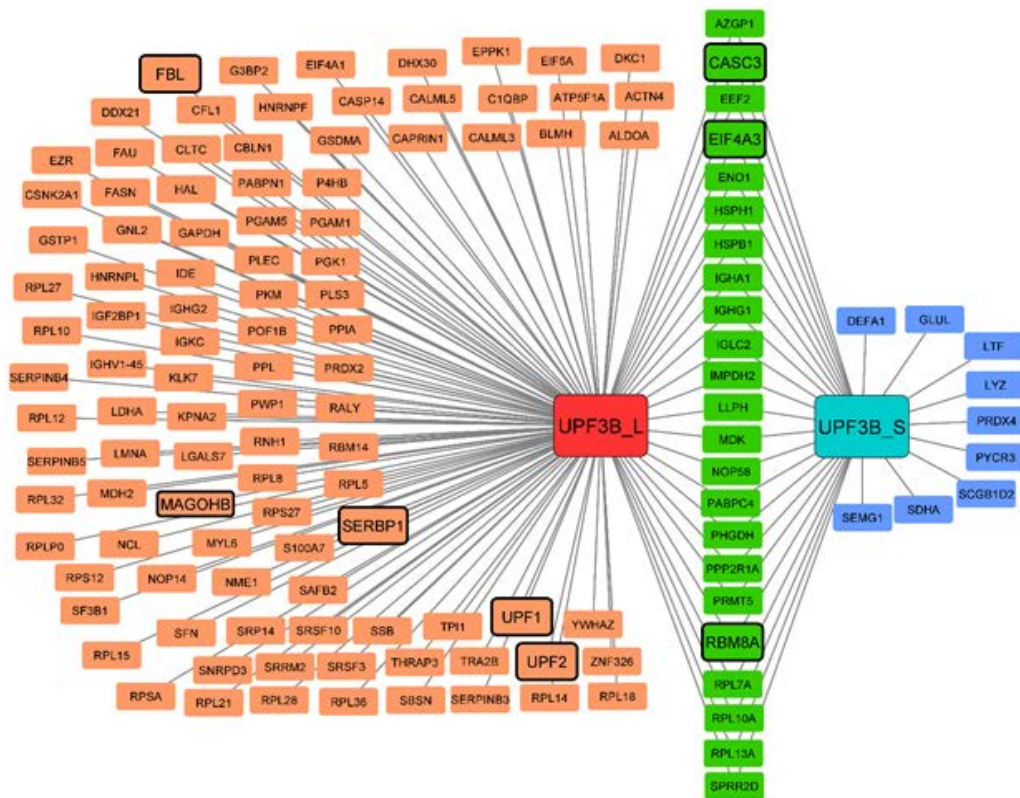
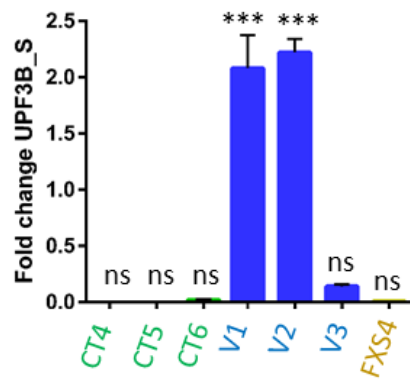
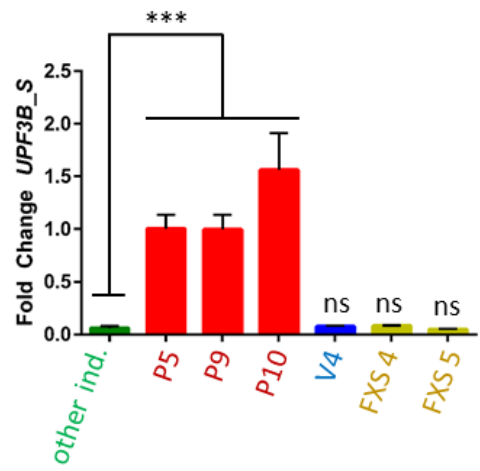


Figure 5

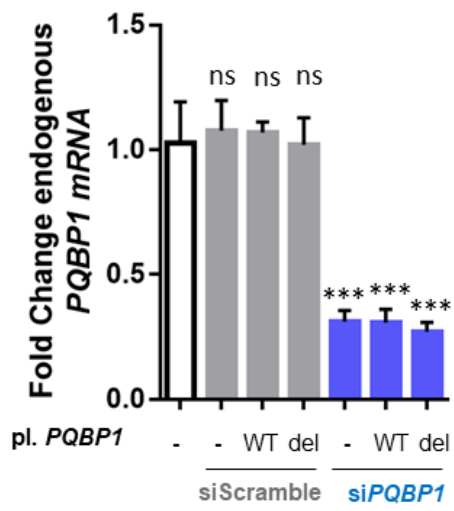
A.



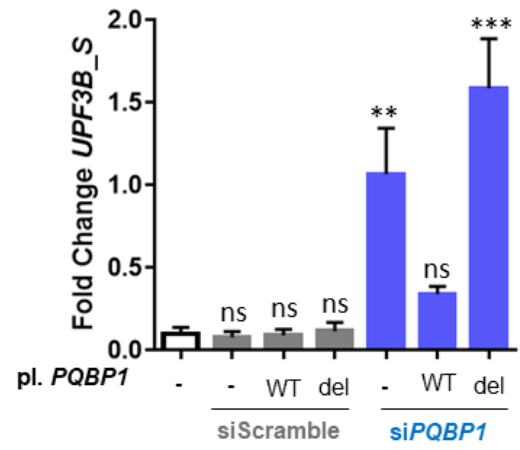
B.



C.



D.



## Supplementaries

### Figure S1. Transcriptomic analysis of controls LCL or carrying *PQBP1* variants and investigation of *APP* and *TSPAN5* mRNA levels.

(A) Principal Component Analysis of transcriptomic analysis performed on LCL from control individuals (CT1-3) and individuals carrying *PQBP1* variant (P1-5). (B) Volcano plot showing changes in gene expression between control individuals and individuals carrying *PQBP1* variants. P-value ( $-\log_{10}$ ) is expressed for each gene in function of  $\log_2$ fold change ( $\log_2$ FC). Genes not significantly deregulated are shown in black; while significantly deregulated genes (adjusted p-value  $<0.05$ ) are shown in red. (C) RT-qPCR analysis of *APP* mRNA level (normalized on GAPDH and YWHAZ) in LCL from the first set of control individuals (Cont, in green), individuals with pathogenic variants in *PQBP1* (Patho, in red) and individuals affected by X-fragile syndrome (FXS, in yellow) (n=1 per cell line). (D) Western Blot analysis on total protein extract from controls and individuals carrying *PQBP1* variants (LCL) using APP antibody and beta-Tubulin antibody as a control. (E) RT-qPCR analysis of *APP* mRNA level (normalized on GAPDH and YWHAZ) in LCL from additional control individuals (CT1-4, in green), individuals varying *PQBP1* pathogenic variants (P6-7, in red) or VUS (V1-3, in blue) and individual affected by X-fragile syndrome (FSX4, in yellow) (n=3 per cell line); control LCL from the first set were used a calibrator. (F) RT-qPCR analysis of *TSPAN5* mRNA level (normalized on GAPDH and YWHAZ) in LCL from the first set of control individuals (Cont, in green) and individuals varying *PQBP1* pathogenic variants (Patho, in red). (G) RT-qPCR analysis of *TSPAN5* mRNA (normalized on GAPDH and YWHAZ) in LCL from additional control individuals (CT4-6, in green), individuals carrying *PQBP1* pathogenic variants (P6-7, in red) or VUS (V1-3, in blue) and in one individual affected by X-fragile syndrome (FXS4, in yellow) (n=3 per cell line); control LCL from the first set were used a calibrator. Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; \*\*\*: adjusted p-value $<0.001$ ; Errors bars represent SEM (standard error of the mean).

### Figure S2. Expression of *UPF3B* isoforms

(A) Visualization of DEXseq results for the *UPF3B* gene. The normalized counts are represented for control individuals and individuals carrying *PQBP1* pathogenic variants. Exons found significantly upregulated (adjusted p-value  $<0.05$ ) in patients after DEXseq analysis are



labelled with a single \*. Genomic positions are given according to hg38 version of the genome. **(B)** RT-qPCR analysis of *UPF3B* long isoform (*UPF3B\_L*) mRNA level (normalized on *GAPDH* and *YWHAZ*) in LCL from control individuals (Cont, in green), individuals with pathogenic variants in *PQBPI* (Patho, in red) and individuals affected by X-fragile syndrome (FXS, in yellow) (n =1 per cell line). **(C)** RT-qPCR analysis of *UPF3B\_L* mRNA level (normalized on *GAPDH* and *YWHAZ*) in both hNSC cells lines (hNSC-1 and hNSC-2) after a transient knock-down (KD) of *PQBPI* (n=3 per line). For **(B)** and **(C)**, multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; Errors bars represent SEM (standard error of the mean).

**Figure S3. *PQBPI* siRNA transfection leads to decrease of *PQBPI* expression at the mRNA and protein level.**

**(A)** RT-qPCR analysis of *PQBPI* mRNA level (normalized on *GAPDH* and *YWHAZ*) performed on both hNSC lines (hNSC-1 and hNSC-2) transfected with *Scramble* siRNA (siScramble) or *PQBPI* siRNA (siPQBPI) (n=3 per line); Multiple comparisons tests were performed using one-way ANOVA test with Dunnet's correction: ns: not significant; \*\*: adjusted p-value<0.01; Errors bars represent SEM (standard error of the mean). **(B)** Western Blot analysis on total protein extract from hNSCs-1 line treated with lipofectant alone (INF) or transfected with *Scramble* or *PQBPI* siRNA.

**Figure S4. *SIX3* expression is not altered in LCL from individuals carrying pathogenic variant in *PQBPI***

RT-qPCR analysis of *SIX3* mRNA level (normalized on *GAPDH* and *YWHAZ*) in LCL from control individuals (Cont, in green) and individuals carrying pathogenic variants in *PQBPI* (Patho, in red) (n=1 per line). Student's t-test was performed; ns : not significant ; Errors bars represent SEM (standard error of the mean).

**Supplementary tables**

Sup Table 1. List of the samples and variants analyzed in this study

Sup Table 2. List of the primers used for RT-qPCR analysis

Sup Table 3. Genes differentially expressed in LCL from individuals with pathogenic variants in *PQBPI* identified by DEseq analysis

Sup Table 4. Exons differentially expressed in LCL from individuals with pathogenic variants in *PQBPI* identified by DEXseq analysis

Sup Table 5. Genes differentially expressed in hNSC-1 after *PQBPI* knock-down (KD) identified by DEseq analysis

Sup Table 6. Protein interactors of UPF3B\_L and UPF3B\_L identified by Immunoprecipitation coupled to Mass Spectrometry

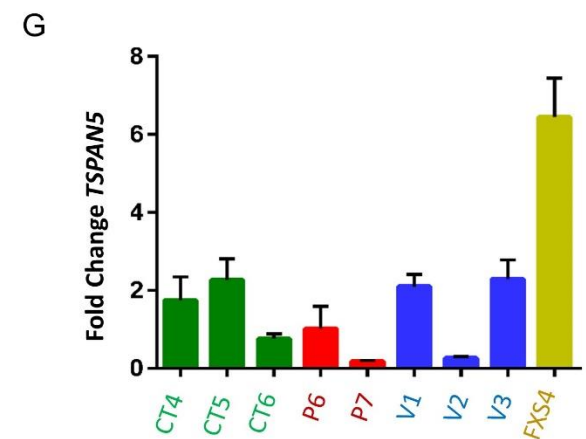
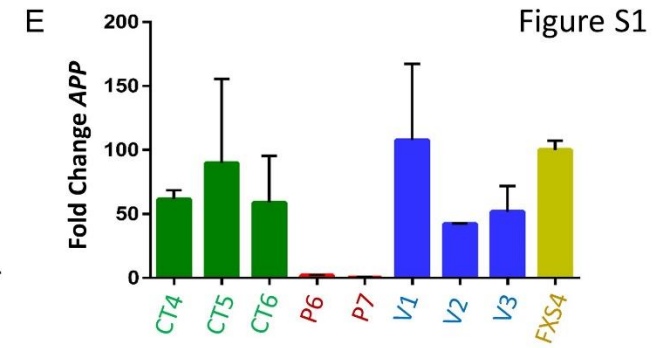
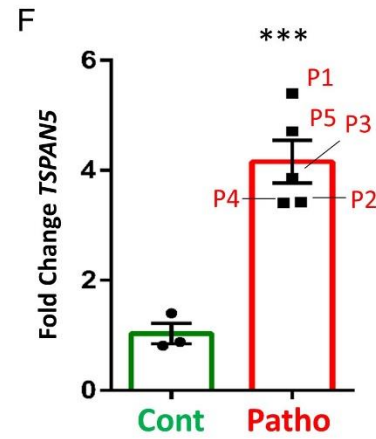
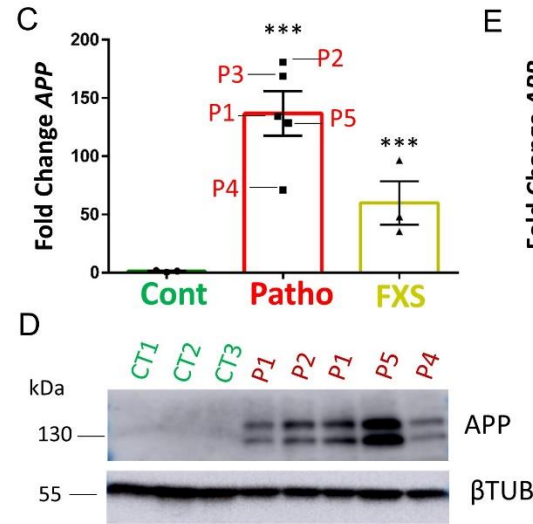
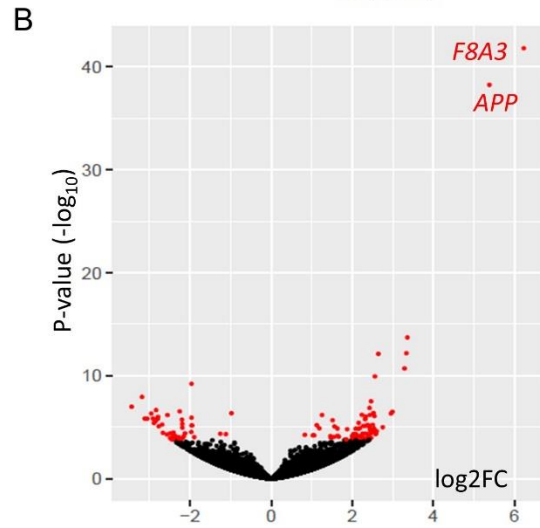
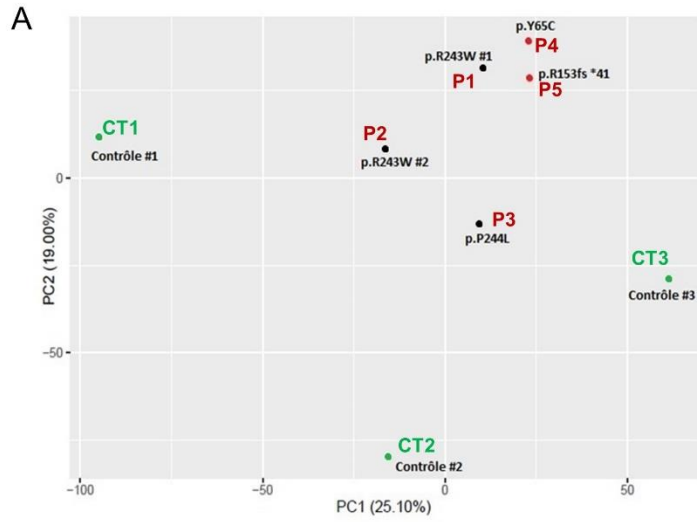


Figure S2

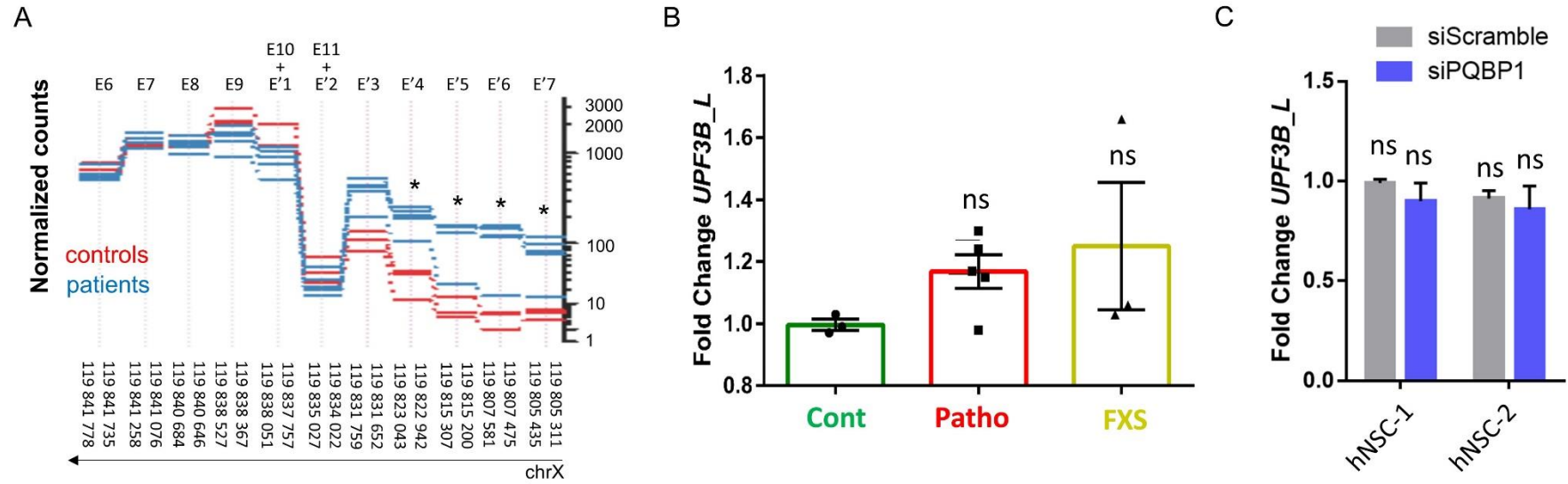
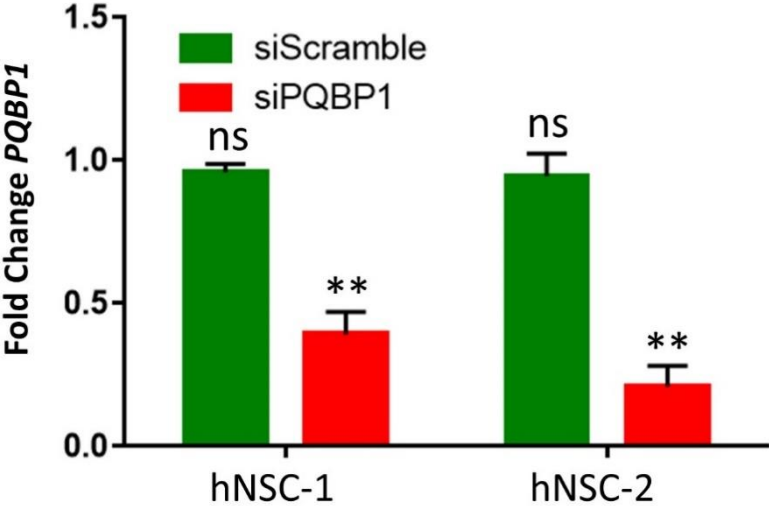


Figure S3

A



B

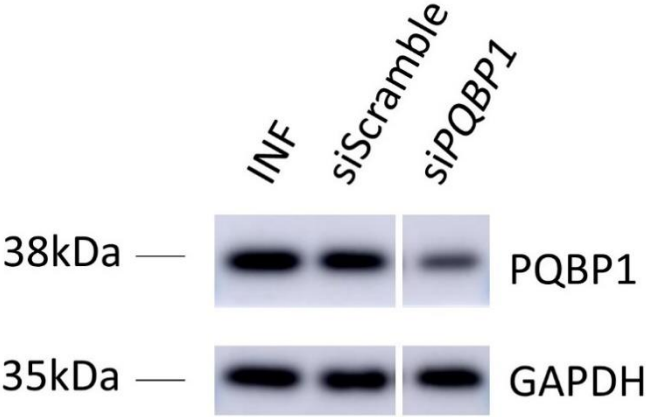


Figure S4

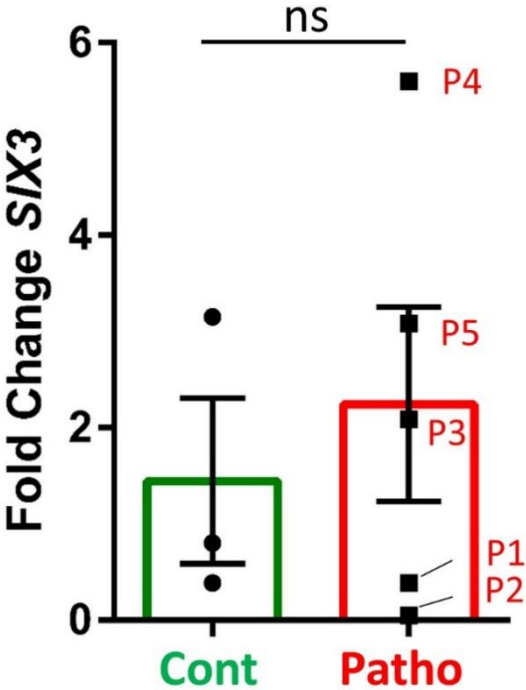


Table S1

Individual	variant	reporting	LCL	RNASeq	PaxGene	Plasmid
P1	c.727C>T, p.Arg243Trp	Hu et al, 2016	Y	Y	N	Y
P2	c.727C>T, p.Arg243Trp	Hu et al, 2016	Y	Y	N	Y
P3	c.731C>T, p.Pro244Leu	Redin et al., 2014	Y	Y	N	Y
P4	c.194A>G, p.Tyr65Cys	Lubs et al., 2006	Y	Y	N	Y
P5	c.459_462delAGAG, p.Arg153fs	Kalscheuer et al., 2003	Y	Y	N	Y
<b>P6</b>	c.461_462dup, p.Glu155fs	Kalscheuer et al., 2003	Y	N	N	Y
<b>P7</b>	c.461_462dup, p.Glu155fs	Kalscheuer et al., 2003	Y	N	N	Y
P8	c.459_462delAGAG, p.Arg153fs	Kalscheuer et al., 2003	Y	N	N	Y
P9	c.459_462delAGAG, p.Arg153fs	<i>this report</i>	N	N	Y	Y
P10	c.727C>T, p.Arg243Trp	<i>this report</i>	N	N	Y	Y
V1	c.717delT, p.Phe240fs	Hu et al, 2016	Y	N	N	Y
V2	c.717delT, p.Phe240fs	Hu et al, 2016	Y	N	N	Y
V3	c.29G>C, p.Arg10Pro	<i>this report</i>	Y	N	N	Y
V4	c.475C>T, p.Arg159Trp	<i>this report</i>	N	N	Y	Y
FXS1	<i>FMR1</i> : del exon 17	Quartier et al. 2017	Y	N	N	N
FXS2	<i>FMR1</i> : CGG expansion	-	Y	N	N	N
FXS3	<i>FMR1</i> :del exon 17	Quartier et al. 2017	Y	N	N	N
FXS4	<i>FMR1</i> : c.990+1G>A	Quartier et al. 2017	Y	N	N	N
C1	-	-	Y	Y	N	N
C2	-	-	Y	Y	N	N
C3	-	-	Y	Y	N	N
C4	-	-	Y	N	N	N
C6	-	-	Y	N	N	N
C5	-	-	Y	N	N	N
-	c.541C>T, p.Arg181Trp	gnomAD	-	-	-	Y

Table S2

Gene/isoform	primer Foward 5'-> 3'	primer Reverse 5'-> 3'
UPF3B_L	ATGTTCTGAAGCTGCTCCT	GGGAAGATAAGCAGGATCGC
UPF3B_S	GGTTATGTCATCAGGGTGCG	TTGATGGCAACAATATCCCCT
YWHAZ	ACTTTTGGTACATTGTGGCTTCAA	CCGCCAGGACAAACCAGTAT
GAPDH	AATCCCATCACCATCTTCCA	TGGACTCCACGACGTACTCA
APP	TCCATTTCATCATGGTGTGGT	GTTCTGCTGCATCTTGGACA
AMBRA1	CTTTGTCCCAAGGCGTTTCT	AGAGTGAGTAGCCAAGCCAG
APC2	CCTGGTGCAGTGTGTCCT	GCCTTGTCTCTGTGCTGTG
ASB1	CATCTGATTCTTCGCTGCC	CTCACTGGAATCAACCGCAG
CELF2	TGTCACATTTTCTACAAGGGCA	TGTCAGCAAACCTCACCCAG
CPS1	TTCGGAGGACAGCTGTTGAT	GCGAGATTTCTGCACAGCTT
FIGNL2	CACTGCCTCCAACCTCCTAA	TTGAGGAAGGAGGCGTCCG
FOXJ1	CCTGTCGGCCATCTACAAGT	GTTTCAGAGACAGGTTGTGGC
FXD6	CTGTGGTCTCTTCTCGGTT	CATTGGCGGTGATGAGGTTT
GMFB	GGTTGTTTGTGATGTTGCCG	AATGCCCTCAAGCTCCTCAT
HSDL2	GGCACAATCTATACTGCTGCT	CTTTCTCCACTGCAGCACTG
KCNJ2	GACCCAGACAACCATAGGCT	CATGACTGCGCCAATGATGA
LRFN4	CCAGTACAACAGCAGCGAAG	AGGCAGAGGTCATAGTCAGC
MAPK8IP2	GTCAACAGCACATCTCGGTC	GATGAACCTGAACACAGCCC
MID1IP1	ATCTGCGACACCTACAACCA	CTGGTCCATGTTGTTACGG
MREG	ATCAGCAGGCCAAAGACTCA	TACTTCCCTTCGAACCTGCC
NPAS3	ATCAGGCACAGTCACTTGGA	TAGCTATGGTGGCACTGGAC
NR2F1	CAATACTGCCGCTCAAGAA	CTGGGTTGGAGGCATTCTTC
NRXN2	GCTGTCTCCAGTGGCCAAGAA	ACTTCCAGAGGGATGGGCGAAA
PLEKHA6	CAATGCACCTGTCACCAAGG	AGTAGAAGAGGCAGCGATCC
PRDM1	CCTGGGATTCTGGTGTGAT	AACGTGTGCCCTTTGGTATG
RBPM2	ACTCTACTTGCTCTTCCGGC	CACGGCTGTCAAAGATCACA
RNF167	TATGGTGCCGCTGTAGTACA	AGATCTGCTGCTGGATTTCC
SARM1	ATCGGGGCTTTCTACCTCTG	GCCGACTTAGTGCCATTGG
SERPINE1	CCGCTCTTCCACAAATCAG	GGGCGTGGTGAACCTCAGTAT
SMIM13	CCACACAGAATCAGATCCGC	ACTGCCTTTTCATCCTTACA
SPP1	ACTACCATGAGAATTGCAGTGA	GAACTTCCAGAATCAGCCTGT
SPTBN2	TGTGGTGCCAGATGAAGACT	GTCTGGCCGGTGTATATGC
SRDSP1	GCAGACCCACTCAGTGTGTTG	GGTCATCTGTTCCAAACGCA
STMM3	GCACCATTTCCGCTACAAG	TGGTAGACGGTATTGGGGTG
SYT2	GGAGGAGGAGAAAGAGCCAG	AGAACGCCACAGTAAGCT



Table S3

Gene	Description	Normalized reads								Log2 Fold Change	Adjusted p-value	
		C1	C2	C3	P1_p.Arg243Trp	P2_p.Arg243Trp	P3_p.Pro244Leu	P4_p.Tyr65Cys	P5_p.Arg153fs			
F8A3	coagulation factor VIII-associated 3		2	2	3	609	610	256	268	649	6,23	6,2E-38
APP	amyloid beta precursor protein	281	162	83	11013	14784	18819	7923	9383		5,38	1,1E-34
IFNG	interferon, gamma	12	39	9	351	325	370	177	294		3,36	2,5E-10
ASB9	ankyrin repeat and SOCS box containing 9	16	4	16	120	181	103	261	276		3,33	5,9E-09
F13A1	coagulation factor XIII A chain	2741	3299	5600	21396	19615	24103	35873	44559		2,64	5,9E-09
IL12RB2	interleukin 12 receptor subunit beta 2	34	53	77	567	463	350	448	180		2,56	6,6E-07
EIF5AL1	eukaryotic translation initiation factor 5A-like 1	79	129	160	33	21	25	27	36		-1,96	3,0E-06
MARK1	microtubule affinity regulating kinase 1	180	229	63	5	21	1	9	8		-3,18	5,0E-05
TMEM37	transmembrane protein 37	44	16	65	267	182	220	390	474		2,46	1,2E-04
ADGRV1	adhesion G protein-coupled receptor V1	22660	137	58	42	66	36	82	33		-3,44	3,6E-04
CCNA1	cyclin A1	4	5	6	53	17	24	36	45		2,43	4,4E-04
LILRA4	leukocyte immunoglobulin like receptor A4	17	22	14	160	150	36	187	244		2,52	9,6E-04
SCIMP	SLP adaptor and CSK interacting membrane protein	21477	21020	18608	12391	8910	10157	9729	9653		-0,98	9,6E-04
DNAJC12	DnaJ heat shock protein family (Hsp40) member C12	1373	675	51	41	60	54	10	37		-2,96	9,6E-04
EPHB1	EPH receptor B1	1514	416	3028	17304	6577	10761	12928	17192		2,43	1,1E-03
TRIM9	tripartite motif containing 9	35	26	27	355	66	155	193	135		2,24	1,1E-03
EFCAB2	EF-hand calcium binding domain 2	797	660	545	1634	1608	1805	1222	1978		1,25	1,1E-03
PPARG	peroxisome proliferator activated receptor gamma	38	50	91	243	274	239	905	1154		2,55	1,4E-03
TUBA8	tubulin alpha 8	8	10	13	70	42	73	143	31		2,33	1,4E-03
FOS	Fos proto-oncogene, AP-1 transcription factor subunit	1393	1644	134	63	53	191	33	53		-2,79	1,5E-03
CTB-133G6.1		124	391	143	61	43	53	37	44		-1,96	1,7E-03
HEMGN	hemogen	94	59	1	2	3	3	3	3		-2,92	1,9E-03
DIRAS1	DIRAS family GTPase 1	415	1866	103	55	108	32	20	59		-2,80	2,0E-03
IFNA13	interferon, alpha 13	4	5	4	27	38	14	40	34		2,32	2,1E-03
MTTP	microsomal triglyceride transfer protein	736	399	697	2596	1465	1579	2178	1569		1,51	2,4E-03
CORO2A	coronin 2A	20	127	4	3	6	1	1	3		-2,85	3,0E-03
RASSF8	Ras association domain family member 8	27	30	49	168	75	154	277	375		2,16	4,1E-03
TUSC1	tumor suppressor candidate 1	16	54	102	0	3	0	3	6		-2,89	4,1E-03
SRL	sarcalumenin	6	8	4	0	1	0	0	0		-2,69	5,6E-03
CXCL9	C-X-C motif chemokine ligand 9	16	17	9	53	35	267	73	150		2,38	5,7E-03
GALNT10	polypeptide N-acetylgalactosaminyltransferase 10	6850	6002	7398	18198	10724	14428	18015	14538		1,12	5,9E-03
FAM167A	family with sequence similarity 167 member A	805	758	367	62	203	109	213	89		-1,96	6,0E-03
SMPDL3B	sphingomyelin phosphodiesterase acid like 3B	7	19	1	96	62	33	67	84		2,32	6,4E-03
ANXA3	annexin A3	12	33	20	273	160	35	130	386		2,42	6,4E-03
TSPAN5	tetraspanin 5	407	292	485	980	939	2101	985	1409		1,56	6,8E-03
DOC2A	double C2 domain alpha	23	38	24	8	4	4	2	4		-2,19	8,6E-03
CLIC2	chloride intracellular channel 2	1199	321	314	2663	1865	7310	4298	2891		2,15	9,6E-03
CACNA1E	calcium voltage-gated channel subunit alpha1 E	80	128	105	665	318	201	1291	498		2,08	9,6E-03
PLXDC2	plexin domain containing 2	17	12	86	193	341	118	984	887		2,53	9,8E-03
ITGA1	integrin subunit alpha 1	250	619	273	1520	817	1881	2973	1306		1,87	1,0E-02
RAB3C	RAB3C, member RAS oncogene family	4	11	13	101	76	153	645	20		2,60	1,5E-02
KCNJ2	potassium voltage-gated channel subfamily J member 2	4	1	2	5	8	28	81	40		2,51	1,5E-02
PCDHGB4	protocadherin gamma subfamily B, 4	67	19	44	7	15	11	4	5		-1,97	1,8E-02
HYDIN	HYDIN, axonemal central pair apparatus protein	9	41	13	1	5	1	1	2		-2,41	2,0E-02
MYRIP	myosin VIIA and Rab interacting protein	1	15	17	96	129	135	32	347		2,48	2,0E-02
INPP5J	inositol polyphosphate-5-phosphatase J	4	10	50	3	2	0	1	2		-2,49	2,2E-02
HEBP1	heme binding protein 1	7	54	17	109	470	81	149	418		2,33	2,2E-02
TNFSF15	tumor necrosis factor superfamily member 15	10	6	35	286	59	46	114	189		2,26	2,2E-02

