



**HAL**  
open science

# Etudes des mécanismes à l'origine de l'excès de garçons dans la déficience intellectuelle et les troubles du spectre autistique

Angelique Quartier

## ► To cite this version:

Angelique Quartier. Etudes des mécanismes à l'origine de l'excès de garçons dans la déficience intellectuelle et les troubles du spectre autistique. Biologie du développement. Université de Strasbourg, 2017. Français. NNT : 2017STRAJ003 . tel-01558286

**HAL Id: tel-01558286**

**<https://theses.hal.science/tel-01558286>**

Submitted on 7 Jul 2017

**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.

*ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ*  
Institut de Génétique et de Biologie Moléculaire et Cellulaire

**THÈSE** présentée par :  
**Angélique QUARTIER**

Soutenue publiquement le 13 janvier 2017

Pour obtenir le grade de : **Docteur de l'Université de Strasbourg**  
Discipline/S spécialité : Sciences de la vie et de la santé

---

**Etude des mécanismes à l'origine de l'excès de  
garçons dans la déficience intellectuelle et les  
troubles du spectre autistique**

*Mechanisms involved in the male excess observed in  
intellectual disability and autism spectrum disorders*

---

THÈSE dirigée par :

**M. MANDEL Jean-Louis**

Professeur, Collège de France, Paris

co-encadrée par :

**Mme PITON Amélie**

Maitre de conférences, Praticienne Hospitalier, Strasbourg

MEMBRES DU JURY :

**RAPPORTEURS**

**Mme FRANCIS Fiona**

Professeure, Institut du Fer à Moulin (IFM), Paris

**M. BOURGERON Thomas**

Professeur, Institut Pasteur, Paris

**EXAMINATEURS**

**M. CATO Andrew**

Professeur, Karlsruhe Institute of Technology (KIT), Karlsruhe

**M. METZGER Daniel**

Directeur de Recherches, Institut de Génétique et de Biologie Moléculaire  
et Cellulaire (IGBMC), Illkirch



*"La science ?*

*Après tout, qu'est-elle, sinon une longue et systématique curiosité ? "*

André Maurois

*"Grandir n'est pas s'enrichir de quelque chose de nouveau,  
mais découvrir ce que l'on a déjà à l'intérieur"*

Alexandre Jollien





# Remerciements

---

*Je me suis toujours demandé pourquoi les remerciements d'une thèse étaient aussi longs...  
Maintenant, je comprends...*

## La Science...

Je tiens tout d'abord à remercier l'ensemble des membres de mon jury de m'avoir fait l'honneur de juger mes travaux de recherche et tout particulièrement les **Professeurs Fiona Francis** et **Thomas Bourgeron** d'avoir accepté d'être rapporteurs de ceux-ci. Je sais à quel point votre temps est précieux et votre emploi du temps surchargé donc je vous remercie d'accorder un peu de votre attention à ce travail et d'y apporter votre expertise dans les domaines de la neurogénèse, du développement cortical et de la génétique de l'autisme. Je remercie également vivement le **Professeur Andrew Cato** et le **Docteur Daniel Metzger** qui, certes, n'auront pas l'obligation de lire les 300 pages de ce manuscrit mais avec qui je suis certaine d'avoir une discussion emplie de critiques constructives et de conseils bienveillants. Enfin, un merci tout particulier à la **Docteure Alexandra Benchoua**, membre invitée de mon jury, qui nous a transmis tous les secrets de ces cellules si capricieuses et nous a ensuite beaucoup aidé lorsqu'elles se montraient trop mystérieuses ☺

Je me dois bien sûr de remercier chaleureusement mes 2 directeurs de thèse (enfin... mon directeur et ma co-encadrante...) le **Professeur Jean-Louis Mandel** et la **Docteure (+HDR ☺) Amélie Piton**. Jean-Lou, merci de m'avoir accordé ta confiance pour mener à bien ce projet passionnant qui devait traîner dans un coin de ta tête depuis un petit moment ! J'ai beaucoup d'admiration pour toi et ton travail et je suis heureuse d'avoir été ta dernière (ou avant-dernière ?) doctorante. Amélie, que dire... merci pour tout ! Ce projet s'est avéré plus compliqué que prévu mais tu n'as jamais cessé de me soutenir et de trouver des solutions pour contourner ou surmonter les difficultés. Tu as toujours été présente et en même temps tu m'as laissé mon autonomie pour me permettre d'apprendre à réfléchir par moi-même... Je n'aurais pas pu imaginer avoir de meilleurs encadrants que vous et vous remercie du fond du cœur pour tout ce que vous m'avez transmis.

Un grand merci à tous les membres de mon équipe, les formidables "Mandel"... Tout d'abord les anciens désormais partis à l'aventure post-doctorale aux "States", **Claire et Ricardos**... Vous étiez les doctorants de l'équipe quand moi, petite master fraîchement diplômée, je suis arrivée. Merci de m'avoir fait découvrir ce qu'est une thèse (toujours de bons conseils pour les manips') et ses à-côtés (vive les soirées à thème chez Claire !). **Francesca**, merci pour ton soutien aussi bien pendant les moments déprimants que lors de nos réussites (ce satané CRISPR, ce déménagement qui tombe au mauvais moment... mais le chocolat résout tout !). Merci **Christel** de m'avoir embarqué par hasard scientifique dans l'aventure *DCC*. Merci à **Nicolas H.** pour ta fidèle assistance technique, à **Jean** pour notre courte co-habitation et à **Florent et Timothée** pour nos repas et pauses-café souvent peu diététiques. Une pensée pour les stagiaires que j'ai pu encadrer ou co-encadrer d'avoir été si dociles : **Grace, Léna, Inès**... c'était un plaisir de vous initier à la recherche. Bien sûr je n'oublie pas la seconde partie de notre équipe, les "Moine" : **Hervé, Eric, Andréa**, merci pour ces moments partagés. Chers "Mandel", travailler en votre compagnie a été très enrichissant aussi bien scientifiquement que socialement... Merci d'avoir joué le jeu lorsque j'organisais des Secret Santa ou des sorties restau et d'avoir été les cobayes de mes expériences culinaires ou des invendus de Laurent... Je suis certaine que ces 4 années ne se seraient pas passées aussi bien sans vous !

Je tiens ensuite à remercier les autres membres du département de Médecine Translationnelle et Neurogénétique de l'IGBMC, en particulier l'équipe "Puccio" avec laquelle nous avons des liens étroits... **Charline, Frankoïz, Nadège, Aurélie, Odile, Laura, Alain** et j'en oublie ! Merci d'avoir rendu mes pauses-déjeuner si conviviales et divertissantes. Merci à **Jamel, Hélène, Ekat, Vicky, Loïc, Johan** et tous les membres de l'équipe "Chelly" avec qui nous ne formons maintenant plus qu'une... Et parce que j'avais souvent l'impression de passer ma vie en salle de culture, un petit mot spécial pour ceux qui y étaient parfois présents aussi souvent que moi : **Tiphaine, Arielle et Hichem**... C'est toujours plus sympa de travailler en musique et dans la bonne humeur (même les week-ends !). Enfin merci à **Karim** dont les conversations-débats et les blagues qui faisaient parfois rougir m'ont manqué pendant ma dernière année.

Je tiens également à remercier l'ensemble des services techniques de l'IGBMC sans qui ce projet n'aurait pas pu avancer. A la plateforme de séquençage haut-débit : **Bernard, Muriel et Serge** à qui je venais réclamer mes échantillons et surtout les délais de séquençage...

A la plateforme de bio-informatique : **Stéphanie** et **Céline** qui ont été formidables dans l'analyse (et la ré-ré-analyse) de mes données. A la plateforme de criblage cellulaire : **Laurent** et **Anne** qui ont tout essayé afin de trouver un phénotype quantifiable chez mes bien-aimées cellules. Merci également au service de culture cellulaire qui a toujours répondu présent.

Merci à **Mims** et **Myriam** dont l'efficacité administrative n'est plus à prouver. Un grand merci également à **Maïté** qui est capable de trouver l'article introuvable et qui m'a appris comment dompter la Xerox en toutes circonstances...C'est également grâce à elle que l'impression de ce manuscrit a pu être aussi simple ! Merci à **Nicolas B.** pour son aide logistique qui m'a accompagné lorsque j'organisais toute sorte de choses scientifiques ou non...

### Les Amis...

Parce que la thèse, ce n'est pas simplement travailler de 8h à 18h au labo, merci à tous ceux qui m'entourent et me font passer de si bons moments. Et oui **Charline**, tu as du te reconnaître ! Merci de m'avoir suivi dans mes idées farfelues du genre "opération vide ton frigo" ou "viens, on va dans les châteaux gonflables !" et pour toutes les séances de psychothérapie (parfois même pendant tes dissections !)... Parfois, râler c'est indispensable pour se sentir bien dans sa peau et dans sa thèse ! **Léa**, quant à toi, tu as été le fil rouge de ma thèse... D'OpenLab à ma thèse en 180s, nous avons tenté plein d'aventures ensemble ce qui nous a permis de nous évader un peu de notre thèse... et ça fait du bien ! Merci aussi pour toutes ses discussions au détour d'un couloir (et qui sait pourquoi, la plupart du temps devant les toilettes ?!)

Grâce à mes directeurs/co-encadrants de thèse très compréhensifs, j'ai pu participer à de nombreuses activités extra-scientifiques comme l'organisation du "Brocanthon" au profit du Téléthon ou la vulgarisation scientifique dans des écoles ou lors de la fête de la science. J'ai également fait partie d'une association pour doctorants et post-doctorants de l'IGBMC, la SPB, au sein de laquelle j'ai rencontré des gens géniaux, notamment **David** et **Justine** avec qui j'ai partagé le bureau... Merci à vous deux pour cette merveilleuse aventure si épanouissante ! Je vous souhaite sincèrement le meilleur à venir ! Quelques lignes également pour tous les collègues de l'association avec qui j'ai passé de si bon moments : **Michel, Perrine, Karima, Valério, Simone, Alexey, Iskander, Samantha, Lorraine, Claire, Laura, Mélanie, Rachel, Alice, Raphaël, Jérôme, Alexis...**

Merci également à tous mes collègues d'OpenLab avec qui j'ai partagé l'amour d'enseigner la science : **Eléa, Daphné, Gaétan Romumucœur, Romain...** on est toujours prêts à se retrouver autour d'une bière... D'ailleurs, c'est quand la prochaine ? ☺

## **La Famille...**

On ne choisit pas sa famille et c'est là que je me dis que j'ai quand même beaucoup de chance ! **Mes parents, ma sœur Mélanie (et Olivier), mes grands-parents et toute ma famille du Ch'Nord, ma belle-famille...** même si vous êtes loin, vous avez tous été là pour moi et vous êtes indispensables pour me permettre de garder les pieds sur terre et ne jamais oublier d'où je viens... Je vous ai très peu parlé dans les détails de mon travail car, à mes yeux, vous étiez là pour me faire déconnecter et vous avez brillamment rempli cette mission (même si vous étiez toujours curieux de savoir ce que je pouvais bien faire dans mon labo avec mes neurones en boîte pour soigner l'autisme ☺).

A **Laurent**, mon conjoint, confident, ami, amant, bouc émissaire... Je sais que je t'en ai parfois fait baver en ramenant ma mauvaise humeur à la maison et heureusement, tu étais toujours là pour me changer les idées en me parlant de pâtisserie ! C'est souvent compliqué de partager la vie d'une scientifique mais ça fait 9 ans que ça dure et je suis heureuse que tu m'aies fait franchir les caps les plus importants de ma vie (déménagement(s), PACS, bébé...). Faisons que ça dure... Je t'aime.

Pour finir, à ma fille **Chloé** dont les biberons et les couches ont rythmé l'écriture de ce manuscrit... Si tes pleurs m'ont parfois interrompu, ton sourire n'a cessé de me rappeler qu'une très belle aventure m'attend après celle qui est en train de s'achever... Même si je suis fière du travail que j'ai accompli pendant ces quatre années, je suis encore plus fière des progrès que je te vois accomplir chaque jour et tu resteras à ce jour ma plus belle réussite !

*Ma mémoire est bien trop petite pour contenir toutes les personnes qui m'ont si gentiment aidé au cours de ma thèse. Si vous êtes l'une de celles-ci et que vous n'avez pas personnellement été cité ci-dessus, je m'en excuse et vous envoie mes mille mercis !*

# Résumé

---

La déficience intellectuelle (DI) et les troubles du spectre autistique (ASD) sont deux troubles neurodéveloppementaux (NDD) présentant de nombreux chevauchements génétiques et phénotypiques ainsi qu'un biais de sexe important, avec plus de garçons atteints (ratio garçons:filles de 1,4:1 pour la DI et de 4:1 pour les ASD). Au sein de notre laboratoire, le taux de diagnostic des patients souffrant de DI et/ou d'ASD est significativement plus élevé chez les filles que chez les garçons. J'ai plus particulièrement analysé les conséquences fonctionnelles de mutations pathogènes (dans le gène *FMRI* impliqué dans le syndrome de l'X Fragile) ainsi que des variations faux-sens potentiellement causales dans les gènes *PQBPI* (syndrome de Renpenning) et *NLGN3*. Bien que les trois exemples étudiés soient des variations présentes dans des gènes du chromosome X chez des patients mâles, nous n'avons pas observé de différence significative entre filles et garçons au niveau de la proportion de mutations pathogènes sur le chromosome X (5,3% *versus* 7,6%) sur l'ensemble de notre cohorte, confirmant ainsi que les mutations causales totalement pénétrantes sur ce chromosome ne peuvent pas expliquer la totalité de l'excès de garçons atteints de DI ou d'ASD. Nous avons donc choisi d'étudier une autre des hypothèses, plus environnementale, qui pourrait rendre le cerveau masculin plus susceptible au développement de NDD : le rôle des androgènes au cours du développement du cerveau. J'ai étudié l'effet de ces hormones masculines dans des précurseurs neuronaux humains (hNSCs) et observé que les androgènes augmentent la prolifération des hNSCs et les protègent contre la mort cellulaire en conditions stressantes. J'ai également mis en évidence que les androgènes, *via* leur récepteur (le récepteur aux androgènes), régulent une centaine de gènes dans les hNSCs avec, parmi eux, un enrichissement en gènes connus pour être différentiellement exprimés chez les individus souffrant d'ASD (dont *NRCAM* et *FAM107A*). La régulation de ces gènes par les androgènes pendant le développement du cerveau pourrait ainsi participer à la sensibilité accrue du cerveau masculin, exposé à d'autres facteurs génétiques et environnementaux, à développer une NDD.

# Summary

---

Intellectual disability (ID) and autism spectrum disorders (ASD) are two common neurodevelopmental disorders (NDD) with many genetic and phenotypic overlaps. Another common feature is the existence of a gender bias, very strong for ASD (4 males for 1 female) and notable for ID (1.4:1). In our team, the diagnostic yield of patients affected by ID with or without ASD is significantly higher in girls than in boys. I studied more specifically the functional consequences of pathogenic mutations (in the *FMR1* gene involved in the Fragile X syndrome) as well as missense variations that were predicted as potentially pathogenic in *PQBPI* (Renpenning syndrome) and *NLGN3* genes. Although these three examples are variations in genes on the X chromosome identified in male patients, we did not observe a significant difference between girls and boys in the proportion of pathogenic mutations on the X chromosome (5.3% versus 7.6%) in our cohort. This confirms that rare and fully penetrant mutations on this chromosome cannot explain the totality of the male bias observed in ID and ASD. We therefore studied another hypothesis, more environmental, which could make the male brain more susceptible to develop NDD: the role of androgens during brain development. I studied the effect of these male hormones in human neuronal precursors (hNSCs) and observed that androgens increase the proliferation of hNSCs and protect them against cell death under stressful conditions. I also showed that androgens, *via* their receptor (the androgen receptor), regulate a hundred genes in hNSCs with, among them, an enrichment of genes known to be differentially expressed in individuals with ASD (*e.g NRCAM* and *FAM107A*). The regulation of these genes by androgens during brain development could thus contribute to the increased sensitivity of the male brain, exposed to other genetic and environmental factors, to develop an NDD.