ADGRG2	adhesion G protein-coupled receptor G2	2	7	5	10	46	13	89	41	2,26	2,3E-02
IL17RB	interleukin 17 receptor B	664	1577	516	2845	4569	1834	11708	7265	2,09	2,3E-02
SH3D21	SH3 domain containing 21	319	626	383	206	208	197	190	181	-1,12	2,3E-02
CTNNA2	catenin alpha 2	11092	369	207	92	93	179	513	168	-2,57	2,5E-02
CFAP69	cilia and flagella associated protein 69	26	12	44	165	112	42	257	271	2,08	2,6E-02
NAPEPLD	N-acyl phosphatidylethanolamine phospholipase D	722	530	545	1223	1033	971	1049	1124	0,83	2,6E-02
CCDC73	coiled-coil domain containing 73	37	31	30	69	57	88	60	80	1,06	2,8E-02
PIK3R3	phosphoinositide-3-kinase regulatory subunit 3	2263	2046	3104	4170	5165	4418	6041	6199	1,03	2,8E-02
GDA	guanine deaminase	25	13	156	957	205	549	1957	353	2,39	2,9E-02
COL6A6	collagen type VI alpha 6	3	8	2	0	0	0	1	0	-2,42	3,0E-02
IL12B	interleukin 12B	12	59	40	552	143	334	185	72	2,14	3,0E-02
GBP5	guanylate binding protein 5	547	208	558	2442	465	3878	2174	3245	2,00	3,1E-02
CSRP2	cysteine and glycine rich protein 2	237	203	274	944	330	589	776	979	1,46	3,1E-02
GPM6A	glycoprotein M6A	173	2170	386	59	54	181	135	18	-2,36	3,1E-02
C4BPB	complement component 4 binding protein beta	58	65	154	652	352	109	752	805	2,02	3,3E-02
SLC2A13	solute carrier family 2 member 13	297	72	300	1211	686	703	817	758	1,67	3,4E-02
TMSB15A	thymosin beta 15a	0	6	3	35	35	32	12	24	2,22	3,4E-02
MCC	mutated in colorectal cancers	168	140	479	3340	377	850	2634	1786	2,12	3,5E-02
RGS13	regulator of G-protein signaling 13	118	241	163	566	640	248	1799	1885	2,03	3,5E-02
RMDN2	regulator of microtubule dynamics 2	113	98	46	240	293	191	480	218	1,55	3,5E-02
ABCA1	ATP binding cassette subfamily A member 1	115	410	368	52	39	32	54	130	-1,90	3,5E-02
DSG4	desmoglein 4	30	13	2	2	3	0	2	2	-2,28	3,5E-02
RBP1	retinol binding protein 1	1	1	1	19	4	9	42	144	2,46	4,0E-02
MAK	male germ cell associated kinase	16	19	50	103	74	93	80	92	1,47	4,0E-02
NEFH	neurofilament, heavy polypeptide	171	1384	471	35	198	31	61	123	-2,18	4,1E-02
SNCAIP	synuclein alpha interacting protein	23	48	3	2	4	0	2	4	-2,31	4,2E-02
AMIGO2	adhesion molecule with Ig-like domain 2	1	0	1	10	7	4	3	19	2,30	4,3E-02
JPH3	junctophilin 3	2	10	2	0	1	0	0	0	-2,44	4,3E-02
PYHIN1	pyrin and HIN domain family member 1	98	51	38	2310	12	1237	604	952	2,36	4,8E-02
SCHIP1	schwannomin interacting protein 1	49	40	124	568	77	3410	834	274	2,34	4,8E-02
DDAH1	dimethylarginine dimethylaminohydrolase 1	17	6	19	120	145	5	243	1011	2,39	5,0E-02

Table S4

ensembl_gene_id	exon_nb	gene_name	description	Normalized reads										Chr	start	end	strand	log2 fold change	Adjusted p-value
				C1	C2	C3	P1_p.Arg243Trp	P2_p.Arg243Trp	P3_p.Pro244Leu	P4_p.Tyr65Cys	P5_p.Arg153fs								
ENSG00000125351	E004	UPF3B	UPF3 regulator of nonsense transcripts homolog B (yeast)	37	42	13	212	234	257	103	198	chrX	119822942	119823043	-	1,06	0,00		
ENSG00000157680	E052	DGKI	diacylglycerol kinase iota	5	0	18	4	5	2	0	2	chr7	137846462	137846536	-	-1,81	0,00		
ENSG00000139537	E010	CCDC65	coiled-coil domain containing 65	12	14	7	0	0	1	0	0	chr12	48918144	48918274	+	-2,80	0,00		
ENSG00000140968	E004	IRF8	interferon regulatory factor 8	41	41	30	396	215	310	67	292	chr16	85900928	85901246	+	1,07	0,00		
ENSG00000163584	E001	RPL22L1	ribosomal protein L22 like 1	1	0	0	443	8	574	426	612	chr3	170864875	170865793	-	4,38	0,00		
ENSG00000157388	E035	CACNA1D	calcium voltage-gated channel subunit alpha1 D	0	0	2	43	15	23	26	34	chr3	53730442	53730484	+	1,77	0,00		
ENSG00000134216	E005	CHIA	chitinase, acidic	0	94	35	0	2	0	0	0	chr1	111310445	111310460	+	-2,03	0,00		
ENSG00000125351	E002	UPF3B	UPF3 regulator of nonsense transcripts homolog B (yeast)	6	5	1	121	118	161	15	149	chrX	119807475	119807581	-	2,18	0,00		
ENSG00000168872	E026	DDX19A	DEAD-box helicase 19A	96	50	72	1068	1474	512	316	1564	chr16	70370386	70371371	+	1,05	0,00		
ENSG00000140968	E005	IRF8	interferon regulatory factor 8	6	12	9	157	116	75	20	127	chr16	85902689	85902814	+	1,56	0,00		
ENSG00000134216	E004	CHIA	chitinase, acidic	0	88	33	0	2	0	0	0	chr1	111310400	111310444	+	-1,99	0,00		
ENSG00000058866	E028	DGKG	diacylglycerol kinase gamma	12	8	5	112	53	177	36	7	chr3	186279134	186279850	-	1,26	0,00		
ENSG00000113269	E001	RNF130	ring finger protein 130	14	8	7	8	9	6	4	7	chr5	179911651	179920426	-	-1,42	0,00		
ENSG00000171291	E004	ZNF439	zinc finger protein 439	3	20	29	6	6	3	3	3	chr19	11848931	11849062	+	-1,24	0,00		
ENSG00000119778	E030	ATAD2B	ATPase family, AAA domain containing 2B	10	9	6	0	2	1	1	0	chr2	23832208	23832602	-	-1,84	0,00		
ENSG00000173809	E024	TDRD12	tudor domain containing 12	1	95	2	0	0	0	0	0	chr19	32807549	32807648	+	-7,64	0,00		
ENSG00000125351	E003	UPF3B	UPF3 regulator of nonsense transcripts homolog B (yeast)	6	14	4	156	131	161	26	160	chrX	119815200	119815307	-	1,73	0,00		
ENSG00000166377	E074	ATP9B	ATPase phospholipid transporting 9B (putative)	1	6	2	18	21	8	19	17	chr18	79361739	79361992	+	1,26	0,00		
ENSG00000139437	E035	TCHP	trichoplein keratin filament binding	0	0	1	8	4	8	9	9	chr12	109980905	109981179	+	2,32	0,00		
ENSG00000125351	E001	UPF3B	UPF3 regulator of nonsense transcripts homolog B (yeast)	7	3	6	70	95	77	14	117	chrX	119805311	119805435	-	1,76	0,00		
ENSG00000109452	E049	INPP4B	inositol polyphosphate-4-phosphatase type II B	0	0	0	3	5	0	31	18	chr4	142431260	142431385	-	8,93	0,00		
ENSG00000134216	E006	CHIA	chitinase, acidic	0	78	31	0	2	0	0	0	chr1	111310461	111310464	+	-1,92	0,00		
ENSG00000082074	E010	FYB	FYN binding protein	127	15	0	1	1	0	9	1	chr5	39125998	39126135	-	-1,13	0,00		
ENSG00000126091	E036	ST3GAL3	ST3 beta-galactoside alpha-2,3-sialyltransferase 3	1	5	4	79	58	23	15	43	chr1	43920123	43920217	+	1,62	0,00		
ENSG00000073417	E002	PDE8A	phosphodiesterase 8A	6	10	16	3	3	3	2	0	chr15	84980513	84980685	+	-1,41	0,00		
ENSG00000100029	E043	PES1	pescadillo ribosomal biogenesis factor 1	0	1	4	17	12	11	9	15	chr22	30605414	30605460	-	1,35	0,00		
ENSG00000198836	E033	OPA1	OPA1, mitochondrial dynamin like GTPase	0	0	0	25	12	6	4	12	chr3	193644106	193644151	+	5,31	0,00		
ENSG00000134216	E007	CHIA	chitinase, acidic	0	68	30	0	2	0	0	0	chr1	111310465	111310467	+	-1,85	0,00		
ENSG00000198799	E019	LRIG2	leucine rich repeats and immunoglobulin like domains 2	18	6	5	0	0	0	3	2	chr1	113107758	113107826	+	-1,73	0,00		
ENSG00000157388	E001	CACNA1D	calcium voltage-gated channel subunit alpha1 D	2	2	5	0	0	0	0	0	chr3	53328963	53329444	+	-7,63	0,00		
ENSG00000163362	E002	C1orf106	chromosome 1 open reading frame 106	0	6	50	0	0	7	3	5	chr1	200891074	200891340	+	-1,19	0,00		
ENSG00000182446	E062	NPLOC4	NPL4 homolog, ubiquitin recognition factor	7	1	1	26	34	17	23	35	chr17	81629864	81629875	-	1,59	0,00		
ENSG00000148219	E026	ASTN2	astrotactin 2	3	5	8	0	1	0	0	0	chr9	116863583	116863733	-	-2,58	0,00		
ENSG00000131018	E179	SYNE1	spectrin repeat containing nuclear envelope protein 1	33	5	2	0	0	0	0	0	chr6	152430482	152430709	-	-9,26	0,00		
ENSG00000138363	E018	ATIC	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase	20	5	12	0	1	0	2	1	chr2	215326139	215326139	+	-1,99	0,01		
ENSG00000203666	E021	EFCAB2	EF-hand calcium binding domain 2	7	10	5	2	0	1	3	0	chr1	245075574	245075611	+	-1,90	0,01		
ENSG00000131018	E184	SYNE1	spectrin repeat containing nuclear envelope protein 1	9	4	2	0	0	0	0	0	chr6	152435391	152435925	-	-8,59	0,01		
ENSG00000187118	E009	CMC1	C-X9-C motif containing 1	20	6	9	0	1	2	2	2	chr3	28291801	28291847	+	-1,63	0,01		
ENSG00000152700	E020	SAR1B	secretion associated Ras related GTPase 1B	0	1	4	9	26	13	15	22	chr5	134613111	134613137	-	1,41	0,01		
ENSG00000177675	E006	CD163L1	CD163 molecule like 1	21	0	3	0	1	0	4	0	chr12	7357487	7357606	-	-1,84	0,01		
ENSG00000162735	E010	PEX19	peroxisomal biogenesis factor 19	5	7	13	62	51	23	73	68	chr1	160279635	160279800	-	1,12	0,01		
ENSG00000163781	E001	TOPBP1	topoisomerase (DNA) II binding protein 1	3	4	5	28	19	14	21	16	chr3	133598175	133598462	-	1,15	0,01		
ENSG00000035681	E026	NSMAF	neutral sphingomyelinase activation associated factor	0	0	0	5	7	2	7	10	chr8	58594738	58594747	-	6,10	0,01		
ENSG00000240563	E005	L1TD1	LINE1 type transposase domain containing 1	44	61	3	5	12	133	5	1	chr1	62211667	62212328	+	-3,51	0,01		
ENSG00000058335	E031	RASGRF1	Ras protein specific guanine nucleotide releasing factor 1	2	1	2	42	4	6	20	12	chr15	79015327	79015409	-	1,30	0,01		
ENSG00000119125	E001	GDA	guanine deaminase	0	1	6	1	0	1	0	0	chr9	72114595	72114833	+	-3,12	0,01		
ENSG00000058335	E032	RASGRF1	Ras protein specific guanine nucleotide releasing factor 1	0	3	3	46	7	7	30	11	chr15	79017770	79017906	-	1,29	0,01		
ENSG00000184702	E019	SEPT5	septin 5	5	5	7	0	0	0	0	0	chr22	19719893	19720114	+	-7,09	0,01		
ENSG00000103226	E046	NOMO3	NODAL modulator 3	1	1	5	11	20	10	51	23	chr16	16280189	16280209	+	1,67	0,01		
ENSG00000101935	E007	AMMECR1	Alport syndrome gene 1	42	11	9	2	2	8	1	0	chrX	110219480	110219613	-	-1,39	0,01		
ENSG00000163719	E027	MTMR14	myotubularin related protein 14	5	11	7	30	28	23	66	34	chr3	9683245	9683728	+	1,02	0,01		
ENSG00000130558	E013	OLFM1	olfactomedin 1	7	1	1	0	0	0	0	0	chr9	135095060	135095222	+	-6,19	0,01		
ENSG00000232653	E018	GOLGA8N	golgin A8 family member N	28	13	0	0	0	0	0	0	chr15	32598068	32598177	+	-10,24	0,01		
ENSG00000076356	E005	PLXNA2	plexin A2	3	45	4	99	92	57	20	104	chr1	208029043	208030303	-	1,27	0,01		
ENSG00000121671	E024	CRY2	cryptochrome circadian clock 2	7	3	7	38	68	12	67	99	chr11	45870613	45870841	+	1,51	0,01		
ENSG00000119403	E025	PHF19	PHD finger protein 19	1	5	3	19	33	5	22	28	chr9	120868357	120868622	-	1,43	0,01		
ENSG00000105443	E012	CYTH2	cytohesin 2	0	1	0	12	10	5	4	13	chr19	48472721	48473297	+	2,28	0,01		
ENSG00000061918	E004	GUCY1B3	guanylate cyclase 1 soluble subunit beta	0	6	31	0	0	0	15	19	chr4	155759031	155759058	+	-1,16	0,01		
ENSG00000146842	E012	TMEM209	transmembrane protein 209	1	0	0	6	9	8	4	16	chr7	130181720	130182057	-	2,32	0,01		
ENSG00000104529	E090	EEF1D	eukaryotic translation elongation factor 1 delta	5	16	23	4	3	1	4	3	chr8	143597112	143597347	-	-1,04	0,01		
ENSG00000035681	E025	NSMAF	neutral sphingomyelinase activation associated factor	2	1	3	17	11	7	14	20	chr8	58594301	58594737	-	1,37	0,01		
ENSG00000002933	E025	TMEM176A	transmembrane protein 176A	0	0	6	1	0	0	8	0	chr7	150804473	150804485	+	-2,56	0,01		
ENSG00000189367	E001	KIAA0408	KIAA0408	0	0	0	34	7	3	21	38	chr6	127450353	127451216	-	5,55	0,01		
ENSG00000197959	E008	DNM3	dynamamin 3	18	15	5	3	2	1	2	0	chr1	172032402	172032500	+	-1,15	0,01		
ENSG00000131018	E183	SYNE1	spectrin repeat containing nuclear envelope protein 1	60	5	2	0	0	0	0	0	chr6	152433946	152435390	-	-9,61	0,02		

ENSG00000188243	E004	COMMD6	COMM domain containing 6	1	2	0	14	19	5	24	12 chr13	75529709	75529823	-	1,76	0,02
ENSG00000170627	E016	GTSF1	gametocyte specific factor 1	0	4	0	0	71	0	0	3 chr12	54464703	54465066	-	1,84	0,02
ENSG00000178104	E103	PDE4DIP	phosphodiesterase 4D interacting protein	0	0	0	11	5	4	6	3 chr1	149008449	149008495	+	10,58	0,02
ENSG00000136930	E007	PSMB7	proteasome subunit beta 7	5	5	5	0	0	0	1	0 chr9	124396760	124396847	-	-2,42	0,02
ENSG00000204628	E042	RACK1	receptor for activated C kinase 1	61	44	25	7	6	10	9	5 chr5	181241437	181241442	-	-1,11	0,02
ENSG00000139537	E004	CCDC65	coiled-coil domain containing 65	0	2	2	0	8	8	6	13 chr12	48904149	48904446	+	1,23	0,02
ENSG00000131018	E177	SYNE1	spectrin repeat containing nuclear envelope protein 1	7	4	4	0	0	0	1	0 chr6	152428205	152428392	-	-2,35	0,02
ENSG00000167104	E003	BPIFB6	BPI fold containing family B member 6	14	14	0	0	0	0	0	0 chr20	33032984	33033083	+	-5,68	0,02
ENSG00000275835	E002	TUBGCP5	tubulin gamma complex associated protein 5	2	0	0	8	5	7	5	10 chr15	22996845	22996958	-	1,72	0,02
ENSG00000152061	E020	RABGAP1L	RAB GTPase activating protein 1 like	11	16	5	1	0	3	1	0 chr1	174275100	174275189	+	-1,44	0,02
ENSG00000141027	E043	NCOR1	nuclear receptor corepressor 1	6	0	3	18	29	12	20	26 chr17	16065695	16065885	-	1,34	0,02
ENSG00000232653	E017	GOLGA8N	golgin A8 family member N	24	12	0	0	0	0	0	0 chr15	32597926	32597972	+	-10,17	0,02
ENSG00000072121	E057	ZFYVE26	zinc finger FYVE-type containing 26	12	3	5	0	1	0	0	0 chr14	67815529	67815769	-	-2,71	0,02
ENSG00000123146	E012	ADGRE5	adhesion G protein-coupled receptor E5	2	3	0	27	29	10	10	17 chr19	14391391	14391492	+	1,67	0,02
ENSG00000175567	E010	UCP2	uncoupling protein 2	6	5	10	22	52	19	37	54 chr11	73977018	73977064	-	1,16	0,02
ENSG00000165458	E002	INPPL1	inositol polyphosphate phosphatase like 1	6	7	10	1	0	1	1	1 chr11	72223944	72223960	+	-1,54	0,03
ENSG00000165105	E001	RASEF	RAS and EF-hand domain containing	11	5	4	0	0	0	2	2 chr9	82979585	82982782	-	-1,42	0,03
ENSG00000234745	E019	HLA-B	major histocompatibility complex, class I, B	3	44	133	3	0	0	3	0 chr6	31356958	31357085	-	-2,72	0,03
ENSG00000068878	E060	PSME4	proteasome activator subunit 4	11	6	4	1	0	2	0	1 chr2	53942159	53942224	-	-1,61	0,03
ENSG00000103226	E047	NOMO3	NODAL modulator 3	2	1	3	6	17	8	34	13 chr16	16280210	16280330	+	1,52	0,03
ENSG00000116586	E011	LAMTOR2	late endosomal/lysosomal adaptor, MAPK and MTOR activator 2	2	3	5	16	15	6	13	25 chr1	156058278	156058314	+	1,01	0,03
ENSG00000152822	E020	GRM1	glutamate metabotropic receptor 1	28	11	7	0	2	0	1	2 chr6	146434886	146437595	+	-1,13	0,03
ENSG00000111596	E034	CNOT2	CCR4-NOT transcription complex subunit 2	46	19	34	6	0	14	9	5 chr12	70301831	70301865	+	-1,07	0,03
ENSG00000138029	E001	HADHB	hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase	14	5	3	0	0	0	1	0 chr2	26243170	26243269	+	-2,69	0,03
ENSG00000165533	E015	TTC8	tetratricopeptide repeat domain 8	8	5	4	0	0	2	0	0 chr14	88836908	88837066	+	-1,94	0,03
ENSG00000127249	E016	ATP13A4	ATPase 13A4	0	1	0	23	14	4	16	10 chr3	193435676	193435744	-	1,68	0,03
ENSG00000172932	E038	ANKRD13D	ankyrin repeat domain 13D	9	12	21	38	129	18	53	100 chr11	67300303	67300499	+	1,05	0,03
ENSG00000168631	E008	DPCR1	diffuse panbronchiolitis critical region 1	25	33	16	10	17	1	1	5 chr6	30949411	30952038	+	-1,31	0,03
ENSG00000159873	E004	CCDC117	coiled-coil domain containing 117	14	17	9	1	3	5	1	3 chr22	28773035	28773223	+	-1,09	0,03
ENSG00000213366	E031	GSTM2	glutathione S-transferase mu 2	5	4	4	4	7	2	18	7 chr1	109673203	109673525	+	1,00	0,03
ENSG00000173809	E030	TDRD12	tudor domain containing 12	0	99	4	2	0	1	1	0 chr19	32815563	32815620	+	-1,24	0,03
ENSG00000148660	E022	CAMK2G	calcium/calmodulin dependent protein kinase II gamma	5	1	1	14	23	4	40	25 chr10	73837014	73837467	-	1,64	0,03
ENSG00000162733	E004	DDR2	discoidin domain receptor tyrosine kinase 2	23	5	2	7	1	0	3	1 chr1	162632470	162632631	+	-1,19	0,04
ENSG00000087263	E016	OGFOD1	2-oxoglutarate and iron dependent oxygenase domain containing 1	5	4	6	38	41	13	23	40 chr16	56458513	56458547	+	1,23	0,04
ENSG00000165891	E001	E2F7	E2F transcription factor 7	9	5	4	0	0	1	1	0 chr12	77021247	77021247	-	-1,86	0,04
ENSG00000168237	E001	GLYCK	glycerate kinase	1	11	8	0	0	0	1	1 chr3	52287089	52287218	+	-2,07	0,04
ENSG00000134962	E003	KLB	klotho beta	8	23	2	0	3	0	0	2 chr4	39437727	39437995	+	-1,30	0,04
ENSG00000099204	E043	ABLIM1	actin binding LIM protein 1	1	5	7	31	15	13	22	4 chr10	114575538	114575599	-	1,27	0,04
ENSG00000183696	E004	UPP1	uridine phosphorylase 1	6	5	12	3	1	2	2	2 chr7	48088727	48088908	+	-1,24	0,04
ENSG00000130733	E038	YIPF2	Yip1 domain family member 2	1	11	5	0	0	0	0	0 chr19	10928658	10928681	-	-5,75	0,04
ENSG00000184307	E013	ZDHHC23	zinc finger DHHC-type containing 23	0	0	0	16	8	0	4	14 chr3	113957574	113957851	+	5,51	0,04
ENSG00000089558	E010	KCNH4	potassium voltage-gated channel subfamily H member 4	1	0	0	9	22	0	28	6 chr17	42166297	42166546	-	1,99	0,04
ENSG00000077147	E021	TM9SF3	transmembrane 9 superfamily member 3	8	6	3	0	0	0	0	1 chr10	96576345	96576633	-	-2,24	0,04
ENSG00000230453	E011	ANKRD18B	ankyrin repeat domain 18B	0	0	0	3	11	0	9	14 chr9	33536887	33536945	+	6,69	0,04
ENSG00000230453	E012	ANKRD18B	ankyrin repeat domain 18B	0	0	0	3	11	0	9	14 chr9	33539449	33539502	+	6,69	0,04
ENSG00000180773	E010	SLC36A4	solute carrier family 36 member 4	17	8	5	0	2	5	0	2 chr11	93166017	93167530	-	-1,22	0,04
ENSG00000103540	E029	CCP110	centriolar coiled-coil protein 110	6	3	2	23	27	10	34	23 chr16	19545891	19546411	+	1,24	0,04
ENSG00000114473	E035	IQCG	IQ motif containing G	5	4	2	25	27	7	25	28 chr3	197956641	197959528	-	1,29	0,05
ENSG00000139537	E005	CCDC65	coiled-coil domain containing 65	1	0	2	0	6	8	4	11 chr12	48904447	48904468	+	1,31	0,05
ENSG00000091542	E004	ALKBH5	alkB homolog 5, RNA demethylase	1	1	0	20	30	1	13	15 chr17	18206561	18206814	+	2,24	0,05
ENSG00000155085	E011	AK9	adenylate kinase 9	8	0	1	14	14	3	11	10 chr6	109506419	109506547	-	1,06	0,05
ENSG00000058866	E029	DGKG	diacylglycerol kinase gamma	1	6	0	64	31	78	27	2 chr3	186279851	186279973	-	1,76	0,05

Table S5

## Normalized reads

Gene name	Description	INF#1	siScr#1	siPQBP1#1	INF#2	siScr#2	siPQBP1#2	log2(FC) PQBP1 vs INF	Adjusted p-value PQBP1 vs INF
PQBP1	polyglutamine binding protein 1	3323	3511	1187	3574	3460	1795	-0,82	3,7E-08
FOXJ1	forkhead box J1	256	247	161	433	424	241	-0,61	1,1E-05
GMFB	glia maturation factor beta	28356	26751	11146	26984	26350	17648	-0,68	1,1E-05
CELF2	CUGBP, Elav-like family member 2	6264	5498	2538	5175	5316	3547	-0,67	1,1E-05
MAPK8IP2	mitogen-activated protein kinase 8	310	314	208	701	640	437	-0,53	1,2E-05
APC2	adenomatosis polyposis coli 2	2014	2124	1109	4635	4919	3126	-0,58	1,2E-05
CPM	carboxypeptidase M	1997	1984	3126	3620	4123	5827	0,55	6,3E-05
MID1IP1	MID1 interacting protein 1	2014	2065	1289	1718	1761	1276	-0,46	1,6E-04
SYT2	synaptotagmin 2	361	342	228	478	417	282	-0,54	6,1E-04
FXYD6	FXYD domain containing ion transp	8060	8831	5235	15564	15586	11412	-0,45	1,2E-03
SMIM13	small integral membrane protein 1	1581	1183	703	1359	1192	1067	-0,55	1,8E-03
H2AFX	H2A histone family member X	12146	13159	9204	12967	13104	8709	-0,42	3,2E-03
PLEKHA6	pleckstrin homology domain conta	584	572	400	286	267	206	-0,43	3,2E-03
AMBRA1	autophagy and beclin 1 regulator 1	3254	3121	1907	3613	3762	2873	-0,45	3,7E-03
SPP1	secreted phosphoprotein 1	49197	49961	67796	63809	71583	96956	0,44	7,4E-03
RBPMS2	RNA binding protein with multiple :	2224	2358	1363	1901	2013	1435	-0,44	8,0E-03
KCNJ2	potassium voltage-gated channel s	169	156	243	102	114	186	0,50	8,5E-03
MREG	melanoregulin	955	1017	627	1340	1315	910	-0,46	1,1E-02
SARM1	sterile alpha and TIR motif containi	4592	4493	3183	7525	7345	5996	-0,37	1,1E-02
HSDL2	hydroxysteroid dehydrogenase like	4360	4067	2183	4695	4457	3673	-0,48	1,1E-02
ASB1	ankyrin repeat and SOCS box conta	8303	7910	4668	7844	8209	6470	-0,44	1,1E-02
MT-CO2	mitochondrially encoded cytochr	61262	62707	198540	209951	201618	275531	0,55	1,1E-02
SPTBN2	spectrin beta, non-erythrocytic 2	5718	5330	4105	10423	10737	8276	-0,36	1,1E-02
RNF167	ring finger protein 167	5537	5974	3922	6018	6334	4519	-0,39	1,1E-02
LRFN4	leucine rich repeat and fibronectin	1231	1321	851	1373	1495	932	-0,44	1,1E-02
SAPCD2	suppressor APC domain containing	406	434	315	783	782	504	-0,42	1,1E-02
NR2F1	nuclear receptor subfamily 2 group	2288	2356	1715	7450	7609	5374	-0,38	1,1E-02
PRDM1	PR domain 1	2304	2179	3008	1969	2292	3495	0,46	1,2E-02

STMN3	stathmin 3	10780	11183	8162	17428	16926	13196	-0,35	1,3E-02
NRXN2	neurexin 2	679	706	453	568	562	366	-0,46	1,4E-02
SERPINE1	serpin family E member 1	6443	6347	8526	1817	2449	3702	0,49	1,5E-02
CPS1	carbamoyl-phosphate synthase 1	6352	5998	3419	4346	4248	3433	-0,46	1,6E-02
LRRRC4B	leucine rich repeat containing 4B	1677	1732	1254	1714	1728	1197	-0,39	1,7E-02
SRPX	sushi repeat containing protein, X-I	845	879	1245	658	782	995	0,44	1,8E-02
PMP22	peripheral myelin protein 22	7217	6987	4797	1825	1632	1336	-0,42	2,0E-02
LRRRC61	leucine rich repeat containing 61	620	655	434	554	534	379	-0,42	2,0E-02
NPAS3	neuronal PAS domain protein 3	634	604	334	929	968	753	-0,45	2,0E-02
MAN1C1	mannosidase alpha class 1C memb	88	86	34	59	57	37	-0,51	2,6E-02
MT-ND3	mitochondrially encoded NADH:ub	24598	28264	50416	39841	42424	54449	0,48	2,8E-02
ZNF521	zinc finger protein 521	15050	14584	9604	9906	9837	8019	-0,39	2,8E-02
TM4SF18	transmembrane 4 L six family mem	5057	4381	6631	103	113	274	0,49	2,8E-02
UHMK1	U2AF homology motif (UHM) kinas	1169	977	574	1013	1046	795	-0,47	3,1E-02
N4BP3	NEDD4 binding protein 3	99	105	56	134	138	100	-0,44	3,1E-02
CDO1	cysteine dioxygenase type 1	2873	3107	1703	4110	3874	3120	-0,42	3,1E-02
CC2D2A	coiled-coil and C2 domain containi	1865	1680	1216	921	884	748	-0,38	3,1E-02
GDF15	growth differentiation factor 15	967	1060	1674	1358	1549	1894	0,45	3,1E-02
CRADD	CASP2 and RIPK1 domain containin	436	468	238	497	489	347	-0,46	3,1E-02
BSN	bassoon presynaptic cytomatrix pr	148	119	78	201	210	154	-0,46	3,1E-02
EDN1	endothelin 1	1199	1070	1649	316	382	557	0,45	3,2E-02
CCND3	cyclin D3	3980	4336	2782	3889	3742	2849	-0,39	3,2E-02
ABCD1	ATP binding cassette subfamily D nr	603	639	372	1318	1304	946	-0,43	3,4E-02
AHDC1	AT-hook DNA binding motif contain	13040	12524	8966	6888	7606	5541	-0,36	3,5E-02
DYRK2	dual specificity tyrosine phosphory	10261	9625	6485	9674	9503	7831	-0,39	4,0E-02
CCDC113	coiled-coil domain containing 113	253	250	112	313	311	203	-0,48	4,2E-02
JPH2	junctional protein 2	961	945	657	991	1075	808	-0,35	4,6E-02
TNRC6C	trinucleotide repeat containing 6C	2193	2162	1174	2409	2404	2037	-0,42	4,8E-02
CBL	Cbl proto-oncogene	5033	3940	2521	5493	5529	4534	-0,45	4,8E-02
ACADSB	acyl-CoA dehydrogenase, short/br	1483	1385	803	1593	1692	1240	-0,44	4,8E-02
NRP1	neuropilin 1	885	894	1174	417	471	680	0,41	4,9E-02

Table S6

Gene		PSM			long vs empty		short vs empty		Interaction
		empty	long	short	ratio	delta	ratio	delta	
UPF3B	Regulator of nonsense transcripts 3B	3	205	33	68,3	202	11	30	both
UPF3Bv004	UPF3Bv004 forme 2	2	41	87	20,5	39	43,5	85	both
EPPK1	Epiplakin	3	71	1	23,7	68	0,3	-2	UPF3B_L
PABPC4	Polyadenylate-binding protein 4	0	39	12	INF	39	INF	12	both
SFN	14-3-3 protein sigma	0	34	2	INF	34	INF	2	UPF3B_L
NCL	Nucleolin	4	33	7	8,3	29	1,75	3	UPF3B_L
DDX21	Nucleolar RNA helicase 2	0	32	1	INF	32	INF	1	UPF3B_L
UPF2	Regulator of nonsense transcripts 2	0	30	0	INF	30	-	-	UPF3B_L
LMNA	Prelamin-A/C	2	29	3	14,5	27	1,5	1	UPF3B_L
YWHAZ	14-3-3 protein zeta/delta	0	28	2	INF	28	INF	2	UPF3B_L
RPLP0	60S acidic ribosomal protein P0	4	28	1	7	24	0,25	-3	UPF3B_L
HSPH1	Heat shock protein 105 kDa	0	26	17	INF	26	INF	17	both
CASC3	Protein CASC3	0	26	8	INF	26	INF	8	both
CASP14	Caspase-14	3	26	5	8,67	23	1,67	2	UPF3B_L
LGALS7	Galectin-7	0	26	1	INF	26	INF	1	UPF3B_L
CLTC	Clathrin heavy chain 1	0	23	1	INF	23	INF	1	UPF3B_L
DHX30	ATP-dependent RNA helicase DHX30	0	22	0	INF	22	-	-	UPF3B_L
HSPB1	Heat shock protein beta-1	1	21	6	21	20	6	5	both
RPL14	60S ribosomal protein L14	4	21	5	5,3	17,0	1,3	1	UPF3B_L
RPL10A	60S ribosomal protein L10a	2	21	4	10,5	19,0	2	2	both
ENO1	Alpha-enolase	3	20	10	6,7	17,0	3,3	7	both
EEF2	Elongation factor 2	3	20	9	6,7	17,0	3	6	both
SERPINB4	Serpin B4	0	20	1	INF	20	INF	1	UPF3B_L
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	4	19	4	4,75	15	0	0	UPF3B_L
IGHG1	Immunoglobulin heavy constant gamma 1	0	18	5	INF	18	INF	5	both
RPS12	40S ribosomal protein S12	4	18	5	4,5	14	1,25	1	UPF3B_L
SERPINB3	Serpin B3	3	18	3	6	15	0	0	UPF3B_L
MAGOHB	Protein mago nashi homolog 2	1	18	2	18	17	2	1	UPF3B_L
PLEC	Plectin	4	18	1	4,5	14	0,25	-3	UPF3B_L
RBM8A	RNA-binding protein 8A	1	16	5	16	15	5	4	both
IGF2BP1	Insulin-like growth factor 2 mRNA-binding protein 1	4	16	4	4	12	1	0	UPF3B_L
RPSA	40S ribosomal protein SA	4	16	4	4	12	1	0	UPF3B_L
POF1B	Protein POF1B	0	16	0	INF	16	-	-	UPF3B_L
RPL7A	60S ribosomal protein L7a	2	15	4	7,5	13	2	2	both
IGKC	Immunoglobulin kappa constant	3	15	4	5	12	1,33	1	UPF3B_L

PPL	Periplakin	0	15	0	INF	15	-	-	UPF3B_L
EIF4A3	Eukaryotic initiation factor 4A-III	2	14	7	7	12	3,5	5	both
PKM	Pyruvate kinase PKM	3	14	4	4,7	11	1,33	1	UPF3B_L
RPL27	60S ribosomal protein L27	2	14	3	7	12	1,5	1	UPF3B_L
EZR	Ezrin	1	14	2	14	13	2	1	UPF3B_L
TPI1	Triosephosphate isomerase	2	14	2	7	12	1	0	UPF3B_L
RPL12	60S ribosomal protein L12	3	14	2	4,7	11	0,67	-1	UPF3B_L
LDHA	L-lactate dehydrogenase A chain	2	14	0	7	12	-	-	UPF3B_L
ATP5F1A	ATP synthase subunit alpha, mitochondrial	0	13	1	INF	13	INF	1	UPF3B_L
IDE	Insulin-degrading enzyme	0	13	0	INF	13	-	-	UPF3B_L
PPIA	Peptidyl-prolyl cis-trans isomerase A	3	12	5	4	9	1,67	2	UPF3B_L
PRDX2	Peroxiredoxin-2	4	12	4	3	8	1	0	UPF3B_L
RPS27	40S ribosomal protein S27	2	12	3	6	10	1,5	1	UPF3B_L
SAFB2	Scaffold attachment factor B2	3	12	3	4	9	1	0	UPF3B_L
UPF1	Regulator of nonsense transcripts 1	0	12	0	INF	12	-	-	UPF3B_L
P4HB	Protein disulfide-isomerase	0	12	0	INF	12	-	-	UPF3B_L
DKC1	H/ACA ribonucleoprotein complex subunit DKC1	0	12	0	INF	12	-	-	UPF3B_L
CALML5	Calmodulin-like protein 5	4	12	0	3	8	-	-	UPF3B_L
AZGP1	Zinc-alpha-2-glycoprotein	1	11	4	11	10	4	3	both
LLPH	Protein LLP homolog	2	11	4	5,5	9	2	2	both
RPL10	60S ribosomal protein L10	2	11	3	5,5	9	1,5	1	UPF3B_L
RPL18	60S ribosomal protein L18	2	11	2	5,5	9	1	0	UPF3B_L
FAU	40S ribosomal protein S30	3	11	2	3,67	8	0,67	-1	UPF3B_L
TRA2B	Transformer-2 protein homolog beta	1	11	1	11	10	1	0	UPF3B_L
RPL28	60S ribosomal protein L28	3	11	0	3,67	8	-	-	UPF3B_L
PRMT5	Protein arginine N-methyltransferase 5	2	10	14	5	8	7	12	both
HNRNPF	Heterogeneous nuclear ribonucleoprotein F	3	10	3	3,33	7	1	0	UPF3B_L
GSDMA	Gasdermin-A	0	10	1	INF	10	INF	1	UPF3B_L
RPL15	60S ribosomal protein L15	1	10	1	10	9	1	0	UPF3B_L
SBSN	Suprabasin	3	10	1	3,33	7	0,33	-2	UPF3B_L
SF3B1	Splicing factor 3B subunit 1	0	10	0	INF	10	-	-	UPF3B_L
RPL21	60S ribosomal protein L21	2	10	0	5	8	-	-	UPF3B_L
NOP58	Nucleolar protein 58	3	9	6	3	6	2	3	both
C1QBP	1 Q subcomponent-binding protein	2	9	2	4,5	7	1	0	UPF3B_L
RPL8	60S ribosomal protein L8	4	9	2	2,25	5	0,5	-2	UPF3B_L
PGK1	Phosphoglycerate kinase 1	0	9	1	INF	9	INF	1	UPF3B_L
PLS3	Plastin-3	0	9	1	INF	9	INF	1	UPF3B_L
SERPINB5	Serpin B5	0	9	0	INF	9	-	-	UPF3B_L
CSNK2A1	Casein kinase II subunit alpha	0	9	0	INF	9	-	-	UPF3B_L



THRAP3	Thyroid hormone receptor-associated protein 3	2	9	0	4,5	7	-	-	UPF3B_L
RPL13A	60S ribosomal protein L13a	1	8	4	8	7	4	3	both
SRP14	Signal recognition particle 14 kDa protein	4	8	4	2	4	1	0	UPF3B_L
IGHV1-45	Immunoglobulin heavy variable 1-45	4	8	4	2	4	1	0	UPF3B_L
S100A7	Protein S100-A7	3	8	2	2,67	5	0,67	-1	UPF3B_L
FBL	rRNA 2'-O-methyltransferase fibrillarin	2	8	1	4	6	0,5	-1	UPF3B_L
ALDOA	Fructose-bisphosphate aldolase A	0	8	0	INF	8	-	-	UPF3B_L
CFL1	Cofilin-1	1	8	0	8	7	-	-	UPF3B_L
MDK	Midkine	2	7	4	3,5	5	2	2	both
KPNA2	Importin subunit alpha-1	0	7	2	INF	7	INF	2	UPF3B_L
GNL2	Nucleolar GTP-binding protein 2	1	7	1	7	6	1	0	UPF3B_L
G3BP2	Ras GTPase-activating protein-binding protein 2	0	7	0	INF	7	-	-	UPF3B_L
CAPRIN1	Caprin-1	0	7	0	INF	7	-	-	UPF3B_L
MDH2	Malate dehydrogenase, mitochondrial	0	7	0	INF	7	-	-	UPF3B_L
RPL5	60S ribosomal protein L5	0	7	0	INF	7	-	-	UPF3B_L
SRRM2	Serine/arginine repetitive matrix protein 2	0	7	0	INF	7	-	-	UPF3B_L
PGAM1	Phosphoglycerate mutase 1	0	7	0	INF	7	-	-	UPF3B_L
HAL	Histidine ammonia-lyase	0	7	0	INF	7	-	-	UPF3B_L
EIF4A1	Eukaryotic initiation factor 4A-I	1	7	0	7	6	-	-	UPF3B_L
BLMH	Bleomycin hydrolase	2	7	0	3,5	5	-	-	UPF3B_L
MYL6	Myosin light polypeptide 6	2	6	3	3	4	1,5	1	UPF3B_L
HNRNPL	Heterogeneous nuclear ribonucleoprotein L	3	6	3	2	3	1	0	UPF3B_L
FASN	Fatty acid synthase	0	6	1	INF	6	INF	1	UPF3B_L
PWP1	Periodic tryptophan protein 1 homolog	0	6	1	INF	6	INF	1	UPF3B_L
SRSF3	Serine/arginine-rich splicing factor 3	2	6	1	3	4	0,5	-1	UPF3B_L
SNRPD3	Small nuclear ribonucleoprotein Sm D3	3	6	1	2	3	0,33	-2	UPF3B_L
RPL32	60S ribosomal protein L32	3	6	1	2	3	0,33	-2	UPF3B_L
GSTP1	Glutathione S-transferase P	0	6	0	INF	6	-	-	UPF3B_L
PABPN1	Polyadenylate-binding protein 2	0	6	0	INF	6	-	-	UPF3B_L
RPL36	60S ribosomal protein L36	1	6	0	6	5	-	-	UPF3B_L
SSB	Lupus La protein	1	6	0	6	5	-	-	UPF3B_L
IGHG2	Immunoglobulin heavy constant gamma 2	1	6	0	6	5	-	-	UPF3B_L
RBM14	RNA-binding protein 14	1	6	0	6	5	-	-	UPF3B_L
EIF5A	Eukaryotic translation initiation factor 5A-1	2	6	0	3	4	-	-	UPF3B_L
SERBP1	Plasminogen activator inhibitor 1	2	6	0	3	4	-	-	UPF3B_L
IGHA1	Immunoglobulin heavy constant alpha 1	2	5	9	2,5	3	4,5	7	both
SPRR2D	Small proline-rich protein 2D	2	5	4	2,5	3	2	2	both
CBLN1	Cerebellin-1	2	5	2	2,5	3	1	0	UPF3B_L
PGAM5	Ser/thr protein phosphatase PGAM5	0	5	2	INF	5	INF	2	UPF3B_L

ACTN4	Alpha-actinin-4	0	5	0	INF	5	-	-	UPF3B_L
NME1	Nucleoside diphosphate kinase A	0	5	0	INF	5	-	-	UPF3B_L
ZNF326	DBIRD complex subunit ZNF326	0	5	0	INF	5	-	-	UPF3B_L
RNH1	Ribonuclease inhibitor	0	5	0	INF	5	-	-	UPF3B_L
KLK7	Kallikrein-7	0	5	0	INF	5	-	-	UPF3B_L
SRSF10	Serine/arginine-rich splicing factor 10	0	5	0	INF	5	-	-	UPF3B_L
NOP14	Nucleolar protein 14	2	5	0	2,5	3	-	-	UPF3B_L
RALY	RNA-binding protein Raly	2	5	0	2,5	3	-	-	UPF3B_L
CALML3	Calmodulin-like protein 3	0	5	0	INF	5	-	-	UPF3B_L
PHGDH	D-3-phosphoglycerate dehydrogenase	0	4	11	INF	4	INF	11	both
PPP2R1A	Ser/thr phosphatase 2A	0	4	6	INF	4	INF	6	both
IGLC2	Immunoglobulin lambda constant 2	2	4	5	2	2	2,5	3	both
PRDX4	Peroxiredoxin-4	4	3	11	0,75	-1	2,75	7	UPF3B_S
LYZ	Lysozyme C	3	3	9	0	0	3	6	UPF3B_S
IMPDH2	Inosine-5'-monophosphate dehydrogenase 2	0	3	6	INF	3	INF	6	both
DEFA1	Neutrophil defensin 1	1	2	5	2	1	5	4	UPF3B_S
LTF	Lactotransferrin	0	1	22	INF	1	INF	22	UPF3B_S
GLUL	Glutamine synthetase	0	1	5	INF	1	INF	5	UPF3B_S
SEMG1	Semenogelin-1	0	0	9	-	-	INF	9	UPF3B_S
SDHA	Succinate dehydrogenase subunit	0	0	8	-	-	INF	8	UPF3B_S
PYCR3	Pyrroline-5-carboxylate reductase 3	0	0	5	-	-	INF	5	UPF3B_S
SCGB1D2	Secretoglobin family 1D member 2	0	0	5	-	-	INF	5	UPF3B_S

# Discussion de l'article 3

Le syndrome de Renpenning est une forme de DI syndromique causée par des variations perte de fonction dans le gène *PQBPI*. Ce travail a permis d'investiguer le rôle de *PQBPI* dans la régulation de la transcription dans un modèle de LCL et de hNSCs. Nous avons pu mettre en évidence l'augmentation de l'expression d'une isoforme du gène *UPF3B* spécifiquement associée à une perte d'expression de *PQBPI*. L'expression de cette isoforme, également retrouvée spécifiquement chez les patients Renpenning, pourra permettre d'évaluer les conséquences de VSI identifiés dans *PQBPI*. Nous avons également étudié le rôle potentiel de cette isoforme ainsi que son interactome.

## ***La surexpression d'UPF3B\_S est associée à une perte de fonction de PQBPI***

Les analyses réalisées dans les lignées LCL ont révélé une grande variabilité d'expression des gènes testés et indépendante de la présence de variants perte de fonction dans *PQBPI*, comme par exemple pour le gène *APP*. En effet, nous avons pu mettre en évidence la surexpression d'*APP* chez les patients Renpenning au niveau ARNm mais également au niveau protéique dans une première série de LCL. Cependant, cette surexpression n'a pas pu être confirmée lors des analyses de la deuxième série et une étude parue en 2017 a également écarté la présence d'un lien entre *PQBPI* et la régulation transcriptionnelle d'*APP* (Chapuis et al., 2017). Malgré les variabilités observées lors des analyses transcriptomiques des lignées LCL, il reste important d'étudier les événements observés lors de cette analyse. En effet, l'analyse par exon réalisée sur la première série de LCL a révélé la surexpression d'une isoforme du gène *UPF3B* (*UPF3B\_S*) chez les patients Renpenning. Par la suite, nous avons pu valider cette augmentation chez les patients *PQBPI* dans la deuxième série de LCL. Dans un second temps, nous avons étudié les événements d'épissage alternatifs détectés dans la lignée hNSC (hNSC-1 = SA001) partiellement inactivée pour *PQBPI*. Ainsi, 133 événements furent identifiés mais aucun exon d'*UPF3B\_S* ne fut retrouvé comme sur représenté dans les hNSCs inactivées. Etant donné le faible niveau d'expression de cette isoforme, nous avons tout de même investigué son expression par RT-qPCR et avons pu mettre en évidence qu'*UPF3B\_S* était surexprimé en cas d'inactivation de *PQBPI* dans la lignée hNSC-1 mais également dans la deuxième lignée de hNSC (hNSC-2), ainsi que dans des lignées HeLa et HEK293 traitées par siRNA *PQBPI*. Nous avons donc pu mettre en évidence un phénomène robuste de régulation d'*UPF3B\_S* par *PQBPI*, retrouvé dans les cellules LCL mais également dans 4 autres types cellulaires. Enfin, nous avons

également pu tester l'expression de l'ARN *UPF3B\_S* sur des extraits de sang de patients (Paxgene) et pu mettre en évidence que l'isoforme était surexprimée chez les patients Renpenning par rapport à des individus atteints d'autres formes de DI (dont un patient FXS). Le mécanisme par lequel la perte de *PQBP1* régulerait l'expression d'*UPF3B\_S* n'est à l'heure actuelle pas élucidé. En effet, même si l'augmentation d'*UPF3B\_S* est détectable après inactivation de *PQBP1*, nous ne pouvons conclure quant à une régulation directe ou indirecte. Le rôle d'*UPF3B\_S* est à l'heure actuelle inconnu et l'ARNm semble exprimé chez l'Homme adulte principalement dans le testicule (GTEx Consortium, 2013). Il n'est cependant pas exclu qu'*UPF3B\_S* soit exprimé de manière transitoire durant le développement embryonnaire ou dans un contexte physiologique particulier.

### **Caractérisation de l'intéractomes et rôle d'*UPF3B\_S***

La protéine codée par l'isoforme majoritaire du gène *UPF3B* (*UPF3B\_L*) joue un rôle majeur dans le recrutement de la machinerie du NMD en interagissant avec les facteurs UPF2 et RBM8A et est impliquée dans un syndrome de DI liée à l'X (Tarpey et al., 2007). La protéine *UPF3B\_S* ne possède en commun avec *UPF3B\_L* que le domaine de liaison à RBM8A. De ce fait, nous avons émis l'hypothèse d'une inhibition compétitive du processus de NMD par *UPF3B\_S*, qui serait incapable de recruter UPF2. L'analyse de l'intéractome d'*UPF3B\_S* a mis en évidence un nombre plus faible d'interacteurs pour *UPF3B\_S* et ce résultat semble cohérent avec l'absence de la partie N-terminale de la protéine comparée à *UPF3B\_L*. Parmi les 23 protéines interagissant à la fois avec *UPF3B\_S* et *UPF3B\_L*, nous avons comme attendu identifié le facteur RBM8A. Comme attendu également, nous avons pu confirmer l'absence d'interaction entre *UPF3B\_S* et le facteur UPF2. Au vu de ces résultats, nous avons souhaité investiguer les conséquences de l'expression d'*UPF3B\_S* sur la régulation du système NMD mais n'avons pas pu mettre en avant d'effet évident. De plus, nous n'avons observé aucun enrichissement en ARNm décrit dans la littérature comme ciblés par le NMD dans les analyses transcriptomiques de LCL et de hNSCs inactivées pour *PQBP1*. Le rôle d'*UPF3B\_S* reste donc à l'heure actuelle inconnu mais des effets autres que sur le NMD pourraient être investigués. En effet, *UPF3B* est également connu pour interagir avec le facteur de traduction ERF3A et régule la terminaison de la traduction en facilitant la dissociation du ribosome (Neu-Yilik et al., 2017). C'est pourquoi nous souhaitons étudier les conséquences de la surexpression d'*UPF3B\_L* et *UPF3B\_S* sur la régulation globale de la traduction en évaluant dans un premier temps la synthèse protéique globale grâce à l'incorporation de la puromycine. Parallèlement, nous savons également que les protéines UPF3A et UPF3B sont décrites comme ayant des rôles

redondants et que la protéine UPF3A est potentiellement impliquée dans la régulation du cycle cellulaire (Shum et al., 2016). Or, lors de ce travail, nous avons également mis en évidence un effet d'une inactivation de *PQBP1* dans deux modèles de hNSCs sur leur capacité proliférative. De plus, un rôle de *PQBP1* dans la régulation du cycle cellulaire a déjà été rapporté en 2015 (Ito et al., 2015). Les auteurs ont pu montrer un allongement du cycle cellulaire et notamment de la phase M dans des NSPCs murins partiellement inactivés pour *Pqbp1*, cohérent avec la microcéphalie observée chez les patients Renpenning. Des expériences préliminaires que nous avons réalisées suggèrent une augmentation du nombre de cellules positives à la Caspase-3 clivée, indiquant une potentielle mort des cellules par apoptose, mais ces résultats doivent encore être répliqués. Nous souhaitons également tester l'impact de la surexpression des isoformes *UPF3B\_S* et *UPF3B\_L* sur la prolifération cellulaire. Enfin, les 9 protéines détectées par l'analyse de spectrométrie de masse et interagissant spécifiquement avec *UPF3B\_S* sont impliquées dans la réponse immunitaire (*DEFA1*, *LTF*, *LYZ*, *SEMG1*) et dans la régulation de la synthèse des acides aminés (*PRDX4*, *GLUL*, *PYRC3*, *SDHA*, *SCGB1D2*). De plus, une analyse réalisée dans des cultures primaires de monocytes humains a révélé que les lignées issues de patients Renpenning présentaient une réponse immunitaire diminuée en cas d'infection par le Virus d'Immunodéficience Humaine (VIH) (Yoh et al., 2016). Il serait donc pertinent de tester l'implication d'*UPF3B\_S* sur la régulation de ces deux processus. D'une part en exposant par exemple des cellules immunitaires issues de patients Renpenning ou transfectées par un ARN interférent *PQBP1* à un virus et d'autre part, en analysant la composition globale de ces cellules en acides aminés. Pour finir, il reste également important d'envisager l'absence de rôle physiologique de l'isoforme *UPF3B\_S*. En effet, *UPF3B\_S* reste très peu exprimé voire indétectable dans les cellules exprimant normalement la protéine *PQBP1*. De ce fait, l'expression d'*UPF3B\_S* engendrée par l'absence de *PQBP1* pourrait n'être qu'un artefact de transcription, sans conséquence dans la physiopathologie du syndrome de Renpenning.

### ***UPF3B\_S*, un outil pour la caractérisation de VSI dans *PQBP1***

Comme nous l'avons précédemment énoncé, l'augmentation d'expression d'*UPF3B\_S* associée à une perte de fonction de *PQBP1* a pu être confirmée dans 5 types cellulaires différents. Ainsi, nous avons émis l'hypothèse d'utiliser *UPF3B\_S* comme marqueur biologique du syndrome de Renpenning. Comme preuve de concept, nous avons pu montrer que l'expression de la protéine *PQBP1* sauvage permettait de ramener l'expression d'*UPF3B\_S* à un niveau basal, contrairement à la protéine *PQBP1* porteuse d'une variation perte de fonction. De ce fait, notre prochain objectif est de tester la capacité de toutes les protéines *PQBP1*

porteuses de variations perte de fonction ou de VSI à restaurer l'expression d'*UPF3B\_S* suite à l'inactivation de la protéine *PQBP1* endogène. Nous souhaitons également vérifier la présence d'une potentielle corrélation entre les différents types et la localisation de variations dans *PQBP1* avec le niveau d'expression d'*UPF3B\_S*. En effet, nous avons pu observer dans les lignées LCL que le niveau d'expression d'*UPF3B\_S* du patient porteur du variant faux-sens p.Tyr65Cys semblait inférieur à celui des autres patients. Ainsi, il est possible que ce variant, situé dans le domaine WW ait un impact moins important que les variations de décalage du cadre de lecture ou les variants présents dans le domaine C-terminal de la protéine. De plus, la validation d'expression d'*UPF3B\_S* dans le sang des patients Renpenning pourrait permettre de l'utiliser en routine diagnostique pour tester les VSI identifiés dans *PQBP1*. Cependant, afin d'utiliser *UPF3B\_S* comme biomarqueur du syndrome, ces résultats doivent être confirmés chez d'autres patients Renpenning. De plus, malgré qu'*UPF3B\_S* ne semble pas exprimé chez les patients FXS il reste essentiel de vérifier que son expression est spécifique du syndrome de Renpenning en testant l'ARN de patients atteints de plusieurs autres formes de DI.

## Conclusion

Ce travail nous a permis de mieux comprendre le rôle de *PQBP1* dans la régulation de l'expression des gènes dans un modèle de LCL issus de leucocytes de patients Renpenning mais également dans un modèle de précurseurs neuronaux humains. Ces résultats permettront une meilleure compréhension de la physiopathologie du syndrome et dévoilent plusieurs pistes qui pourront potentiellement, à terme, permettre une meilleure prise en charge des patients Renpenning. Nous avons également pu mettre en évidence que l'augmentation d'expression de l'isoforme *UPF3B\_S* est un marqueur robuste associé à une diminution d'expression ou une perte de fonction de *PQBP1* et ce, dans cinq modèles cellulaires différents. Nous avons également commencé les investigations concernant le rôle potentiel de cette isoforme. A l'heure actuelle, il semble que la protéine *UPF3B\_S* n'ait pas d'impact sur la régulation du système NMD mais nous projetons d'étudier les conséquences de sa surexpression sur la régulation de la traduction ainsi que la prolifération et le cycle cellulaire. Ainsi, sous réserve de résultats complémentaires actuellement en cours, nous proposons d'utiliser l'expression d'*UPF3B\_S* comme biomarqueur du syndrome de Renpenning. En effet, *UPF3B\_S* étant détectable dans l'ARN extrait de sang de patients Renpenning, la mesure de son expression pourrait être appliquée en routine diagnostique. De plus, dans le cas où *UPF3B\_S* serait effectivement un

biomarqueur robuste, le suivi de son expression pourrait permettre de monitorer l'effet de traitements potentiels.





# DISCUSSION GÉNÉRALE

---



## La caractérisation des variants de signification inconnue, un enjeu majeur en génétique

J'ai pu lors de ces études proposer différentes approches afin de caractériser les conséquences de variants identifiés dans les gènes *DYRK1A*, *PQBPI* et *NLGN3*. J'ai utilisé dans le cadre de *DYRK1A* un modèle de surexpression de protéines mutantes dans des cellules en culture. Cette approche permet d'évaluer facilement et relativement rapidement l'impact de n'importe quel variant situé dans la séquence codante mais n'est pas approprié dans le cas de variants suspectés d'affecter l'épissage de variants situés dans les régions introniques ou UTR. Il reste en effet nécessaire de disposer de constructions de type « minigène » ou d'échantillons / lignées cellulaires issus des patients pour l'étude de ce type de variations. La surexpression ne reflète cependant pas les concentrations de protéine mutante dans un contexte physiologique. En effet, ces modèles de surexpression ne respectent pas les balances stœchiométriques, les interactions avec d'autres facteurs et peut également moduler certaines voies cellulaires ou encore avoir un effet toxique. Il reste donc primordial de confirmer au préalable l'absence d'effet d'une augmentation d'expression de la protéine sauvage sur les paramètres testés et d'utiliser des variants contrôles dont l'effet est connu. Les tests développés peuvent être très variables et dépendent du type de variants étudiés ainsi que des rôles des protéines concernées.

Malgré la pertinence du modèle hNSC pour l'étude des mécanismes physiopathologiques impliqués dans les TND, nous n'avons malheureusement pas pu utiliser ces lignées dans le cadre de la caractérisation des VSI en raison de leur incapacité à être transfectées par des plasmides. Ainsi, il serait idéal de générer des lignées stables porteuses des différentes variations d'intérêt par édition du génome des lignées hNSCs ou par génération de lignées iPSCs différenciées par la suite en hNSC. Le choix du modèle utilisé lors de la caractérisation de VSI est généralement contraint par la quantité de variants à étudier. Ainsi, il n'est à l'heure actuelle pas envisageable d'éditer des lignées hNSC ou de générer des lignées d'iPSCs pour la caractérisation de nombreux VSI en routine diagnostique. De la même manière, même si la génération de modèles animaux Knock-In pour les variations d'intérêt reste très pertinente, comme ce fut le cas pour la souris *Dyrk1a* p.Thr588Asn, l'utilisation de ce type de modèle reste à l'heure actuelle limitée au cadre de la recherche et n'est pas compatible avec une utilisation diagnostique. Des modèles de caractérisation de variant à haut débit sont cependant développés dans différents contextes. Par exemple, une étude a pu caractériser l'effet d'une modification de la quasi-totalité des nucléotides codant pour les domaines critiques de BRCA1 (RING et BRCT) qui lui confèrent son rôle de suppresseur de tumeur, testant ainsi une

combinaison de près de 4000 variants (Findlay et al., 2018). Le gène *BRCA1* code pour une protéine impliquée dans la réparation de l'ADN et des variants perte de fonction dans ce gène conduisent à un risque accru de cancer du sein et de l'ovaire. Les auteurs ont utilisé pour ce screen la lignée HAP1 en raison de son caractère haploïde (sauf pour 330 gènes du chromosome 15). Ainsi, 2181 gènes furent identifiés comme nécessaires à la viabilité de ces cellules dans une étude précédente dont le gène *BRCA1* (Blomen et al., 2015). De ce fait, en transfectant un plasmide codant pour l'enzyme Cas9 ainsi qu'une banque d'ARN guide dirigés contre les positions nucléotidiques d'intérêt, Findlay et al. a pu utiliser la survie des cellules HAP1 comme readout et caractériser l'effet de l'ensemble de ces variants sur la fonction de *BRCA1*. Il serait donc tout à fait possible d'adapter cette technologie à l'ensemble des 2181 gènes caractérisés dans la lignée HAP1 parmi lesquels on peut retrouver *DYRK1A*. On notera que *PQBPI* n'a malheureusement pas été testé dans l'article initial de 2015 mais il est tout à fait possible qu'il fasse parti des gènes de viabilité. Il serait donc techniquement possible de tester l'ensemble des changements possibles affectant les séquences codantes des gènes *DYRK1A* et *PQBPI* en commençant par exemple par les domaines les plus importants comme le domaine kinase de *DYRK1A* et les domaines WW et CTD de *PQBPI*. Il est également intéressant de noter que parmi les 1500 gènes confirmés comme associé à la DI (Kochinke et al., 2016) 325 d'entre eux font partie des gènes de viabilité de la lignée HAP1. Cependant, outre son intérêt fondamental, la mise en place d'un screening complet pour l'ensemble de ces gènes reste à l'heure actuelle inenvisageable en raison de son coût. Cependant, au vu du nombre restreint de positions nucléotidiques concernées par les VSI, il serait tout à fait pertinent d'imaginer une plateforme sur laquelle déposer les variants à tester afin de ne considérer que les positions d'intérêt.