# Sommaire

---

<i>Remerciements</i> .....	<i>i</i>
<i>Résumé</i> .....	<i>v</i>
<i>Summary</i> .....	<i>vi</i>
<i>Sommaire</i> .....	<i>vii (here !)</i>
<i>Index des figures</i> .....	<i>xii</i>
<i>Index des tables</i> .....	<i>xiii</i>
<i>Index des annexes</i> .....	<i>xiv</i>
<i>Liste des abréviations</i> .....	<i>xv</i>

<b><u>Introduction</u></b> .....	<b>1</b>
<b>Chapitre 1. Les maladies neurodéveloppementales (NDD)</b> .....	<b>3</b>
<b>Partie 1. Définition et étiologie</b> .....	<b>3</b>
<b>I. Définitions</b> .....	<b>3</b>
1. La déficience intellectuelle (DI).....	3
2. Les troubles du spectre autistique (ASD).....	4
3. Chevauchement clinique.....	6
<b>II. Les causes génétiques</b> .....	<b>7</b>
1. La génétique de la DI.....	7
i. Anomalies chromosomiques et CNV.....	8
ii. Les formes monogéniques de DI.....	9
a. Les DI autosomiques dominantes.....	10
b. Les DI autosomiques récessives.....	11
c. Les DI liées au chromosome X (XLID).....	12
iii. Le modèle de l'oligogénisme.....	12
2. La génétique de l'autisme.....	13
i. Mise en évidence de la part génétique.....	13
ii. Les causes multifactorielles.....	14
a. Etude d'association gène candidat.....	15
b. Etude d'association pangénomique (GWAS).....	15
c. Etude de CNV.....	16
iii. Les formes monogéniques d'autisme.....	18
a. Cas des ASD syndromiques.....	18
b. Les ASD liés au chromosome X.....	18
c. Les ASD autosomiques dominants.....	19
d. Les ASD autosomiques récessifs.....	20
iv. Le modèle de l'oligogénisme.....	21
v. Bases de données pour la recherche sur l'autisme.....	21
3. Chevauchement génétique.....	22
i. Des centaines de gènes identifiés.....	22
ii. ... convergeant vers des voies biologiques communes.....	24



III. Les causes environnementales.....	26
1. La DI "acquise".....	26
i. Les expositions toxiques.....	26
ii. Les infections.....	27
iii. Les traumatismes.....	27
2. Cas des ASD.....	28
i. Les facteurs obstétriques.....	28
ii. Les infections et les dysfonctionnements immunitaires.....	29
iii. L'exposition à des xénobiotiques.....	29
a. Les médicaments.....	30
b. Les métaux lourds.....	30
c. Les polluants atmosphériques.....	30
d. La nutrition et le mode de vie maternel.....	31
iv. Les hormones.....	31
<b>Partie 2. Hypothèses pouvant expliquer l'excès de garçons observé dans les NDD...</b>	<b>33</b>
I. La sous-estimation des filles affectées.....	33
II. Effets protecteurs chez les femmes.....	34
1. Seuil de tolérance aux événements génétiques et pénétrance incomplète...35	
2. Avoir deux chromosomes X.....	36
i. Effet de compensation par le second chromosome X.....	36
a. Cas de la DI.....	37
b. Cas des ASD.....	37
ii. Biais d'inactivation du chromosome X.....	38
iii. Dosage des gènes du chromosome X.....	39
iv. L'empreinte du chromosome X.....	40
III. Facteurs de susceptibilité chez l'homme.....	42
1. L'implication du chromosome Y.....	43
i. Rôle de SRY dans le cerveau et liens avec la DI et les ASD.....	43
ii. Autres gènes candidats.....	44
2. L'implication des hormones masculines.....	45
i. La détermination sexuelle.....	46
ii. Production des hormones testiculaires et rétrocontrôles.....	48
iii. Liens entre l'AMH, l'inhibine B et les ASD.....	50
iv. L'hormone testiculaire majeure : la testostérone.....	51
<b>Chapitre 2. Les androgènes.....</b>	<b>53</b>
I. Biosynthèse, production, transport et métabolisme.....	53
1. Biosynthèse de la testostérone.....	53
2. Transport et métabolisme.....	54
3. Production et taux de testostérone au cours de la vie.....	56
4. Cas des neurostéroïdes.....	59
II. Le récepteur aux androgènes (AR).....	60
1. Généralités.....	60
i. Classification.....	60

ii. Organisation de l'AR.....	61
a. Le domaine de transactivation NH2 terminal (NTD).....	62
b. Le domaine de liaison à l'ADN (DBD).....	62
c. Le domaine de liaison du ligand COOH terminal (LBD).....	63
d. La région charnière.....	64
iii. Les isoformes de l'AR.....	64
iv. Modifications post-traductionnelles.....	66
2. Pathologies associées à l'AR.....	67
i. L'insensibilité aux androgènes (AIS).....	68
ii. Le syndrome de Kennedy.....	68
iii. Le cancer de la prostate.....	69
3. Mécanisme d'action de la testostérone <i>via</i> l'AR.....	70
i. Activation et translocation de l'AR.....	71
ii. Reconnaissance des éléments de réponses aux androgènes (ARE) par l'AR.....	72
iii. Régulation de la transcription (action génomique).....	73
iv. Actions non-génomiques.....	74
v. Rôles des co-régulateurs.....	77
4. Effets pléiotropiques des androgènes.....	77
i. Caractères sexuels primaires et secondaires.....	78
ii. Rôles dans le tissu musculaire.....	79
iii. Rôles dans le tissu osseux.....	79
iv. Rôles dans les systèmes nerveux central et périphérique.....	79
III. Rôles des androgènes dans le cerveau et liens avec la DI et les ASD.....	80
1. Expression de l'AR et rôles des androgènes dans le cerveau.....	80
i. Les comportements reproducteurs et non-reproducteurs.....	80
ii. Implication dans le dimorphisme sexuel.....	81
iii. Localisation cellulaire/subcellulaire de l'AR et conséquences fonctionnelles.....	82
iv. Interaction avec la voie des œstrogènes.....	83
2. Implication des androgènes prénataux dans la DI et les ASD.....	84
i. Indices venant des individus souffrant d'anomalies de production des androgènes.....	84
ii. Le projet "Testostérone Fœtale" de Cambridge.....	85
iii. La théorie du "cerveau masculin extrême".....	86
iv. Implication des stéroïdes, du gène de l'AR et de ses co-régulateurs dans la DI et les ASD.....	87

**Objectifs de la thèse.....91**

<b><u>Résultats</u></b> .....	<b>95</b>
<b>Chapitre 1. Mutations causales dans des gènes du chromosome X</b> .....	<b>97</b>
<u>Matériels et méthodes</u> .....	98
<u>Partie 1. Identification de mutations causales dans <i>FMR1</i></u> .....	100
I. Le syndrome de l'X-Fragile.....	100
II. Résultats, discussion et perspectives.....	101
<b><u>Publication : Intragenic <i>FMR1</i> disease-causing variants: a significant mutational mechanism leading to Fragile-X Syndrome</u></b> .....	<b>103</b>
<u>Partie 2. Identification d'une mutation potentiellement causale dans <i>PQBPI</i></u> .....	151
I. Le syndrome de Renpenning.....	151
II. Résultats.....	152
1. Identification d'une mutation faux-sens potentiellement causale dans <i>PQBPI</i> chez deux frères souffrant de DI et d'ASD.....	152
2. Etudes fonctionnelles.....	154
III. Discussion et perspectives.....	158
<u>Partie 3. Identification d'une mutation potentiellement causale dans <i>NLGN3</i></u> .....	161
I. <i>NLGN3</i> , un gène d'autisme ?.....	161
II. Résultats.....	164
1. Identification d'une mutation faux-sens potentiellement causale dans <i>NLGN3</i> chez trois jeunes garçons souffrant de DI et d'ASD.....	164
2. Etudes fonctionnelles.....	166
III. Discussion et perspectives.....	168
<u>Partie 4. Discussion du chapitre 1</u> .....	169
<b>Chapter 2. Effects of the androgens on human neural precursors</b> .....	<b>177</b>
I. Preliminary results.....	177
1. Choice of the cellular model.....	177
2. Choice of androgens doses and cytotoxicity.....	178
3. Microarray on neurons differentiated from hNSCs.....	179
II. Results.....	181
<b><u>Paper: Genes and mechanisms regulated by androgens as possible contributors to the male excess observed in autism</u></b> .....	<b>183</b>
III. Discussion and perspectives.....	261
1. The case of testosterone-only regulated genes.....	261
2. Cellular model variability.....	263
3. DHT/androgens-regulated genes and their expression within the brain.....	264
4. Pathway analysis.....	266
i. Cell proliferation and survival.....	266
ii. Other pathways related to brain development and ID/ASD.....	267
5. Link between genes regulated by androgens and genes involved in ID and/or ASD.....	268
6. Non-genomic actions of androgens in hNSCs.....	269

<b><u>General discussion</u></b> .....	<b>271</b>
I. Monogenic <i>versus</i> multifactorial.....	272
II. Link between genes and environment.....	274
III. Clarifying recent findings and guiding future research directions.....	275
1. ID and ASD definition and diagnosis.....	276
2. "How is the risk to develop ID and ASD linked to sex".....	276
3. A sex-specific developmental trajectory.....	277
IV. Potential implications in therapeutics and opening to other diseases.....	277
<b><u>Annexes</u></b> .....	<b>281</b>
<b><u>Annexe 1. Paper: Efficient strategy for the molecular diagnosis of intellectual disability using targeted high throughput sequencing</u></b> .....	<b>283</b>
<b><u>Annexe 2. Paper: Mutations in <i>DCC</i> Cause Agenesis of the Corpus Callosum</u></b> .....	<b>299</b>
<b><u>References</u></b> .....	<b>319</b>

# Index des figures

---

<b>Figure 1.</b> Augmentation des gènes impliqués dans la DI isolée ou syndromique.....	9
<b>Figure 2.</b> Localisation des CNV identifiés dans les analyses pangénomiques de 9 études concernant plus de 5000 patients autistes.....	17
<b>Figure 3.</b> Fréquence de la DI, des ASD et de l'épilepsie chez les individus porteurs de variants perte de fonction dans des gènes candidats pour les NDD.....	23
<b>Figure 4.</b> Schéma récapitulatif de la différenciation sexuelle chez l'Homme.....	47
<b>Figure 5.</b> Contrôle de la production d'hormones par les testicules.....	49
<b>Figure 6.</b> Evolution du taux de production des différentes hormones masculines depuis le développement <i>in utero</i> jusqu'à l'âge adulte.....	50
<b>Figure 7.</b> Biosynthèse des stéroïdes.....	54
<b>Figure 8.</b> Répartition de la testostérone plasmatique.....	55
<b>Figure 9.</b> Niveau de testostérone plasmatique à différentes périodes de la vie chez les sujets mâles.....	57
<b>Figure 10.</b> Structure du récepteur aux androgènes humain.....	61
<b>Figure 11.</b> Structure du domaine de liaison à l'ADN de l'AR.....	63
<b>Figure 12.</b> Structure de l'AR et de ses principales isoformes.....	66
<b>Figure 13.</b> Modifications post-traductionnelles de l'AR.....	67
<b>Figure 14.</b> Structures moléculaires de la testostérone et de la dihydrotestostérone.....	71
<b>Figure 15.</b> Les différents modes d'action du récepteur aux androgènes.....	76
<b>Figure 16.</b> Principaux effets physiologiques des androgènes sur le développement et le fonctionnement de l'organisme humain mâle.....	78
<b>Figure 17.</b> Identification d'une mutation faux-sens potentiellement causale dans <i>PQBP1</i> .....	154
<b>Figure 18.</b> Etude de l'effet de variations dans <i>PQBP1</i> sur l'interaction avec U5-15kD.....	156
<b>Figure 19.</b> Etude de l'effet de variations dans <i>PQBP1</i> sur l'épissage de deux transcrits cibles.....	157
<b>Figure 20.</b> Structure du complexe trans-synaptique neurexine-neuroligine.....	162
<b>Figure 21.</b> Identification d'une mutation faux-sens potentiellement causale dans <i>NLGN3</i> .....	165
<b>Figure 22.</b> Effets de variations dans <i>NLGN3</i> sur la stabilité et l'expression de la protéine.....	167
<b>Figure 23.</b> Androgens cytotoxicity.....	179
<b>Figure 24.</b> Factorial correspondence analysis.....	264
<b>Figure 25.</b> Multifactorial sex-differential liability model.....	274

# Index des tables

---

<b><u>Table 1.</u></b> Prévalence des syndromes génétiques connus les plus communément associés aux ASD.....	5
<b><u>Table 2.</u></b> Dosages de la testostérone plasmatique et cérébrale à différentes périodes de la vie chez les sujets mâles et femelles.....	58
<b><u>Table 3.</u></b> Classification des ARE.....	73
<b><u>Table 4.</u></b> Co-régulateurs de l'AR impliqués dans des cas de DI et/ou ASD.....	89
<b><u>Table 5.</u></b> Mutations causales et potentiellement causales identifiées au sein de notre cohorte de patients souffrant de DI isolée ou syndromique.....	172
<b><u>Table 6.</u></b> Microarrays results summary performed on neurons differentiated 17 days from hNSCs.....	180
<b><u>Table 7.</u></b> Evidence for the effect of sex steroids in male biased neuropsychiatric conditions...	279

# Index des annexes

---

**Annexe 1. Paper: Efficient strategy for the molecular diagnosis of intellectual disability using targeted high throughput sequencing.....283**

**Annexe 2. Paper: Mutations in *DCC* Cause Agenesis of the Corpus Callosum.....299**

# Liste des abréviations

---

*Ce manuscrit étant rédigé en français et en anglais, la liste des abréviations comprend également une traduction anglais/français.*

- ACC** Agenesis of the corpus callosum *ou* Agénésie du corps calleux
- ADHD** Attention deficit hyperactivity disorders  
*ou* Troubles de déficit de l'attention avec ou sans l'hyperactivité
- ADI-R** Autism diagnostic interview revised  
*ou* Examen revisité pour le diagnostic de l'autisme
- ADOS** Autism diagnostic observation schedule  
*ou* Echelle d'observation pour le diagnostic de l'autisme
- AIS** Androgen insensibility syndrome  
*ou* Syndrome d'insensibilité aux androgènes
- AMH** Anti-müllerian hormone *ou* Hormone anti-müllérienne
- AQ** Autism spectrum quotient *ou* Quotient autistique
- AR** Androgen receptor *ou* Récepteur aux androgènes
- ARE** Androgen response element *ou* Élément de réponse aux androgènes
- ASD** Autism spectrum disorders *ou* Troubles du spectre autistique
- Boîte D** Boîte distale
- Boîte P** Boîte proximale
- CAH** Congenital adrenal hyperplasia  
*ou* Hyperplasie congénitale des glandes surrénales
- CE** Cryptic exon *ou* Exon cryptique
- array-CGH** Array comparative genomic hybridization  
*ou* Puce d'hybridation génomique comparative
- ChIP-chip** Chromatin immunoprecipitation coupled with DNA chip  
*ou* Immunoprécipitation de la chromatine couplée à puce d'expression
- ChIP-seq** Chromatin immunoprecipitation coupled with sequencing  
*ou* Immunoprécipitation de la chromatine couplée à du séquençage
- CNV** Copy number variation *ou* Variation du nombre de copie
- DAG** Diacylglycérol
- DBD** DNA binding domain *ou* Domaine de liaison à l'ADN
- DECIPHER** Database of genomic variation and phenotype in humans using ensembl resources *ou* Base de données des variations génomiques et des phénotypes chez l'Homme à l'aide des ressources ensembl
- DHEA** Déhydroépiandrostérone
- DHT** Dihydrotestostérone
- DI** Déficience intellectuelle
- DR3** Direct repeat separated by 3pb  
*ou* Répétition directe de demi-sites séparés de 3pb
- DSM-V** The diagnostic and statistical manual of mental disorders, V<sup>th</sup> edition  
*ou* Manuel diagnostic et statistique des troubles mentaux, 5<sup>ème</sup> édition
- DZ** Dizygotic *ou* Dizygote



**EMB** Extreme male brain *ou* Cerveau masculin extrême  
**EQ** Empathy quotient *ou* Quotient d'empathie  
**ERR** ER linked receptor *ou* Récepteur lié aux ER  
**ESR/ER** Estrogen receptor *ou* Récepteur aux œstrogènes  
**ExAC** Exome aggregation consortium  
**FSH** Follicle stimulating hormone *ou* Hormone folliculo-stimulante  
**FXPOI** Fragile X primary ovarian insufficiency  
*ou* Insuffisance ovarienne associée à l'X Fragile  
**FXS** Fragile X syndrome *ou* Syndrome de l'X Fragile  
**FXTAS** Fragile X tremor and ataxia syndrome  
*ou* Syndrome de tremblements et d'ataxie associé à l'X Fragile  
**GABA** Gamma-aminobutyric acid *ou* Acide  $\gamma$ -aminobutyrique  
**GABA<sub>A</sub>** Récepteur GABA de type A  
**GnRH** Gonadotropin releasing hormone  
*ou* Hormone de libération des gonadotrophines hypophysaires  
**GR** Glucocorticoid receptor *ou* Récepteur aux glucocorticoïdes  
**GRE** Glucocorticoid response element *ou* Élément de réponse aux glucocorticoïdes  
**GWAS** Genome-wide association study *ou* Etude d'association pangénomique  
**HA** Hemagglutinine humaine  
**hCG** Human chorionic gonadotropin *ou* Hormone chorionique gonadotrope humaine  
**hESCs** Human embryonic stem cells *ou* Cellules souches embryonnaires humaines  
**hNSCs** Human neural stem cells *ou* Cellules souches neurales humaines  
**ID** Intellectual disability  
**IP3** Inositol triphosphate  
**IR3** Inverted repeat separated by 3pb  
*ou* Répétition inversée de demi-sites séparés de 3pb  
**ISH** *In situ* hybridization *ou* Hybridation *in situ*  
**LBD** Ligand binding domain *ou* Domaine de liaison du ligand COOH terminal  
**LH** Luteinizing hormone *ou* Hormone lutéinisante  
**MAPK** Mitogen-activated protein kinase *ou* Protéine kinase activée par un mitogène  
**MR** Mineralocorticoid receptor *ou* Récepteur aux minéralocorticoïdes  
**MTT** 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide  
**MZ** Monozygotic *ou* Monozygote  
**NDD** Neurodevelopmental disorders *ou* Maladies neurodéveloppementales  
**NLGN** Neuroligine  
**NLS** Nuclear localization signal *ou* Signal de localisation nucléaire  
**NMDA** N-méthyl-D-aspartate  
**NR3** Nuclear receptors subfamily 3 *ou* Sous-famille des récepteurs nucléaires 3  
**NRXN** Neurexine  
**NTD** NH2 terminal domain *ou* Domaine de transactivation NH2 terminal  
**PGR** Progesterone receptor *ou* Récepteur à la progestérone  
**PI3K** Phosphoinositide 3-kinase  
**PKA** Protéine kinase A  
**PKC** Protéine kinase C