## **La signature de méthylation, un outil pour l'interprétation des VSI et une meilleure compréhension des mécanismes moléculaires**

Comme énoncé précédemment, le développement de signatures de méthylations associées aux syndromes de DI se généralise mais il n'existe à l'heure actuelle pas de relation constante entre la fonction du gène affecté et la présence d'une telle signature dans le sang (Aref-Eshghi et al., 2020). Nous avons pu mettre en évidence une signature de méthylation spécifique associée à une perte de fonction de *DYRK1A*, malgré le fait que *DYRK1A* soit également impliqué dans la régulation de nombreux processus autres que le remodelage de la chromatine. Cette signature s'avère très utile pour tester facilement et rapidement l'effet perte-de-fonction potentiel de n'importe quel VSI identifié dans ce gène. A l'heure actuelle, aucune

implication dans un processus épigénétique n'est rapportée pour *PQBP1*. Cependant, la protéine PQBP1 est capable de s'associer à différents ARNm, présente dans les complexes ribonucléoprotéiques nucléaires (Wan et al., 2015) et certaines études rapportent un rôle de modulation épigénétique de ces granules (Anderson et Kedersha, 2009). De plus, PQBP1 est un modulateur de l'activité transcriptionnelle et il reste probable que des changements épigénétiques indirects soient associés à cette régulation transcriptionnelle. Ainsi, il est aujourd'hui impossible de prédire avec certitude la présence d'une signature de méthylation associée au syndrome de Renpenning mais il serait particulièrement intéressant d'évaluer le profil de méthylation des patients de la même manière que nous avons pu le faire pour *DYRK1A*. Lors de la caractérisation des VSI *DYRK1A* nous avons pu confirmer la robustesse et l'intérêt de cette signature pour la reclassification des variants. L'identification de cette signature chez un patient porteur d'un variant n'affectant apparemment pas la structure mais juste l'activité de l'enzyme suggère que cette signature est associée à une incapacité de *DYRK1A* à phosphoryler ses cibles. Dans certains cas les signatures de méthylation ont permis d'identifier des variants GoF comme par exemple pour le gène *EZH2* (Choufani et al., 2020). Malgré l'identification d'un profil miroir évoquant un GoF (p.Gly486Asp), nous n'avons pas été en mesure de valider ce gain de fonction en étudiant l'activité kinase dans nos cellules. Afin de valider une hypothèse de GoF, il reste nécessaire de disposer de sang de patient porteur d'un variant ayant un effet GoF certain. En effet, l'analyse du profil d'un patient porteur d'une duplication du gène concerné pourrait ne pas être pertinente si le variant entraîne une suractivité de la protéine ne correspondant pas à 1,5 fois son activité initiale. L'étude du profil de méthylation dans un modèle animal ou cellulaire porteur de la variation reste également contestable en raison des différences entre les espèces, les contextes physiologiques et les types cellulaires. Enfin, il reste également difficile d'évaluer les conséquences sur les profils de méthylation dans le cas de variant à effet dominant négatif. De ce fait, il reste possible que le profil de méthylation ne permette pas de caractériser l'effet de variants dans ces cas particuliers et ces limites peuvent être étendues à d'autres gènes que *DYRK1A*. En plus d'être un outil puissant pour étudier les VSI, les signatures de méthylation associées à ces gènes peuvent également permettre de mieux comprendre leurs rôles et la physiopathologie de ces syndromes. En effet, la méthylation des îlots CpG, notamment dans les zones promotrices, peut être associée aux processus de régulation transcriptionnelle et permettre ainsi de prédire de manière indirecte des modifications d'expression pour les gènes concernés. Dans le cas de *DYRK1A*, l'analyse a permis de mettre en évidence une diminution de méthylation de la région promotrice du gène *HIST1H3E* qui code pour une protéine composant les nucléosomes. Les conséquences de cette

hypométhylation restent cependant inconnues et il reste primordial de confirmer le niveau d'expression de l'ARNm correspondant. Nous n'avons cependant pas observé de recoupement entre les gènes différentiellement exprimés identifiés dans les lignées de fibroblastes ou dans les hNSCs inactivées pour *DYRK1A* et les quelques régions promotrices retrouvées comme différentiellement méthylées dans le sang des patients. Il est probable que cette absence de chevauchement soit due aux à la différence de types cellulaires étudiés. Ainsi, il serait très intéressant d'étudier le profil de méthylation de lignées stables de précurseurs neuronaux humains inactivées pour *DYRK1A* et *PQBPI*. Ces études pourraient permettre la comparaison des signatures de méthylation observées dans ces types cellulaires et les analyses transcriptomiques réalisées par notre équipe dans les hNSCs. Dans un second temps, il serait également pertinent d'observer les modifications de méthylation apparaissant lors de la différenciation de ces précurseurs en neurones, que ce soit dans la condition contrôle ou dans le cas de perte de fonction de nos gènes d'intérêt. Ainsi, ces analyses pourraient permettre de mieux comprendre le rôle de ces gènes durant le développement du cerveau humain. De plus, l'utilisation de ces signatures comme biomarqueur pourrait permettre de suivre l'évolution des patients pendant l'administration d'un traitement potentiel.

### **Des études d'interactomes pour mieux caractériser les rôles de ces protéines**

Nous avons pu mettre en évidence lors de l'étude des interactomes de *DYRK1A* et d'*UPF3B\_S* plusieurs protéines impliquées dans le neurodéveloppement. Ces caractérisations peuvent permettre de mieux comprendre le rôle de ces protéines dans la physiopathologie des syndromes. En effet, la découverte de nouveaux acteurs ou de nouvelles voies cellulaires impliqués est essentielle au développement de potentielles cibles thérapeutiques. La caractérisation de l'interactome nécessite de passer par une expérience de co-immunoprécipitation suivie d'une analyse par spectrométrie de masse. Là encore, la génération de lignées neuronales stables est nécessaire car la surexpression des protéines mutantes pourrait affecter l'interactome de *DYRK1A*. De plus, il reste important de rappeler que les analyses de spectrométrie ne permettent d'évaluer que les 2000 à 3000 protéines les plus représentées dans l'échantillon et ne sont de ce fait pas représentatives des interactions faibles ou des partenaires plus faiblement exprimés. De plus, la plupart des approches connues sont basées sur une observation qualitative binaire d'absence ou de présence de l'interaction et ne décrivent pas de manière quantitative les pertes ou gains partiels d'interaction. Il serait donc intéressant de passer par des approches nouvelles, permettant d'évaluer de manière quantitative les interactions de

DYRK1A et PQBP1 en fonction des différents variants en utilisant par exemple la méthode adaptée de « benchtop holdup », développée dans l'équipe de Gilles Travé à l'GBMC (Bonhour et al., 2020). Cette approche permet notamment de mapper les interactions sur l'ensemble de la protéine étudiée, sauvage ou porteuse des variants d'intérêt et de calculer les constantes d'associations correspondantes. En revanche, la mise en place de cette technologie nécessite une purification préalable de la protéine ou des régions peptidiques d'intérêt.

## **Etudes transcriptomiques, les limites des lignées de patients**

Dans le cadre de ma thèse j'ai eu l'occasion d'étudier les résultats d'analyses transcriptomiques réalisées dans un premier temps sur des lignées cellulaires directement issues de cellules de patients : cellules lymphoblastoïdes immortalisées (LCL) pour *PQBP1* et fibroblastes pour *DYRK1A*. J'ai pu observer une très grande variabilité de l'expression des gènes candidats testés et n'ai en effet pu valider que deux gènes dont les expressions sont significativement modulées chez les patients. Ces résultats mettent en lumière la difficulté de mise au point d'une signature transcriptomique dans ces types cellulaires. Les cellules issues de patients sont un modèle intéressant car elles permettent de prendre en compte les différences inter individus pouvant avoir un impact physiologique comme le fond génétique, l'âge, le sexe, l'état de santé général ou l'ethnie. Cependant, la prise en compte de l'ensemble de ces facteurs entraîne une variabilité dans les résultats pouvant empêcher la détection de variations d'expression (Jong et al., 2019, Review). A ces variations inter-individuelles s'ajoutent les variations induites lors du prélèvement (influence de la zone biopsiée et la profondeur de la biopsie), de la mise en culture ou de l'immortalisation (sélection clonale), mais également lors de culture des cellules (nombre de passage, confluence ou absence de synchronisation). Ces variabilités induites sont difficilement mesurables mais une étude rapporte par exemple un biais de transcription concernant 20% d'ARNm sur 18 000 gènes testés dans des lignées LCL provenant de plusieurs populations (projet HapMap) et 4% d'ARNm au sein des différentes populations (Li et al., 2010). Les auteurs rapportent ainsi que les gènes comportant le plus de variabilité sont impliqués dans des processus tels que la structure anatomique, la chimiotaxie ou la réponse immunitaire. A l'inverse, les gènes les moins variables concernent les processus de neuropeptides de signalisation, de processus neurologiques et de traduction. Ces résultats laissent supposer que l'identification de gènes différentiellement exprimés dans ces types cellulaires peut être largement biaisée par ce type de variabilité et que les gènes impliqués dans les processus neuronaux auront plus de chance d'être validés. Le développement de lignées

isogéniques peut théoriquement permettre de réduire les biais engendrés par les variations inter individuelles. Cependant la génération de lignées isogéniques n'est pas dépourvue de variations. En effet, une étude récente rapporte une variabilité aux niveaux transcriptomique et épigénétique dans 5 lignées LCL isogéniques dérivées du même individu (Ozgyin et al., 2019). De plus, les auteurs démontrent également une variabilité intra lignée liée au caractère polyclonal de ces lignées. Au vu de ces observations, il est donc probable que les lignées générées à partir de matériel de patient diffèrent naturellement par le biais de ces processus ainsi que par sélection. En effet, l'utilisation de cultures polyclonales peut conduire à une hétérogénéité au sein de la culture par le processus de sélection clonale. De plus, l'apparition de néo mutations au fur et à mesure des passages pourrait également sélectionner les populations les mieux adaptées aux conditions de culture.

## **Etudes transcriptomiques, les limites du modèle hNSC**

J'ai dans un second temps étudié les résultats d'analyses transcriptomiques dans la lignée hNSC-1 (SA001) partiellement inactivée pour les gènes *DYRK1A* ou *PQBPI*. Nous avons ainsi pu mettre en évidence quelques dizaines de gènes dérégulés dans cette lignée et validés dans la seconde lignée de hNSC (hNSC-2 = GMO1869). Le faible nombre de gènes dérégulés identifiés peut s'expliquer par le nombre de réplicats (n=2 par condition). Ainsi il est probable que certaines régulations plus fines n'aient pas pu être mises en évidence en raison du manque de puissance statistique. Cependant, la validation d'un certain nombre de ces gènes dans la lignée hNSC-2 souligne la robustesse de ces régulations dans le cas des précurseurs neuronaux. Ces résultats nous ont permis de mettre en évidence plusieurs gènes impliqués dans des processus essentiels au neurodéveloppement tels que des facteurs neurotrophiques, des protéines régulant le cycle cellulaire ou des protéines régulant la différenciation et l'intégrité synaptique. La caractérisation de ces dérégulations contribue à une meilleure compréhension de la physiopathologie de ces deux syndromes et permet d'ouvrir ou de confirmer d'autres voies d'investigation. Il serait par exemple intéressant d'étudier la ciliogénèse dans les cellules inactivées pour *PQBPI* en raison de l'implication de *FOXJ1* dans ce processus et de son association à la dyskinesie ciliaire (Wallmeier et al., 2019) ou d'investiguer la régulation de *AMBRA1*, une protéine de l'autophagie impliquée dans la régulation du cycle cellulaire.

Je n'ai pas pu observer de chevauchement entre les gènes différentiellement exprimés dans les lignées de hNSC inactivées et les lignées de patients. De la même manière, je n'ai pas observé de chevauchement avec les expériences de transcriptomique ou de Chip-Seq réalisées



sur d'autres modèles comme chez la souris cKO *Pqbp1* (Wang) ou dans les divers modèles animaux *DYRK1A* comme : la levure *Saccharomyces cerevisiae* (Malcher et al., 2011), les cellules HeLa (Di Vona et al., 2015), les cellules B et T murines (Thompson et al., 2015), le poisson zèbre (Cho et al., 2019) ou le xénope (Willsey et al., 2020). Comme la plupart des gènes (Gtex Consortium 2013), nous pouvons supposer que *DYRK1A* et *PQBPI* ont des rôles tissu-spécifiques et il n'est pas surprenant d'observer des différences entre les différents modèles.

Enfin, les syndromes de Renpenning et *DYRK1A* sont deux troubles du neurodéveloppement associés à de la DI et un phénotype de microcéphalie. Les analyses transcriptomiques ont mis en évidence deux gènes dérégulés en commun (*NRXN2* et *PMP22*) mais ces gènes n'ont pas été validés dans la seconde lignée de hNSC. Il n'est cependant pas exclu que ces syndromes partagent des mécanismes physiopathologiques communs mais le séquençage d'échantillons additionnels est nécessaire à l'investigation de cette hypothèse.

Les lignées hNSC sont un modèle très pertinent d'étude des troubles neurodéveloppementaux. En effet, ces cellules modélisent les processus ayant lieu dans le cerveau embryonnaire humain et peuvent être différenciées en neurones. Enfin, les deux lignées hNSCs utilisées expriment *DYRK1A* et *PQBPI*, confirmant le rôle prépondérant de ces gènes dans le développement du cerveau. Malgré ces paramètres, les hNSCs présentent également plusieurs inconvénients. En effet, malgré leur capacité d'auto-renouvellement, les hNSCs présentent des variabilités intra (effet batch important) et extra-lignées que nous avons pu mettre en évidence avec la validation partielle des gènes différentiellement exprimés. Cependant, une analyse bio-informatique spécifique a permis de prendre en compte les variations liées à l'effet batch lors de nos analyses transcriptomiques, mais ce type de correction visant à lisser les résultats diminue également le nombre de gènes significatifs. Il apparaît également clair que le processus de génération de ces cellules impacte de manière significative leurs régulations physiologiques. Les lignées hNSCs correspondent également à une unique population cellulaire et ne permette de ce fait pas de reproduire les processus dépendant des autres types cellulaires (interactions cellule-cellule, hormones, etc...) ou de l'environnement comme la matrice extra cellulaire. J'ai utilisé durant ces projets la technologie d'interférence ARN qui reste la principale limite en raison de son action transitoire. De ce fait, il nous est impossible d'évaluer les conséquences de l'inactivation de *DYRK1A* et de *PQBPI* sur la différenciation de ces cellules en neurones. De plus, l'inactivation par siRNA est hétérogène au sein de la culture et son efficacité peut varier d'un batch à un autre, c'est pourquoi nous avons réalisé l'ensemble de nos expériences en triplicats techniques et biologiques.

Ainsi d'autres modèles pourraient être développés dans le but de palier à ces limites. Il est par exemple possible de générer des lignées hNSCs stables porteuses de variations dans les gènes d'intérêt grâce à la technologie CRISPR (à partir de l'édition d'iPSCs). De la même manière, il est également possible de générer des lignées d'iPSCs à partir de cellules de patients (comme des fibroblastes) et de différencier ces iPSCs en hNSCs. La génération de ce type de lignées stables peut par exemple permettre d'étudier les conséquences d'inactivation de gènes d'intérêt sur les processus de différenciation en neurones. Par exemple, *DYRK1A* et *PQBPI* ont tous deux été décrits comme impliqués dans les processus de différenciation des progéniteurs neuronaux et dans la régulation de la croissance dendritique dans des modèles murins. Ainsi, l'utilisation de telles lignées cellulaires permettrait d'investiguer ces phénotypes dans un modèle de cellules embryonnaires humaines et de mieux caractériser le rôle de ces protéines dans le développement du cerveau humain. Par exemple, une équipe a récemment réalisé des lignées iPSCs issues de lymphocytes de patients atteints du syndrome de Rubinstein-Taybi (RT), un syndrome de DI causé par des variations perte de fonction dans les gènes *CREBBP* et *EP300* (deux acétyl-transférases) (Calzari et al., 2020). Les auteurs ont étudié les conséquences de ces variations dans ces lignées différenciées en progéniteurs neuronaux et en neurones corticaux et mettent en évidence plusieurs gènes et processus dérégulés dans les lignées mutées. Cependant les analyses en composante principale révèlent une variabilité importante au sein des groupes contrôles et des lignées RT. Ces variabilités sont en partie dues aux cellules de patients utilisées lors de l'immortalisation, qui comme vu précédemment présentent une variabilité inter-individuelle importante. Dans une étude récente, une équipe rapporte une variabilité importante des lignées iPSCs issues de fibroblastes associée notamment aux zones où les biopsies ont été réalisées (Sacco et al., 2018). Les auteurs décrivent ainsi des différences significatives dans différents processus comme les capacités de prolifération et de croissance, l'expression de marqueurs du mésenchyme ou la différenciation en chondrocytes, ostéoblastes et adipocytes. Afin de limiter les variations inter-individuelles, il semble préférable de s'orienter sur un modèle de lignées isogéniques en procédant à une édition du génome. Cependant, les modèles hNSCs ou iPSCs ne permettent pas de refléter l'environnement présent lors du développement du cerveau. Ainsi, l'utilisation des modèles organoïdes peut permettre de pallier ces défauts en mimant l'environnement *in vivo* présent lors du développement. Par exemple, différents modèles organoïdes furent utilisés afin d'étudier les conséquences de variations perte de fonction dans les gènes *CHD8* (O'Roak et al., 2012) et *UBE3A* (Kishino et al., 2019) deux gènes impliqués dans des syndromes de DI. La première équipe rapporte une étude transcriptomique des organoïdes *CHD8*<sup>+/-</sup> soulignant un enrichissement de gènes impliqués dans la neurogénèse, la

différenciation neuronale et la voie Wnt/ $\beta$ -caténine. La seconde met en évidence qu'*UBE3A* régule l'ubiquitination de canaux potassiques et que l'abolition de cette régulation conduit à une augmentation de l'excitabilité des neurones, observation en accord avec le phénotype d'épilepsie présent chez les patients. Les possibilités apportées par ce type de modèle sont immenses mais ces techniques restent coûteuses et chronophages. Pour finir, il est également possible de générer des modèles portant des variations particulières, autres que les classiques perte-de-fonction. Par exemple, afin de caractériser les rôles nucléaires des protéines DYRK1A et PQBP1, il pourrait être pertinent de générer des lignées mutantes dans lesquelles les domaines NLS ont été délétés.



# **ANNEXES**

---



# Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment

## Résumé du projet



Le gène *NLGN3* code pour une protéine membranaire présente notamment au niveau des régions post-synaptiques. Un seul variant faux-sens pathogène avait été publié dans une famille avec troubles du spectre autistique (TSA) et nous avons identifié deux nouveaux variants faux-sens dans deux familles avec plusieurs garçons atteints de DI et de TSA. En collaboration avec Angélique Quartier, j'ai pu montrer que ces variants entraînaient une incapacité des protéines NLGN3 mutantes à se localiser à la membrane, et conduisait à leur rétention dans le réticulum endoplasmique, induisant un stress du réticulum, comme ce qui avait été montré pour la mutation publiée initialement. Ces travaux ont permis de confirmer l'implication du gène *NLGN3* dans une forme de TND et d'améliorer le diagnostic moléculaire du syndrome associé à ce gène.

## Contribution

J'ai pu lors de ce projet étudier les conséquences des variations sur la localisation membranaire de la protéine NLGN3 ainsi que la mise en évidence de sa rétention dans le réticulum endoplasmique en imageant les marquages réalisés par immunofluorescence sur cellules transfectées. J'ai également testé par RT-qPCR la présence d'une induction du stress UPR suite à la transfection de cellules par les différentes constructions codant pour les protéines NLGN3 sauvage et mutantes.

## RESEARCH ARTICLE

# Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment

Angélique Quartier<sup>1,2,3,4\*</sup> | Jérémie Courraud<sup>1,2,3,4\*</sup> | Thuong Thi Ha<sup>5,6</sup> |  
 George McGillivray<sup>7</sup> | Bertrand Isidor<sup>8</sup> | Katherine Rose<sup>9</sup> | Nathalie Drouot<sup>1,2,3,4</sup> |  
 Marie-Armel Savidan<sup>1,2,3,4</sup> | Claire Feger<sup>10</sup> | Hélène Jagline<sup>1,2,3,4</sup> | Jamel Chelly<sup>1,2,3,4,10</sup> |  
 Marie Shaw<sup>5,6</sup> | Frédéric Laumonier<sup>11,12</sup> | Jozef Gecz<sup>5,6</sup>  | Jean-Louis Mandel<sup>1,2,3,4,13</sup> |  
 Amélie Piton<sup>1,2,3,4,10</sup> 

<sup>1</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France

<sup>2</sup>Centre National de la Recherche Scientifique, Illkirch, France

<sup>3</sup>Institut National de la Santé et de la Recherche Médicale, Illkirch, France

<sup>4</sup>Université de Strasbourg, Illkirch, France

<sup>5</sup>School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

<sup>6</sup>Adelaide Medical School and Robinson Research Institute, University of Adelaide, Adelaide, South Australia, Australia

<sup>7</sup>Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia

<sup>8</sup>Service de Génétique Médicale, CHU de Nantes, Nantes, France

<sup>9</sup>Monash Genetics, Monash Health, Clayton, Victoria, Australia

<sup>10</sup>Molecular Genetic Unit, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

<sup>11</sup>UMR 1253, iBrain, Université de Tours, Inserm, Tours, France

<sup>12</sup>Service de Génétique, Centre Hospitalier Universitaire de Tours, Tours, France

<sup>13</sup>University of Strasbourg Institute of Advanced Studies, Strasbourg, France

## Correspondence

Amélie Piton, PhD, Laboratoire "Mécanismes génétiques des maladies neurodéveloppementales", IGBMC, 1, rue Laurent Fries, 67400, Illkirch, France.  
 Email: piton@igbmc.fr

## Funding information

NHMRC, Grant/Award Numbers: APP1041920, APP1091593; Agence Nationale de la Recherche; Fondation de France, Grant/Award Number: 2012-33662; Association pour le Développement de la Neurogénétique; Association Paul et Liba Mandel; Association APLM, Grant/Award Number: ANR-10-LABX-0030-INRT; the Agence Nationale de la Recherche under the frame program Investissements d'Avenir, Grant/Award Number: ANR-10-IDEX-0002-02; Australian NHMRC, Grant/Award Numbers: APP1041920, APP1091593; "Association pour le Développement de la Neurogénétique" to Frédéric Laumonier

## Abstract

The X-linked *NLGN3* gene, encoding a postsynaptic cell adhesion molecule, was involved in a nonsyndromic monogenic form of autism spectrum disorder (ASD) by the description of one unique missense variant, p.Arg451Cys (Jamain et al. 2003). We investigated here the pathogenicity of additional missense variants identified in two multiplex families with intellectual disability (ID) and ASD: c.1789C>T, p.Arg597Trp, previously reported by our group (Redin et al. 2014) and present in three affected cousins and c.1540C>T, p.Pro514Ser, identified in two affected brothers. Over-expression experiments in HEK293 and HeLa cell lines revealed that both variants affect the level of the mature *NLGN3* protein, its localization at the plasma membrane and its presence as a cleaved form in the extracellular environment, even more drastically than what was reported for the initial p.Arg451Cys mutation. The variants also induced an unfolded protein response, probably due to the retention of immature *NLGN3* proteins in the endoplasmic reticulum. In comparison, the c.1894A>G, p.Ala632Thr and c.1022T>C, p.Val341Ala variants, present in males from the general population, have no effect. Our report of two missense variants affecting the normal

\*These authors should contributed equally to this study.



localization of NLGN3 in a total of five affected individuals reinforces the involvement of the *NLGN3* gene in a neurodevelopmental disorder characterized by ID and ASD.

#### KEYWORDS

autism spectrum disorder, intellectual disability, missense variants, neuroligin-3, unfolded protein response

## 1 | BACKGROUND

Autism spectrum disorders (ASD) are characterized by deficits in social communication, restricted interests, and/or stereotyped repetitive behaviors, often associated with impairment in language development and intellectual disability (ID). The role of genetics in ASD is important and if the genetic basis is at least partly multifactorial (Sandin et al., 2017), a large number of "monogenic" syndromes associating ASD with ID have been identified in the last decades (Gonzalez-Mantilla, Moreno-De-Luca, Ledbetter, & Martin, 2016). One of these genes, the neuroligin-3 (*NLGN3*), was involved in ASD 15 years ago with the identification of a unique missense variant, c.1351C>T; Arg451Cys, in two affected brothers (Jamain et al., 2003). This variant affects a conserved residue located in the extracellular esterase-homology domain of *NLGN3* and was shown to alter the transport of the protein and causes its retention in the endoplasmic reticulum (ER). The greatly reduced protein fraction reaching the plasma membrane has a decreased binding affinity for its presynaptic partner *NRXN1* suggesting that the Arg451Cys variant is a loss-of-function variant (Chih, Afridi, Clark, & Scheiffele, 2004; Comoletti et al., 2004). This mutation has also been introduced into the mouse, the knock-in (KI) animals presenting no obvious brain abnormalities, even if some subtle changes in the size of brain structure have been reported afterward (Ellegood, Lerch, & Henkelman, 2011; Tabuchi et al., 2007). An increase in inhibitory synaptic transmission was observed in cortex while an increase in excitatory synaptic transmission and long-term potentiation was detected in hippocampus (Tabuchi et al., 2007). As the authors did not find any alterations of inhibitory synapses in *Nlgn3* knock-out (KO) mice, they suggested that the Arg451Cys variant ultimately may behave as a gain-of-function mutation. Only modest impairments in social interactions were reported in this KI mice (Etherton et al., 2011; Tabuchi et al., 2007) but not reproduced in a second KI model (Chadman et al., 2008), suggesting an influence of the genetic background in the expressivity of the mutation in mice (Jaramillo et al., 2018; Jaramillo, Liu, Pettersen, Birnbaum, & Powell, 2014). Besides this unique mutation identified in human, other arguments also supported the role of *NLGN3* in ASD as *NLGN3* expression was found to be decreased in (a) lymphoblastoid cell lines from individuals with ASD ( $n = 35$ ) compared to control individuals ( $n = 35$ ; Yasuda et al., 2011), (b) some brain tissues from mice treated in utero with valproic acid (environmental rodent model of autism; Kolozsi, Mackenzie, Roulet, deCatanzaro, & Foster, 2009).

The neuroligins are a family of postsynaptic cell-adhesion molecules that are ligands for neuroligins (NRXN), another class of synaptic cell-adhesion molecules, located at the presynaptic side

(Yamagata, Sanes, & Weiner, 2003). These interactions promote synaptic connections between neurons. They also interact by their short intracellular domain with proteins of the postsynaptic density (Chubykin et al., 2005). If neuroligins are not required for synapse formation in vivo (Varoqueaux et al., 2006), they contribute to the proper synaptic functioning. *NLGN3* is located both at excitatory and inhibitory synapses (Chubykin et al., 2005). A recent study describes a novel role of *NLGN3* as a secreted protein (Venkatesh et al., 2015). The authors demonstrated that the extracellular domain of neuronal *NLGN3* can be cleaved. The soluble *NLGN3* generated promotes high-grade glioma proliferation by inducing the PI3K-mTOR pathway activity and activating the *NLGN3* expression in glioma cells.

During the resequencing of several hundred candidate genes, our group recently reported a promising missense variant in *NLGN3*, Arg597Trp, identified in two cousins with ID and autistic features (Redin et al., 2014) but its functional consequences were not assessed. We report here the follow-up study of this family and the identification of an additional family with two brothers carrying a missense Pro514Ser variant. We characterized the functional consequences of these missense variants, showing that they affect *NLGN3* localization and function.

## 2 | METHODS

### 2.1 | Patients and identification/follow-up of missense variants in *NLGN3*

The genetic investigations were approved by the local Ethics Committee of the Strasbourg University Hospital (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale (CCPPRB)). For all patients, written informed consent for genetic testing was obtained from their legal representative. For Family 1, cosegregation analysis was performed in relatives of patient III.3 to test the presence of the variant c.1789C>T, Arg597Trp by amplification using (*NLGN3*\_Foward 5'-CCCCAGGACACCAAGTTCAT-3' and *NLGN3*\_Reverse 5'-TGGTGGTGGACGTATAGTGG-3') and Sanger sequencing (GATC, Germany) in DNA extracted from blood. For family 2, whole exome sequencing (WES) was performed on the whole nuclear family (the two affected boys and their parents, as well as the unaffected sister). Briefly, blood DNAs were captured using the Roche NimbleGen SeqCap EZ exome v3 kit and sequenced on the Illumina HiSeq. 2000 platform using 2 × 101 bp paired-end sequencing. Reads were mapped with BWA-MEM 0.7.8-r455 and variant calling was performed using GATK 3.1-1-g07a4bf8 software packages, using default parameters and following GATK Best

Practices v3. Variants were annotated using an in-house annotation pipeline and filtered using population databases (dbSNP137, 1000genomes, EVS & cg69), cosegregation in the family and predictions of pathogenicity (truncating variants and missense with CADD score above 25 & predicted to be damaging by PolyPhen2 were selected). Variants were submitted to ClinVar. All the NLGN3 variants are named according to the transcript variant NM\_018977.3 that encodes a protein of 828 amino acids (NP\_061850.2). This is not the longest isoform (NP\_851820.1, 848 amino acids, encoded by the longest transcript NM\_181303.1 that includes an additional exon after exon 1) but we observed that NM\_018977.3 was the most expressed in neural stem cells (data from transcriptomic studies we performed in human neuronal stem cells (Quartier et al., 2018)).

## 2.2 | Predictions of variants' effects

All the NLGN3 variants are named according to the isoform NM\_018977.3 that encodes a protein of 828 amino acids. The correspondence in the longest isoform NM\_181303.1 (848 amino acids) is given in Figure 1. The potential effect of missense variants were predicted using SIFT (<http://sift.bii.a-star.edu.sg/>), PolyPhen2 (<http://genetics.bwh.harvard.edu/pph2/>), SNAP2 (<https://www.rostlab.org/services/SNAP/>), MutPred2 (<http://mutpred.mutdb.org/>) (Pejaver, Mooney, & Radivojac, 2017), REVEL (<https://sites.google.com/site/revelgenomics/>), CADD (<https://cadd.gs.washington.edu/>). The presence of the different variants in the general population was checked by using ExAC database (<http://exac.broadinstitute.org/>) and later gnomAD database (<http://gnomad.broadinstitute.org/>) excluding individuals with a neurological/neurodevelopmental phenotype ("non-neuro" gnomAD). 3D modeling was predicted from protein sequence encoded by NM\_018977.3 (Uniprot accession number: Q9NZ94-2) using Raptor (Källberg et al., 2012) and visualized using PyMol.

## 2.3 | Characterization of the X-inactivation profile in carrier females

X-inactivation status was measured on blood DNA at the HUMARA and FRAXA loci (Allen, Zoghbi, Moseley, Rosenblatt, & Belmont, 1992). PCR amplification of both loci was performed on undigested or *HpaII*-digested (digestion of unmethylated allele) DNA using specific FAM-labeled primers. Amplification products were migrated on an ABI PRISM 3,500 Genetic Analyzer (Applied Biosystems, CA).

## 2.4 | NLGN3 cloning and site-directed mutagenesis

NLGN3-001 (NM\_018977.3) clone was purchased from Sino Biological Inc. (<http://www.sinobiological.com>; Cat. number: HG11160). To introduce the HA-tag (just after signal peptide) and the missense variants in NLGN3 open reading frame, site-directed mutagenesis was performed using PFU-Turbo enzyme (18 cycles with 12 min of elongation) with the following couples of primers: HA-tag forward 5'-tggcctgagggccag-taccTACCCATACGATGTTCCAGATTACGCTcaggcccccag-3' and reverse 5'-actgtgggtgctggggcctgAGCGTAATCTGGAACATCGTATGGGTAggtactg

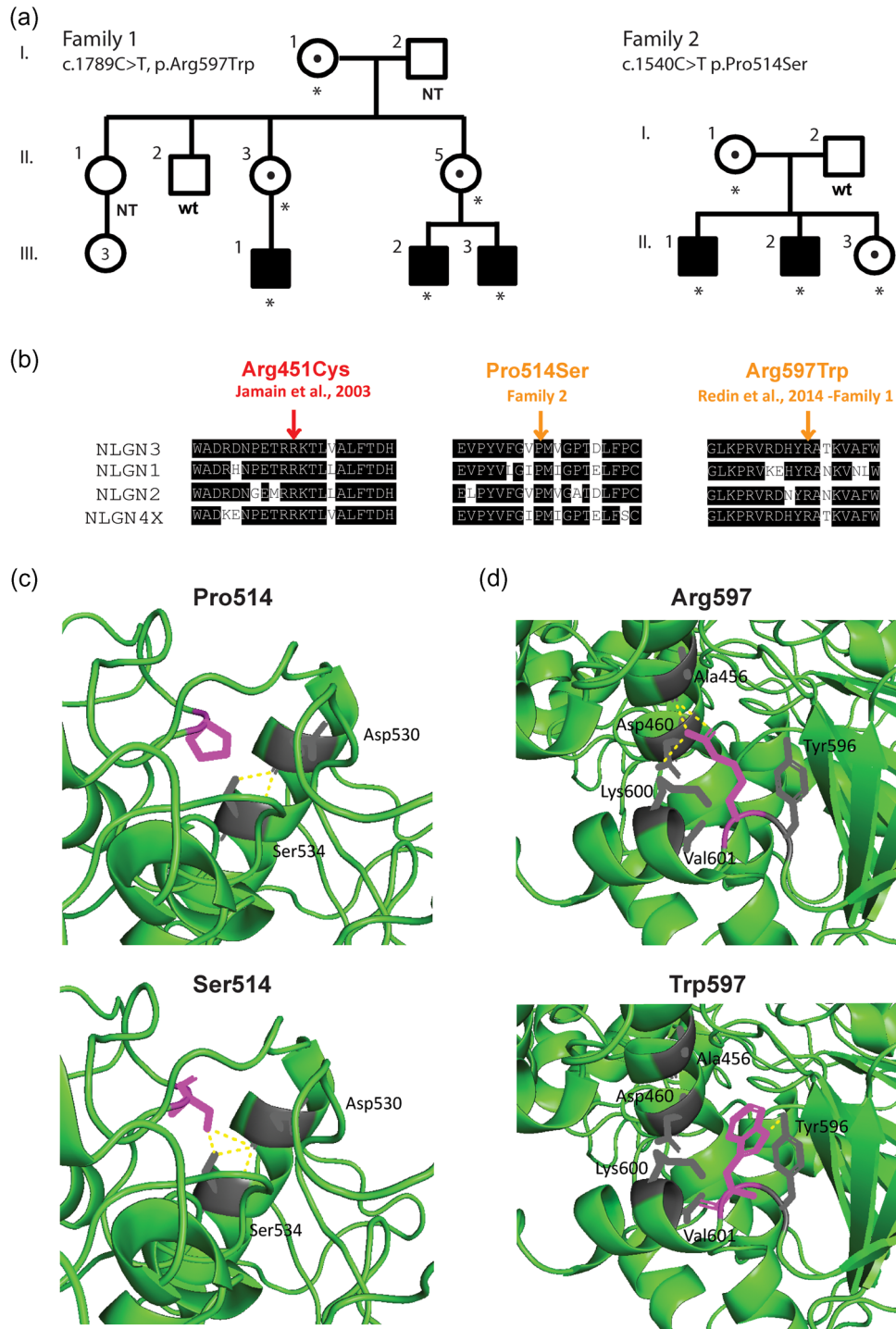
gcc-3', Val341Ala forward 5'-aaagagtccaaggagctggCagagcaggac-3' and reverse 5'-ctggctggatgtcctgctctGccagctcctt-3', Arg451Cys forward 5'-gtgacaacctgagaccgcTgtaaacact-3' and reverse 5'-gagtgccaccagtgttta-cAgcgggtctca-3', Arg597Trp 5'-caagggtccgagatcattacTgggcccataa-3' and reverse 5'-aaagccaccttagtgcccAgtaatgatct-3', Pro514Ser forward 5'-taccctatgttttggggtTctatgtagg-3' and reverse 5'-gtcagtggggctaccata-gAaaccccaaaa-3' and Thr632Ala forward 5'-ccaaagtccgctccggatGccacc-cacag-3' and reverse 5'-gatgtgggagctgtgggtggCatccggaggc-3'. The PCR products were digested with *DpnI* and used to transform competent bacteria NEB 5-alpha Competent *E. coli* (NEB) by heat-shock. After bacterial amplification, plasmids were extracted using NucleoSpin® Plasmid NucleoBond® Xtra Midi (Macherey-Nagel) and sequenced (GATC, Germany) to check the presence of each variation.

## 2.5 | Cell culture and transient transfection

Human HEK293 and HeLa cells were grown in Dulbecco's modified Eagle medium supplemented by 1 g/L glucose with gentamycin and respectively 10% and 5% Fetal Calf Serum in a 37°C, 5% CO<sub>2</sub> humidified incubator with medium renewed every 2 days. Cells were transfected at 60–70% of confluence in six-well plates using Lipofectamine® 2000 DNA transfection reagent (Invitrogen®) in Opti-MEM according to the manufacturer's instructions with 2 µg of each NLGN3 plasmid. Cells were stopped 24/48 hr after transfection for RNA or protein extraction. For four of the seven series of experiments, extracellular medium was also collected.

## 2.6 | Western blot and immunofluorescence

For Western blot analysis, experiments were performed in replicates (4 series of HEK293 and 3 series of HeLa cells). Transfected cells were lysed in radioimmunoprecipitation assay buffer with protease inhibitor cocktail (Roche). Proteins were separated after denaturation on a 10% acrylamide gel and transferred onto a polyvinylidene fluoride membrane. HA-tagged-NLGN3 proteins were visualized using an in-house mouse anti-HA antibody (1:1,000), and their level normalized with results from beta-actin or GAPDH staining using specific antibodies (in-house for beta-actin; #MAB374, Millipore for GAPDH). To visualize the secreted form of NLGN3, extracellular media were collected and an equal amount of medium was submitted to sodium dodecyl sulfate polyacrylamide gel electrophoresis after protein denaturation as described above. For immunofluorescence experiments, transfected cells, previously plated on coverslip in 24-wells plates, were fixed with 4% paraformaldehyde/PBS for 20 min, washed three times with PBS and incubated for 45 min in blocking buffer (PBS, 10% fetal calf serum, 1% bovine serum albumin, and 0.2% Triton X-100). Cells were then incubated with primary rat anti-HA antibody (1:1,000, ref 11867423001, Sigma Aldrich) and secondary anti-rat antibody coupled with fluorescence. 4',6-Diamidino-2-phenylindole was used to stain nucleus and a mouse anti-KDEL (1:500, ab12223, Abcam) to stain the ER. Fluorescence was visualized on an inverted confocal microscope (SP2UV, Leica, Wetzlar, Germany).



**FIGURE 1** Cosegregating missense variants identified in NLGN3 in two multiplex families with ID and autism. (a) Pedigrees of two families carrying the missense variants c.1789C>T, p.Arg597Trp and c.1540C>T, p.Pro514Ser. Open circles represent females, open squares represent unaffected males, closed squares represent affected males, wt = wild-type, \* = mutation carrier, NT: not tested. The patient previously described in Redin et al. (Redin et al., 2014) is the Individual III.1. (b) Evolutionary conservation of amino acids at the different positions affected by the missense changes identified (451, 514, and 597) across members of the human neuroligin family. (c,d) View of predicted 3D structure of NLGN3, with the predicted structural effects of the two missense variants Pro514Ser and Arg597Trp. Side chain changes are predicted to affect putative polar contacts (hydrogen bonds) with adjacent amino acids (yellow dot lines). ID, intellectual disability

## 2.7 | Measure of unfolded protein response by RT-qPCR

To measure UPR induced by the expression of WT or variant NLGN3 protein, we performed RNA extraction on several series of HEK293 (*n* = 5) and of HeLa cells (*n* = 5) 24 hr after plasmid transfection, using the RNeasy mini kit (Qiagen, Valencia, CA) including a DNase treatment. A total of 500 ng to 1 µg RNA was reverse-transcribed into cDNA using random hexamers and SuperScript IV reverse transcriptase according to the manufacturer's recommendation. Real-time PCR (qPCR) were performed on LightCycler 480 II (Roche) using the QuantiTect SYBR Green PCR Master Mix (Qiagen) and primers to amplify *CHOP* (*CHOP\_Foward* 5'-GACCTGCAAGAGGTCCTGTC-3' and *CHOP\_Reverse* 5'-CTCCTCCTCAGTCAGCCAAG-3'), spliced isoform of *XPB1* (*XPB1\_spliced\_Foward* 5'-GGAGTTAAGACAGCGCTTGG-3' and *XPB1\_spliced\_Reverse* 5'-GCCTGCACCTGCTGCGGA-3'), *ATF4* (*ATF4\_Foward* 5'--3' and *ATF4\_Reverse* 5'-AGGTCATCTGGCATGGTTTC-3'), and *BiP* (*BiP\_Foward* 5'-TGTTCAACCAATTATCAGCAAATC-3' and *BiP\_Reverse* 5'-TTCTGCTGTATCCTCTTACCAGT-3'). All qPCR reactions were performed in triplicate. Reaction specificity was controlled by post-amplification melting curve analysis. The relative expression of gene-of-interest vs *GAPDH* and *YWHAZ* was calculated using the 2<sup>-ΔΔC<sub>t</sub></sup> method. Expression of the spliced form of *XPB1*, of *CHOP*, and *BiP* induced by overexpression of NLGN3 variants was compared with what is obtained for overexpression of the wild-type NLGN3. To avoid effect due to the overexpression of the protein itself, we normalized the results with NLGN3 messenger RNA (mRNA) expression instead of *GAPDH/YWHAZ*.

## 2.8 | Statistical analyses

For each statistical analysis, normal distribution was first checked using Shapiro-Wilk and/or Kolmogorov-Smirnov tests. When normality is demonstrated, a *t* test is performed for comparison of two means, with Welch's correction to take into account the inequality of variances. For comparisons of more than two groups, Brown-Forsythe and Welsh analysis of variance tests were performed in case of normal distribution to take into account the inequality of variances. If the analysis of variance is significant, Dunnett's T 3 multiple comparisons test were performed to compare each NLGN3 variant with the wild-type. When the data do not follow a normal distribution, a Kruskal-Wallis' analysis of variance test was performed instead. If the analysis of variance is significant, Dunn's multiple comparison test was performed to compare each NLGN3 variant with the wild-type.

## 3 | RESULTS

### 3.1 | Missense variations identified in NLGN3 in individuals with ID and ASD or autistic features

The first missense variant Arg451Cys was identified in two Swedish brothers with ASD, one with typical autism and ID and one with Asperger syndrome (Jamain et al., 2003). After the discovery of this

mutation, the coding regions of *NLGN3* were extensively sequenced afterwards (in almost 1,500 probands with ASD in total, see Table 1). These studies screened cohorts of very different geographical origins but failed to identify promising nonsynonymous variant in *NLGN3* (Talebizadeh et al., 2006; Avdjieva-Tzavella et al., 2012; Blasi et al., 2006; Gauthier et al., 2005; Mikhailov et al., 2014; Pampanos et al., 2009; Steinberg et al., 2012; ; Vincent et al., 2004; Volaki et al., 2013; Wermter, Kamp-Becker, Strauch, Schulte-Körne, & Remschmidt, 2008; Xu et al., 2014; Yanagi et al., 2012; Ylisaukko-oja et al., 2005). The screening of a Chinese cohort identified a novel missense variant Gly406Ser (Xu et al., 2014), but functional studies performed recently could not identify any functional consequences on NLGN3 function (Xu et al., 2017). Another missense variant was identified by WES performed in a cohort of simplex and multiplex families with ASD, in one boy with ID, hyperactivity and dysmorphic features (c.1022T>C, Val341Ala), but the authors did not reach a conclusion about its pathogenicity as no functional validation was performed (Yu, Coulter, Okamura-Ikeda, & Walsh, 2013). This variant was afterward found in the hemizygous state in several boys from the exome aggregation consortium project (ExAC) and its updated version gnomAD, which was not in favor of its pathogenicity. We previously identified and reported a missense variant in *NLGN3*, c.1789C>T, Arg597Trp present at the hemizygote state in one boy with ID and ASD and in his maternal cousin with the similar phenotype (Redin et al., 2014). We performed a follow-up segregation analysis in this family and found that this variant was present in a third boy, III.3, cousin of III.1, and brother of III.2, whereas it was absent from an unaffected uncle II.2 (Figure 1a, Table 2). All three cousins present with ID, ASD or autistic traits, sleep disorder, but no

**TABLE 1** Number of individuals with ASD previously screened for mutations in *NLGN3*

Publication	Year	N =	Population
Vincent et al. (2004)	2004	196	Canadian
Ylisaukko-oja et al. (2005)	2005	30	Finnish
Gauthier et al. (2005)	2005	96	Canadian
Blasi et al. (2006)	2006	124	IMGSAC
Talebizadeh et al. (2006)	2006	10	American
Wermter et al. (2008)	2008	107	Germany
Pampanos et al. (2009)	2009	169	Grece
Yanagi et al. (2012)	2012	62	Japanese
Avdjieva-Tzavella et al. (2012)	2012	20	Bulgarian
Steinberg et al. (2012)	2012	144	AGRE
Volaki et al. (2013)	2013	40	Greek
Xu et al. (2014)	2014	318*	Chinese
Mikhailov et al. (2014)	2014	143	Thai
Total		1459	

Abbreviations: ASD, autism spectrum disorders; IMGSAC, International Molecular Genetic Study of Autism Consortium; N, number of individuals with ASD screened.

\*A missense variant c.1276G>A, Gly406Ser (Gly426Ser according to NM\_181303.1) was identified.



dysmorphism or other anomalies, except a mild microcephaly that was reported in the two cousins III.2 and III.3. Both presented with mild ID while Individual III.1 presented with severe ID with no speech acquisition at 6 years of age. No bias of X-inactivation status was observed in blood in the mother II.3 and the aunt II.5, and only a potential slight bias (FRAXA: 76/24; HUMARA: 80/20) has been identified for the grand-mother I.1. In a second family, composed of two brothers affected by nonsyndromic ID, autistic traits and language impairment, a family-WES analysis identified a missense variant c.1540C>T, Pro514Ser in *NLGN3*, present in both brothers and inherited from the mother. The variant is carried by the sister who does not present X-inactivation bias in blood. The two brothers of this second family had normal early infant development with language development until 18 months of age and then presented with significant speech regression. They subsequently had a severe disability with autistic traits and no neuropsychology assessment was possible.

### 3.1.1 | *NLGN3* is relatively intolerant to missense changes

These two missense variants, like the originally published pathogenic variant c.1351C>T, Arg451Cys (Jamain et al., 2003), affect highly conserved amino acids and are predicted to be deleterious by all the different programs (Table 2, Figure 1b). They all alter amino acids located in the extracellular domain of *NLGN3*. The change of the arginine to a tryptophan in position 597 and of a proline to a serine in position 514 modify the hydrogen bonds with adjacent amino acids (Figure 1d), might therefore affect *NLGN3* protein structure. We reviewed all the nonsynonymous variants present in the general population using the gnomAD database (v2.1). There is no loss-of-function variant in these general populations, if we except a putative splice variant affecting a noncanonical exon-specific to the longest isoform but falling in intronic region in NM\_018977.3 (c.457+639G>A). A putative truncating variant falling in this exon was also reported previously during the resequencing of X-Linked ID families (c.457+666C>T in the NM\_018977.3 transcript; Hu et al., 2016). Two and half times less missense variants were observed compared with what is expected (151/376), suggesting that *NLGN3* gene is relatively intolerant to missense changes (Z-score = 4.21; this metrics is available in the gene constraint section in gnomAD v2.1). In total, 139 nonsynonymous changes (136 missense variants and 2 in-frame indels) are reported in individuals from the gnomAD cohorts excluding patients with neurodevelopmental or neurological conditions (gnomAD “non-neuro” cohorts; Table S1, Figure 2a). Half of them only (71/139) are present at the hemizygous state in male individuals (9 only in three or more males). Prediction of deleterious effect of these missense variants, based on CADD score, shows that they range from very low effect (CADD = 7) to very strong effect (CADD = 34). It appears, therefore, difficult to discriminate, among the variants identified in patients with neurodevelopmental conditions, those

being truly pathogenic only based on in silico predictions and without performing additional functional studies.

### 3.1.2 | The two novel missense mutations affect the level of mature *NLGN3* proteins

We decided to undertake a study of the effect of the two missense variants Pro514Ser and Arg597Trp on *NLGN3* function, to determine their pathogenicity, compared with the effect of the unique pathogenic mutation validated to date Arg451Cys (Table 2). We also included two rare missense variants reported in males from gnomAD cohorts, as negative controls, Val341Ala (first selected because it was observed in a patient with ASD (Yu et al., 2013) but finally reported in 7 males from “non-neuro” gnomAD cohorts) and Thr632Ala (reported in 138 males from “non-neuro” gnomAD cohorts and found in one man from a cohort of control individuals previously sequenced; Blasi et al., 2006; Figure 2a). *NLGN3* expression level was analyzed in HeLa and HEK293 cell lines. A drastic decrease in the level of *NLGN3*, significant for Pro514Ser and Arg597Trp but not for Arg451Cys, was observed for variant proteins. This decrease concerns especially the mature glycosylated protein (upper band). On the contrary, the Val341Ala and Thr632Ala variants lead to an expression level similar to wild-type (Figure 2b, upper panel).

### 3.1.3 | Missense variants affect *NLGN3* membrane localization and its secretion

By immunostaining, we observed that wild-type *NLGN3* and variant *NLGN3* carrying the variants Val341Ala or Thr632Ala are localized both at the plasma membrane and in the cytoplasm when over-expressed in HEK293 (Figure 3) and HeLa cells (data not shown). On the contrary, we did not detect any membrane localization for the variant *NLGN3* with Pro514Ser and Arg597Trp variants. For the Arg451Cys variant, a weak staining was observed at the membrane but reduced when compared with the wild-type protein. This suggests that part of the *NLGN3* proteins carrying the variant Arg451Cys are not properly addressed to the membrane, consistent with the previous report (Chih et al., 2004). Furthermore, the Arg451Cys variant was found to affect the correct folding of the extracellular domain leading to a partial retention of the protein in the ER, decreasing the trafficking of *NLGN3* to the cell surface (Chih et al., 2004; De Jaco, Dubi, Comoletti, & Taylor, 2010). We confirmed this retention in ER for *NLGN3* proteins with Pro514Ser and Arg597Trp variants by performing a costaining with KDEL, a marker of ER (Figure 3). It was recently shown that the extracellular domain of neuronal *NLGN3* is cleaved and secreted into the extracellular medium (Venkatesh et al., 2015). We confirmed that we can detect a secreted *NLGN3* protein in the extracellular medium of HeLa and HEK293 cells transfected with wild-type, Val341Ala, or Thr632Ala *NLGN3* cDNA (Figure 2b, lower panel). Consistent with the absence of membrane localization, we could not detect secreted forms of the *NLGN3* proteins carrying Pro514Ser and Arg597Trp variants, and

**TABLE 2** Summary of the missense variants tested

Nomenclature (NM_018977.3) according to NM_181303.1	p.Arg451Cys chrX:70387358C>T c.1351C>T	p.Pro514Ser chrX:70387547C>T c.1540C>T	p.Arg597Trp chrX:70389249C>T c.1789C>T	p.Val341Ala chrX:70387029T>C c.1022T>C	p.Thr632Ala chrX:70389354A>G c.1894A>G
Reference dbSNP	rs121917893	-	rs878853147	c.1082T>C, p.Val361Ala rs749067360	c.1954A>G, p.Thr652Ala rs144914894
GnomAD <sup>a</sup>	0	0	0	7	138
Grantham	180	74	101	64	58
SIFT	deleterious (0)	deleterious (0.02)	deleterious (0)	deleterious (0.05)	tolerated (0.49)
PolyPhen2	probably_damaging (1)	probably_damaging (0.97)	probably_damaging (1)	benign (0.13)	benign (0.00)
SNAP2	effect (37)	effect (40)	effect (67)	neutral (-23)	neutral (-51)
MutPred2	OI, TP, U	OI, TP, S	OI, RSA, DB, TP	TP	-
CADD	31	25.9	32	23.7	16.89
REVEL	0.767	0.810	0.769	0.437	0.125

Note: In silico predictions of functional consequences of the different missense changes tested in this study, assessed using several bioinformatics programs including SIFT, PolyPhen2, SNAP2, and MutPred2, CADD, and REVEL (the two last ones compile information from the other programs). MutPred2 predictions include OI, altered ordered interface; TP, altered transmembrane protein; U, loss of ubiquitylation; S, loss of sulfation; RSA, loss of relative solvent accessibility; DB, altered DNA Binding. <sup>a</sup>: Number of males from the “non-neuro” gnomAD cohorts who are hemizygous for the variant.

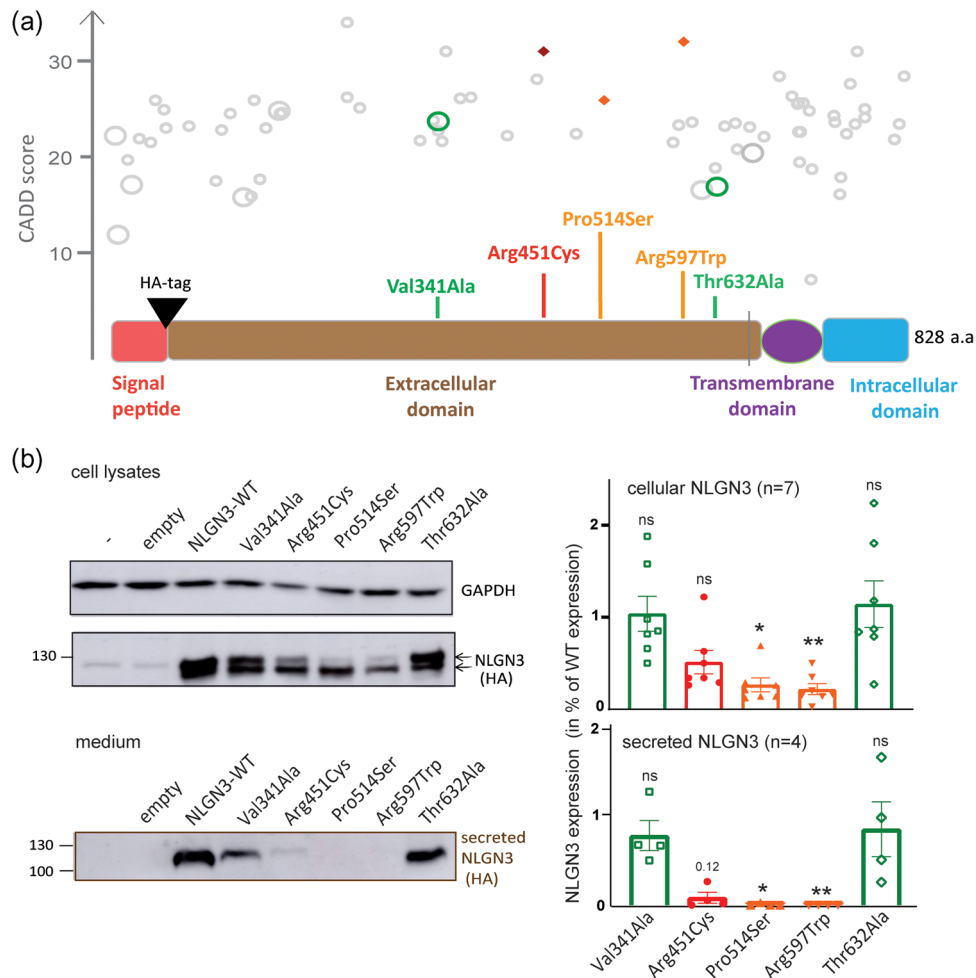
only a residual amount of secreted protein was detected for the Arg451Cys variant.

### 3.1.4 | Activation of the UPR response

It was previously shown that the Arg451Cys NLGN3 mutation is partially retained in the ER subsequently causing ER stress and leading to activation of the UPR (unfolded protein response) cell response (Ulbrich et al., 2016). This stress was presumed to participate in the pathogenicity of the mutation (Fujita et al., 2010). We, therefore, tested whether some UPR markers are induced after overexpression of wild-type and variant NLGN3 proteins. mRNA levels for the spliced transcripts of XBP1 and for CHOP and BiP, markers of the UPR stress, are induced after treatment with thapsigargin, a drug inhibiting the ER calcium pump, in HEK293 cells (Figure 4a). As overexpression of the NLGN3 plasmids, including wild-type one, lead to a slight increase of UPR targets (data not shown), we normalized the UPR response with the level of the NLGN3 expression (Figure 4b). We observed that Arg451Cys, Pro514Ser, and Arg597Trp missense variants significantly induced the expression of genes of the UPR in HEK293 cells (Figure 4c). The UPR response was more pronounced for the two novel mutations as compared with the initial Arg451Cys mutation, which is consistent with the fact that the expression of the mature NLGN3 protein and the protein localization are more altered by the Pro514Ser and Arg597Trp changes than by the Arg451Cys change (Figure 2b and Figure 3).

## 4 | DISCUSSION

The NLGN3 and NLGN4X genes were the first genes described as involved in a nonsyndromic monogenic form of ASD (Jamain et al., 2003). Mutations in NLGN4X were, thereafter, found in individuals with XLID with or without ASD (Laumonnier et al., 2004), however, the role of NLGN3 in NDD was still restricted in ASD with the p.Arg451Cys mutation to date. Extensive functional studies have been performed, in cellular and animal models, all based on the single p.Arg451Cys missense mutation reported. In the meantime, the coding regions of NLGN3 were screened for mutation in numerous cohorts of ASD individuals, without finding additional convincing pathogenic variants. Systematic trio exome sequencing of ASD cohorts was also deceptive, as only two de novo missense variants are cataloged in the denovo db database, from the study by Iossifov et al. (2014): p.Arg195Trp, with a low CADD score (18.3) in a patient with ASD, and p.Arg797Gln, CADD = 20.7, in an unaffected sib. With two novel missense variants affecting the membrane localization of NLGN3 and the expression of the mature glycosylated form of the protein, identified in a total of five affected individuals from two multiplex families and cosegregating with affected status in males, we confirm that missense mutations in NLGN3 can indeed be responsible for a neurodevelopmental disorder. All seven individuals reported now with a pathogenic variant in NLGN3 have had autism or autistic traits. The severity of cognitive impairment is variable in patients, ranging from no ID to severe ID, even in individuals carrying the same mutation, as similarly found for NLGN4X mutations. Sleep



**FIGURE 2** Schematic representation of NLGN3 protein and expression of HA-tagged wild-type and variant NLGN3 proteins in HEK293 and HeLa cells. (a) Schematic representation of the HA-tagged NLGN3 protein (encoded by the NM\_018977.3 transcript) with its different extracellular, transmembrane, and intracellular domains. Missense variants reported at the hemizygotic state in males in “non-neuro” gnomAD populations are indicated by gray circles: variants present in one or two males are indicated by small circles while larger circles indicate variants present in three or more males. The different missense changes studied here are indicated: the initial Arg451Cys pathogenic variant reported by Jamain et al. (2003) (in red); the two variants Arg597Trp and Pro514Ser (in orange); the variant Val341Ala initially reported in an individual with ASD by Yu et al., 2013 but also present in seven males from the “non-neuro” gnomAD general population cohorts and the variant Thr632Ala present in 138 males and classified as certainly benign (in green). (b) Expression of NLGN3 proteins in HEK293 and HeLa cells transiently transfected with NLGN3 constructs was detected by SDS-PAGE and immunoblotting using an anti-HA antibody, revealing the immature form (lower band) and the mature glycosylated form (higher band). Quantifications were performed using the expression of reference proteins (GAPDH or actin) on a total of  $n = 7$  series of cells ( $n = 4$  HEK293 and  $n = 3$  HeLa cells). Expression of the secreted cleaved form of NLGN3 was analyzed in the same manner on extracellular cell media for four of these series of cells. Analysis of variance Kruskal–Wallis tests and Dunn’s multiple comparisons test were performed to compare the expression of variant NLGN3 with wild-type protein: ns: not significant;  $*p < 0.05$ ;  $**p < 0.01$ ; error bars represent SEM. ASD, autism spectrum disorders; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; SEM, standard error of the mean

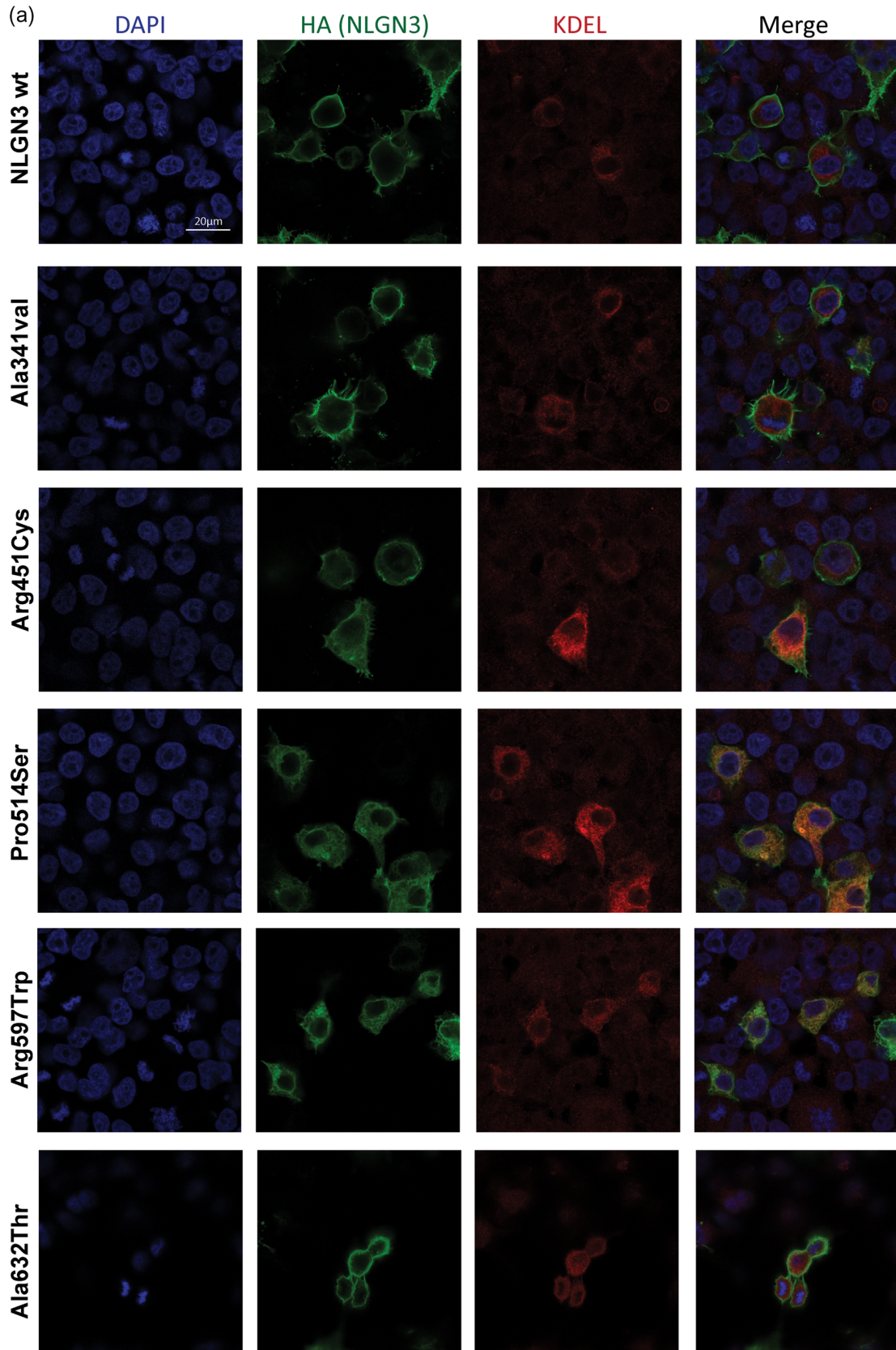
disorders have been described in the three cousins of family 1, and interestingly, sleep alterations were observed in the rat model Nlgn3 KO, with an increase of rapid eye movement (REM) and a decrease in non-REM sleep periods as compared with wild-type (Thomas, Schwartz, Saxe, & Kilduff, 2017).