<b>polyA</b>	Répétition d'alanines (A)
<b>polyG</b>	Répétition de glycines (G)
<b>polyP</b>	Répétition de prolines (P)
<b>polyQ</b>	Répétition de glutamines (Q)
<b>PRE</b>	Progesterone response element <i>ou</i> Élément de réponse à la progestérone
<b>QI</b>	Quotient intellectuel
<b>R1881</b>	Métribolone
<b>RENS</b>	Renpenning syndrome <i>ou</i> Syndrome de Renpenning
<b>RNA-seq</b>	RNA sequencing <i>ou</i> Séquençage de l'ARN
<b>SAGE</b>	Serial analyses of gene expression <i>ou</i> Analyse en série de l'expression génique
<b>SHBG</b>	Sex hormone-binding globulin <i>ou</i> $\beta$ -globuline liant les hormones stéroïdiennes
<b>siRNA</b>	small interfering RNA <i>ou</i> ARN interférent
<b>SLI</b>	Specific language impariment <i>ou</i> Troubles spécifiques du langage
<b>SNP</b>	Single nucleotide polymorphism <i>ou</i> Polymorphisme d'un seul nucléotide
<b>SQ</b>	Systematizing quotient <i>ou</i> Quotient de systématiation
<b>Tau</b>	Transcription activation unit <i>ou</i> Unités activatrices de la transcription
<b>TH</b>	Tyrosine hydroxylase
<b>VABS</b>	Vineland adaptative behavioral scale <i>ou</i> Echelle de maturité sociale de Vineland
<b>WISC</b>	Wechsler intelligence scale for children <i>ou</i> Echelle d'intelligence de Wechsler pour enfants et adolescents
<b>WPPSI</b>	Wechsler preschool and primary scale of intelligence <i>ou</i> Echelle d'intelligence de Wechsler pour la période pre-scolaire et primaire
<b>XCI</b>	X-chromosome inactivation <i>ou</i> Inactivation du chromosome X
<b>XLID</b>	X-linked intellectual disability <i>ou</i> Déficience intellectuelle liée au chromosome X
<b>Xm</b>	Maternal X chromosome <i>ou</i> Chromosome X maternel
<b>Xp</b>	Paternal X chromosome <i>ou</i> Chromosome X paternel

### **Nomenclatures :**

- Tout au long de ce manuscrit, les quatre principaux nucléotides sont appelées A, T, C et G.
- Les acides aminés sont décrits selon la nomenclature standard utilisant le code une ou trois lettre.
- La nomenclature de l'HGNC (HUGO Gene Nomenclature Committee) est utilisée pour décrire les gènes, dont le symbole est écrit en italique. Le symbole de la protéine résultante est écrit en police normale.
- Les variations sont nommées selon les recommandations de l'HGVS (Human Genome Variation Society). De plus, elles sont classées selon les cinq catégories de pathogénicité recommandées par l'ACMG (American College of Medical Genetics and Genomics) (Richards *et al.* 2015) : pathogène, potentiellement pathogène, signification inconnue, potentiellement bénigne et bénigne. Le terme mutation est utilisé pour les variations pathogènes ou potentiellement pathogènes.



# Introduction

---



# **Chapitre 1. Les maladies neurodéveloppementales**

## **Partie 1. Définition et étiologie**

Les maladies neurodéveloppementales (NDD) sont des troubles impliquant une anomalie dans le développement du système nerveux. Cette anomalie peut intervenir à toute étape du développement du cerveau, que ce soit la prolifération, la différenciation, la migration des neurones ou bien l'établissement des connexions synaptiques. On regroupe au sein des NDD un large spectre de désordres cognitifs, neurologiques ou psychiatriques dont les principaux sont la déficience intellectuelle (DI ou ID), les troubles du spectre autistique (ASD), l'épilepsie, les troubles du déficit de l'attention avec ou sans hyperactivité (ADHD), les troubles spécifiques du langage (SLI), le syndrome de Gilles de la Tourette et la schizophrénie. Ensemble, elles touchent de 5 à 10% de la population. Les NDD peuvent être causées par des facteurs génétiques et environnementaux avec vraisemblablement de complexes interactions entre ces facteurs incluant l'épigénétique. Cette partie se focalisera principalement sur la DI et les ASD en donnant tout d'abord leurs définitions cliniques puis en abordant leurs causes génétiques et environnementales.

### **I. Définitions**

#### **1. La déficience intellectuelle (DI)**

La déficience intellectuelle (anciennement nommée retard mental) est un ensemble de NDD caractérisées par un fonctionnement intellectuel inférieur à la moyenne (quotient intellectuel (QI) < 70) et par un déficit au niveau de l'apprentissage et des capacités d'adaptation aux exigences de la vie quotidienne (autonomie personnelle, responsabilité sociale) (définition de "The diagnostic and statistical manual of mental disorders, V<sup>th</sup> edition" ou DSM-V) (Black & Grant 2014). Le fonctionnement intellectuel peut être mesuré par de nombreux outils. Par exemple, il existe des échelles standardisées mesurant le QI de façon dépendante (les échelles de Wechsler WISC ou WPPSI) ou indépendante

(les matrices progressives de Raven) des capacités verbales. Les comportements adaptatifs sont quant à eux mesurés par l'analyse des capacités de communication, d'autonomie, de socialisation et de motricité (notamment *via* l'échelle de maturité sociale de Vineland ou VABS).

La DI se manifeste pendant l'enfance (avant 18 ans) et affecte près de 2% de la population, dont seulement 15% présentent des formes modérées à profondes ( $QI < 50$ ), les 85% restant étant des personnes présentant des DI plus légères ( $70 < QI < 50$ ). Cette pathologie peut être isolée ou syndromique, c'est-à-dire associée à d'autres symptômes comme c'est le cas pour le syndrome de Down (ou trisomie 21 causée par la présence d'un chromosome 21 surnuméraire et caractérisée par une déficience intellectuelle variable, souvent légère, une hypotonie musculaire, une laxité articulaire et une dysmorphie faciale), le syndrome de l'X Fragile (causé par une expansion du triplet CGG dans la région 5' non traduite du gène *FMRI* et caractérisé par une DI modérée à profonde accompagnée d'une dysmorphie faciale et d'une macro-orchidie), le syndrome de Renpenning (causé par des mutations dans le gène *PQBPI* et caractérisé par un déficit intellectuel moyen, une maigreur, une microcéphalie et une petite taille) et bien d'autres.

## **2. Les troubles du spectre autistique (ASD)**

Les ASD (qui seront fréquemment abrégés par "autisme" dans la suite de ce manuscrit) touchent les jeunes enfants, généralement avant l'âge de trois ans. L'autisme affecte environ 1% de la population (Fombonne *et al.* 2012) et sa prévalence semble augmenter ces dernières décennies (un enfant sur 68 d'après les estimations de 2014 du centre pour le contrôle et la prévention des maladies des Etats-Unis contre un enfant sur 150 en 2002 et même un sur 5000 en 1975). Les causes de cette augmentation ne sont pas très claires et sont notamment attribuées à une meilleure reconnaissance de la pathologie, un élargissement des critères diagnostic, au développement des services proposés et à une augmentation de l'incidence (voire vraisemblablement une combinaison de tous ces facteurs).

L'autisme est une condition neurodéveloppementale hautement hétérogène d'un point de vue clinique (d'où l'utilisation du terme "spectre"). C'est en fait un ensemble de pathologies caractérisées par des altérations des interactions et de la communication sociale (communication verbale et non verbale), des intérêts et activités restreints accompagnés ou non de comportements stéréotypés répétés (définition du DSM-V) (Black & Grant 2014).

Le diagnostic des ASD est clinique, basé sur un entretien avec les parents (l'examen revisité pour le diagnostic de l'autisme ou ADI-R) (Lord *et al.* 1994) et/ou sur une observation clinique de l'enfant (l'échelle d'observation pour le diagnostic de l'autisme ou ADOS) (Lord *et al.* 2000). Il existe ensuite d'autres instruments développés afin d'évaluer de façon quantitative les symptômes autistiques dans chaque domaine (communication, comportements répétitifs, etc.).

En plus des symptômes principaux qui permettent le diagnostic des ASD, il existe de nombreuses comorbidités associées aussi bien cognitives (DI), neurologiques (épilepsie, ataxie, etc.), psychiatriques (schizophrénie, ADHD, etc.) que somatiques (dysmorphie, macrocéphalie, obésité, problèmes musculaires, etc.). Tout comme la DI, les ASD peuvent être isolés (présence des symptômes principaux seuls) ou syndromiques (c'est-à-dire associés à des comorbidités). Ces ASD syndromiques représentent 10 à 15% de l'ensemble des ASD et sont pour la plupart bien caractérisés. C'est notamment le cas des ASD associés à des syndromes génétiques connus. Les plus fréquents sont le syndrome de l'X Fragile, le syndrome de Rett, la sclérose tubéreuse de Bourneville, les syndromes de Prader-Willi/Angelman et le syndrome de Cowden (résumé des prévalences dans la **Table 1** ci-dessous). Il est à noter que l'ensemble de ces syndromes sont associés à de la DI (plus ou moins sévère).

Syndrome	Gène(s) impliqué(s)	Prévalence des ASD au sein du syndrome	Prévalence du syndrome au sein des ASD
Syndrome de l'X Fragile	<i>FMR1</i>	21 à 50% des garçons 1 à 3% des filles	2 à 5%
Syndrome de Rett	<i>MECP2</i> <i>CDKL5</i>	25 à 40%	2%
Sclérose tubéreuse de Bourneville	<i>TSC1</i> <i>TSC2</i>	24 à 60%	1 à 4%
Syndromes Prader-Willi/Angelman	Délétion 15q11.2-q13, <i>UBE3A</i>	50 à 81%	1%
Syndrome de Cowden	<i>PTEN</i>	Non évaluée	1%

**Table 1. Prévalence des syndromes génétiques connus les plus communément associés aux ASD.**

Adapté de (Moss & Howlin 2009; Kelleher & Bear 2008)



### 3. Chevauchement clinique

Chez les patients souffrant d'ASD, le nombre et le type de symptômes, le degré de gravité de ces derniers, l'âge de leur apparition et le niveau de fonctionnement varient d'une personne à l'autre. Il en est de même pour les pathologies coexistantes. La DI en fait notamment partie et est présente chez 26 à 70% des enfants avec autisme ce qui en fait la comorbidité la plus fréquente dans les ASD. La grande variabilité observée entre les différentes études est due aux échelles utilisées pour mesurer le QI, celles basées sur la communication ou les capacités verbales ayant tendance à diminuer le score de QI (Chakrabarti & Fombonne 2001; Yeargin-Allsopp *et al.* 2003). Inversement, 30% des patients souffrant de DI présentent également des traits autistiques comme c'est très fréquemment le cas dans le syndrome de l'X Fragile. Le QI est néanmoins très variable au sein des autistes : ils peuvent présenter une DI sévère ou avoir à l'inverse des capacités intellectuelles préservées comme c'est le cas chez les autistes de "hauts niveaux" ou ceux souffrant du syndrome d'Asperger (qui n'ont de plus pas de retard dans l'apparition du langage). Les ASD et la DI semblent ainsi être deux conditions fortement liées d'un point de vue clinique en coexistant au sein de certains syndromes et patients.

Outre leur coexistence fréquente au sein d'un même patient, il existe un autre point commun clinique entre les ASD et la DI. En effet, il est clairement établi que les ASD présentent un biais de sexe important, le ratio garçon/fille étant en moyenne de 4/1 (de 1,3 à 16/1 selon les études) (Fombonne *et al.* 2012). Après analyse détaillée, il s'avère que ce ratio est directement proportionnel au QI des individus : il est de 2/1 lorsqu'on parle des ASD avec DI, de 6/1 pour les ASD sans DI (autisme de "haut niveau") et même de 10/1 pour le syndrome d'Asperger (Newschaffer *et al.* 2007). Si l'on considère seulement la DI, l'excès de garçons est plus modéré mais néanmoins non négligeable : un excès de 30 à 40% de garçons est observé (ratio garçon/fille de 1,3 à 1,4/1) (Leonard & Wen 2002).

De façon plus générale, au sein des maladies neuropsychiatriques, l'excès de garçons se retrouve principalement dans les troubles à apparition précoce pendant l'enfance ce qui inclue notamment les NDD. Ainsi, l'excès de garçons est retrouvé dans d'autres NDD comme les ADHD, les SLI ou le syndrome de Gilles de la Tourette. A l'inverse, l'excès de filles est plus fréquemment retrouvé dans les désordres émotionnels apparaissant à l'adolescence et les maladies auto-immunes (Rutter *et al.* 2003).

*In fine*, il est important de noter que les critères de diagnostic de ces troubles sont sans cesse en évolution comme en témoigne les nombreux changements de classification opérés dans la dernière édition du DSM-V (Black & Grant 2014). Les ASD et la DI sont, d'après leurs définitions, deux entités cliniques distinctes mais peuvent parfois coexister chez un même individu ce qui rend la réalité beaucoup plus complexe et le diagnostic d'autant plus difficile.

## **II. Les causes génétiques**

Comme il vient de l'être décrit, la DI et les ASD sont des pathologies possédant une très forte hétérogénéité clinique et présentent, en accord avec cela, une étiologie extrêmement complexe. Celle-ci inclue des causes génétiques (qui seront développées dans cette partie) ainsi que des causes environnementales (qui seront développées dans la partie suivante) le tout étant fortement interconnecté. Il existe ainsi 2 modèles qui coexistent pour expliquer l'étiologie de ces pathologies :

- **Le modèle monogénique** : Il concerne les maladies syndromiques et/ou monogéniques à pénétrance élevée (exemple du syndrome de l'X Fragile).
- **Le modèle oligogénique/multifactoriel** : Il concerne les syndromes génétiquement complexes, à pénétrance incomplète. Ils résultent de l'association de plusieurs événements génétiques (modèle oligogénique) ou de l'association d'évènements génétiques et de facteurs environnementaux (modèle multifactoriel).

### **1. La génétique de la DI**

Les causes génétiques de la DI incluent les anomalies chromosomiques (nombre ou structure des chromosomes), les variations du nombre de copie (ou CNV pour Copy Number Variant) et les désordres monogéniques causés par des altérations affectant un seul gène (mutations ponctuelles, délétions, insertions).

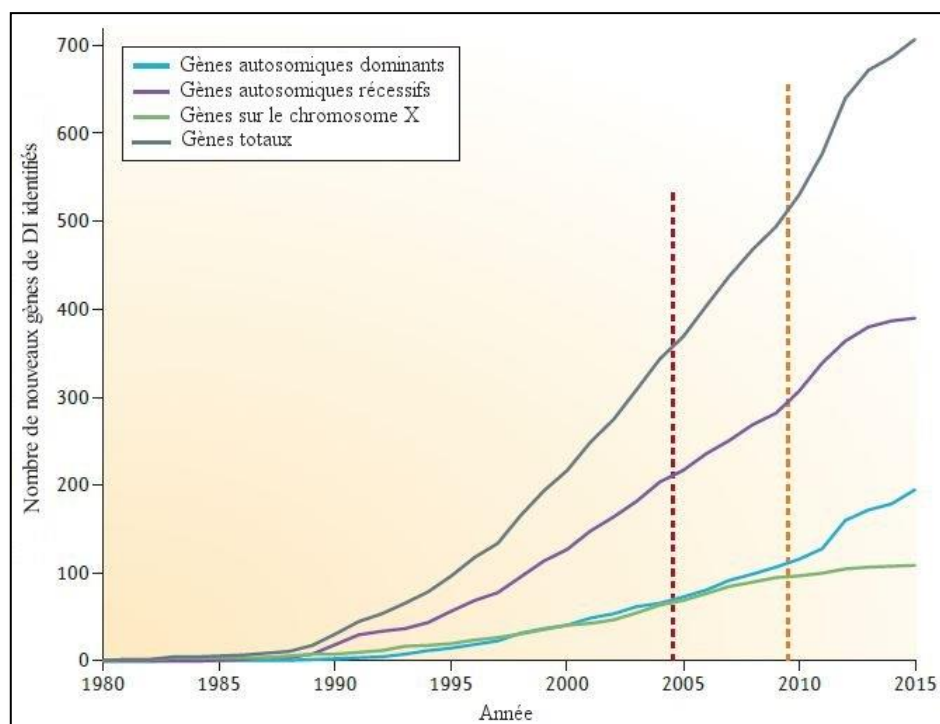
### i. Anomalies chromosomiques et CNV

Les aberrations chromosomiques et les CNV sont très importants dans l'étiologie de la DI car elles sont responsables de très nombreux cas, notamment chez les patients atteints de DI syndromique (Cooper *et al.* 2011). Celles-ci impliquent généralement plusieurs gènes et regroupent :

- **Les anomalies du nombre de chromosomes** (chromosome surnuméraire ou manquant). La première cause chromosomique identifiée comme étant responsable de DI est la présence d'un troisième chromosome 21 (trisomie 21 ou syndrome de Down) (Lejeune *et al.* 1959). C'est également la cause génétique de DI la plus fréquente car elle représente deux tiers des cas causés par des aberrations chromosomiques (Ropers 2010).
- **Les variations structurales balancées ou non** (duplications, délétions, translocations, inversions). Ces variations sont beaucoup plus rares et dispersées tout au long du génome. Leur pathogénicité dépend de leur contenu génique et est également proportionnelle à leur taille (plus elles sont grandes, plus la probabilité d'être pathogène est importante). Enfin, ces variations structurales sont la plupart du temps *de novo* (néovariation se produisant dans une cellule sexuelle d'un des deux parents).
- **Les CNV submicroscopiques** regroupant les délétions et duplications comprises entre 1kb et 5Mb, non détectables par un caryotype. Jusqu'à 20% des cas de DI seraient dus à ces CNV submicroscopiques et certains apparaissent comme étant récurrents. A ce jour, plus de 60 régions présentent des CNV récurrents associées à la DI. Elles sont principalement réparties sur 8 chromosomes (chromosomes 1, 3, 7, 15, 16, 17, 22 et X). Citons comme exemples la délétion 7q11.23 (syndrome de Williams-Beuren) ou la délétion 17p11.2 (syndrome de Smith-Magenis). De plus, la localisation télomérique de ces CNV semble être une caractéristique redondante (Cooper *et al.* 2011).