We demonstrated here that NLGN3 proteins carrying Pro514Ser and Arg597Trp variants do not reach the plasma membrane in the cells we used for overexpression studies. This model indicates that NLGN3 proteins with these variants would be prevented from interacting with neuroligin proteins in the human brain. An induction

of the UPR response, illustrated by an increase of the spliced isoform of XBP1 and an increase of expression of CHOP and BiP, was observed in HEK293 cells. The activation of the UPR stress and the loss of the mature form of NLGN3 strongly suggest that variant proteins are stacked in the ER, which was confirmed by costaining experiments with an ER marker, KDEL.

At this stage, we can only speculate about the underlying mechanism of action of these variants. On aggregate we found that the variants we studied: (a) prevent NLGN3 to reach the cell membrane which may result in loss of interaction between the pre- and postsynaptic





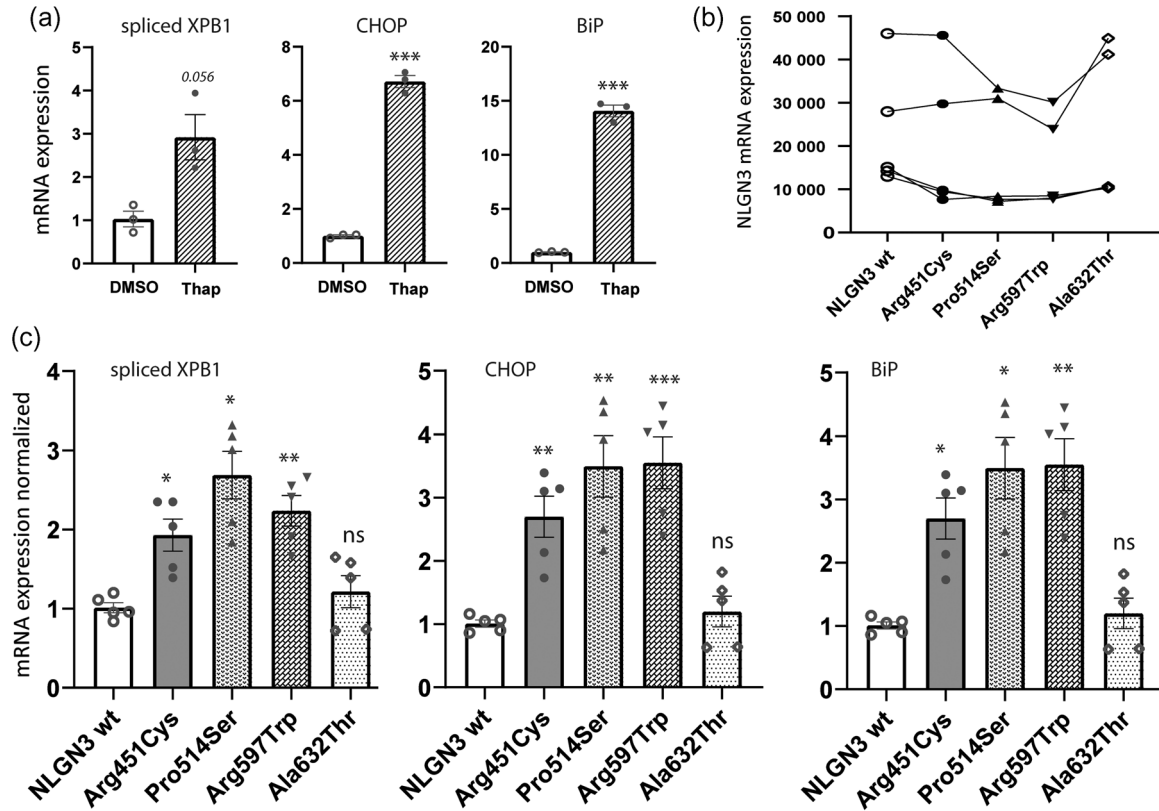
**FIGURE 3** Cellular localization of HA-tagged wild-type and variant NLGN3 proteins in HEK293 cells. HEK293 cell lines were transiently transfected with NLGN3 constructs, HA-tagged wild-type (WT) or variant NLGN3.) Immunofluorescence experiments using an anti-HA antibody revealed the cellular localization of WT and variant NLGN3 proteins (green fluorescence) in HEK293 cells. The DAPI staining indicates the position of the nuclei and the KDEL staining shows the endoplasmic reticulum (ER). DAPI, 4',6-diamidino-2-phenylindole



compartments; that (b) they are not secreted which may have an effect on glial cells proliferation; and that (c) they induce ER stress and resulting cellular response (UPR). The two first observations suggest that these variants would have a loss-of-function effect, whereas the third one (activation of UPR) would rather evoke a gain-of-function effect. The interpretation of this latter result is limited by the fact that this UPR stress is observed *in vitro* in a context of overexpression. However, a link between NLGN3 missense variants and UPR stress is supported by a recently published observation showing that alterations of synaptic function found in the Purkinje cells of the Arg451Cys KI mice (increased frequency of miniature excitatory currents) are rescued by inhibiting the UPR response (Trobiani et al., 2018).

Up to date, the three pathogenic mutations described in NLGN3 are missense changes. Interpretation of missense variants in NLGN3 remains challenging. Indeed, some missense variants found in males in the general population (gnomAD non-neuro cohorts) or having no observed effect on protein function (Xu et al., 2014) are located in

the vicinity of the three pathogenic mutations (Table S1). *In silico* prediction tools are not sufficient to discriminate between the benign and pathogenic variants. At the time of writing, 12 missense variants were reported in Clinvar. If we except the initial Arg451Cys pathogenic variant reported by Jamain et al. (2003) as "risk factor", our Arg597Trp variant (that we previously reported it as "likely pathogenic") and the Thr632Ala (reported as "Benign"/"Likely benign"), all the 9 remaining variants are labeled as "Uncertain significance" (Table S2). The fact that the NLGN3 gene is located on the X chromosome is a bias against detecting *de novo* events (that are more likely to occur one generation above, from the maternal grandfather) and reporting missense variants in publications of large exome or genome cohort studies. Testing the X-inactivation status of mothers do not seem to be informative. Segregation analysis, including testing of maternal grand-parents and affected or unaffected males (brothers, maternal uncles) might be informative but are not always feasible, especially in large cohort studies. Therefore,



**FIGURE 4** Expression of markers of the unfolded protein response (UPR) response in HEK293 cells. (a) Expression of CHOP, BiP, and spliced XPB1 mRNA (normalized by the expression of two reference genes, GAPDH and YWHAZ) are induced in HEK293 cells after a treatment with thapsigargin (5 hr, 1  $\mu$ M,  $n = 3$ ). *t* test with Welch's correction, \*\* $p < 0.001$ . Error bars represent SEM (b) Increase of NLGN3 mRNA after a transient transfection of HEK293 cells with plasmids containing HA-tagged wild-type (WT) or variant NLGN3 proteins compared with Lipofectamine only (Lipo) ( $n = 5$  for each condition; results normalized by GAPDH and YWHAZ expression). Kruskal–Wallis' analysis of variance test did not reveal a significant difference between the conditions; error bars represent SEM. (c) Transfection of the variant Arg451Cys, Pro514Ser, and Arg597Trp NLGN3 proteins in HEK293 cells lead to increased the mRNA expression level of the UPR markers CHOP, BiP, and spliced transcript of XPB1 (results normalized by NLGN3 mRNA expression) compared with cells transfected with WT NLGN3. As Brown–Forsythe and Welch's analysis of variance revealed the significant difference, and multiple comparisons tests were performed using Dunnett's T3 test: ns: not significant; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; error bars represent SEM. mRNA, messenger RNA; SEM, standard error of the mean

functional testing will remain mandatory to conclude about the pathogenicity of missense variants for diagnostic purposes, until more is known about structure/function relations in this transmembrane protein.

## 5 | CONCLUSIONS

In conclusion, our study confirms that several missense variants in NLGN3 are involved in a non-syndromic form of ID associated with autistic manifestations. The three pathogenic variants studied lead to an absence/decrease of membrane localization of NLGN3, a retention in the ER and an induction of a cellular response related to ER stress.

## ACKNOWLEDGMENTS

The authors thank the families for their participation in this study. They thank Didier Devys for interpreting X-inactivation studies, and Paola Rossolillo, from the IGBMC molecular biology platform. This work was funded by Association APLM. This study was also supported by grant ANR-10-LABX-0030-INRT, a French state fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02, Australian NHMRC grants APP1091593 and APP1041920 to Jozef Gecz and grants from Fondation de France (2012-33662) and "Association pour le Développement de la Neurogénétique" to Frédéric Laumonier.

## CONFLICT OF INTERESTS

The authors declare that they have no conflict of interests.

## ORCID

Jozef Gecz  <http://orcid.org/0000-0002-7884-6861>

Amélie Piton  <http://orcid.org/0000-0003-0408-7468>

## REFERENCES

- Allen, R. C., Zoghbi, H. Y., Moseley, A. B., Rosenblatt, H. M., & Belmont, J. W. (1992). Methylation of *HpaII* and *HhaI* sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *American Journal of Human Genetics*, 51(6), 1229–1239.
- Avdjieva-Tzavella, D. M., Todorov, T. P., Todorova, A. P., Kirov, A. V., Hadjidekova, S. P., Rukova, B. B., ... Toncheva, D. I. (2012). Analysis of the genes encoding neuroligins NLGN3 and NLGN4 in Bulgarian patients with autism. *Genetic Counseling (Geneva, Switzerland)*, 23(4), 505–511.
- Blasi, F., Bacchelli, E., Pesaresi, G., Carone, S., Bailey, A. J., & Maestrini, E., International Molecular Genetic Study of Autism Consortium (IMGSAC). (2006). Absence of coding mutations in the X-linked genes neuroligin 3 and neuroligin 4 in individuals with autism from the IMGSAC collection. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, 141B(3), 220–221.
- Chadman, K. K., Gong, S., Scattoni, M. L., Boltuck, S. E., Gandhy, S. U., Heintz, N., & Crawley, J. N. (2008). Minimal aberrant behavioral phenotypes of neuroligin-3 R451C knockin mice. *Autism Research*, 1(3), 147–158.
- Chih, B., Afridi, S. K., Clark, L., & Scheiffele, P. (2004). Disorder-associated mutations lead to functional inactivation of neuroligins. *Human Molecular Genetics*, 13(14), 1471–1477.
- Chubykin, A. A., Liu, X., Comoletti, D., Tsigelny, I., Taylor, P., & Südhof, T. C. (2005). Dissection of synapse induction by neuroligins: Effect of a neuroligin mutation associated with autism. *The Journal of Biological Chemistry*, 280(23), 22365–22374.
- Comoletti, D., De Jaco, A., Jennings, L. L., Flynn, R. E., Gaietta, G., Tsigelny, I., ... Taylor, P. (2004). The Arg451Cys-neuroligin-3 mutation associated with autism reveals a defect in protein processing. *The Journal of Neuroscience*, 24(20), 4889–4893.
- Ellegood, J., Lerch, J. P., & Henkelman, R. M. (2011). Brain abnormalities in a neuroligin3 R451C knockin mouse model associated with autism. *Autism Research*, 4(5), 368–376.
- Etherton, M., Földy, C., Sharma, M., Tabuchi, K., Liu, X., Shamloo, M., ... Südhof, T. C. (2011). Autism-linked neuroligin-3 R451C mutation differentially alters hippocampal and cortical synaptic function. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33), 13764–13769.
- Fujita, E., Dai, H., Tanabe, Y., Zhiling, Y., Yamagata, T., Miyakawa, T., ... Momoi, T. (2010). Autism spectrum disorder is related to endoplasmic reticulum stress induced by mutations in the synaptic cell adhesion molecule, CADM1. *Cell Death & Disease*, 1, e47.
- Gauthier, J., Bonnel, A., St-Onge, J., Karemera, L., Laurent, S., Mottron, L., ... Rouleau, G. A. (2005). NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, 132B(1), 74–75.
- Gonzalez-Mantilla, A. J., Moreno-De-Luca, A., Ledbetter, D. H., & Martin, C. L. (2016). A cross-disorder method to identify novel candidate genes for developmental brain disorders. *JAMA Psychiatry*, 73(3), 275–283.
- Hu, H., Haas, S. A., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A. P. M., ... Kalscheuer, V. M. (2016). X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Molecular Psychiatry*, 21(1), 133–148.
- Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., ... Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515(7526), 216–221.
- De Jaco, A., Dubi, N., Comoletti, D., & Taylor, P. (2010). Folding anomalies of neuroligin3 caused by a mutation in the alpha/beta-hydrolase fold domain. *Chemico-Biological Interactions*, 187(1-3), 56–58.
- Jamain, S., Quach, H., Betancur, C., Råstam, M., Colineaux, C., Gillberg, I. C., ... Paris Autism Research International Sibpair Study (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genetics*, 34(1), 27–29.
- Jaramillo, T. C., Escamilla, C. O., Liu, S., Peca, L., Birnbaum, S. G., & Powell, C. M. (2018). Genetic background effects in Neuroligin-3 mutant mice: Minimal behavioral abnormalities on C57 background. *Autism Research*, 11(2), 234–244.
- Jaramillo, T. C., Liu, S., Pettersen, A., Birnbaum, S. G., & Powell, C. M. (2014). Autism-related neuroligin-3 mutation alters social behavior and spatial learning. *Autism Research*, 7(2), 264–272.
- Kolozsi, E., Mackenzie, R. N., Roulet, F. I., deCatanzaro, D., & Foster, J. A. (2009). Prenatal exposure to valproic acid leads to reduced expression of synaptic adhesion molecule neuroligin 3 in mice. *Neuroscience*, 163(4), 1201–1210.
- Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*, 7(8), 1511–1522.
- Laumonier, F., Bonnet-Brilhault, F., Gomot, M., Blanc, R., David, A., Moizard, M.-P., ... Briault, S. (2004). X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member

- of the neuroligin family. *American Journal of Human Genetics*, 74(3), 552–557.
- Mikhailov, A., Fennell, A., Plong-on, O., Sripo, T., Hansakunachai, T., Roongpraiwan, R., ... Limprasert, P. (2014). Screening of NLGN3 and NLGN4X genes in Thai children with autism spectrum disorder. *Psychiatric Genetics*, 24(1), 42–43.
- Pampanos, A., Volaki, K., Kanavakis, E., Papandreou, O., Youroukos, S., Thomaidis, L., ... Kitsiou-Tzeli, S. (2009). A substitution involving the NLGN4 gene associated with autistic behavior in the Greek population. *Genetic Testing and Molecular Biomarkers*, 13(5), 611–615.
- Pejaver, V., Mooney, S. D., & Radivojac, P. (2017). Missense variant pathogenicity predictors generalize well across a range of function-specific prediction challenges. *Human Mutation*, 38(9), 1092–1108.
- Quartier, A., Chatrousse, L., Redin, C., Keime, C., Haumesser, N., Maglott-Roth, A., ... Piton, A. (2018). Genes and pathways regulated by androgens in human neural cells, potential candidates for the male excess in autism spectrum disorder. *Biological Psychiatry*, 84, 239–252.
- Redin, C., Gérard, B., Lauer, J., Herenger, Y., Muller, J., Quartier, A., ... Piton, A. (2014). Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *Journal of Medical Genetics*, 51(11), 724–736.
- Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Hultman, C., Larsson, H., & Reichenberg, A. (2017). The heritability of autism spectrum disorder. *Journal of the American Medical Association*, 318(12), 1182–1184.
- Steinberg, K. M., Ramachandran, D., Patel, V. C., Shetty, A. C., Cutler, D. J., & Zwick, M. E. (2012). Identification of rare X-linked neuroligin variants by massively parallel sequencing in males with autism spectrum disorder. *Molecular Autism*, 3(1), 8.
- Tabuchi, K., Blundell, J., Etherton, M. R., Hammer, R. E., Liu, X., Powell, C. M., & Südhof, T. C. (2007). A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science*, 318(5847), 71–76.
- Talebizadeh, Z., Lam, D. Y., Theodoro, M. F., Bittel, D. C., Lushington, G. H., & Butler, M. G. (2006). Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism. *Journal of Medical Genetics*, 43, e21.
- Thomas, A. M., Schwartz, M. D., Saxe, M. D., & Kilduff, T. S. (2017). Sleep/wake physiology and quantitative electroencephalogram analysis of the neuroligin-3 knockout rat model of autism spectrum disorder. *Sleep*, 40(10).
- Trobiani, L., Favaloro, F. L., Di Castro, M. A., Di Mattia, M., Cariello, M., Miranda, E., ... De Jaco, A. (2018). UPR activation specifically modulates glutamate neurotransmission in the cerebellum of a mouse model of autism. *Neurobiology of Disease*, 120, 139–150.
- Ulbrich, L., Favaloro, F. L., Trobiani, L., Marchetti, V., Patel, V., Pascucci, T., ... De Jaco, A. (2016). Autism-associated R451C mutation in neuroligin3 leads to activation of the unfolded protein response in a PC12 Tet-On inducible system. *The Biochemical Journal*, 473(4), 423–434.
- Varoqueaux, F., Aramuni, G., Rawson, R. L., Mohrmann, R., Missler, M., Gottmann, K., ... Brose, N. (2006). Neuroligins determine synapse maturation and function. *Neuron*, 51(6), 741–754.
- Venkatesh, H. S., Johung, T. B., Caretti, V., Noll, A., Tang, Y., Nagaraja, S., ... Monje, M. (2015). Neuronal activity promotes glioma growth through neuroligin-3 secretion. *Cell*, 161(4), 803–816.
- Vincent, J. B., Kolozsvari, D., Roberts, W. S., Bolton, P. F., Gurling, H. M. D., & Scherer, S. W. (2004). Mutation screening of X-chromosomal neuroligin genes: No mutations in 196 autism probands. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, 129B(1), 82–84.
- Volaki, K., Pampanos, A., Kitsiou-Tzeli, S., Vrettou, C., Oikonomakis, V., Sofocleous, C., & Kanavakis, E. (2013). Mutation screening in the Greek population and evaluation of NLGN3 and NLGN4X genes causal factors for autism. *Psychiatric Genetics*, 23(5), 198–203.
- Wermter, A.-K., Kamp-Becker, I., Strauch, K., Schulte-Körne, G., & Remschmidt, H. (2008). No evidence for involvement of genetic variants in the X-linked neuroligin genes NLGN3 and NLGN4X in probands with autism spectrum disorder on high functioning level. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, 147B(4), 535–537.
- Xu, X., Hu, Z., Zhang, L., Liu, H., Cheng, Y., Xia, K., & Zhang, X. (2017). Not all neuroligin 3 and 4X missense variants lead to significant functional inactivation. *Brain and Behavior*, 7(9), e00793.
- Xu, X., Xiong, Z., Zhang, L., Liu, Y., Lu, L., Peng, Y., ... Hu, Z. (2014). Variations analysis of NLGN3 and NLGN4X gene in Chinese autism patients. *Molecular Biology Reports*, 41(6), 4133–4140.
- Yamagata, M., Sanes, J. R., & Weiner, J. A. (2003). Synaptic adhesion molecules. *Current Opinion in Cell Biology*, 15(5), 621–632.
- Yanagi, K., Kaname, T., Wakui, K., Hashimoto, O., Fukushima, Y., & Naritomi, K. (2012). Identification of four novel synonymous substitutions in the X-linked genes neuroligin 3 and neuroligin 4X in Japanese patients with autistic spectrum disorder. *Autism Research and Treatment*, 2012, 724072–724075.
- Yasuda, Y., Hashimoto, R., Yamamori, H., Ohi, K., Fukumoto, M., Umeda-Yano, S., ... Takeda, M. (2011). Gene expression analysis in lymphoblasts derived from patients with autism spectrum disorder. *Molecular Autism*, 2(1), 9.
- Ylisaakko-oja, T., Rehnström, K., Auranen, M., Vanhala, R., Alen, R., Kempas, E., ... Järvelä, I. (2005). Analysis of four neuroligin genes as candidates for autism. *European Journal of Human Genetics: EJHG*, 13(12), 1285–1292.
- Yu, T. W., Chahrour, M. H., Coulter, M. E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., ... Walsh, C. A. (2013). Using whole-exome sequencing to identify inherited causes of autism. *Neuron*, 77(2), 259–273.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Quartier A, Courraud J, Thi Ha T, et al. Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment. *Human Mutation*. 2019;1–12. <https://doi.org/10.1002/humu.23836>

# Sex-specific impact of prenatal androgens on social brain default mode subsystems

Michael V. Lombardo, Bonnie Auyeung, Tiziano Prampero, Angélique Quartier, Jérémie Courraud, Rosemary J. Holt, Jack Waldman, Amber N. V. Ruigrok, Natasha Mooney, Richard A. I. Bethlehem, Meng-Chuan Lai, Prantik Kundu, Edward T. Bullmore, Jean-Louis Mandel, Amélie Piton, Simon Baron-Cohen

Published online: 13 August 2018

Early-onset neurodevelopmental conditions (e.g., autism) affect males more frequently than females. Androgens may play a role in this male-bias by sex-differentially impacting early prenatal brain development, particularly neural circuits that later develop specialized roles in social cognition. Here, we find that increasing prenatal testosterone in humans is associated with later reduction of functional connectivity between social brain default mode (DMN) subsystems in adolescent males, but has no effect in females. Since testosterone can work directly via the androgen receptor (AR) or indirectly via the estrogen receptor through aromatase conversion to estradiol, we further examined how a potent non-aromatizable androgen, dihydrotestosterone (DHT), acts via the AR to influence gene expression in human neural stem cells (hNSC)—particularly for genes of high-relevance for DMN circuitry. DHT dysregulates a number of genes enriched for syndromic causes of autism and intellectual disability and for genes that in later development are expressed in anatomical patterns that highly correspond to the cortical midline DMN subsystem. DMN-related and DHT-affected genes (e.g., MEF2C) are involved in a number of synaptic processes, many of which impact excitation-inhibition balance. Androgens have male-specific prenatal influence over social brain circuitry in humans and may be relevant towards explaining some component of male-bias in early-onset neurodevelopmental conditions.

## Contribution

Pour ce travail j'ai réalisé les expériences de RT-qPCR afin de quantifier l'expression de l'ARNm MEF2C dans des ARN extraits de lignées de cellules hNSCs traitées par la DHT.

# Conditional switching of KIF2A mutation provides new insights into cortical malformation pathogenesis

Johan G. Gilet, Ekaterina L. Ivanova, Daria Trofimova, Gabrielle Rudolf, Hamid Meziane, Loic Broix, Nathalie Drouot, Jeremie Courraud, Valerie Skory, Paul Vouilleminot, Maria Osipenko, Nadia Bahi-Buisson, Binnaz Yalcin, Marie-Christine Birling, Maria-Victoria Hinckelmann, Benjamin H. Kwok, John S. Allingham and Jamel Chelly

Accepted: December 20, 2019

By using the Cre-mediated genetic switch technology, we were able to successfully generate a conditional knock-in mouse, bearing the KIF2A<sup>p.His321Asp</sup> missense point variant, identified in a subject with malformations of cortical development. These mice present with neuroanatomical anomalies and microcephaly associated with behavioral deficiencies and susceptibility to epilepsy, correlating with the described human phenotype. Using the flexibility of this model, we investigated RosaCre-, NestinCre- and NexCre-driven expression of the mutation to dissect the pathophysiological mechanisms underlying neurodevelopmental cortical abnormalities. We show that the expression of the p.His321Asp pathogenic variant increases apoptosis and causes abnormal multipolar to bipolar transition in newborn neurons, providing therefore insights to better understand cortical organization and brain growth defects that characterize KIF2A-related human disorders. We further demonstrate that the observed cellular phenotypes are likely to be linked to deficiency in the microtubule depolymerizing function of KIF2A.

## Contribution

Pour ce travail j'ai réalisé et analysé les expériences de PCR quantitative afin de valider l'excision/inversion au niveau de l'exon 10 de *Kif2a* dans des extraits d'ADN génomique de cortex de souris sauvage, KIF2A<sup>+H321D</sup> Rosa<sup>Cre</sup> et KIF2A<sup>+H321D</sup> Nestin<sup>Cre</sup>.

# ***De Novo* Frameshift Variants in the Neuronal Splicing Factor NOVA2 Result in a Common C-Terminal Extension and Cause a Severe Form of Neurodevelopmental Disorder**

Francesca Mattioli, Gaele Hayot, Nathalie Drouot, Bertrand Isidor, Jérémie Courraud, Maria-Victoria Hinckelmann, Frederic Tran Mau-Them, Chantal Sellier, Alica Goldman, Aida Telegrafi, Alicia Boughton, Candace Gamble, Sebastien Moutton, Angélique Quartier, Nolwenn Jean, Paul Van Ness, Sarah Grotto, Sophie Nambot, Ganka Douglas, Yue Cindy Si, Jamel Chelly, Zohra Shad, Elisabeth Kaplan, Richard Dineen, Christelle Golzio, Nicolas Charlet-Berguerand, Jean-Louis Mandel, and Amélie Piton

Published online: 19 mars 2020

The neuro-oncological ventral antigen 2 (NOVA2) protein is a major factor regulating neuron-specific alternative splicing (AS), previously associated with an acquired neurologic condition, the paraneoplastic opsoclonus-myoclonus ataxia (POMA). We report here six individuals with *de novo* frameshift variants in *NOVA2* affected with a severe neurodevelopmental disorder characterized by intellectual disability (ID), motor and speech delay, autistic features, hypotonia, feeding difficulties, spasticity or ataxic gait, and abnormal brain MRI. The six variants lead to the same reading frame, adding a common proline rich C-terminal part instead of the last KH RNA binding domain. We detected 41 genes differentially spliced after *NOVA2* down regulation in human neural cells. The *NOVA2* variant protein shows decreased ability to bind target RNA sequences and to regulate target AS events. It also fails to complement the effect on neurite outgrowth induced by *NOVA2* down regulation in vitro and to rescue alterations of retinotectal axonal pathfinding induced by loss of NOVA2 ortholog in zebrafish. Our results suggest a partial loss-of-function mechanism rather than a full heterozygous loss-of-function, although a specific contribution of the novel C-terminal extension cannot be excluded.

## **Contribution**

J'ai pour ce travail réalisé la culture des hNSCs SA001 ainsi que les transfections de siRNA dans ce modèle. J'ai également extraits les ARN totaux utilisés pour les expériences de RT-qPCR et réalisé avec Nathalie Drouot la préparation des bibliothèques pour le séquençage d'ARN.

# Biallelic *PDE2A* variants: a new cause of syndromic paroxysmal dyskinesia

Diane Doummar, Christel Dentel, Romane Lyautey, Julia Metreau, Boris Keren, Nathalie Drouot, Ludivine Malherbe, Viviane Bouilleret, Jérémie Courraud, Maria Paola Valenti-Hirsch, Lorella Minotti, Blandine Dozieres-Puyravel, Séverine Bär, Julia Scholly, Elise Schaefer, Caroline Nava, Thomas Wirth, Hala Nasser, Marie de Salins, Anne de Saint Martin, Marie Thérèse Abi Warde, Philippe Kahane, Edouard Hirsch, Mathieu Anheim, Sylvie Friant, Jamel Chelly, Cyril Mignot, Gabrielle Rudolf

Published online: 28 May 2020

Cause of complex dyskinesia remains elusive in some patients. A homozygous missense variant leading to drastic decrease of PDE2A enzymatic activity was reported in one patient with childhood-onset choreodystonia preceded by paroxysmal dyskinesia and associated with cognitive impairment and interictal EEG abnormalities. Here, we report three new cases with biallelic *PDE2A* variants identified by trio whole-exome sequencing. Mitochondria network was analyzed after Mitotracker™Red staining in control and mutated primary fibroblasts. Analysis of retrospective video of patients' movement disorder and refinement of phenotype was carried out. We identified a homozygous gain of stop codon variant c.1180C>T; p.(Gln394\*) in *PDE2A* in siblings and compound heterozygous variants in young adult: a missense c.446C>T; p.(Pro149Leu) and splice-site variant c.1922+5G>A predicted and shown to produce an out of frame transcript lacking exon22. All three patients had cognitive impairment or developmental delay. The phenotype of the two oldest patients, aged 9 and 26, was characterized by childhood-onset refractory paroxysmal dyskinesia initially misdiagnosed as epilepsy due to interictal EEG abnormalities. The youngest patient showed a proven epilepsy at the age of 4 months and no paroxysmal dyskinesia at 15 months. Interestingly, analysis of the fibroblasts with the biallelic variants in *PDE2A* variants revealed mitochondria network morphology changes. Together with previously reported case, our three patients confirm that biallelic *PDE2A* variants are a cause of childhood-onset refractory paroxysmal dyskinesia with cognitive impairment, sometimes associated with choreodystonia and interictal baseline EEG abnormalities or epilepsy.