## ii. Les formes monogéniques de DI

Comme abordé précédemment, les anomalies chromosomiques et CNV comprennent pour la plupart plusieurs gènes dont la contribution exacte dans la pathologie reste parfois incertaine. Cependant, certaines formes de DI peuvent être considérées comme des désordres monogéniques car elles sont la conséquence d'une mutation (délétion, duplication ou mutation ponctuelle) dans un gène unique. Plus de 700 gènes ont été impliqués dans les formes monogéniques de DI (**Figure 1**) et elles peuvent suivre tous les modes de transmission connus : lié au chromosome X, autosomique dominant ou autosomique récessif.



**Figure 1. Augmentation des gènes impliqués dans la DI isolée ou syndromique.**

Vue graphique de l'augmentation de la découverte de gènes de DI isolée ou syndromique au cours du temps en fonction du mode de transmission. Les lignes pointillées verticales représentent l'introduction des puces génomiques (en rouge) et des technologies de séquençage nouvelle génération (en orange) pour la détection de nouveaux gènes de DI. Avec ce graphique, il est clair que l'identification de nouveaux gènes de DI n'a pas encore atteint de plateau, à l'exception peut-être pour les XLID (Vissers *et al.* 2016).

### a. Les DI autosomiques dominantes

Dans ce mode de transmission, le gène en cause est porté par un autosome et la présence de l'allèle muté à l'état hétérozygote suffit pour que la maladie se manifeste. Ce type de DI est retrouvé si un parent porteur transmet la mutation à sa descendance. Dans ce cas, la pénétrance de la mutation est souvent incomplète, c'est-à-dire que la mutation n'aura pas les mêmes conséquences phénotypiques chez tous les individus porteurs. La plupart du temps, le degré d'atteinte du parent porteur est faible voire nul alors que celui de l'enfant est plus important (par exemple des mutations dans le gène *SLC2A1* peuvent être héritées d'un parent asymptomatique) (Leen *et al.* 2010). L'autre possibilité est la présence d'un mosaïcisme germinale (mutations présentes seulement dans certaines cellules d'un parent dont les cellules germinales). Le dernier cas de figure concerne les mutations apparaissant *de novo* (mutations apparaissant dans les gamètes ou, plus rarement, dans l'œuf fécondé). Si cette néovariation est à l'origine d'un allèle pathologique dominant et/ou que l'allèle sain restant ne permet pas une expression suffisante du gène concerné afin de maintenir sa fonction (haploinsuffisance), un enfant peut être malade bien qu'aucun de ses parents ne soit atteint.

Ces néomutations ne sont pas à négliger car elles sont particulièrement fréquentes chez l'Homme (on considère que 50 à 100 nouvelles variations apparaissent chez chaque individu, dont en moyenne une touchant les parties codantes) (Veltman & Brunner 2012). Même si ce type de mutation avait déjà été détecté par séquençage de gènes candidats chez des patients souffrant de DI, le séquençage d'exome en trio (parents et enfant atteint) a permis à l'hypothèse d'un mode de transmission autosomique dominant de la DI de prendre de l'importance (Vissers *et al.* 2010). Ainsi des mutations *de novo* dans les gènes *STXBPI*, *SYNGAP1*, *SCN2A* et *TCF4* semblent être récurrentes dans les cas sporadiques de DI (Hamdan, Gauthier, *et al.* 2009; Hamdan, Piton, *et al.* 2009; Amiel *et al.* 2007; de Ligt *et al.* 2012; Rauch *et al.* 2012). A l'heure actuelle, près de 150 gènes ont été impliqués dans la DI selon un mode de transmission autosomique dominant et pénétrance complète et ce nombre ne cesse d'augmenter chaque année. Les mutations autosomiques dominantes à pénétrance incomplète (mutations héritées d'un parent asymptomatique) sont quant à elles plus difficiles à étudier par définition, laissant un large champ d'étude encore inexploré.

### b. Les DI autosomiques récessives

Dans ce mode de transmission, le gène en cause est porté par un autosome et la présence de l'allèle muté à l'état homozygote est nécessaire pour que la maladie se manifeste. Ces formes de DI concernent particulièrement celles causées par des désordres métaboliques dont la prévalence est estimée à 1% en Europe. Elles résultent d'un déficit enzymatique dans l'une des voies métaboliques permettant la synthèse ou la dégradation de biomolécules indispensables au bon fonctionnement de l'organisme ayant pour conséquence une carence d'un composé nécessaire ou, à l'inverse, une accumulation d'un produit toxique. On peut citer pour exemples :

- les maladies du métabolisme des acides aminés comme la phénylcétonurie causée par une mutation dans le gène *PAH* codant pour la phénylalanine hydroxylase dont la déficience entraîne un excès toxique de phénylalanine,
- les maladies lysosomales comme la maladie de Farber causée par une mutation dans le gène *ASAH1* codant pour la céramidase acide hydrolase lysosomale dont un déficit est responsable de l'accumulation intracellulaire toxique de céramide,
- les maladies énergétiques, peroxymales, de glycosylation, etc.

La prévalence de ces maladies métaboliques est plus importante dans les régions de forte consanguinité. Ainsi, la phénylcétonurie a une prévalence à variabilité géographique considérable : 1:10000 naissances vivantes en Europe contre 1:4000 en Turquie (car 25 à 30% des mariages y sont consanguins) (Tadmouri *et al.* 2009).

La consanguinité augmentant fortement les probabilités d'homozygotie des variants (et donc l'émergence de désordres récessifs), de nombreuses études ont été réalisées sur de grandes familles consanguines dont plusieurs membres étaient atteints de DI (syndromique ou non). Une récente étude suggère même que la sévérité de la DI est directement liée au pourcentage du génome présent à l'état homozygote (Gandin *et al.* 2015). Les premières données ont ainsi mis en évidence les gènes *PRSS12* (Molinari *et al.* 2002), *CRBN* (Higgins *et al.* 2004) et *CC2D1A* (Basel-Vanagaite *et al.* 2006) auxquels se sont ajoutés 50 puis 30 autres gènes grâce à l'étude de respectivement 136 puis 121 familles consanguines (Najmabadi *et al.* 2011; Riazuddin *et al.* 2016). Jusqu'à maintenant, ce n'est pas moins de 360 gènes qui ont été associés ou sont candidats à une forme autosomique récessive de DI.

### c. Les DI liées au chromosome X (XLID)

Les hommes n'ayant qu'un chromosome X, la présence d'un allèle muté à l'état hémizygotique est suffisante pour que la maladie se manifeste (la question de dominance ou de récessivité ne se pose alors pas). Les mutations des gènes du chromosome X sont ainsi l'une des hypothèses qui expliquerait l'excès de garçons de 30 à 40% rapporté dans la DI et ont fait l'objet d'études intensives.

La recherche de gènes impliqués dans la DI sur le chromosome X (gènes de XLID pour X-Linked Intellectual Disability) a commencé dans les années 1960 avec l'étude de grandes familles montrant un mode de transmission clairement lié au chromosome X (seuls les garçons étaient atteints, les mères étant vectrices). Depuis, de nombreux consortiums internationaux réunissant des équipes du monde entier ont permis d'identifier pas moins de 100 gènes du chromosome X supposés impliqués dans la DI (Lubs *et al.* 2012). A titre d'exemple, la première implication d'une anomalie du chromosome X dans un cas de DI fût le syndrome de l'X Fragile en 1966. Ce syndrome dû à une mutation dans *FMR1* (Oberlé *et al.* 1991) est la seconde cause de DI après le syndrome de Down et la première cause monogénique de DI (Coffee *et al.* 2009). D'autres gènes du chromosome X sont mutés de façon récurrente dans les cas de DI isolées ou syndromiques comme c'est le cas d'*ARX*, *MECP2*, *SLC16A2* ou *ILIRAPL1*. Il est intéressant de noter que de plus en plus de cas de XLID sont rapportés chez des femmes (exemples des gènes *DDX3X*, *NAA10*, etc.), point qui sera abordé dans une prochaine partie. Cependant, même si le nombre de gènes de XLID ne cesse d'augmenter (leur implication dans la DI est revue dans Piton *et al.*) (Piton *et al.* 2013), l'identification de nouveaux gènes commence à atteindre un plateau (**Figure 1**) et l'expansion du triplet CGG de la partie 5' non traduite du gène *FMR1* reste l'anomalie la plus récurrentement identifiée.

### iii. Le modèle de l'oligogénisme

Malgré l'important travail fourni tout au long de ces années pour comprendre la génétique de la DI, de nombreux cas restent non résolus, et ce même après analyse par séquençage du génome entier. Ainsi, il existe l'hypothèse "oligogénique" ou "hits multiples" impliquant l'association de plusieurs facteurs dans plusieurs gènes. Ce modèle est notamment supporté par les récentes études de CNV montrant que certains d'entre eux

possèdent une pénétrance plus importante s'ils sont accompagnés d'autres grands CNV. Par exemple, la microdélétion récurrente de 520kb du locus 16p12.1 est associée de manière significative au retard de développement chez l'enfant dans deux études patients/contrôles. Un quart des enfants atteints possède également un second large CNV et les présentations cliniques de ces derniers sont plus nombreuses et/ou plus sévères que celles des enfants n'ayant que la microdélétion 16p12.1 (Girirajan *et al.* 2011). Ces données suggèrent que le phénotype associé à la microdélétion 16p12.1 possède une expressivité variable dépendante de l'environnement génétique. Cette théorie de l'oligogénisme implique de rechercher à l'avenir dans les études par séquençage (d'exome ou de génome entier) plusieurs hits (CNV et/ou variants rares) afin d'évaluer sa contribution réelle dans l'étiologie de la DI.

## 2. La génétique de l'autisme

### i. Mise en évidence de la part génétique

Dès la première description clinique de l'autisme par Léo Kanner en 1943, celui-ci suggère le caractère inné de l'autisme mais évoque également le rôle défavorable de l'environnement familial dans lequel la majorité des enfants observés avaient grandi, ceux-ci étant souvent élevés par des mères manquant d'amour maternel (il parlait alors de "mères réfrigérateurs"). Cette idée a ensuite été défendue dans les années 1960-1970 par le psychologue Bruno Bettelheim et est toujours ancrée dans l'imaginaire populaire. Par la suite, il a été difficile de faire accepter par l'opinion publique que les facteurs génétiques jouent un rôle majeur dans l'apparition de l'autisme.

Ainsi, la part génétique de l'autisme a été reportée pour la toute première fois en 1977 sur la base d'une étude comparant la concordance phénotypique entre 11 paires de jumeaux monozygotiques (MZ) et 10 paires de jumeaux dizygotiques (DZ) (Folstein & Rutter 1977). En effet, la différence entre le taux de concordance des jumeaux MZ (qui ont le même génome) et DZ (qui possèdent 50% de leur génome en commun) reflète la part des facteurs génétiques impliqués dans la pathologie étudiée. Dans le début des années 1990, l'augmentation de la taille des cohortes étudiées et l'utilisation des instruments standards de diagnostic a permis d'atteindre le taux de 60 à 90% de concordance pour les MZ contre 0 à 10% pour les DZ, supportant la forte composante génétique de l'autisme (Steffenburg *et al.* 1989; Bailey *et al.* 1995; Hallmayer *et al.* 2011).



Parallèlement à ces études de jumeaux, des études familiales ont permis d'évaluer que le taux de récurrence chez les frères et sœurs des individus autistes varie de 2 à 8% (Bailey *et al.* 1998; Muhle *et al.* 2004) et atteint même les 14% pour les fratries de filles autistes (Ritvo *et al.* 1989) ce qui est de 20 à 50 fois plus important que la prévalence dans la population générale (Ozonoff *et al.* 2011). Ce risque relatif est cependant bien plus faible que pour une maladie monogénique. L'ensemble de ces études (de jumeaux et familiales) permettent ainsi clairement d'établir qu'il existe une susceptibilité génétique à l'autisme.

Néanmoins, étant donné que la concordance des MZ est inférieure à 100% et que le degré d'atteinte ainsi que la sévérité des symptômes varie sensiblement au sein des paires concordantes, il semble que les facteurs environnementaux et les interactions complexes entre gènes et environnement soient également importants dans l'étiologie de l'autisme. A l'inverse de la DI, l'autisme a tout d'abord été considéré comme étant un désordre complexe ayant majoritairement des causes multifactorielles, même si quelques formes d'autisme syndromique monogénique étaient déjà connues.

## **ii. Les causes multifactorielles**

L'autisme est une maladie complexe avec une très forte hétérogénéité clinique laissant supposer une forte hétérogénéité génétique. Le modèle généralement accepté pour expliquer l'origine de l'autisme était l'implication de variations génétiques fréquentes (CNV ou SNP pour Single Nucleotide Polymorphism) qui ne sont pas pathogènes à elles seules mais constituent un facteur de prédisposition à la maladie. Ainsi, les premières études sur l'autisme recherchaient des variants communs à effet modéré, tout d'abord par des études de liaison sur génome entier qui mettent en évidence la transmission préférentielle de loci aux membres atteints des différentes familles. Durant ces dernières décennies, 14 études ont été menées identifiant plus de 90 loci sur 34 régions mais moins de 10 s'avèrent réellement significatifs (Yang & Gill 2007). Parallèlement, des études d'association (gène candidat ou sur génome entier comparant individus atteints et contrôles) et de CNV ont également été menées (pour revue voir De la Torre-Ubieta *et al.*) (De la Torre-Ubieta *et al.* 2016).

### a. Etude d'association gène candidat

Les approches gènes candidats se sont concentrées sur des gènes supposés impliqués dans l'étiologie de la pathologie de par leur localisation (régions déterminées par les études d'association, de CNV ou de réarrangements chromosomiques), leur expression dans le cerveau ou leur rôle dans son développement et/ou la fonction cognitive. Les études étaient néanmoins souvent discordantes, avec un manque de réplification entre elles. Par exemple, plusieurs études montrent une association entre des polymorphismes présents dans le gène *FOXP2* et les ASD (Gong *et al.* 2004; Li *et al.* 2005) alors que d'autres ne mettent en évidence aucune association significative (Newbury *et al.* 2002; Gauthier *et al.* 2003; Marui *et al.* 2005).

### b. Etude d'association pangénomique (GWAS)

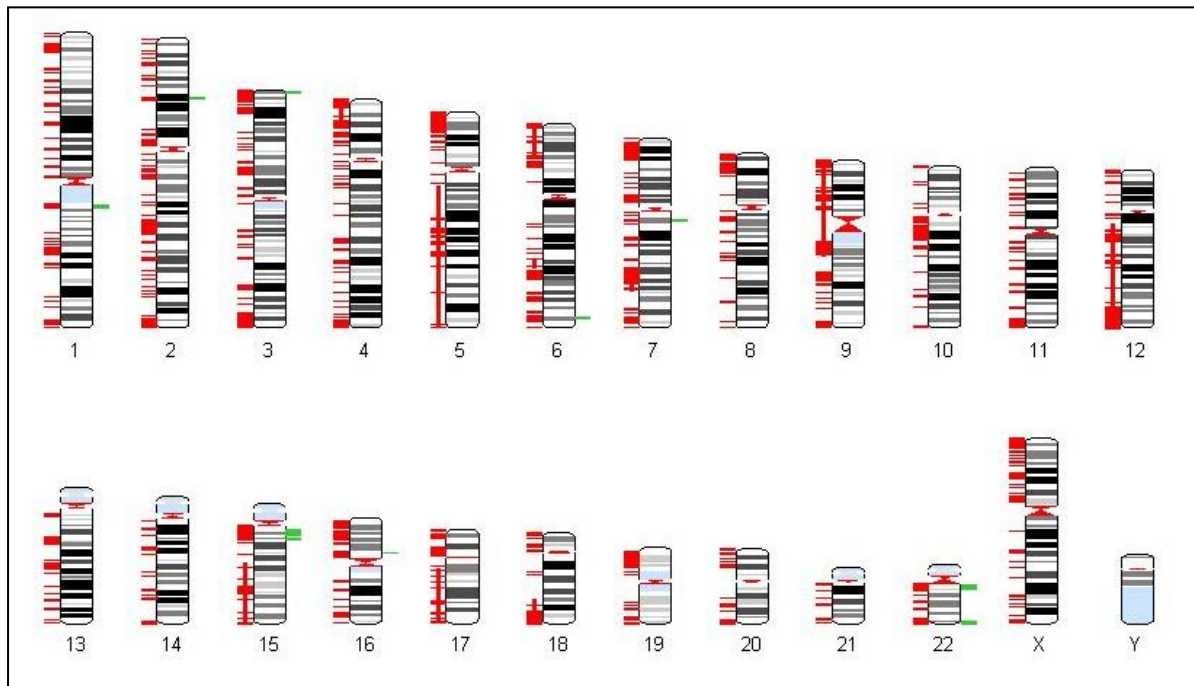
Les études d'association sur l'ensemble du génome (GWAS pour Genome-Wide Association Study) ont recherché des variations génétiques (SNP) présents dans la population générale mais enrichis chez les sujets souffrant d'ASD, suggérant leur rôle en tant que facteur de risque dans l'autisme. Jusqu'à maintenant, les études par GWAS ont également été très décevantes en ne mettant en évidence que peu de SNP suggestifs (c'est-à-dire avec un signal clairement significatif) (Kuo *et al.* 2015) à proximité ou dans des gènes tels que *CDH9* et *CDH10* (Wang, Zhang, *et al.* 2009; Ma *et al.* 2009), *SEMA5A* (Weiss *et al.* 2009) ou *MACROD2* (Anney *et al.* 2010). De plus, ces études ont été peu répliquées et les études de réplifications sont parfois même négatives (exemple du polymorphisme rs4141463 de *MACROD2* significativement associé dans l'étude d'Anney *et al.* mais pas dans l'étude de Curran *et al.*) (Anney *et al.* 2010; Curran *et al.* 2011). Même si ces études ont été réalisées sur de grands effectifs (plus de 1000), d'autres études sur de plus larges cohortes sont probablement nécessaires pour avoir une puissance statistique satisfaisante pour une étude pangénomique. Ce manque de résultats convaincants reflète l'hétérogénéité génétique des ASD, et ce même pour les analyses réalisées sur des cohortes cliniquement homogènes (Chaste *et al.* 2015). Désormais, de nouvelles techniques d'analyses sont mises au point afin de contourner les limitations dues à l'hétérogénéité clinique et génétique intrinsèque aux ASD (intégration de données cliniques détaillées ainsi que des données de séquençage du génome (Jiménez-Barrón *et al.* 2015) ou introduction d'un facteur de risque additionnel représentant l'environnement ou le modèle oligogénique (Lu & Cantor 2012)).