## Contribution

J'ai pour ce travail cultivé les lignées de fibroblastes contrôles utilisées pour les analyses mitotracker.

# RÉFÉRENCES

---





- Abdel-Salam, Ghada M H et al. « Phenotypic and Molecular Insights into PQBP1-related Intellectual Disability ». *American Journal of Medical Genetics* 176(11):2446-50
- Abekhoukh, Sabiha et al. 2013. « Dyrk1A, a Serine/Threonine Kinase, Is Involved in ERK and Akt Activation in the Brain of Hyperhomocysteinemic Mice ». *Molecular Neurobiology* 47(1): 105-16.
- Álvarez, Mónica, Xavier Estivill, et Susana de la Luna. 2003. « DYRK1A Accumulates in Splicing Speckles through a Novel Targeting Signal and Induces Speckle Disassembly ». *Journal of Cell Science* 116(15): 3099-3107.
- Alvarez, Altafaj et al. 2007 « DYRK1A autophosphorylation on Serine Residue 520 Modulate Its Kinase Activity via 14-3-3 Binding ». *Molecular Biology of the Cell* 18(1): 1167-78
- Amendola, Laura M. et al. 2016. « Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium ». *The American Journal of Human Genetics* 98(6): 1067-76.
- Anderson, Paul, et Nancy Kedersha. 2009. « RNA Granules: Post-Transcriptional and Epigenetic Modulators of Gene Expression ». *Nature Reviews Molecular Cell Biology* 10(6): 430-36.
- Anjum, Rana, et John Blenis. 2008. « The RSK Family of Kinases: Emerging Roles in Cellular Signalling ». *Nature Reviews Molecular Cell Biology* 9(10): 747-58.
- Aranda, Sergi, Ariadna Laguna, et Susana de la Luna. 2011. « DYRK Family of Protein Kinases: Evolutionary Relationships, Biochemical Properties, and Functional Roles ». *The FASEB Journal* 25(2): 449-62.
- Arbones, Maria L. et al. 2019. « DYRK1A and Cognition: A Lifelong Relationship ». *Pharmacology & Therapeutics* 194: 199-221.
- Aref-Eshghi, Erfan et al. 2020. « Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders ». *The American Journal of Human Genetics* 106(3): 356-70.
- Arqué, Glòria et al. 2008. « Impaired Spatial Learning Strategies and Novel Object Recognition in Mice Haploinsufficient for the Dual Specificity Tyrosine-Regulated Kinase-1A (Dyrk1A) » éd. Wim E. Crusio. *PLoS ONE* 3(7): e2575.
- Arqué, Glòria, María Martínez de Lagrán, Maria L. Arbonés, et Mara Dierssen. 2009. « Age-Associated Motor and Visuo-Spatial Learning Phenotype in Dyrk1A Heterozygous Mutant Mice ». *Neurobiology of Disease* 36(2): 312-19.
- Arranz, Juan et al. 2019. « Impaired Development of Neocortical Circuits Contributes to the Neurological Alterations in DYRK1A Haploinsufficiency Syndrome ». *Neurobiology of Disease* 127: 210-22.
- Arron, Joseph R. et al. 2006. « NFAT Dysregulation by Increased Dosage of DSCR1 and DYRK1A on Chromosome 21 ». *Nature* 441(7093): 595-600.

- Barallobre, M J et al. 2014. « DYRK1A Promotes Dopaminergic Neuron Survival in the Developing Brain and in a Mouse Model of Parkinson's Disease ». *Cell Death & Disease* 5(6): e1289-e1289.
- Barford, David. 2011. « Structure, Function and Mechanism of the Anaphase Promoting Complex (APC/C) ». *Quarterly Reviews of Biophysics* 44(2): 153-90.
- Beaulieu, Chandree L et al. 2013. « Intellectual Disability Associated with a Homozygous Missense Mutation in THOC6 ». *Orphanet Journal of Rare Diseases* 8(1): 62.
- Benavidespiccione, R et al. 2005. « Alterations in the Phenotype of Neocortical Pyramidal Cells in the Dyrk1A<sup>+/-</sup> Mouse ». *Neurobiology of Disease* 20(1): 115-22.
- Bessone et al. 2021 « DYRK1A role in microtubule-based axonal transport regulates the retrograde dynamics of APP vesicles in human neurons ». *BioRxiv Preprint*.
- Beunders, Gea et al. 2013. « Exonic Deletions in AUTS2 Cause a Syndromic Form of Intellectual Disability and Suggest a Critical Role for the C Terminus ». *The American Journal of Human Genetics* 92(2): 210-20.
- Blackburn, Alexandria T. M. et al. 2019. « DYRK1A-Related Intellectual disability: A Syndrome Associated with Congenital Anomalies of the Kidney and Urinarytract ». *Genetics in Medicine* 21(12): 2755-64.
- Boissart, C et al. 2013. « Differentiation from Human Pluripotent Stem Cells of Cortical Neurons of the Superficial Layers Amenable to Psychiatric Disease Modeling and High-Throughput Drug Screening ». *Translational Psychiatry* 3(8): e294-e294.
- van Bon, B W M et al. 2016. « Disruptive de Novo Mutations of DYRK1A Lead to a Syndromic Form of Autism and ID ». *Molecular Psychiatry* 21(1): 126-32.
- Bonhoure, Anna et al. 2020. « Benchtop Holdup Assay for Quantitative Affinity-Based Analysis of Sequence Determinants of Protein-Motif Interactions ». *Analytical Biochemistry* 603: 113772.
- Boni, Jacopo et al. 2020. « The DYRK Family of Kinases in Cancer: Molecular Functions and Therapeutic Opportunities ». *Cancers* 12(8): 2106.
- Bronicki, Lucas M et al. 2015. « Ten New Cases Further Delineate the Syndromic Intellectual Disability Phenotype Caused by Mutations in DYRK1A ». *European Journal of Human Genetics* 23(11): 1482-87.
- Buntinx, Cans, Colleaux, Courbois, Debbané, et al. 2016 « Déficiences intellectuelles. » Rapport de recherche, *Institut national de la santé et de la recherche médicale* Éditions EDP Sciences ISSN : 1264-1782
- Calzari, Luciano et al. 2020. « Transcriptome Analysis of iPSC-Derived Neurons from Rubinstein-Taybi Patients Reveals Deficits in Neuronal Differentiation ». *Molecular Neurobiology* 57(9): 3685-3701.
- Chakrabarti, L., Z. Galdzicki, et T. F. Haydar. 2007. « Defects in Embryonic Neurogenesis and Initial Synapse Formation in the Forebrain of the Ts65Dn Mouse Model of Down

- Syndrome ». *Journal of Neuroscience* 27(43): 11483-95.
- Chakraborty, A. et al. 2011. « Casein Kinase-2 Mediates Cell Survival through Phosphorylation and Degradation of Inositol Hexakisphosphate Kinase-2 ». *Proceedings of the National Academy of Sciences* 108(6): 2205-9.
- Chen, Brianna et al. 2021. « DYRK1A Negatively Regulates CDK5-SOX2 Pathway and Self-Renewal of Glioblastoma Stem Cells ». *International Journal of Molecular Sciences* 22(8): 4011.
- Chen, Chun-Kan et al. 2014. « Activity-Dependent Facilitation of Synaptojanin and Synaptic Vesicle Recycling by the Minibrain Kinase ». *Nature Communications* 5(1): 4246.
- Chen, Jia-Yun, Jia-Ren Lin, Feng-Chiao Tsai, et Tobias Meyer. 2013. « Dosage of Dyrk1a Shifts Cells within a P21-Cyclin D1 Signaling Map to Control the Decision to Enter the Cell Cycle ». *Molecular Cell* 52(1): 87-100.
- Chiurazzi, Pietro, et Filomena Pirozzi. 2016. « Advances in Understanding – Genetic Basis of Intellectual Disability ». *F1000Research* 5: 599.
- Cho, Hyun-Ju et al. 2019. « Vascular Defects of *DYRK1A* Knockouts Are Ameliorated by Modulating Calcium Signaling in Zebrafish ». *Disease Models & Mechanisms*: dmm.037044.
- Choufani, Sanaa et al. 2020. « DNA Methylation Signature for EZH2 Functionally Classifies Sequence Variants in Three PRC2 Complex Genes ». *The American Journal of Human Genetics* 106(5): 596-610.
- Contestabile, Andrea et al. 2007. « Cell Cycle Alteration and Decreased Cell Proliferation in the Hippocampal Dentate Gyrus and in the Neocortical Germinal Matrix of Fetuses with down Syndrome and in Ts65Dn Mice ». *Hippocampus* 17(8): 665-78.
- Cossée, Mireille et al. 2006. « Exonic Microdeletions in the X-Linked PQBP1 Gene in Mentally Retarded Patients: A Pathogenic Mutation and in-Frame Deletions of Uncertain Effect ». *European Journal of Human Genetics* 14(4): 418-25.
- Courcet, Jean-Benoît et al. 2012. « The *DYRK1A* Gene Is a Cause of Syndromic Intellectual Disability with Severe Microcephaly and Epilepsy ». *Journal of Medical Genetics* 49(12): 731-36.
- Courraud, Chater-Diehl, Durand et al. 2021 « Integrative approach to interpret *DYRK1A* variants, leading to a frequent neurodevelopmental disorder ». *Genetic in Medicine* 10.1038/s41436-021-01263-1. Online ahead of print.
- Dang, T et al. 2018. « Autism-Associated Dyrk1a Truncation Mutants Impair Neuronal Dendritic and Spine Growth and Interfere with Postnatal Cortical Development ». *Molecular Psychiatry* 23(3): 747-58.
- De Toma, Ilario et al. 2019. « *DYRK1A* Overexpression Alters Cognition and Neural-Related Proteomic Pathways in the Hippocampus That Are Rescued by Green Tea Extract and/or Environmental Enrichment ». *Frontiers in Molecular Neuroscience* 12: 272.

- Degoutin, Joffrey L. et al. 2013. « Riquiqui and Minibrain Are Regulators of the Hippo Pathway Downstream of Dachshous ». *Nature Cell Biology* 15(10): 1176-85.
- Depienne, Christel et al. 2009. « Sporadic Infantile Epileptic Encephalopathy Caused by Mutations in PCDH19 Resembles Dravet Syndrome but Mainly Affects Females » éd. Miriam H. Meisler. *PLoS Genetics* 5(2): e1000381.
- Dibbens, Leanne M et al. 2008. « X-Linked Protocadherin 19 Mutations Cause Female-Limited Epilepsy and Cognitive Impairment ». *Nature Genetics* 40(6): 776-81.
- Ding, Shaohong et al. 2012. « Regulation of Alternative Splicing of Tau Exon 10 by 9G8 and Dyrk1A ». *Neurobiology of Aging* 33(7): 1389-99.
- Di Vona, Chiara et al. 2015. « Chromatin-Wide Profiling of DYRK1A Reveals a Role as a Gene-Specific RNA Polymerase II CTD Kinase ». *Molecular Cell* 57(3): 506-20.
- Di Vona, et al. 2021. « The DYRK1A protein kinase regulates ribosome mass and translation rates through transcriptional control of ribosomal protein genes expression ». *BioRxiv* Preprint
- Doummar et al. 2020 « Biallelic PDE2A variants: a new cause of syndromic paroxysmal dyskinesia ». *European Journal of Human genetics* 28(10):1403-13
- Duchon, Arnaud, et Yann Herault. 2016. « DYRK1A, a Dosage-Sensitive Gene Involved in Neurodevelopmental Disorders, Is a Target for Drug Development in Down Syndrome ». *Frontiers in Behavioral Neuroscience* 10.  
<http://journal.frontiersin.org/Article/10.3389/fnbeh.2016.00104/abstract> (28 août 2021).
- Earl, Rachel K. et al. 2017. « Clinical Phenotype of ASD-Associated DYRK1A Haploinsufficiency ». *Molecular Autism* 8(1): 54.
- Ehe, Ben K. et al. 2017. « Identification of a DYRK1A-Mediated Phosphorylation Site within the Nuclear Localization Sequence of the Hedgehog Transcription Factor GLI1 ». *Biochemical and Biophysical Research Communications* 491(3): 767-72.
- Ernst, Julia et al. 2020. « Ocular Findings of Albinism in *DYRK1A*- Related Intellectual Disability Syndrome ». *Ophthalmic Genetics* 41(6): 650-55.
- Evers, Jochem M.G. et al. 2017. « Structural Analysis of Pathogenic Mutations in the *DYRK1A* Gene in Patients with Developmental Disorders ». *Human Molecular Genetics*: ddw409.
- Fernández, Virginia, Cristina Llinares-Benadero, et Víctor Borrell. 2016. « Cerebral Cortex Expansion and Folding: What Have We Learned? » *The EMBO Journal* 35(10): 1021-44.
- Ferron, Sacri R. et al. 2010. « Regulated Segregation of Kinase Dyrk1A during Asymmetric Neural Stem Cell Division Is Critical for EGFR-Mediated Biased Signaling ». *Cell Stem Cell* 7(3): 367-79.
- Findlay, Gregory M. et al. 2018. « Accurate Classification of BRCA1 Variants with

- Saturation Genome Editing ». *Nature* 562(7726): 217-22.
- Firth, Helen V. et al. 2009. « DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources ». *The American Journal of Human Genetics* 84(4): 524-33.
- Fischbach, K F, et M Heisenberg. « NEUROGENETICS AND BEHAVIOUR IN INSECTS ». *The company of Biologist limited* 112,65-93.
- Flynn, Maureen, Ying S. Zou, et Aubrey Milunsky. 2011. « Whole Gene Duplication of the PQBP1 Gene in Syndrome Resembling Renpenning ». *American Journal of Medical Genetics Part A* 155(1): 141-44.
- Fotaki, V. et al. 2004. « Haploinsufficiency of Dyrk1A in Mice Leads to Specific Alterations in the Development and Regulation of Motor Activity. » *Behavioral Neuroscience* 118(4): 815-21.
- Fotaki, Vassiliki et al. 2002. « *Dyrk1A* Haploinsufficiency Affects Viability and Causes Developmental Delay and Abnormal Brain Morphology in Mice ». *Molecular and Cellular Biology* 22(18): 6636-47.
- Gartel, Andrei L, et Angela L Tyner. 2002 « The Role of the Cyclin-Dependent Kinase Inhibitor P21 in Apoptosis ». *Molecular cancer and therapeutics* 1,639-49
- Germanaud, D et al. 2011. « The Renpenning Syndrome Spectrum: New Clinical Insights Supported by 13 New PQBP1-Mutated Males ». *Clinical Genetics* 79(3): 225-35.
- Ghanem, Louis, et Richard A Steinman. 2006. « P21Waf1 Inhibits Granulocytic Differentiation of 32Dcl3 Cells ». *Leukemia Research*: 8:1285-92
- Gilet et al. 2020 « Conditional switching of KIF2A mutation provides new insights into cortical malformation pathogeny ». *Human Molecular Genetic* 29(5):766-784
- Gilissen, Christian et al. 2014. « Genome Sequencing Identifies Major Causes of Severe Intellectual Disability ». *Nature* 511(7509): 344-47.
- Gong, Siew-Ging. 2014. « Isoforms of Receptors of Fibroblast Growth Factors: FGFR ISOFORMS ». *Journal of Cellular Physiology* 229(12): 1887-95.
- Graaf, Katrin et al. 2006. « The protein kinase DYRK1A phosphorylates the splicing factor SF3b1/SAP155 at Thr434, a novel in vivo phosphorylation site ». *BMC Biochemistry* 7(1): 7.
- Grau et al. 2014 «DYRK1A-mediated phosphorylation of GluN2A at Ser1048 regulates the surface expression and channel activity of GluN1/GluN2A receptors ». *Frontiers in Cellular Neuroscience* 17;8:331.
- Guard, Steven E. et al. 2019. « The Nuclear Interactome of DYRK1A Reveals a Functional Role in DNA Damage Repair ». *Scientific Reports* 9(1): 6539.
- Guedj, Fayçal et al. 2012. « DYRK1A: A Master Regulatory Protein Controlling Brain Growth ». *Neurobiology of Disease* 46(1): 190-203.

- Guimera, J. 1996. « A Human Homologue of Drosophila Minibrain (MNB) Is Expressed in the Neuronal Regions Affected in Down Syndrome and Maps to the Critical Region ». *Human Molecular Genetics* 5(9): 1305-10.
- Guimera, Jordi, Caty Casas, Xavier Estivill, et Melanie Pritchard. 1999. « Human Minibrain Homologue (MNBH/DYRK1): Characterization, Alternative Splicing, Differential Tissue Expression, and Overexpression in Down Syndrome ». *Genomics* 57(3): 407-18.
- Guo, Xiumei, Jason G. Williams, Thaddeus T. Schug, et Xiaoling Li. 2010. « DYRK1A and DYRK3 Promote Cell Survival through Phosphorylation and Activation of SIRT1 ». *Journal of Biological Chemistry* 285(17): 13223-32.
- Hämmerle, B. et al. 2003. « Expression Patterns and Subcellular Localization of the Down Syndrome Candidate Protein MNB/DYRK1A Suggest a Role in Late Neuronal Differentiation: *Mnb* Expression during Neuronal Differentiation ». *European Journal of Neuroscience* 17(11): 2277-86.
- Hämmerle, Barbara et al. 2011. « Transient Expression of *Mnb/Dyrk1a* Couples Cell Cycle Exit and Differentiation of Neuronal Precursors by Inducing *P27KIP1* Expression and Suppressing NOTCH Signaling ». *Development* 138(12): 2543-54.
- Hämmerle, Barbara, Carina Elizalde, et Francisco J. Tejedor. 2008. « The Spatio-Temporal and Subcellular Expression of the Candidate Down Syndrome Gene *Mnb/Dyrk1A* in the Developing Mouse Brain Suggests Distinct Sequential Roles in Neuronal Development ». *European Journal of Neuroscience* 27(5): 1061-74.
- Hämmerle, Christoph H. F., Ronald E. Jung, et Andreas Feloutzis. 2002. « A Systematic Review of the Survival of Implants in Bone Sites Augmented with Barrier Membranes (Guided Bone Regeneration) in Partially Edentulous Patients: Implants in Regenerated Bone ». *Journal of Clinical Periodontology* 29: 226-31.
- Han, Tao et al. 2019. « Interplay between C-Src and the APC/C Co-Activator Cdh1 Regulates Mammary Tumorigenesis ». *Nature Communications* 10(1): 3716.
- Hashimoto, S. et al. 2011. « MED23 Mutation Links Intellectual Disability to Dysregulation of Immediate Early Gene Expression ». *Science* 333(6046): 1161-63.
- Hille, Susanne et al. 2016. « *Dyrk1a* Regulates the Cardiomyocyte Cell Cycle via D-Cyclin-Dependent Rb/E2f-Signalling ». *Cardiovascular Research* 110(3): 381-94.
- Himpel, Sunke et al. 2000. « Specificity Determinants of Substrate Recognition by the Protein Kinase DYRK1A ». *Journal of Biological Chemistry* 275(4): 2431-38.
- Himpel et al. 2001. « Identification of the Autophosphorylation Sites and Characterization of Their Effects in the Protein Kinase DYRK1A ». : 9.
- Hornbeck, Peter V. et al. 2015. « PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations ». *Nucleic Acids Research* 43(D1): D512-20.
- Hu, H et al. 2016. « X-Exome Sequencing of 405 Unresolved Families Identifies Seven Novel Intellectual Disability Genes ». *Molecular Psychiatry* 21(1): 133-48.

- Hu, Hao et al. 2019. « Genetics of Intellectual Disability in Consanguineous Families ». *Molecular Psychiatry* 24(7): 1027-39.
- Huang, Ju, et Azad Bonni. 2016. « A Decade of the Anaphase-Promoting Complex in the Nervous System ». *Genes & Development* 30(6): 622-38.
- Huang, Yu et al. « Mnb/Dyrk1A Phosphorylation Regulates the Interaction of Dynamin 1 with SH3 Domain-Containing Proteins ». : 13.
- Huttlin, Edward L. et al. 2015. « The BioPlex Network: A Systematic Exploration of the Human Interactome ». *Cell* 162(2): 425-40.
- Huttlin et al. 2017. « Architecture of the Human Interactome Defines Protein Communities and Disease Networks ». *Nature* 545(7655): 505-9.
- Iglesias, Alejandro et al. 2014. « The Usefulness of Whole-Exome Sequencing in Routine Clinical Practice ». *Genetics in Medicine* 16(12): 922-31.
- Ikeuchi, Yoshiho et al. 2013. « The XLID Protein PQBP1 and the GTPase Dynamin 2 Define a Signaling Link That Orchestrates Ciliary Morphogenesis in Postmitotic Neurons ». *Cell Reports* 4(5): 879-89.
- Imafuku, Ichiro et al. 1998. « Polar Amino Acid-Rich Sequences Bind to Polyglutamine Tracts ». *Biochemical and Biophysical Research Communications* 253(1): 16-20.
- Iossifov, Ivan et al. 2012. « De Novo Gene Disruptions in Children on the Autistic Spectrum ». *Neuron* 74(2): 285-99.
- Iossifov et al. 2014. « The Contribution of de Novo Coding Mutations to Autism Spectrum Disorder ». *Nature* 515(7526): 216-21.
- Ishihara, Keiichi et al. 2010. « Enlarged Brain Ventricles and Impaired Neurogenesis in the Ts1Cje and Ts2Cje Mouse Models of Down Syndrome ». *Cerebral Cortex* 20(5): 1131-43.
- Ito, H et al. 2015. « In Utero Gene Therapy Rescues Microcephaly Caused by Pqbp1-Hypofunction in Neural Stem Progenitor Cells ». *Molecular Psychiatry* 20(4): 459-71.
- Ito, Hikaru et al. 2009. « Knock-down of PQBP1 Impairs Anxiety-Related Cognition in Mouse ». *Human Molecular Genetics* 18(22): 4239-54.
- Iwasaki, Yasuno, et Gerald H. Thomsen. 2014. « The Splicing Factor PQBP1 Regulates Mesodermal and Neural Development through FGF Signaling ». *Development* 141(19): 3740-51.
- Jang, Suk Min, Saliha Azebi, Guillaume Soubigou, et Christian Muchardt. 2014. « DYRK1A Phosphorylates Histone H3 to Differentially Regulate the Binding of HP1 Isoforms and Antagonize HP1-mediated Transcriptional Repression ». *EMBO reports* 15(6): 686-94.
- Ji, Jianling et al. 2015. « DYRK1A Haploinsufficiency Causes a New Recognizable Syndrome with Microcephaly, Intellectual Disability, Speech Impairment, and Distinct



- Facies ». *European Journal of Human Genetics* 23(11): 1473-81.
- Jiang, Xiaoling et al. 2015. « Genetic Dissection of the Down Syndrome Critical Region ». *Human Molecular Genetics* 24(22): 6540-51.
- Jin, Nana et al. 2015. « Truncation and Activation of Dual Specificity Tyrosine Phosphorylation-Regulated Kinase 1A by Calpain I ». *Journal of Biological Chemistry* 290(24): 15219-37.
- de Jong, Tristan V., Yuri M. Moshkin, et Victor Guryev. 2019. « Gene Expression Variability: The Other Dimension in Transcriptome Analysis ». *Physiological Genomics* 51(5): 145-58.
- Kalscheuer, Vera M et al. 2003. « Mutations in the Polyglutamine Binding Protein 1 Gene Cause X-Linked Mental Retardation ». *Nature Genetics* 35(4): 313-15.
- Kannan, Madhuvanthi, Shih-Ju Lee, Nicola Schwedhelm-Domeyer, et Judith Stegmüller. 2012. « The E3 Ligase Cdh1-Anaphase Promoting Complex Operates Upstream of the E3 Ligase Smurf1 in the Control of Axon Growth ». *Development* 139(19): 3600-3612.
- Karczewski, Konrad J. et al. 2020. « The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans ». *Nature* 581(7809): 434-43.
- Kentrup, Heiner et al. 1996. « Dyrk, a Dual Specificity Protein Kinase with Unique Structural Features Whose Activity Is Dependent on Tyrosine Residues between Subdomains VII and VIII ». *Journal of Biological Chemistry* 271(7): 3488-95.
- Kim, Doyeun et al. 2004. « Regulation of Dyrk1A Kinase Activity by 14-3-3 ». *Biochemical and Biophysical Research Communications* 323(2): 499-504.
- Kim, Jung-Hyun, Liping Li, et Linda M.S. Resar. 2021. « Doubling up on Function: Dual-Specificity Tyrosine-Regulated Kinase 1A (DYRK1A) in B Cell Acute Lymphoblastic Leukemia ». *Journal of Clinical Investigation* 131(1): e142627.
- Kim, M.-Y. et al. 2006. « A Repressor Complex, AP4 Transcription Factor and Geminin, Negatively Regulates Expression of Target Genes in Nonneuronal Cells ». *Proceedings of the National Academy of Sciences* 103(35): 13074-79.
- Kim, Oc-Hee et al. 2017. « Zebrafish Knockout of Down Syndrome Gene, DYRK1A, Shows Social Impairments Relevant to Autism ». *Molecular Autism* 8(1): 50.
- Kinstrie, R. et al. 2010. « Characterization of a Domain That Transiently Converts Class 2 DYRKs into Intramolecular Tyrosine Kinases ». *Science Signaling* 3(111): ra16-ra16.
- Kircher, Martin et al. 2014. « A General Framework for Estimating the Relative Pathogenicity of Human Genetic Variants ». *Nature Genetics* 46(3): 310-15.
- Kishino, Tatsuya, Marc Lalonde, et Joseph Wagstaff. 1997. « UBE3A/E6-AP Mutations Cause Angelman Syndrome ». 15: 4.
- Kochinke, Korinna et al. 2016. « Systematic Phenomics Analysis Deconvolutes Genes

- Mutated in Intellectual Disability into Biologically Coherent Modules ». *The American Journal of Human Genetics* 98(1): 149-64.
- Kuhn, Christian et al. 2009. « DYRK1A Is a Novel Negative Regulator of Cardiomyocyte Hypertrophy ». *Journal of Biological Chemistry* 284(25): 17320-27.
- Kurabayashi, N., et K. Sanada. 2013. « Increased Dosage of DYRK1A and DSCR1 Delays Neuronal Differentiation in Neocortical Progenitor Cells ». *Genes & Development* 27(24): 2708-21.
- Laguna, Ariadna et al. 2008. « The Protein Kinase DYRK1A Regulates Caspase-9-Mediated Apoptosis during Retina Development ». *Developmental Cell* 15(6): 841-53.
- Laguna et al. 2013. « Triplication of DYRK1A Causes Retinal Structural and Functional Alterations in Down Syndrome ». *Human Molecular Genetics* 22(14): 2775-84.
- Laham, Amina Jamal, Maha Saber-Ayad, et Raafat El-Awady. 2021. « DYRK1A: A down Syndrome-Related Dual Protein Kinase with a Versatile Role in Tumorigenesis ». *Cellular and Molecular Life Sciences* 78(2): 603-19.
- Lee, Brittany J. et al. 2006. « Rules for Nuclear Localization Sequence Recognition by Karyopherin $\beta$ 2 ». *Cell* 126(3): 543-58.
- Lee, Kyu-Sun et al. 2020. « A Novel de Novo Heterozygous DYRK1A Mutation Causes Complete Loss of DYRK1A Function and Developmental Delay ». *Scientific Reports* 10(1): 9849.
- Lee, Sang Bae et al. 2016. « An ID2-Dependent Mechanism for VHL Inactivation in Cancer ». *Nature* 529(7585): 172-77.
- Lee, Ye Hyung et al. 2021. « Protein Phosphatase PPM1B Inhibits DYRK1A Kinase through Dephosphorylation of PS258 and Reduces Toxic Tau Aggregation ». *Journal of Biological Chemistry* 296: 100245.
- Lee, Youngkyun et al. 2009. « Negative Feedback Inhibition of NFATc1 by DYRK1A Regulates Bone Homeostasis ». *Journal of Biological Chemistry* 284(48): 33343-51.
- Lejeune, Turpin, Gautier. 1962 « Chromosomic diagnosis of mongolism ». *Arch Fr Pediatr.* 1959;16:962-3.
- Leonard, Helen, et Xingyan Wen. 2002. « The Epidemiology of Mental Retardation: Challenges and Opportunities in the New Millennium ». *Mental Retardation and Developmental Disabilities Research Reviews* 8(3): 117-34.
- Lepagnol-Bestel, Aude-Marie et al. 2009. « DYRK1A Interacts with the REST/NRSF-SWI/SNF Chromatin Remodelling Complex to Deregulate Gene Clusters Involved in the Neuronal Phenotypic Traits of Down Syndrome ». *Human Molecular Genetics* 18(8): 1405-14.
- Li, Chan et al. 2013. « Sox2 Transcriptionally Regulates Pqbp1, an Intellectual Disability-Microcephaly Causative Gene, in Neural Stem Progenitor Cells » éd. Debra L. Silver. *PLoS ONE* 8(7): e68627.

- Li, Dan, Rebecca A. Jackson, Permeen Yusoff, et Graeme R. Guy. 2010. « Direct Association of Sprouty-Related Protein with an EVH1 Domain (SPRED) 1 or SPRED2 with DYRK1A Modifies Substrate/Kinase Interactions ». *Journal of Biological Chemistry* 285(46): 35374-85.
- Li, Jingjing et al. 2010. « Gene Expression Variability within and between Human Populations and Implications toward Disease Susceptibility » éd. Donna K. Slonim. *PLoS Computational Biology* 6(8): e1000910.
- Li, Shanshan et al. 2018. « DYRK1A Interacts with Histone Acetyl Transferase P300 and CBP and Localizes to Enhancers ». *Nucleic Acids Research* 46(21): 11202-13.
- de Ligt, Joep et al. 2012. « Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability ». *New England Journal of Medicine* 367(20): 1921-29.
- Litovchick, L. et al. 2011. « DYRK1A Protein Kinase Promotes Quiescence and Senescence through DREAM Complex Assembly ». *Genes & Development* 25(8): 801-13.
- Liu, Heng et al. 2017. « NFATc1 Phosphorylation by DYRK1A Increases Its Protein Stability » éd. Sakamuri V. Reddy. *PLOS ONE* 12(2): e0172985.
- Liu, Xian, Lin-Xia Dou, Junhai Han, et Zi Chao Zhang. 2020. « The Renpenning Syndrome-Associated Protein PQBP1 Facilitates the Nuclear Import of Splicing Factor TXNL4A through the Karyopherin B2 Receptor ». *Journal of Biological Chemistry* 295(13): 4093-4100.
- Llorian, Miriam et al. 2004. « SIPP1, a Novel Pre-mRNA Splicing Factor and Interactor of Protein Phosphatase-1 ». *Biochemical Journal* 378(1): 229-38.
- Lochhead, Pamela A., Gary Sibbet, Nick Morrice, et Vaughn Cleghon. 2005. « Activation-Loop Autophosphorylation Is Mediated by a Novel Transitional Intermediate Form of DYRKs ». *Cell* 121(6): 925-36.
- Lombardo et al. 2018 « Sex-specific impact of prenatal androgens on social brain default mode subsystems » *Molecular psychiatry* 25(9):2175-88
- Lonsdale, John et al. 2013. « The Genotype-Tissue Expression (GTEx) Project ». *Nature Genetics* 45(6): 580-85.
- Lu, Mei et al. 2011a. « REST Regulates DYRK1A Transcription in a Negative Feedback Loop ». *Journal of Biological Chemistry* 286(12): 10755-63.
- Lu et al. 2019. « DYRK1A Aggravates  $\beta$  Cell Dysfunction and Apoptosis by Promoting the Phosphorylation and Degradation of IRS2 ». *Experimental Gerontology* 125: 110659.
- Lubs, H. 2006. « Golabi-Ito-Hall Syndrome Results from a Missense Mutation in the WW Domain of the PQBP1 Gene ». *Journal of Medical Genetics* 43(6): e30-e30.
- Luco, Stephanie M. et al. 2016. « Case Report of Novel DYRK1A Mutations in 2 Individuals with Syndromic Intellectual Disability and a Review of the Literature ». *BMC Medical Genetics* 17(1): 15.