### c. Etude de CNV

Les remaniements chromosomiques de grande taille (variations structurales et CNV) sont retrouvés dans 2 à 5% des cas d'ASD (Reddy 2005). Une étude globale de CNV rares (fréquence < 1% dans la population générale) réalisée chez près de 1000 patients souffrant d'ASD révèle une fréquence totale des CNV rares plus élevée chez les patients (avec en moyenne 3,6 gènes affectés par un CNV chez les contrôles contre 4,3 chez les patients) (Pinto *et al.* 2010). Ils sont répartis sur l'ensemble du génome, touchant tous les chromosomes. Les CNV *de novo* sont particulièrement enrichis chez les autistes (avec une fréquence de 10% contre 1% environ chez les témoins) (Sebat *et al.* 2007). Ainsi, les études estiment que 8% des cas sporadiques d'autisme portent des CNV *de novo* contre seulement 2% de leurs frères et sœurs non atteints (Krumm *et al.* 2015).

Outre les CNV rares hérités ou *de novo*, la combinaison de plusieurs études de CNV (ou de cohortes de réplication) ont mis en évidence des CNV récurrents chez les patients souffrant d'ASD. La **Figure 2** regroupe l'ensemble des CNV identifiés dans neuf études à large échelle (Sebat *et al.* 2007; Szatmari *et al.* 2007; Marshall *et al.* 2008; Cuscó *et al.* 2009; Glessner *et al.* 2009; Gregory *et al.* 2009; van der Zwaag *et al.* 2009; Pinto *et al.* 2010; Bremer *et al.* 2011) ainsi que les plus récurrents. Ils sont également répartis sur l'ensemble du génome, touchant tous les chromosomes à l'exception du chromosome Y.

Néanmoins, le cas des CNV est beaucoup plus complexe puisque si certains sont seulement des facteurs de prédisposition, d'autres ont clairement été identifiés comme responsables de cas d'autisme. Avec une prévalence proche de 1% chacune, la duplication maternelle 15q11-q13 ainsi que les CNV du locus 16p11.2 sont les CNV les plus fréquemment décrits chez les patients autistes (Veenstra-VanderWeele & Cook 2004; Weiss *et al.* 2008). De plus, la région 15q11-q13 comprend la région 15q11.2-q13 dont la délétion est responsable des syndromes de Prader-Willi/Angelman (la région critique étant localisée entre les points de cassure BP2 et BP3). D'autres CNV récurrents tels que les loci 7q11.23, 22q11.2 ou 22q13, dont les délétions sont responsables respectivement des syndromes de Williams-Beuren, DiGeorge et Phelan-McDermid, sont également associés aux ASD (Vorstman *et al.* 2006) et à la DI.



**Figure 2. Localisation des CNV identifiés dans les analyses pangénomiques de 9 études concernant plus de 5000 patients autistes** (Sebat *et al.* 2007; Szatmari *et al.* 2007; Marshall *et al.* 2008; Cuscó *et al.* 2009; Glessner *et al.* 2009; Gregory *et al.* 2009; van der Zwaag *et al.* 2009; Pinto *et al.* 2010; Bremer *et al.* 2011). Les lignes rouges sur la gauche de chaque chromosome représentent plus de 1200 CNV validés identifiés dans ces études. Les lignes vertes sur la droite des chromosomes 1, 2, 3, 6, 7, 15, 16 et 22 représentent les 10 CNV rares récurrents identifiés au moins dans trois des neuf études. Issu de (Croft *et al.* 2011)

*In fine*, les variations structurales ainsi que les CNV rares hérités ou *de novo* ont un rôle prépondérant dans l'étiologie génétique de l'autisme. Les études de CNV à grande échelle ont permis d'estimer que plusieurs centaines de gènes et loci (200 à 400) sont associés aux ASD (Levy *et al.* 2011; Sanders *et al.* 2011). De plus, les CNV rares qui ont été décrits permettent de délimiter des régions chromosomiques potentiellement impliquées dans les ASD ce qui facilite la mise en évidence de nouveaux gènes candidats (même si l'implication de certains de ces CNV dans la pathologie reste parfois incertaine car rarement répliquée).

### iii. Les formes monogéniques d'autisme

En parallèle du modèle multifactoriel, il est apparu plus tardivement un modèle monogénique impliquant des évènements génétiques rares, fortement délétères et associés à une pénétrance élevée. Ces évènements génétiques peuvent être des variations structurales (tels que les réarrangements chromosomiques), des CNV (comme il l'a été abordé précédemment) ou des mutations affectant un seul gène (insertion, délétion et mutations ponctuelles).

#### a. Cas des ASD syndromiques

Même si le modèle multifactoriel a pendant longtemps été considéré comme le modèle génétique de l'autisme, il avait déjà été identifié avant 2000 quelques cas de mutations pénétrantes pour des formes syndromiques d'ASD. Ceux-ci représentent 10 à 15% de l'ensemble des ASD et sont pour la plupart associés à des syndromes génétiques connus. Les plus fréquents sont le syndrome de l'X Fragile, le syndrome de Rett, la sclérose tubéreuse de Bourneville, les syndromes de Prader-Willi/Angelman et le syndrome de Cowden (voir la section "Définitions" de cette partie). Ces syndromes associés aux ASD sont ainsi l'archétype des formes monogéniques d'autisme car, à l'exception des syndromes de Prader-Willi/Angelman qui sont majoritairement causés par une délétion 15q11.2-q13 comprenant plusieurs gènes, une anomalie génétique dans un unique gène est nécessaire et suffisante pour entraîner la pathologie (la plupart du temps *de novo*, avec une notion de prémutation pour le syndrome de l'X Fragile). Il faut néanmoins noter que la sclérose tubéreuse de Bourneville (causée notamment par une mutation hétérozygote dans *TSC1*) fait office d'exception puisqu'elle peut être héritée d'un parent apparemment asymptomatique, l'expressivité de cette maladie étant dans ce cas variable (Sybert & Hall 1979; Cassidy *et al.* 1983).

#### b. Les ASD liés au chromosome X

Comme il l'a été abordé, les ASD sont parfois inclus dans des syndromes génétiques connus. Or, parmi les cinq plus fréquents déjà évoqués, deux sont causés par des gènes présents sur le chromosome X (Syndrome de l'X Fragile : *FMRI*; Syndrome de Rett : *MECP2* et *CDKL5*). Par la suite, les études de CNV ont permis de mettre en évidence des loci de susceptibilité au sein desquelles des études laborieuses de séquençage par approche gène candidat ont été réalisées. Des mutations dans les gènes candidats *ARX*

(localisé en Xp21.3) (Turner *et al.* 2002), *CDKL5* (localisé en Xp22.13) (Russo *et al.* 2009; Szafranski *et al.* 2015) et *PTCHD1* (localisé en Xp22.11) (Chaudhry *et al.* 2015) ont été identifiées chez des patients souffrant d'ASD mais, à l'exception de *PTCHD1* qui pourrait compter pour 1% des cas (Noor *et al.* 2010), elles restent extrêmement rares (Chaste *et al.* 2007; White *et al.* 2010) et ne concernent là encore que des cas d'autisme syndromique.

C'est en 2003 que le séquençage de gènes candidats a permis d'identifier avec succès les premières mutations tronquantes à forte pénétrance dans des cas d'ASD non syndromiques dans les gènes *NLGN3* et *NLGN4X* (Jamain *et al.* 2003). Une délétion de 2pb dans le gène *NLGN4X* a également été trouvée dans une famille comportant des cas de DI accompagnée ou non de traits autistiques (Laumonnier *et al.* 2004). Dès lors la recherche de mutations rares s'est accélérée, notamment grâce à l'arrivée des nouvelles technologies de séquençage haut débit permettant de séquencer plusieurs centaines de gènes simultanément (séquençage ciblé, d'X-exome, d'exome). Ainsi, en 2011, le séquençage systématique de 111 gènes du chromosome X codant pour des protéines synaptiques chez 142 patients souffrant d'autisme a permis d'identifier huit variants *de novo* dont seulement deux ont été confirmés (une délétion décalant le cadre de lecture du gène *ILIRAPL1* et une mutation faux-sens dans *MECP2*) (Piton *et al.* 2011). Il semble donc que les mutations dans des gènes du chromosome X soient beaucoup moins fréquentes dans les ASD que dans la DI.

### c. Les ASD autosomiques dominants

Tout comme pour la DI, ce type de transmission est retrouvé dans les cas de mutations à pénétrance incomplète ou de mutations apparaissant *de novo* (avec parfois un mosaïcisme germlinal parental). A l'heure actuelle, très peu de mutations à pénétrance incomplète ou résultant d'un mosaïcisme germlinal parental ont été décrites. Cependant, comme c'était le cas pour les CNV, les mutations rares *de novo* sont fortement associées à l'autisme (Sanders *et al.* 2012). Celles-ci sont deux à trois fois plus fréquentes chez les patients souffrant d'ASD, en particulier dans les cas sporadiques et seraient responsables de 10% des cas d'autisme (Iossifov *et al.* 2014). Ainsi, cinq des six premières études d'exomes publiées se sont focalisées sur les variants *de novo* (O'Roak *et al.* 2011; Neale *et al.* 2012; O'Roak *et al.* 2012; Sanders *et al.* 2012; Iossifov *et al.* 2012). Des centaines de mutations tronquantes ont alors été identifiées avec une pathogénicité fortement probable (de par leur effet sur la protéine codée par le gène muté). A l'inverse de la DI, il a été flagrant de constater

la très grande hétérogénéité des gènes impliqués. En effet, dans ces cinq études d'exomes, seuls cinq gènes sont retrouvés de manière récurrente (chez deux patients seulement) : *CHD8*, *DYRK1A*, *NTNG1*, *POGZ* et *SNC2A*. Ainsi, contrairement à la DI, parmi les milliers de patients autistes étudiés par séquençage d'exomes ou de génomes, seulement quelques gènes ont été retrouvés mutés plus d'une fois laissant supposer un degré d'hétérogénéité génétique bien plus important (Michaelson *et al.* 2012).

#### d. Les ASD autosomiques récessifs

L'implication du mode de transmission autosomique récessif dans les ASD a été supportée par la recherche de variations récessives rares dans des gènes candidats. Une étude sur plus de 1400 trios parents-enfant atteint réalisée par le "Autism Genome Project" a révélé que les haplotypes homozygotes (régions du génome retrouvés à l'état homozygote) sont significativement enrichis chez les patients autistes comparés à leurs parents contrôles, en particulier dans les régions possédant des gènes candidats à l'autisme (Casey *et al.* 2012). Une précédente étude avait quant à elle montré l'importance des larges délétions homozygotes (Morrow *et al.* 2008). Parallèlement, une étude de séquençage d'exomes couplé à de la cartographie d'homozygotie effectuée sur des familles consanguines souffrant d'ASD a tenté de mettre en évidence les gènes impliqués suivant un mode autosomique récessif (Chahrouh *et al.* 2012). Celle-ci a révélé quatre nouveaux gènes candidats ayant des mutations à l'état homozygote potentiellement pathogènes et ségrégeant parfaitement avec la maladie au sein des familles (*UBE3B*, *CLTCL1*, *NCKAP5L* et *ZNF18*). Une seconde étude de 163 familles consanguines (cohorte collectée par Morrow *et al.*) (Morrow *et al.* 2008) a pu identifier des mutations homozygotes dans les gènes *AMT*, *PEX7*, *SYNE1*, *VPS13B*, *PAH* et *POMGNT1*, certains d'entre eux présentant également ce type de mutation dans des familles non consanguines (Yu *et al.* 2013). Il est à noter que certains de ces gènes sont retrouvés dans des formes syndromiques d'ASD (*SYNE1*, *VPS13B*, *POMGNT1*) posant la question des critères de recrutement utilisés. Néanmoins, il semble que les variants transmis selon un mode autosomique récessif ne soient pas si rares dans l'autisme, notamment les mutations perte de fonction dont la contribution dans les ASD est estimée à 3% (Lim *et al.* 2013).



#### iv. Le modèle de l'oligogénisme

Même si de plus en plus de formes d'ASD monogéniques sont désormais mises en évidence, elles contribuent à une moindre proportion des cas comparé à la DI, le paradigme de l'oligogénisme étant encore plus important dans les ASD qu'il l'était dans la DI. Les études récentes de GWAS ont souligné la contribution importante de variants communs à effet faible aux ASD, selon le modèle oligogénique (voire polygénique) (Gaugler *et al.* 2014). Une étude a ainsi révélé que les patients autistes présentent plus de variants hétérozygotes (hérités ou *de novo*) dans des gènes de susceptibilité à l'autisme comparés aux contrôles (Schaaf *et al.* 2011). L'association de plusieurs de ces variants au sein d'un même patient pourrait être à l'origine de la pathologie et conforte la théorie de l'oligogénisme. Une autre étude rapporte trois patients souffrant d'ASD chez qui des délétions *de novo* dans *SHANK2* sont associées à des CNV hérités du locus 15q11-q13 (également impliqué dans l'autisme) (Leblond *et al.* 2012). Enfin, la première étude de séquençage d'exomes réalisée chez des patients souffrant d'ASD a notamment mis en évidence la présence d'une mutation *de novo* dans *FOXP1* et d'une mutation héritée dans *CNTNAP2* chez un même patient (O'Roak *et al.* 2011). L'ensemble de ces études supporte ainsi le rôle non négligeable du modèle "multiple hits" dans l'étiologie de l'autisme.

*In fine*, il semble que le modèle multifactoriel et le modèle monogénique soient tous deux impliqués dans l'étiologie de l'autisme. Néanmoins, la contribution des formes monogéniques a tendance à être plus importante dans les cas d'ASD avec DI que dans les cas d'ASD de haut niveau (Iossifov *et al.* 2014; Ronemus *et al.* 2014). De plus, une étude suédoise analysant le risque de récurrence relatif chez plus de deux millions d'enfants a permis de montrer que les facteurs environnementaux participeraient au même niveau que les facteurs génétiques, expliquant ainsi la complexité de l'étiologie des ASD et l'hétérogénéité phénotypique (Sandin *et al.* 2014).