- Lyst, Matthew J., et Adrian Bird. 2015. « Rett Syndrome: A Complex Disorder with Simple Roots ». *Nature Reviews Genetics* 16(5): 261-75.
- Macias, Maria J, Silke Wiesner, et Marius Sudol. 2002. « WW and SH3 Domains, Two Different Scaffolds to Recognize Proline-Rich Ligands ». *FEBS Letters*: 8.
- Maenz, Barbara et al. 2008. « Characterization of the Human DYRK1A Promoter and Its Regulation by the Transcription Factor E2F1 ». *BMC Molecular Biology* 9(1): 30.
- Maestre, Carolina et al. 2008. « Cdk5 Phosphorylates Cdh1 and Modulates Cyclin B1 Stability in Excitotoxicity ». *The EMBO Journal* 27(20): 2736-45.
- Malcher, Mario, Sarah Schladebeck, et Hans-Ulrich Mösch. 2011. « The Yak1 Protein Kinase Lies at the Center of a Regulatory Cascade Affecting Adhesive Growth and Stress Resistance in *Saccharomyces Cerevisiae* ». *Genetics* 187(3): 717-30.
- Mameesh, Maha M, Adila Al-Kindy, Majda Al-Yahyai, et Anuradha Ganesh. « Microphthalmos-Anophthalmos-Coloboma (MAC) Spectrum in Two Brothers with Renpenning Syndrome Due To ». *OPHTHALMIC GENETICS*: 8.
- Mao, Junhao et al. 2002. « Regulation of Gli1 Transcriptional Activity in the Nucleus by Dyrk1 ». *Journal of Biological Chemistry* 277(38): 35156-61.
- Martin Lorenzo, Sandra et al. 2021. « Targeting the RHOA Pathway Improves Learning and Memory in Adult Kctd13 and 16p11.2 Deletion Mouse Models ». *Molecular Autism* 12(1): 1.
- Martínez-Garay, Isabel et al. 2007. « A Two Base Pair Deletion in the PQBP1 Gene Is Associated with Microphthalmia, Microcephaly, and Mental Retardation ». *European Journal of Human Genetics* 15(1): 29-34.
- Marti et al. 2003 « DYRK1A expression pattern supports specific roles of this kinase in the adult central nervous system ». *Brain research* 964 250–63
- Masih, Suzena, Amita Moirangthem, et Shubha R. Phadke. 2020. « Renpenning Syndrome in an Indian Patient ». *American Journal of Medical Genetics Part A* 182(2): 293-95.
- Matsumoto, Naomichi et al. « 'Department of Human Genetics, Nagasaki University ». : 3.
- Mattioli, Francesca et al. 2020. « De Novo Frameshift Variants in the Neuronal Splicing Factor NOVA2 Result in a Common C-Terminal Extension and Cause a Severe Form of Neurodevelopmental Disorder ». *The American Journal of Human Genetics* 106(4): 438-52.
- Maulik, Pallab K. et al. 2011. « Prevalence of Intellectual Disability: A Meta-Analysis of Population-Based Studies ». *Research in Developmental Disabilities* 32(2): 419-36.
- Mauro, Maurizio et al. 2012. « P21 Promotes Error-Free Replication-Coupled DNA Double-Strand Break Repair ». *Nucleic Acids Research* 40(17): 8348-60.
- Meissner, Laura E. et al. 2020. « *DYRK1A* Pathogenic Variants in Two Patients with Syndromic Intellectual Disability and a Review of the Literature ». *Molecular*

- Genetics & Genomic Medicine* 8(12).  
<https://onlinelibrary.wiley.com/doi/10.1002/mgg3.1544> (28 août 2021).
- Mera-Rodriguez et al. 2021. «Is Senescence-Associated  $\beta$ -Galactosidase a Reliable in vivo Marker of Cellular Senescence During Embryonic Development? ». *Cell developmental Biology* 9: 623175.
- Mi, Huaiyu et al. 2019. « PANTHER Version 14: More Genomes, a New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools ». *Nucleic Acids Research* 47(D1): D419-26.
- Miyata, Yoshihiko et al. 2014. « The Molecular Chaperone TRiC/CCT Binds to the Trp-Asp 40 (WD40) Repeat Protein WDR68 and Promotes Its Folding, Protein Kinase DYRK1A Binding, and Nuclear Accumulation ». *Journal of Biological Chemistry* 289(48): 33320-32.
- Mizuguchi, Mineyuki et al. 2014. « Mutations in the PQBP1 Gene Prevent Its Interaction with the Spliceosomal Protein U5–15kD ». *Nature Communications* 5(1): 3822.
- Mizuguchi et al. 2016. « Allosteric Modulation of the Binding Affinity between PQBP1 and the Spliceosomal Protein U5-15kD ». *FEBS Letters* 590(14): 2221-31.
- Møller, Rikke S. et al. 2008. « Truncation of the Down Syndrome Candidate Gene DYRK1A in Two Unrelated Patients with Microcephaly ». *The American Journal of Human Genetics* 82(5): 1165-70.
- Murakami, Noriko et al. 2006. « Phosphorylation of Amphiphysin I by Minibrain Kinase/Dual-Specificity Tyrosine Phosphorylation-Regulated Kinase, a Kinase Implicated in Down Syndrome ». *Journal of Biological Chemistry* 281(33): 23712-24.
- Na, Sabrina D., et Thomas G. Burns. 2016. « Wechsler Intelligence Scale for Children-V: Test Review ». *Applied Neuropsychology: Child* 5(2): 156-60.
- Najmabadi, Hossein et al. 2011. « Deep Sequencing Reveals 50 Novel Genes for Recessive Cognitive Disorders ». *Nature* 478(7367): 57-63.
- Neumann, Fernanda et al. 2018. « DYRK1A Inhibition and Cognitive Rescue in a Down Syndrome Mouse Model Are Induced by New Fluoro-DANDY Derivatives ». *Scientific Reports* 8(1): 2859.
- Neu-Yilik, Gabriele et al. 2017. « Dual Function of UPF3B in Early and Late Translation Termination ». *The EMBO Journal* 36(20): 2968-86.
- Nguyen et al. 2017. « Dual-Specificity Tyrosine Phosphorylation-Regulated Kinase 1A (DYRK1A) Inhibitors: A Survey of Recent Patent Literature ». *Expert Opinion on Therapeutic Patents* 27(11): 1183-99.
- Nguyen et al. 2018. « Correction of Cognitive Deficits in Mouse Models of Down Syndrome by a Pharmacological Inhibitor of DYRK1A ». *Disease Models & Mechanisms*: dmm.035634.
- Nguyen, Tuan, et Simone Di Giovanni. 2008. « NFAT Signaling in Neural Development and

- Axon Growth ». *International Journal of Developmental Neuroscience* 26(2): 141-45.
- Nolen, Taylor and Gourisankar Ghosh. 2004 « Regulation of Protein Kinases Controlling Activity through Activation Segment Conformation ». *Cell Press* 15(1): 661-75
- Oberlé, Rousseau, Heitz et al. 1991 « Instability of a 550 Base Pair DNA Segment and Abnormal Methylation in Fragile X Syndrome ». *Science* 24;252(5009):1097-102
- Okamoto, N. et al. 2015. « Targeted Next-Generation Sequencing in the Diagnosis of Neurodevelopmental Disorders: Targeted next-Generation Sequencing in the Diagnosis of Neurodevelopmental Disorders ». *Clinical Genetics* 88(3): 288-92.
- Okazawa, Hitoshi et al. 2002. « Interaction between Mutant Ataxin-1 and PQBP-1 Affects Transcription and Cell Death ». *Neuron* 34(5): 701-13.
- Okazawa, Hitoshi, Marius Sudol, et Tina Rich. 2001. « PQBP-1 (Np/PQ): A Polyglutamine Tract-Binding and Nuclear Inclusion-Forming Protein ». *Brain Research Bulletin* 56(3-4): 273-80.
- Oksenberg, Nir, Laurie Stevison, Jeffrey D. Wall, et Nadav Ahituv. 2013. « Function and Regulation of AUTS2, a Gene Implicated in Autism and Human Evolution » éd. James Noonan. *PLoS Genetics* 9(1): e1003221.
- Okuda, T. 2003. « PQBP-1 Transgenic Mice Show a Late-Onset Motor Neuron Disease-like Phenotype ». *Human Molecular Genetics* 12(7): 711-25.
- Okui, Michiyo et al. 1999. « High-Level Expression of the Mnb/Dyrk1A Gene in Brain and Heart during Rat Early Development ». *Genomics* 62(2): 165-71.
- Ori-McKenney, Cassandra M. et al. 2016. « Phosphorylation of  $\beta$ -Tubulin by the Down Syndrome Kinase, Minibrain/DYRK1a, Regulates Microtubule Dynamics and Dendrite Morphogenesis ». *Neuron* 90(3): 551-63.
- O’Roak, B. J. et al. 2012. « Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders ». *Science* 338(6114): 1619-22.
- Ozgyin, Lilla, Attila Horvath, Zsuzsanna Hevessy, et Balint L. Balint. 2019. « Extensive Epigenetic and Transcriptomic Variability between Genetically Identical Human B-Lymphoblastoid Cells with Implications in Pharmacogenomics Research ». *Scientific Reports* 9(1): 4889.
- Pal, Debjani, et Matthew K. Summers. 2018. « The Roles of Cullin RING Ligases and the Anaphase Promoting Complex/Cyclosome in the Regulation of DNA Double Strand Break Repair ». In *Ubiquitination Governing DNA Repair - Implications in Health and Disease*, éd. Effrossyni Boutou et Horst-Werner Stürzbecher. InTech. <http://www.intechopen.com/books/ubiquitination-governing-dna-repair-implications-in-health-and-disease/the-roles-of-cullin-ring-ligases-and-the-anaphase-promoting-complex-cyclosome-in-the-regulation-of-d> (28 août 2021).
- Park, Joongkyu et al. 2010. « Dyrk1A Phosphorylates P53 and Inhibits Proliferation of Embryonic Neuronal Cells ». *Journal of Biological Chemistry* 285(41): 31895-906.

- Pathak, Ankita et al. 2018. « DYRK1A Kinase Inhibition with Emphasis on Neurodegeneration: A Comprehensive Evolution Story-Cum-Perspective ». *European Journal of Medicinal Chemistry* 158: 559-92.
- Perluigi, Marzia et al. 2014. « Neuropathological Role of PI3K/Akt/MTOR Axis in Down Syndrome Brain ». *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842(7): 1144-53.
- Piton, Amélie et al. 2014. « 20 Ans Après: A Second Mutation in MAOA Identified by Targeted High-Throughput Sequencing in a Family with Altered Behavior and Cognition ». *European Journal of Human Genetics* 22(6): 776-83.
- Pober, Barbara R. 2010. « Williams–Beuren Syndrome ». *New England Journal of Medicine* 362(3): 239-52.
- Pozo, Natividad et al. 2013. « Inhibition of DYRK1A Destabilizes EGFR and Reduces EGFR-Dependent Glioblastoma Growth ». *Journal of Clinical Investigation* 123(6): 2475-87.
- Qi, Yunglong et al. 2005. « PQBP-1 Is Expressed Predominantly in the Central Nervous System during Development: Developmental Expression of PQBP-1 ». *European Journal of Neuroscience* 22(6): 1277-86.
- Qian, Wei et al. 2013. « Dual-Specificity Tyrosine Phosphorylation-Regulated Kinase 1A (Dyrk1A) Enhances Tau Expression ». Alejandra Alonso et Chengxin Gong. *Journal of Alzheimer's Disease* 37(3): 529-38.
- Quartier and Courraud et al. 2019 « Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment ». *Human mutation* 40(11):2021-32
- R. Menon, Vijay et al. 2019. « DYRK1A Regulates the Recruitment of 53BP1 to the Sites of DNA Damage in Part through Interaction with RNF169 ». *Cell Cycle* 18(5): 531-51.
- Rachdi, Latif et al. 2014. « Dyrk1a Haploinsufficiency Induces Diabetes in Mice through Decreased Pancreatic Beta Cell Mass ». *Diabetologia* 57(5): 960-69.
- Radhakrishnan, Aneesha et al. 2016. « A Dual Specificity Kinase, DYRK1A, as a Potential Therapeutic Target for Head and Neck Squamous Cell Carcinoma ». *Scientific Reports* 6(1): 36132.
- Rauch, Anita et al. 2012. « Range of Genetic Mutations Associated with Severe Non-Syndromic Sporadic Intellectual Disability: An Exome Sequencing Study ». *The Lancet* 380(9854): 1674-82.
- Raveau, Matthieu et al. 2018. « DYRK1A-Haploinsufficiency in Mice Causes Autistic-like Features and Febrile Seizures ». *Neurobiology of Disease* 110: 180-91.
- Recasens, Ariadna et al. 2021. « Global Phosphoproteomics Reveals DYRK1A Regulates CDK1 Activity in Glioblastoma Cells ». *Cell Death Discovery* 7(1): 81.
- Redin, Claire et al. 2014. « Efficient Strategy for the Molecular Diagnosis of Intellectual Disability Using Targeted High-Throughput Sequencing ». *Journal of Medical*

- Genetics* 51(11): 724-36.
- Renpenning et al. 1963. « Familial Sex-Linked Mental Retardation ». *Canadian Medical Association Journal* 3;87(18):954-6
- Richards, Aziz, Bick, Das et al. 2015. « Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology ». *Genetics in Medicine* 17(5): 405-23.
- Roewenstrunk, Julia et al. 2019. « A Comprehensive Proteomics-Based Interaction Screen That Links DYRK1A to RNF169 and to the DNA Damage Response ». *Scientific Reports* 9(1): 6014.
- Romanov, Vasily S., et K. Lenhard Rudolph. 2016. « P21 Shapes Cancer Evolution ». *Nature Cell Biology* 18(7): 722-24.
- Roskoski, Robert. 2012. « ERK1/2 MAP Kinases: Structure, Function, and Regulation ». *Pharmacological Research* 66(2): 105-43.
- Ruaud, Lyse et al. 2015. « DYRK1A Mutations in Two Unrelated Patients ». *European Journal of Medical Genetics* 58(3): 168-74.
- Ryoo, Soo-Ryoon et al. 2008. « Dual-Specificity Tyrosine(Y)-Phosphorylation Regulated Kinase 1A-Mediated Phosphorylation of Amyloid Precursor Protein: Evidence for a Functional Link between Down Syndrome and Alzheimer's Disease ». *Journal of Neurochemistry* 104(5): 1333-44.
- Sacco, Anna Maria et al. 2019. « Diversity of Dermal Fibroblasts as Major Determinant of Variability in Cell Reprogramming ». *Journal of Cellular and Molecular Medicine* 23(6): 4256-68.
- Scattoni, Maria Luisa, Shruti U. Gandhi, Laura Ricceri, et Jacqueline N. Crawley. 2008. « Unusual Repertoire of Vocalizations in the BTBR T+tf/J Mouse Model of Autism » éd. Wim E. Crusio. *PLoS ONE* 3(8): e3067.
- Shaikh, Mirja N. et al. 2016. « Minibrain Drives the Dacapo Dependent Cell Cycle Exit of Neurons in the *Drosophila* Brain by Promoting *Asense* and *Prospero* Expression ». *Development*: dev.134338.
- Shen, Weijun et al. 2015. « Inhibition of DYRK1A and GSK3B Induces Human  $\beta$ -Cell Proliferation ». *Nature Communications* 6(1): 8372.
- Shen, Yuqian et al. 2021. « PQBP1 Promotes Translational Elongation and Regulates Hippocampal mGluR-LTD by Suppressing EEF2 Phosphorylation ». *Molecular Cell* 81(7): 1425-1438.e10.
- Shi, Jianhua et al. 2008. « Increased Dosage of Dyrk1A Alters Alternative Splicing Factor (ASF)-Regulated Alternative Splicing of Tau in Down Syndrome ». *Journal of Biological Chemistry* 283(42): 28660-69.
- Sitz, Jan Hendrik et al. 2004. « Dyrk1A Potentiates Steroid Hormone-Induced Transcription



- via the Chromatin Remodeling Factor Arip4 ». *Molecular and Cellular Biology* 24(13): 5821-34.
- Sitz, J.H. et al. 2008. « The down Syndrome Candidate Dual-Specificity Tyrosine Phosphorylation-Regulated Kinase 1A Phosphorylates the Neurodegeneration-Related Septin 4 ». *Neuroscience* 157(3): 596-605.
- Sivakumar, Sushama, et Gary J Gorbsky. 2015. « Spatiotemporal Regulation of the Anaphase-Promoting Complex in Mitosis ». : 30.
- Skaar, Jeffrey R, et Michele Pagano. 2009. « Control of Cell Growth by the SCF and APC/C Ubiquitin Ligases ». *Current Opinion in Cell Biology* 21(6): 816-24.
- Smith, Desmond J. et al. 1997. « Functional Screening of 2 Mb of Human Chromosome 21q22.2 in Transgenic Mice Implicates Minibrain in Learning Defects Associated with Down Syndrome ». *Nature Genetics* 16(1): 28-36.
- Snijders Blok, Lot et al. 2015. « Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling ». *The American Journal of Human Genetics* 97(2): 343-52.
- Song, Woo-Joo et al. 1996. « Isolation of Human and Murine Homologues of TheDrosophilaMinibrain Gene: Human Homologue Maps to 21q22.2 in the Down Syndrome “Critical Region” ». *Genomics* 38(3): 331-39.
- Soppa, Ulf et al. 2014. « The Down Syndrome-Related Protein Kinase DYRK1A Phosphorylates P27<sup>Kip1</sup> and Cyclin D1 and Induces Cell Cycle Exit and Neuronal Differentiation ». *Cell Cycle* 13(13): 2084-2100.
- Sorrell, Henderson, Redondo, et al. 2019. Crystal Structure of DYRK1A with small molecule inhibitor. Deposited on PDB: 2019-06-18 Released: 2019-06-26
- Souchet, Benoit et al. 2014. « Excitation/Inhibition Balance and Learning Are Modified by Dyrk1a Gene Dosage ». *Neurobiology of Disease* 69: 65-75.
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. 2005. Vineland adaptive behavior scales: Second edition (Vineland II), survey interview form/caregiver rating form. Livonia, MN: Pearson Assessments.
- Srouf et al. 2016. « Gain-of-Function Mutations in RARB Cause Intellectual Disability with Progressive Motor Impairment ». *Human Mutation* 37(8):786-93.
- Stark, C. 2006. « BioGRID: A General Repository for Interaction Datasets ». *Nucleic Acids Research* 34(90001): D535-39.
- Stefos, Georgios C., Ulf Soppa, Mara Dierssen, et Walter Becker. 2013. « NGF Upregulates the Plasminogen Activation Inhibitor-1 in Neurons via the Calcineurin/NFAT Pathway and the Down Syndrome-Related Proteins DYRK1A and RCAN1 Attenuate This Effect » éd. Cheng-Xin Gong. *PLoS ONE* 8(6): e67470.
- Stevenson, Roger E. et al. 2005. « Renpenning Syndrome Comes into Focus ». *American Journal of Medical Genetics Part A* 134A(4): 415-21.

- Straub, Jonas et al. 2018. « Missense Variants in RHOBTB2 Cause a Developmental and Epileptic Encephalopathy in Humans, and Altered Levels Cause Neurological Defects in *Drosophila* ». *The American Journal of Human Genetics* 102(1): 44-57.
- Sudol Marius 1996 « The WW module competes with the SH3 domain? ». *Trends in Biochemical Sciences*. 21(5):161-3
- Tadmouri, Ghazi O et al. 2009. « Consanguinity and Reproductive Health among Arabs ». *Reproductive Health* 6(1): 17.
- Takahashi, Kazutoshi et al. 2007. « Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors ». *Cell* 131(5): 861-72.
- Takahashi, Keiko et al. 2009. « Nematode Homologue of PQBP1, a Mental Retardation Causative Gene, Is Involved in Lipid Metabolism » éd. Sotirios Koutsopoulos. *PLoS ONE* 4(1): e4104.
- Talukdar, Husain A. et al. 2016. « Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease ». *Cell Systems* 2(3): 196-208.
- Tamura, T. et al. 2010. « *Drosophila* PQBP1 Regulates Learning Acquisition at Projection Neurons in Aversive Olfactory Conditioning ». *Journal of Neuroscience* 30(42): 14091-101.
- Tanaka, Hikari et al. 2018. « The Intellectual Disability Gene PQBP1 Rescues Alzheimer's Disease Pathology ». *Molecular Psychiatry* 23(10): 2090-2110.
- Tarpey, Patrick S et al. 2007. « Mutations in UPF3B, a Member of the Nonsense-Mediated mRNA Decay Complex, Cause Syndromic and Nonsyndromic Mental Retardation ». *Nature Genetics* 39(9): 1127-33.
- Tatton-Brown, Katrina et al. 2017. « Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability ». *The American Journal of Human Genetics* 100(5): 725-36.
- Tejedor, F et al. 1995. « Minibrain: A New Protein Kinase Family Involved in Postembryonic Neurogenesis in *Drosophila* ». *Neuron* 14(2): 287-301.
- Telley, L. et al. 2019. « Temporal Patterning of Apical Progenitors and Their Daughter Neurons in the Developing Neocortex ». *Science* 364(6440): eaav2522.
- The DDD Study et al. 2014. « Synaptic, Transcriptional and Chromatin Genes Disrupted in Autism ». *Nature* 515(7526): 209-15.
- The UniProt Consortium et al. 2021. « UniProt: The Universal Protein Knowledgebase in 2021 ». *Nucleic Acids Research* 49(D1): D480-89.
- Thompson, Benjamin J. et al. 2015. « DYRK1A Controls the Transition from Proliferation to Quiescence during Lymphoid Development by Destabilizing Cyclin D3 ». *Journal of Experimental Medicine* 212(6): 953-70.
- Trujillano, Daniel et al. 2017. « Clinical Exome Sequencing: Results from 2819 Samples

- Reflecting 1000 Families ». *European Journal of Human Genetics* 25(2): 176-82.
- Tschop, K. et al. 2011. « A Kinase ShRNA Screen Links LATS2 and the PRB Tumor Suppressor ». *Genes & Development* 25(8): 814-30.
- Varjosalo, Markku et al. 2013. « The Protein Interaction Landscape of the Human CMGC Kinase Group ». *Cell Reports* 3(4): 1306-20.
- Vissers, Lisenka E L M et al. 2010. « A de Novo Paradigm for Mental Retardation ». *Nature Genetics* 42(12): 1109-12.
- Vissers, Lisenka E. L. M., Christian Gilissen, et Joris A. Veltman. 2016. « Genetic Studies in Intellectual Disability and Related Disorders ». *Nature Reviews Genetics* 17(1): 9-18.
- Vodermaier, Hartmut C. et al. 2003. « TPR Subunits of the Anaphase-Promoting Complex Mediate Binding to the Activator Protein CDH1 ». *Current Biology* 13(17): 1459-68.
- Wallmeier, Julia et al. 2019. « De Novo Mutations in FOXP1 Result in a Motile Ciliopathy with Hydrocephalus and Randomization of Left/Right Body Asymmetry ». *The American Journal of Human Genetics* 105(5): 1030-39.
- Wan, Didi et al. 2015. « X Chromosome-Linked Intellectual Disability Protein PQBP1 Associates with and Regulates the Translation of Specific MRNAs ». *Human Molecular Genetics* 24(16): 4599-4614.
- Wang, Peng et al. 2015. « A High-Throughput Chemical Screen Reveals That Harmine-Mediated Inhibition of DYRK1A Increases Human Pancreatic Beta Cell Replication ». *Nature Medicine* 21(4): 383-88.
- Wang, Pin, Luanluan Wang, Long Chen, et Xiulian Sun. 2017. « Dual-Specificity Tyrosine-Phosphorylation Regulated Kinase 1A Gene Transcription Is Regulated by Myocyte Enhancer Factor 2D ». *Scientific Reports* 7(1): 7240.
- Wang, Q. et al. 2013. « PQBP1, a Factor Linked to Intellectual Disability, Affects Alternative Splicing Associated with Neurite Outgrowth ». *Genes & Development* 27(6): 615-26.
- Waragai, M. 1999. « PQBP-1, a Novel Polyglutamine Tract-Binding Protein, Inhibits Transcription Activation by Brn-2 and Affects Cell Survival ». *Human Molecular Genetics* 8(6): 977-87.
- Waragai, M. et al. 2000. « PQBP-1/Npw38, a Nuclear Protein Binding to the Polyglutamine Tract, Interacts with U5-15kD/Dim1p via the Carboxyl-Terminal Domain ». *Biochemical and Biophysical Research Communications* 273(2): 592-95.
- Wegiel, Jerzy et al. 2004. « Cell Type- and Brain Structure-Specific Patterns of Distribution of Minibrain Kinase in Human Brain ». *Brain Research* 1010(1-2): 69-80.
- Wegiel, Jerzy, Cheng-Xin Gong, et Yu-Wen Hwang. 2011. « The Role of DYRK1A in Neurodegenerative Diseases: Role of DYRK1A in Neurodegenerative Diseases ». *FEBS Journal* 278(2): 236-45.
- Widowati, Esti Wahyu et al. 2018. « Functional Characterization of DYRK1A Missense

- Variants Associated with a Syndromic Form of Intellectual Deficiency and Autism ». *Biology Open* 7(4): bio032862.
- Willsey, Helen Rankin et al. 2020. « Neurodevelopmental Disorder Risk Gene *DYRK1A* Is Required for Ciliogenesis and Brain Size in *Xenopus* Embryos ». *Development*: dev.189290.
- Woo, Savio L. C. et al. 1983. « Cloned Human Phenylalanine Hydroxylase Gene Allows Prenatal Diagnosis and Carrier Detection of Classical Phenylketonuria ». *Nature* 306(5939): 151-55.
- Yabut, O., J. Domogauer, et G. D'Arcangelo. 2010. « Dyrk1A Overexpression Inhibits Proliferation and Induces Premature Neuronal Differentiation of Neural Progenitor Cells ». *Journal of Neuroscience* 30(11): 4004-14.
- Yang, Eun Jin, Young Soo Ahn, et Kwang Chul Chung. 2001. « Protein Kinase Dyrk1 Activates CAMP Response Element-Binding Protein during Neuronal Differentiation in Hippocampal Progenitor Cells ». *Journal of Biological Chemistry* 276(43): 39819-24.
- Yang, Shin-Sheng et al. 2020. « PQBP1, an Intellectual Disability Causative Gene, Affects Bone Development and Growth ». *Biochemical and Biophysical Research Communications* 523(4): 894-99.
- Yang, Y. et al. 2009. « A Cdc20-APC Ubiquitin Signaling Pathway Regulates Presynaptic Differentiation ». *Science* 326(5952): 575-78.
- Yin, Xiaomin et al. 2012. « Dual-Specificity Tyrosine Phosphorylation-Regulated Kinase 1A (Dyrk1A) Modulates Serine/Arginine-Rich Protein 55 (SRp55)-Promoted Tau Exon 10 Inclusion ». *Journal of Biological Chemistry* 287(36): 30497-506.
- Yoh, Sunnie M. et al. 2015. « PQBP1 Is a Proximal Sensor of the CGAS-Dependent Innate Response to HIV-1 ». *Cell* 161(6): 1293-1305.
- Yu, Dan, Claudia Cattoglio, Yuhua Xue, et Qiang Zhou. 2019. « A Complex between DYRK1A and DCAF7 Phosphorylates the C-Terminal Domain of RNA Polymerase II to Promote Myogenesis ». *Nucleic Acids Research* 47(9): 4462-75.
- Zhang, Xiao-Yan et al. 2017. « Mutations of *PQBP1* in Renpenning Syndrome Promote Ubiquitin-Mediated Degradation of FMRP and Cause Synaptic Dysfunction ». *Human Molecular Genetics* ddx010.

**Jérémie COURRAUD**

**Caractérisation fonctionnelle de deux formes de Déficience Intellectuelle associées à des mutations dans *DYRK1A* et *PQBPI***

**Résumé**

La déficience intellectuelle (DI) est un trouble du neurodéveloppement et la première raison de consultation génétique. Cependant, un certain nombre de variants restent classés en tant que Variants de signification inconnue (VSI) et contribuent au phénomène d'errance diagnostique. De plus, la DI est caractérisée par une hétérogénéité génétique extrême et les mécanismes physiopathologiques associés sont généralement mal connus.

Mon travail de doctorat a consisté à mettre au point des approches expérimentales permettant de reclasser l'ensemble des VSI dans les gènes *DYRK1A* et *PQBPI*. J'ai également investigué les mécanismes cellulaires dérégulés lors de l'inactivation de ces gènes dans un modèle de précurseurs neuronaux.

Mes travaux de doctorat permettront donc d'améliorer le diagnostic moléculaire de la déficience intellectuelle associée à des mutations dans les gènes *DYRK1A* et *PQBPI*. Ils permettront également de mieux comprendre les mécanismes physiopathologiques impliqués dans ces troubles neurodéveloppementaux. Ces connaissances aideront à mieux prendre en charge les individus atteints et pourront permettre d'identifier de nouvelles cibles thérapeutiques.

Mot Clés : Déficience Intellectuelle, Variants de Signification Inconnue, *DYRK1A*, *PQBPI*

**Résumé en anglais**

Intellectual disability (ID) is a neurodevelopmental disorder and first reason for genetic counseling. However, a number of variants remain classified as Variants of Unknown Significance (VUS) and contribute to the phenomenon of diagnostic wandering. In addition, ID is characterized by extreme genetic heterogeneity and the associated pathophysiological mechanisms are generally poorly understood.

My PhD consisted in developing experimental approaches to reclassify all VUS in the *DYRK1A* and *PQBPI* genes. I also investigated the deregulated cellular mechanisms following the inactivation of these genes in a model of neuronal precursors.

This work will therefore improve the molecular diagnosis of intellectual disability associated with mutations in *DYRK1A* and *PQBPI*. Furthermore, it will also provide a better understanding of the pathophysiological mechanisms involved in these neurodevelopmental disorders. Finally, these results will help to better manage affected individuals and may help identify new therapeutic targets.

Key words: Intellectual Disability, Variants of Unknown Significance, *DYRK1A*, *PQBPI*