#### v. Bases de données pour la recherche sur l'autisme

Dans le but de rassembler toutes les mutations pénétrantes et anomalies génétiques récurrentes impliquées dans l'étiologie de l'autisme, une revue exhaustive de C. Betancur a montré qu'il y aurait 103 gènes et 44 loci récurrents reportés chez des patients avec autisme



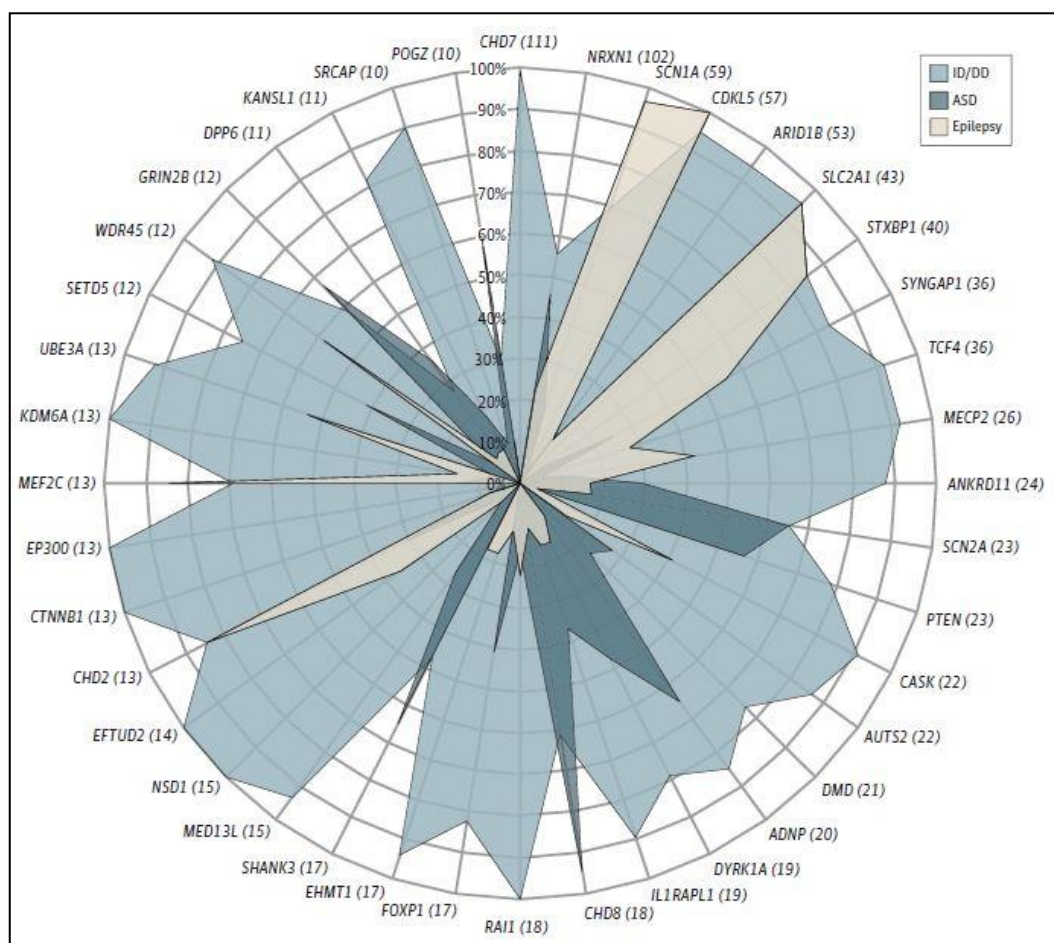
(Betancur, 2011). A ce jour, plus de 1000 gènes ou SNP ont été associés et impliqués dans l'autisme. Afin d'aider les chercheurs à suivre la littérature grandissante à ce sujet, de nombreuses bases de données ont vu le jour (SFARI, la base de donnée de la Simons Foundation [www.sfari.org](http://www.sfari.org), AutismKB [www.autismkb.cbi.pku.edu.cn](http://www.autismkb.cbi.pku.edu.cn) ou AutDB [www.autism.mindspec.org/autdb/Welcome.do](http://www.autism.mindspec.org/autdb/Welcome.do)) (Banerjee-Basu & Packer 2010; Abrahams *et al.* 2013; Xu *et al.* 2012; Basu *et al.* 2009). Celles-ci rassemblent les gènes, CNV et régions de linkage associés ou impliqués dans les ASD et parfois même les animaux modèles. Elles donnent ainsi accès à l'ensemble des données et évidences pour un gène donné et leur donne un score en fonction de la robustesse de leur implication dans les ASD.

### 3. Chevauchement génétique

#### i. Des centaines de gènes identifiés...

L'étiologie de la DI et des ASD est extrêmement complexe : dans les deux cas, des centaines de gènes ont été identifiés, causant autant de variantes du tableau clinique de ces maladies. Même si l'architecture génétique de ces pathologies diffère (apparemment plutôt monogénique pour la DI et plutôt multifactorielle pour les ASD), elles présentent un certain chevauchement génétique. Ceci est illustré par l'étude de CNV de Pinto *et al.* qui a montré une fréquence accrue chez les patients souffrant d'ASD de CNV affectant des loci précédemment impliqués dans la DI (Pinto *et al.* 2010). Les 103 gènes et 44 loci récurrents reportés chez des patients souffrant d'autisme dans la revue de C. Betancur sont également impliqués dans la DI indiquant que ces deux désordres neurodéveloppementaux partagent bien des bases génétiques communes (Betancur, 2011; Guilmatre *et al.* 2009). Il existe ainsi un grand chevauchement entre les gènes impliqués dans la DI et ceux impliqués dans les ASD, et tous les gènes pouvant conduire à des ASD peuvent également conduire à une DI avec ou sans ASD. L'un des exemple les plus connus concerne le gène *NLGN4X* dont la première mutation a été rapportée dans une famille comportant deux frères atteints, l'un souffrant d'autisme avec DI, le second présentant le syndrome d'Asperger (Jamain *et al.* 2003). Par la suite, une mutation tronquante a été identifiée dans une large famille avec 13 hommes affectés par de la DI isolée (10 individus), des ASD isolés (un individu) ou des ASD associés à de la DI (deux individus) (Laumonier *et al.* 2004).

Un autre exemple concerne le gène *IL1RAPL1* dans lequel une grande délétion intragénique des exons 3 à 7 a été identifiée dans une famille avec trois frères atteints avec des degrés différents d'autisme et/ou de DI (Piton *et al.* 2008). De manière générale, il existe des causes génétiques communes entre des conditions apparemment différentes, certains gènes étant tout de même impliqués de façon prépondérante dans la DI (comme *ARID1B*, *CTNNB1*, etc.) et d'autres plutôt dans l'autisme (comme *CHD8*, *GRIN2B*, etc.) (**Figure 3**).



**Figure 3.** Fréquence de la DI, des ASD et de l'épilepsie chez des individus porteurs de variants perte de fonction dans des gènes candidats pour les NDD. Sont représentés les gènes ayant au moins trois variants *de novo* perte de fonction (Gonzalez-Mantilla *et al.* 2016)

## ii. ...convergeant vers des voies biologiques communes

Depuis 2006, le nombre d'études de CNV, de SNP, de séquençage d'exomes et maintenant de génomes chez les patients souffrant de DI et d'ASD ont généré une très grande quantité de données et plusieurs nouveaux gènes sont identifiés comme impliqués dans ces désordres chaque année. Malgré le grand nombre de gènes identifiés dans la DI et les ASD, ceux-ci convergent vers un petit nombre de voies biologiques communes dont les altérations sont associées aux phénotypes observés dans ces conditions. Les modalités selon lesquelles ces voies neurobiologiques sont altérées expliqueraient la variabilité clinique de ces pathologies.

Les processus biologiques dans lesquelles ces gènes sont impliqués impactent incontestablement la maturation du cerveau (prolifération cellulaire, migration des neurones, croissance des neurites, etc.) et plus particulièrement la formation, la fonction et l'homéostasie de la synapse (Guilmatre *et al.* 2009; Toro *et al.* 2010; Gilman *et al.* 2011; Krumm *et al.* 2014; Pinto *et al.* 2014). Pour résumer, l'activité et la plasticité synaptique sont impliquées dans la physiopathologie de la DI et des ASD à des niveaux très divers :

- **La régulation du taux des protéines synaptiques** qui s'effectue :
  - o au niveau transcriptionnel (exemple du gène *MECP2* impliqué dans le syndrome de Rett),
  - o au niveau traductionnel (exemples de *FMRI* ainsi que de la voie PI3K-AKT-mTOR régulée notamment par *NFI*, *TSC1/2* et *PTEN* eux aussi impliqués dans des syndromes avec DI et traits autistiques) (Cuscó *et al.* 2009),
  - o au niveau de leur transport (exemple de *FMRI*),
  - o au niveau de leur dégradation par la voie de l'ubiquitine (exemples des gènes *UBE3A*, *PARK2*, *FBXO10*, *USP7*) (Kelleher & Bear 2008; Glessner *et al.* 2009).
  
- **La structure de la synapse.** Des anomalies génétiques dans des gènes codant pour des protéines d'adhérence cellulaire pré et post-synaptique (*NRXN1*, *NLG1-4*, *CNTNAP2*, *CNTN4*, etc.), des protéines d'échafaudage intracellulaire de la densité post-synaptique (*SHANK1-3*, *CASK*, etc.) ainsi que dans des protéines composant et/ou régulant les vésicules synaptiques (*SYN1*, *DISC1*, etc.) ont été impliquées dans la DI et les ASD (Südhof, 2008;

Glessner *et al.* 2009; Wang, Zhang, *et al.* 2009; Fassio *et al.* 2011; Crepel *et al.* 2010). Le cytosquelette, notamment les protéines modifiant la morphologie du réseau d'actine (telles que FLNA, CTNNA3, DOCK8, SPTAN1, CYFIP1 ou les RhoGTPases CDC42, RHOB, RAC1 régulées par l'activité synaptique), est également un acteur clé dans la régulation des épines dendritiques et la croissance des neurites (Gilman *et al.* 2011).

- **Les récepteurs synaptiques.** De nombreux gènes codant pour des sous-unités de récepteurs au glutamate ou à l'acide  $\gamma$ -aminobutyrique (GABA, les deux principaux neurotransmetteurs du cerveau) ont été associés à la DI et aux ASD (*GRIN2A*, *GRIN2B*, *GRIK2*, *GRIA3*, *GRM1/3*, *GABRQ*, *GABRA*, etc.) mais d'autres types de récepteurs voltage-dépendant sont également impliqués (*SCN2A*, *SCN1A*, *CACNA1C*, *CACNA1H*, etc.). Le rôle de la synapse glutamatergique semble néanmoins prépondérant (Cuscó *et al.* 2009).

D'autre part, il s'avère que l'épigénétique joue un rôle clé dans la DI et les ASD. En effet, il s'agit du second processus biologique (après la synapse) enrichi en gènes associés ou impliqués dans ces pathologies (Krumm *et al.* 2014; Pinto *et al.* 2014). Dans son étude d'exomes de patients souffrant d'autisme sporadique O'Roak *et al.* a montré que 40% des mutations *de novo* identifiées convergent vers le réseau de remodelage de la chromatine (O'Roak *et al.* 2012). Pour exemples citons les gènes codant pour des remodeleurs de la chromatine tels que le complexe SWI/SNF (dont *ARID1B* code pour une sous-unité), *MECP2*, *ATRX*, *CHD*, *ASH1L*, *KDM6B*, *MLL5*, *EHMT1*, etc (Krumm *et al.* 2014). De plus, plusieurs syndromes génétiques montrent bel et bien l'interaction entre génétique et épigénétique au sein de la DI et des ASD : 1) Le syndrome de Rett dont le gène muté *MECP2* code pour une protéine qui lie spécifiquement l'ADN méthylé et régule la transcription; 2) Le syndrome de l'X Fragile qui est causé par une hyperméthylation de l'expansion CGG de *FMRI* qui induit l'extinction du gène et 3) Les syndromes de Prader-Willi/Angelman causés majoritairement par la délétion 15q11.2-q13, région soumise à empreinte parentale dans le cerveau. Si les anomalies génétiques associées à la DI et aux ASD ont fait et font l'objet d'études intensives, les anomalies épigénétiques ont été moins étudiées (notamment à cause de la spécificité tissulaire et temporelle des régulations épigénétiques qui sont donc difficilement analysables à partir d'ADN sanguin) et restent un enjeu important pour la compréhension de l'étiologie des NDD.

La DI et les ASD sont des pathologies dont les causes génétiques sont multiples : aberrations chromosomiques, CNV, mutations ponctuelles voire la combinaison de plusieurs de ces anomalies chez un même individu atteint. Mis à part certains syndromes, la fréquence de ces évènements génétiques au sein de la DI et des ASD est extrêmement faible (parfois même retrouvé chez un seul individu) ce qui laisse suggérer que ces pathologies ne sont pas des entités cliniques simples mais plutôt les manifestations développementales et comportementales de centaines de désordres génétiques distincts. Une très forte hétérogénéité génétique est donc retrouvée (surtout pour les ASD), allant de pair avec l'hétérogénéité clinique. Pourtant, les centaines de gènes impliqués convergent vers des voies biologiques communes montrant une physiopathologie commune entre ces pathologies. A cela s'ajoute un autre niveau de complexité venant de l'environnement puisque certains cas de DI et d'ASD pourraient impliquer à la fois des polymorphismes fréquents (facteur de risque génétique) et des facteurs environnementaux de toutes sortes.

### **III. Les causes environnementales**

#### **1. La DI "acquise"**

D'après une étude sur une cohorte de près de 11000 patients souffrant de DI, les causes environnementales de DI représentent 18% des cas (Stevenson *et al.* 2003), chiffre confirmé par une revue plus récente (Rogers *et al.* 2008). Ces DI dites "acquises" sont issues d'atteintes du système nerveux central se produisant pendant les étapes critiques de son développement (que ce soit pré, péri ou post-natal, c'est-à-dire respectivement pendant la grossesse, l'accouchement ou après la naissance) entraînant des dommages irréversibles. Elles peuvent être d'origine toxique, infectieuse ou traumatique.

##### **i. Les expositions toxiques**

Même s'il existe la barrière placentaire qui protège l'embryon puis le fœtus d'une grande partie des produits toxiques et des pathogènes auxquels la mère est exposée, elle n'est pas totalement étanche puisque c'est au travers du placenta que se font les échanges de substances entre la mère et l'embryon. En fonction de leur poids moléculaire et leur plus ou moins grande solubilité dans le sang, certaines substances toxiques peuvent traverser

la barrière et être tératogènes, c'est-à-dire causer des malformations chez l'embryon (retard de développement, retard mental, anomalies de formation des organes). C'est notamment le cas d'un empoisonnement au plomb qui diminue le QI et les capacités d'apprentissage des fœti et enfants exposés (Beattie *et al.* 1975; Bellinger *et al.* 1992; Koller *et al.* 2004). D'autre part, le syndrome d'alcoolisation fœtal causé par une consommation d'alcool par la mère pendant la grossesse concerne entre 0,6 et 3 naissances sur 1000 selon les communautés et est la cause la plus fréquente de DI liée à une exposition toxique (Lemay *et al.* 2003). Enfin, d'autres types d'expositions toxiques peuvent venir de maladies métaboliques comme par exemple une phénylcétonurie maternelle oubliée ou asymptomatique qui entraîne une augmentation toxique du taux de phénylalanine causant ainsi une DI accompagnée d'une microcéphalie et d'un retard de croissance intra-utérin.

### **ii. Les infections**

Tout comme les toxines, certains pathogènes peuvent passer la barrière placentaire et entraîner des anomalies du développement du cerveau. Ainsi, les infections intra-utérines à la rubéole, au cytomégalo virus et à la toxoplasmose et les infections post-natales de type méningite ou encéphalite entraînent des lésions du cerveau causant des DI. Cependant, de par le suivi médical désormais mis à disposition des futures mères et des nourrissons, l'ensemble de ces traumatismes infectieux ne comptent que pour une très faible proportion des cas de DI.

### **iii. Les traumatismes**

Pendant la grossesse mais surtout lors de l'accouchement, le fœtus peut être soumis à divers traumatismes entraînant une souffrance fœtale. Par la suite, le nourrisson peut également être victime de divers traumatismes physiques. Ensemble, ces événements empêchant le bon fonctionnement du cerveau seraient responsables de 5% à 10% des cas de DI. Citons par exemple la malnutrition (intra-utérine ou post-natale), l'anoxie cérébrale causée par exemple par un prolapsus du cordon ombilical lors de l'accouchement ou encore le syndrome du bébé secoué dont l'enfant peut être victime dans ses premiers mois de vie.

## 2. Cas des ASD

De par leurs conséquences irréversibles sur le système nerveux central en développement, l'ensemble des événements environnementaux responsables des DI "acquises" peuvent également être impliqués dans la survenue d'ASD (par exemple, les enfants souffrant du syndrome d'alcoolisation fœtale présentent dans certains cas des traits autistiques) (Aronson *et al.* 1997; Nanson, 1992; Stevens *et al.* 2013). Cependant, étant donné l'implication plus importante de l'environnement dans les ASD, d'autres facteurs environnementaux viennent s'ajouter à ceux de la DI. Ces événements sont des facteurs de risque qui, couplés à un environnement génétique particulier, entraînent l'apparition du phénotype autistique ou, tout au moins, augmentent son risque d'apparition (même si leur implication est parfois controversée). Comme le propose le modèle multifactoriel, il semble que l'interaction complexe entre gènes et environnement pendant des périodes critiques du développement du système nerveux central (la plus vulnérable étant entre la 3<sup>ème</sup> et la 6<sup>ème</sup> semaine de développement) (Selevan *et al.* 2000; Miller *et al.* 2005) soit très certainement à la base de la complexité de l'autisme. Il sera abordé ici à la fois les événements pré, péri et post-nataux (pour revue, voir (Landrigan, 2010; Ornoy *et al.* 2015; Sealey *et al.* 2016)).

### i. Les facteurs obstétricaux

De nombreuses études ont investigué l'association entre l'autisme et les facteurs obstétricaux (complications lors du travail ou de l'accouchement, facteurs néonataux, etc.). Les facteurs les plus fréquemment associés avec l'autisme sont : une souffrance néonatale, une naissance prématurée, un accouchement par césarienne et un faible score Apgar (Glasson *et al.* 2004; Hultman *et al.* 2002; Larsson *et al.* 2005). Les causes sous-jacentes de ces événements obstétricaux sont rarement connues ainsi que la relation temporelle entre ces facteurs obstétricaux et les manifestations biologiques de l'autisme à proprement parler.



## **ii. Les infections et les dysfonctionnements immunitaires**

Les voies reliant le système immunitaire et l'autisme sont toujours incertaines, notamment à cause de la faible disponibilité des données (nature, lieu et moment du processus non clairement définis). Néanmoins, quelques indices existent concernant un rôle des infections et des facteurs immunitaires pendant la période précoce et critique du neurodéveloppement. Ainsi, l'exposition prénatale à des agents viraux tels que le cytomégalo virus ou la rubéole ont été associés à l'autisme mais cette exposition virale précoce semble être une explication peu probable dans beaucoup de cas (responsable d'environ 0,3% des cas) (Fombonne *et al.* 2012; Libbey *et al.* 2005; Hutton, 2016). Une modeste association avec l'asthme et l'allergie maternelle pendant le seconde semestre de grossesse a été rapportée dans une étude (Croen *et al.* 2005). D'autres hypothèses concernant la contribution d'un traitement antibiotique (Fallon, 2005; Casavant, 2006) ou d'une vaccination rougeole, oreillons, rubéole chez les nouveau-nés ont été avancées mais plusieurs études ont contredit une telle association (Fombonne *et al.* 2006; DeStefano, 2007). Enfin, des liens entre auto-immunité (de la mère pendant la grossesse ou de l'enfant pendant ses premières années de vie) et autisme ont été avancés (Enstrom *et al.* 2009; Castellani *et al.* 2009; Goines & Van de Water, 2010; Chen *et al.* 2016). L'ensemble de ces résultats restent controversés et ne permettent pas encore de clarifier si un dysfonctionnement immunitaire pendant le neurodéveloppement précoce est un facteur de risque à l'autisme mais les déficiences immunitaires sont une des nombreuses caractéristiques des individus souffrant d'autisme.

## **iii. L'exposition à des xénobiotiques**

Nous sommes de plus en plus exposés à des substances xénobiotiques (c'est-à-dire des substances présentes dans un organisme vivant mais qui lui sont naturellement étrangères) ce qui est, par définition, non naturel. D'après certains scientifiques, l'augmentation du nombre de cas d'autisme observée ces dernières années serait due à une exposition de la population humaine à ces divers xénobiotiques (Sealey *et al.* 2016). Parmi ces substances chimiques (synthétiques ou non), on compte notamment les médicaments, les métaux lourds, les pesticides, l'alcool, les drogues, etc.



### a. Les médicaments

Trois médicaments ayant des propriétés tératogènes connues ont été identifiés comme potentiel facteur de risque de l'autisme : le thalidomide (prescrit durant les années 1950-1960 comme sédatif et anti-nauséeux) (Strömmland *et al.* 1994), le misoprostol (fréquemment employé depuis les années 1990 pour le déclenchement du travail lors de l'accouchement) (da Silva Dal Pizzol *et al.* 2006) et l'acide valproïque (antiépileptique également utilisé comme stabilisateur d'humeur dans les cas de schizophrénie et de troubles bipolaires) (Moore *et al.* 2000; Rasalam *et al.* 2005; Christensen *et al.* 2013). Ce dernier est le seul élément environnemental ayant été clairement mis en cause dans les ASD. En effet, de nombreuses études sur des modèles animaux (notamment des rats) démontrent sans ambiguïté la relation de cause à effet entre l'injection d'acide valproïque à des mères gestantes et l'apparition de traits autistiques chez les petits (pour revue voir (Roulet *et al.* 2013)).

### b. Les métaux lourds

Plusieurs métaux lourds comme le plomb, le mercure, et l'aluminium sont connus pour être des neurotoxines et ont des conséquences néfastes sur le neurodéveloppement en entraînant des anomalies du langage, de l'attention et de la mémoire (Mendola *et al.* 2002; Landrigan *et al.* 2004). Des études ont montrées une augmentation du risque d'autisme chez les enfants exposés à ces métaux lourds (Yassa, 2014; Mohamed *et al.* 2015) notamment utilisés comme adjuvants dans les vaccins (Shaw *et al.* 2014). Néanmoins, ces études sont très controversées et la dose à laquelle nous sommes exposés reste extrêmement faible.

### c. Les polluants atmosphériques

D'autres composés chimiques sont suspectés d'être impliqués dans l'étiologie de l'autisme comme c'est le cas des polluants atmosphériques. On y compte par exemple les métaux en suspensions dans l'air tels que le mercure, le cadmium, le nickel, le manganèse mais également les solvants, les particules de diesel (Windham *et al.* 2006; Roberts *et al.* 2013) et les pesticides (comme le glyphosate connu sous le nom commercial de Roundup<sup>®</sup> (Samsel & Seneff 2015) ou le chlorpyrifos-éthyl (de Cock *et al.* 2012; Levin *et al.* 2001)). Une étude a également montré que les solvants contenus dans les parfums et cosmétiques seraient hautement mutagènes, neurotoxiques et neuromodulateurs (Bagasra *et al.* 2013).

#### d. La nutrition et le mode de vie maternel

La nutrition maternelle (ainsi que son mode de vie) pendant la grossesse joue un rôle crucial pour le développement du fœtus. Ainsi, certaines suppléments en vitamines pendant la période périconceptionnelle ou pendant la grossesse (prise d'acide folique (vitamine B9) ou de vitamine D par exemple) sont fortement recommandées pour le bon déroulement de celles-ci. Si l'acide folique permet de réduire le risque de défaut de fermeture du tube neural de plus de 70% (Wald & Sneddon 1991) et la vitamine D est nécessaire pour une bonne croissance fœtale (Thorne-Lyman & Fawzi 2012), certaines études évoquent également une possible association entre une carence en ces vitamines et le risque d'autisme (Neggers, 2014; Surén *et al.* 2013; Cannell, 2008). D'autre part, il a été suggéré que la consommation de drogues illicites telles que la cocaïne augmente le risque d'autisme chez l'enfant à naître (Davis *et al.* 1992).

#### iv. Les hormones

Parmi les substances xénobiotiques précédemment abordées, la plupart sont des perturbateurs endocriniens, c'est-à-dire des molécules aux propriétés hormono-mimétiques affectant l'équilibre hormonal.

D'autre part, les hormones stéroïdiennes ont un rôle prépondérant dans le cerveau. Leur production devient significative à la puberté et permet notamment la sexualisation du cerveau (à savoir une féminisation par les œstrogènes et une masculinisation par les androgènes). Cependant, les androgènes sont également produits temporairement par les testicules récemment formées des fœti mâles et vont agir sur le cerveau pendant les phases critiques de son développement. Les fœti femelles quant à elles ne produisent pas encore d'œstrogènes (et ne produisent évidemment pas d'androgènes) et voient donc leur cerveau se développer en l'absence de toute hormone sexuelle. Etant donné l'excès de garçons observé dans la DI et les ASD, la présence d'androgènes pendant le développement du cerveau *in utero* est assimilé à un facteur de risque de ces pathologies, pouvant expliquer en partie le biais de sexe observé (Knickmeyer & Baron-Cohen 2006). Cet aspect sera développé en détail et fera l'objet d'un chapitre de cette introduction.

Pour résumer, l'ensemble des événements (toxiques, infectieux et traumatiques) responsables de DI « acquise » peuvent également être impliqués dans les ASD. De plus, étant donné l'implication plus importante de l'environnement dans les ASD, d'autres facteurs environnementaux viennent s'ajouter à ceux de la DI. Il existe ainsi un grand nombre de facteurs de risque de l'autisme liés à l'environnement mais l'impact de ces différents facteurs sur la pathogénie cellulaire et moléculaire de l'autisme n'a pas toujours été clairement évalué. Leur rôle dans l'étiologie des ASD reste encore très controversé mais il est clair qu'ils participent grandement à l'hétérogénéité phénotypique observée au sein des personnes autistes et font partie intégrante du modèle multifactoriel.

## **Partie 2. Hypothèses pouvant expliquer l'excès de garçons observé dans les NDD**

Beaucoup de maladies présentent un fort biais de sexe, l'un des sexes étant plus fréquemment affecté que l'autre. Dans certains cas, l'explication est évidente : quand la pathologie affecte un organe spécifique du sexe (cancer de la prostate ou de l'utérus), est due à des mutations récessives dans des gènes du chromosome X (hémophilie A et B, dystrophie musculaire de Duchenne, etc.) ou est associée à des différences environnementales spécifiques du sexe (augmentation du cancer des poumons chez les femmes liée à l'évolution de leur consommation de tabac). Cependant, pour d'autres pathologies, les mécanismes sous-jacents à ce biais de sexe ne sont pas encore clairement établis, ce qui est le cas pour la DI et les ASD.

Comme il l'est résumé dans deux revues (Baron-Cohen *et al.* 2011; Schaafsma & Pfaff 2014), plusieurs hypothèses majeures peuvent contribuer ensemble à cet excès de garçons : 1) le sous ou mauvais diagnostic des femmes; 2) la présence d'effets protecteurs chez les femmes, à la fois génétiques (tolérance aux évènements génétiques, nombre et inactivation du chromosome X, phénomène d'empreinte parentale) et environnementaux et 3) la présence de facteurs de susceptibilité chez les hommes dont par exemple la présence du chromosome Y (et l'implication de certains gènes comme *SRY*) ou le statut hormonal du fœtus masculin. L'ensemble de ces hypothèses seront détaillées dans cette partie mis à part la testostérone qui fera l'objet du second chapitre de cette introduction.

### **I. La sous-estimation des filles affectées**

L'existence de l'excès de garçons chez les enfants affectés par des retards cognitifs (notamment la DI) et qui sont de ce fait dans des institutions spécialisées est connue depuis le début du XX<sup>ème</sup> siècle (et confirmée par la suite par des études épidémiologiques). Cet excès a tout d'abord été attribué à un biais social puisque les femmes affectées auraient été plus souvent gardées à la maison plutôt qu'institutionnalisées. Cependant, cette explication ne serait plus d'actualité et, à partir des années 1960, la découverte de grandes familles ayant une ségrégation de la pathologie clairement liée au chromosome X a permis d'établir de nouvelles hypothèses pour expliquer ce biais de sexe dans la DI (ce qui sera abordé ultérieurement).

L'autre point concerne les ASD : les manifestations du phénotype de l'autisme sont-elles différentes en fonction du sexe ? Cette question a été posée à travers de nombreuses études mais seulement l'une d'entre elle prend en compte l'âge et le QI (Mandy *et al.* 2012). Ainsi, en ce qui concerne le syndrome d'Asperger (dont le QI moyen était de 92,6 dans l'étude), il semble que les filles aient moins de comportements répétés et stéréotypés et donc moins de difficultés à l'école. Elles s'adaptent également mieux à leur environnement et seraient capables de compenser leurs déficits sociaux en "faisant semblant" d'être normales (Dworzynski *et al.* 2012). Ces différences seraient également valables pour les autistes avec DI même si elles restent plus subtiles. Etant donné que les comportements répétés et stéréotypés sont essentiels pour certains cliniciens dans le diagnostic de l'autisme, les filles sont moins reconnues. De plus, leurs traits autistiques peuvent être éclipsés par d'autres symptômes atypiques tels qu'une anorexie mentale (Baron-Cohen *et al.* 2013; Rhind *et al.* 2014) ou des troubles de la personnalité borderline (également appelés hyperémotivité envahissante) (Trubanova *et al.* 2014). Ainsi, les filles autistes sont en moyenne diagnostiquées plus tard que leurs homologues masculins (Begeer *et al.* 2013; Giarelli *et al.* 2010). Ce sont donc peut-être les outils de diagnostic utilisés (notamment l'ADOS et l'ADI-R) qui ne sont pas adaptés pour détecter de façon subtile l'autisme chez les filles (notamment celles souffrant du syndrome d'Asperger).

## **II. Effets protecteurs chez les femmes**

Malgré les nombreuses études montrant un biais de diagnostic des filles souffrant d'autisme, les chercheurs s'accordent sur le fait qu'un réel excès de garçons est observé dans l'autisme mais aussi, de façon plus modeste, dans la DI (Fombonne *et al.* 2012; Leonard & Wen 2002). Les facteurs protecteurs viendraient à la fois des autosomes (notamment car le seuil de tolérance des femmes aux anomalies génétiques est plus élevé) mais aussi des gonosomes (car les femmes possèdent deux chromosomes X ce qui implique différents mécanismes pouvant entraîner une protection face à l'occurrence de diverses pathologies).

## 1. Seuil de tolérance aux évènements génétiques et pénétrance incomplète

Comme il vient de l'être évoqué, le sous-diagnostic des femmes souffrant d'ASD est, en partie, la conséquence de leur phénotype moins sévère. D'un point de vue génétique, les femmes auraient un seuil de tolérance plus élevé que les hommes, c'est-à-dire qu'elles seraient moins sensibles aux atteintes génétiques exprimant de manière moins marquée la pathologie. Ceci implique qu'il leur faudrait une plus grosse "charge" génétique, c'est-à-dire des anomalies génétiques plus dommageables, pour que la pathologie se manifeste. En effet, une étude a mis en évidence que les femmes autistes présentent trois fois plus de CNV autosomiques délétères ainsi qu'un excès de SNP délétères comparé aux hommes autistes. (Jacquemont *et al.* 2014). Ces évènements sont d'autant plus délétères qu'ils surviennent majoritairement *de novo* et les CNV sont également de plus grande taille (ils contiennent en moyenne 15,5 gènes chez les femmes contre deux chez les hommes) (Levy *et al.* 2011). Enfin, le risque de récurrence chez un couple ayant un enfant autiste est respectivement de 7% et 4% si leur premier enfant est une fille ou un garçon ce qui montre bien une charge génétique plus lourde pour les filles autistes (Ozonoff *et al.* 2011).

Une autre façon d'expliquer la nécessité d'une forte "charge" génétique pour entraîner des signes cliniques chez les femmes autistes serait de considérer la pénétrance de ces évènements génétiques. En effet, les études des CNV rares dans les ASD ont amené l'hypothèse qu'une grande majorité des cas d'autisme serait due à des anomalies génétiques *de novo* sur les autosomes et qu'elles auraient une pénétrance réduite chez les femmes. Une étude de risque d'autisme dans des familles multiplex (familles avec plusieurs membres atteints) soutient cette théorie puisqu'elle a mis en évidence que les évènements *de novo* ont une forte pénétrance chez les hommes et une relativement faible pénétrance chez les femmes (Zhao *et al.* 2007). Un exemple concret est celui d'une microdélétion hétérozygote dans *SHANK1* dont les hommes porteurs souffrent d'un autisme de haut niveau alors que les femmes porteuses présentent simplement des signes d'anxiété et de timidité (Sato *et al.* 2012). Néanmoins, une question reste entière : pourquoi la pénétrance est-elle différente selon le sexe ?

*In fine*, il semble que les femmes aient un seuil de tolérance aux atteintes génétiques plus élevé que les hommes et qu'il faille donc des événements génétiques plus nombreux et/ou plus nocifs pour voir se développer la pathologie. De plus, la pénétrance incomplète chez les femmes de certaines anomalies génétiques implique qu'elles présentent un phénotype moins sévère (voire même asymptomatique), permettant ainsi la transmission de ces anomalies génétiques à leur descendance. C'est ainsi que de nombreux cas d'autisme viennent de la transmission d'anomalies génétiques d'une mère asymptomatique à son fils qui, lui, présentera un ASD.

## **2. Avoir deux chromosomes X...**

L'un des facteurs protecteurs les plus flagrants des femmes réside dans leur génotype XX. En effet, le chromosome X joue un rôle majeur dans le développement, la structure et le fonctionnement du cerveau et contient, comparé aux autres chromosomes, un excès de gènes fortement exprimés dans le cerveau (Nguyen & Distèche 2006). Ainsi, il est apparu logique de constater que de nombreux cas de DI (et quelques cas d'ASD) sont causés par des anomalies dans des gènes portés par le chromosome X : le fait que les femmes possèdent deux chromosomes X pourrait être un facteur protecteur. Ainsi, chez les femmes, les mutations délétères sur le chromosome X peuvent être compensées par l'expression du second allèle intact, ce qui n'est pas le cas chez les hommes. La différence d'expression des gènes du chromosome X entre les individus XX et XY a également son importance et des mécanismes épigénétiques tels que l'inactivation du chromosome X (XCI) ou le phénomène d'empreinte peuvent aussi participer à l'effet protecteur des femmes face à la DI et aux ASD ce qui sera également discuté dans cette partie.

### **i. Effet de compensation par le second chromosome X**

La présence, chez les femmes, de deux gonosomes identiques les rend inévitablement moins sensibles aux anomalies génétiques présentes sur l'un d'entre eux. En effet, chez une femme, si un gène du chromosome X est muté entraînant une anomalie de son expression, la seconde copie intacte de ce gène présente sur le second chromosome X peut prendre le relais et permettre l'expression normale dudit gène et ce, malgré la mutation.

Cependant, ce système de compensation n'est pas possible chez les hommes et donc, si une mutation récessive est présente sur un gène du chromosome X, elle se comportera comme une mutation dominante et entraînera le développement de la pathologie associée. Comme il l'a été abordé dans la première partie de ce chapitre, plus de 100 gènes du chromosome X ont été impliqués dans la DI et moitié moins dans les ASD (Betancur, 2011), rendant l'hypothèse de compensation plausible pour chacun d'entre eux.

### a. Cas de la DI

Le syndrome de l'X Fragile est la première cause monogénique de DI. Sa fréquence a conduit à l'hypothèse selon laquelle les XLID (isolées ou syndromiques) pourraient expliquer le sex-ratio homme/femme de 1,3 à 1,4 présent dans la DI. Ceci suggère implicitement que jusqu'à 30 à 40% des hommes souffrant de DI auraient une mutation liée au chromosome X. L'énorme hétérogénéité des XLID avait empêché le test de cette hypothèse à l'époque de son émission puisque la centaine de gènes de XLID identifiée ne pouvait pas être testée rapidement en l'absence du séquençage haut débit. Cependant en 2004, la comparaison de l'apparition d'une mutation récurrente facilement testable dans le gène *ARX* dans des familles avec une claire ségrégation liée au chromosome X et dans des cas sporadiques de DI a permis de montrer que les XLID monogéniques peuvent expliquer tout au plus un tiers de l'excès de garçons, laissant les deux tiers restants inexpliqués (Mandel & Chelly 2004). Depuis, d'autres études vont dans le même sens et estiment que les XLID peuvent expliquer jusqu'à 10% des cas de DI chez les hommes ce qui est insuffisant pour expliquer entièrement le biais de sexe observé (Ropers & Hamel 2005; Lubs *et al.* 2012). En parallèle, de plus en plus de cas de XLID sont rapportées chez des femmes (exemple du gène *DDX3X* dont les mutations expliqueraient de 1 à 3% des DI inexpliquées chez les femmes) (Snijders Blok *et al.* 2015), ce qui implique d'autant moins les mutations causales des gènes du chromosomes X dans l'explication du biais de sexe observé.

### b. Cas des ASD

Les anomalies de structure et de nombre des chromosomes sexuels (syndromes de Turner 45,X0 ou de Klinefelter 47,XXY qui seront abordés dans la partie "dosage des gènes du chromosome X") sont des désordres retrouvés associés à l'autisme. Cependant, aucune région du chromosome X n'a été associée de façon significative aux ASD



(et ce malgré les nombreuses études de liaison et d'association pangénomiques). De plus, comme il l'a été abordé en détail dans la partie "génétique de l'autisme", les cas d'ASD monogéniques liés au chromosome X ne sont pas les plus courants. En effet, les syndromes associés à l'autisme et causés par des gènes du chromosome X tels que le syndrome de l'X Fragile ou le syndrome de Rett ne comptent que pour environ 4 à 5% des cas d'ASD. D'autres gènes du chromosome X ont été impliqués dans les ASD (*ARX*, *CDKL5*, *NLGN3*, *NLGN4X*, *PTCHD1*) mais mis à part *PTCHD1* dont la fréquence des anomalies (CNV *de novo* ou variants faux-sens rares) est estimée à 1% (Noor *et al.* 2010), les autres gènes ne seraient que très rarement impliqués (Chaste *et al.* 2007; White *et al.* 2010; Vincent *et al.* 2004; Blasi *et al.* 2006). Parallèlement, le séquençage systématique de 111 gènes du chromosome X codant pour des protéines synaptiques chez 142 patients souffrant d'autisme a permis de mettre en évidence un faible nombre de mutations validées (seulement deux, et chez des femmes...) et environ 5% de variants faux-sens potentiellement pathogéniques chez les hommes (Piton *et al.* 2011). Ainsi, l'ensemble de ses études indique que des anomalies dans des gènes du chromosome X comptent seulement pour une faible proportion des cas d'ASD et peuvent également être retrouvées chez les femmes ce qui ne permet pas d'expliquer le fort biais de sexe observé dans les ASD, en particulier le ratio homme/femme de 6 pour 1 dans l'autisme de haut niveau.

## ii. Biais d'inactivation du chromosome X

La présence de deux chromosomes X chez les femmes entraîne un problème de dosage génique. Ainsi, pour éviter que les gènes présents sur le chromosome X ne soient exprimés deux fois plus chez les femmes que chez les hommes, un système d'inactivation du chromosome X (XCI) a été mis en place au cours de l'évolution. Ce processus d'inactivation d'un des chromosomes X chez les femmes est aléatoire et consiste en une condensation du chromosome à inactiver médiée tout d'abord par l'expression du gène *XIST* puis maintenue *via* des mécanismes épigénétiques (méthylation de l'ADN ou méthylation, acétylation des histones) (Heard & Distèche 2006). Ce choix d'inactivation de l'un ou l'autre des chromosomes X par la cellule féminine persiste au cours des divisions subséquentes, de sorte que chez la femme adulte, certaines des cellules somatiques contiennent un chromosome X actif provenant du père, tandis que dans les autres cellules, il provient de la mère.

Les femmes sont ainsi mosaïques pour leur chromosome X ce qui laisse suggérer que les effets d'une anomalie génétique sur ce chromosome seraient moins nocifs pour une femme que pour un homme, entraînant un phénotype atténué (car toutes les cellules de son organisme n'exprimeront pas le gène muté).

Cependant, chez certaines femmes, le processus de XCI n'est pas aléatoire. C'est ainsi que lorsqu'il existe une anomalie sur l'un des chromosomes X, il y a une inactivation préférentielle de l'X remanié ou muté. Cette sélection est secondaire par rapport au développement embryonnaire : au début, l'inactivation est aléatoire, puis survient au cours du développement embryonnaire une sélection des populations cellulaires ayant inactivé le chromosome X anormal. Ce phénomène a déjà été identifié dans plusieurs pathologies (Gorski, 1991; Wengler *et al.* 1995; Devriendt *et al.* 1997). Une étude analysant le profil de XCI de mères d'enfants souffrant d'ASD ainsi que de filles autistes a mis en évidence deux mères et une fille possédant des mutations dans des gènes du X impliqués dans les ASD (*NLGN3*, *ATRX* et *MECP2*) et qui présentent un XCI biaisé, inactivant préférentiellement le chromosome porteur de la mutation (Gong *et al.* 2008). Ce cas de figure, s'il reste rare, est tout de même envisageable et fait du XCI un autre phénomène protecteur chez les femmes (en tous cas quand il s'agit de mutations délétères portées par le chromosome X).

### iii. Dosage des gènes du chromosome X

Le XCI est un mécanisme indispensable pour maintenir un dosage génique correct chez les femmes. Toutefois, il est à noter qu'environ 15% des gènes du chromosome X inactivé sont encore exprimés (Carrel & Willard 2005), le dosage de l'expression de ces gènes étant donc subtilement altéré chez les femmes (Berletch *et al.* 2011). De plus, il semble qu'il y ait parmi ces gènes un excès de gènes associés à la DI (22/114 dont notamment *PDHA1*, *HCF1*, *FLNA*, *PGK1*, *PQBPI*...) ce qui est plus qu'attendu par hasard, même en prenant en compte le grand nombre de gènes impliqués dans la DI sur le chromosome X (91/833 à l'époque) (Zhang *et al.* 2013). Cette différence du dosage génique entre hommes et femmes pourrait ainsi être impliquée dans le biais de sexe observé dans certaines pathologies (telles que la DI et les ASD) si les gènes non silencieux sont protecteurs.

Les cas les plus illustratifs d'anomalies du dosage des gènes du chromosome X sont les cas d'aneuploïdies du chromosome X (individus X0, XXY, XXX, XXYY, XXXY, XXXX, etc.). Si l'absence d'un chromosome X (syndrome de Turner : individus femelles 45,X0) est caractérisé par une petite taille et une insuffisance ovarienne, le phénotype d'un ou plusieurs chromosomes X surnuméraires est très hétérogène. Le seul point commun entre ces phénotypes "extra X" est la DI qui peut être plus ou moins sévère (van Rijn & Swaab 2015; Bishop *et al.* 2011; Tartaglia *et al.* 2010; Rومان *et al.* 2002). Par exemple, les individus mâles souffrant du syndrome de Klinefelter (caractérisé par au moins un chromosome X supplémentaire dans le caryotype masculin : XXY, XXYY, XXXY, XXXXY, etc.) présentent un phénotype d'autant plus sévère que le nombre de X supplémentaires est élevé : chaque X supplémentaire est associé à une baisse de QI d'environ 15 points, le langage étant le plus affecté. De plus, les individus "extra X" (notamment XXX, XXY et XXYY) présentent communément des traits autistiques et des problèmes de langage (van Rijn *et al.* 2008; Bruining *et al.* 2009; Bishop *et al.* 2011; Tartaglia *et al.* 2012). Néanmoins, la présence de DI et de traits autistiques chez les personnes "extra X" va à l'encontre des hypothèses précédemment décrites impliquant un effet protecteur du chromosome X. Cependant, en regardant l'incidence des traits autistiques dans différentes aneuploïdies, les études suggèrent que l'augmentation des risques de développer des traits autistiques ne serait pas un simple effet de dosage génique du chromosome X car ils apparaissent souvent dans un contexte de défaut d'apprentissage qui serait le réel facteur de risque chez ces personnes "extra X". Une régulation subtile de l'expression des gènes du chromosome X est donc nécessaire, toute anomalie pouvant entraîner des problèmes cognitifs, comme c'est le cas dans les aneuploïdies du chromosome X.

#### **iv. L'empreinte du chromosome X**

Le phénomène d'empreinte parentale a été mis en évidence dans les années 1980 par des chercheurs qui tentaient de créer artificiellement des embryons de souris issus de la fusion de deux génomes haploïdes mâles ou femelles. En voyant que les embryons possédant un génome uniquement masculin avortaient systématiquement alors que les embryons possédant un génome uniquement féminin se développaient à terme mais étaient plus petits que la normale, ils ont suggéré que les génomes maternels et paternels ne jouaient pas le même rôle au cours du développement. Ils proposent alors qu'il existe une empreinte spécifique du génome se produisant pendant la gamétogénèse et donc que la présence d'un pronucléus mâle

et femelle est essentielle au bon développement de l'embryon (Surani *et al.* 1984). Par la suite, de nombreux gènes soumis à empreinte parentale ont été mis en évidence, leur nombre s'élevant à une centaine chez l'Homme (référencés sur le site [www.geneimprint.com](http://www.geneimprint.com)). Un gène soumis à empreinte parentale ne voit pas sa copie maternelle exprimée de la même manière que sa copie paternelle. En général, l'une des deux copies du gène est totalement éteinte alors que l'autre est active. Ceci est dû à l'apposition de marques épigénétiques (méthylation de l'ADN ou méthylation, acétylation des histones) de façon "parentale spécifique", la copie maternelle n'ayant pas les mêmes marques que la copie paternelle ce qui induit cette différence d'expression génique en fonction de l'origine parentale de la copie. L'un des exemples les plus connus reliant l'empreinte parentale aux ASD concerne les syndromes de Prader-Willi/Angelman causés dans la majorité des cas par une délétion du locus critique 15p11.2-q13 soumis à empreinte parentale (le syndrome de Prader-Willi étant causé par la délétion paternelle et le syndrome d'Angelman par la délétion maternelle de ce locus).

Même si à l'heure actuelle les gènes soumis à empreinte parentale sont exclusivement retrouvés sur les autosomes, plusieurs loci soumis à empreinte ont été mis en évidence sur le chromosome X, le premier étant le Xp11.4-p21.1 qui serait impliqué dans la survie de l'embryon mâle (Naumova *et al.* 1998). En parallèle, Skuse *et al.* remarquent que parmi 80 femmes souffrant du syndrome de Turner, celles ayant hérité sur X paternel (Xp) présentent moins de difficultés sociales que celles ayant hérité du X maternel (Xm) (Skuse *et al.* 1997). Ils avancent ainsi l'hypothèse qu'un locus "sociocognitif" serait présent sur le chromosome X et soumis à empreinte maternelle : les gènes exprimés par le Xp pourraient agir comme un facteur protecteur face aux difficultés sociales dans le syndrome de Turner. Les femmes typiques héritent d'un chromosome X de chaque parent (XmXp) alors que les hommes typiques possèdent seulement le X transmis par leur mère (XmY) ce qui pourrait expliquer les différences sexuelles au niveau social, mais également, par extension, dans les ASD (Skuse *et al.* 1997; Skuse, 2000). De plus, toutes les patientes X0 également diagnostiquées pour des ASD avaient hérité du Xm (Donnelly *et al.* 2000; Skuse, 1999) et il a également été montré que l'origine maternelle du X surnuméraire chez les hommes souffrant du syndrome de Klinefelter était corrélée à des difficultés sociales et de communication plus importantes (Bruining *et al.* 2010), allant dans le sens de cette hypothèse. Néanmoins, il faut garder à l'esprit que 74% des femmes X0 héritent du Xm ce qui augmente les probabilités de retrouver les ASD associés au syndrome de Turner avec Xm, ce qui nuance ces résultats (Jacobs *et al.* 1997).

Pour résumer, le nombre de chromosome X peut jouer un rôle dans la protection des femmes face à la DI et aux ASD de différentes façons. La plus classique concerne les anomalies génétiques récessives délétères qui ne peuvent être compensées chez les hommes alors qu'elles peuvent l'être, au moins en partie, chez les femmes par le chromosome X possédant l'allèle normal du gène. En parallèle, les variations moins délétères (que ce soit les CNV, SNP ou les variants rares) pourraient avoir des conséquences moins importantes chez les femmes que chez les hommes (de par une pénétrance incomplète ou le XCI par exemple). Le dosage de l'expression des gènes du chromosome X ainsi que l'origine parentale de celui-ci seraient également des phénomènes importants. De plus, les mécanismes épigénétiques tels que le XCI ou l'empreinte parentale peuvent être influencés par l'environnement ce qui rend le tableau beaucoup plus complexe. *In fine*, même si de nombreux indices suggèrent que le chromosome X joue un rôle crucial dans la protection des femmes face à la DI et aux ASD, il ne peut pas expliquer à lui seul le biais de sexe observé dans ces pathologies (notamment pour les ASD dont le sex-ratio moyen est de 4 hommes atteints pour une femme).

### **III. Facteurs de susceptibilité chez l'homme**

En miroir des facteurs protecteurs présents chez les femmes, il semble que les hommes présentent des facteurs de susceptibilité à la DI et aux ASD ce qui les rendrait plus vulnérables. Il semblerait ainsi que les CNV *de novo* présents chez les hommes atteints d'autisme soient majoritairement des délétions (39 délétions *versus* 22 duplications). Etant donné que les délétions sont plus susceptibles d'être nocives que les duplications, ces observations révèlent une possible vulnérabilité inhérente au sexe masculin (Levy *et al.* 2011). En parallèle, l'ensemble des points exposés concernant la protection des femmes face à la DI et aux ASD peut également être interprété comme une plus forte susceptibilité des hommes à développer ces pathologies. En effet, le plus fort seuil de tolérance des femmes aux événements génétiques va de pair avec la plus forte sensibilité des hommes à ces mêmes événements. De plus, le fait de n'avoir qu'un chromosome X, qui plus est maternel, est l'une des explications à la plus forte vulnérabilité des hommes *via* les phénomènes de compensation, de XCI et d'empreinte déjà abordés. Néanmoins, il reste deux facteurs de susceptibilité masculins à évoquer, l'un concernant la présence du chromosome Y, le second portant sur le rôle des hormones masculines pendant le développement fœtal, points qui complètent ceux déjà discutés.

## 1. L'implication du chromosome Y

Outre le fait de n'avoir qu'un seul chromosome X, les hommes possèdent également un chromosome Y qui est absent chez les femmes. La présence même de ce chromosome Y semble être à elle seule un facteur de susceptibilité à l'autisme puisqu'il a été montré que les hommes "extra Y" (XYY et XXYY) ont un risque augmenté de présenter des NDD comme par exemple des ADHD, des ASD ou de la DI (Margari *et al.* 2014; Bardsley *et al.* 2013; Tartaglia *et al.* 2012). Les analyses d'haplotypes du chromosome Y sont cependant discordantes puisque l'une ne révèle aucun haplotype spécifique des autistes (quand comparés à une population contrôle) (Jamain *et al.* 2002) alors que la seconde montre une sous ou sur-représentation de certains haplotypes chez les autistes (Serajee & Mahbubul Huq 2009). Ces études ne permettent ainsi pas de conclure clairement quant au rôle du chromosome Y dans les ASD.

D'autre part, l'effet du chromosome Y peut passer par la présence de gènes de susceptibilité à l'autisme portés par ce chromosome, notamment sur la partie "mâle spécifique" (c'est-à-dire celle qui n'a pas son homologue sur le chromosome X). Cependant, l'impact potentiel des gènes du chromosome Y est assez restreint puisqu'il existe, chez les mammifères, un très faible nombre de gènes sur la partie "mâle spécifique" de ce chromosome (seulement 27 protéines différentes semblent être codées par les 78 gènes spécifiques du chromosome Y humain) (Skaletsky *et al.* 2003). De plus, la majorité de ces gènes sont exprimés uniquement dans les cellules germinales masculines ou dans les testicules. Néanmoins, quelques-uns sont exprimés dans le cerveau et sont donc d'intéressants candidats et pourraient être impliqués dans le biais de sexe observé dans la DI et les ASD. C'est notamment le cas de *SRY*, *NLGN4Y*, *PCDH11Y*, *TSPY* et *ZFY*.

### i. Rôle de *SRY* dans le cerveau et liens avec la DI et les ASD

Chez l'Homme, le gène *SRY* est l'acteur clé du système de détermination XY car il est à la base du sexe gonadique des individus (point qui sera expliqué en détail dans la partie suivante). Cependant, *SRY* est également exprimé dans le cerveau fœtal et adulte humain (Clépet *et al.* 1993; Mayer *et al.* 1998) ce qui pourrait l'impliquer dans les différences observées entre les cerveaux masculin et féminin. De plus, il a été montré que *SRY*, qui code pour un facteur de transcription, régulerait directement l'expression de la tyrosine hydroxylase (TH), l'enzyme dopaminergique clé

(Milsted *et al.* 2004) ce qui impacte notamment les comportements moteurs (expériences sur les mouvements volontaires chez des rats) (Dewing *et al.* 2006). Une autre étude réalisée sur des lignées de neuroblastomes humains révèle que *SRY* régulerait également d'autres gènes de la voie de biosynthèse et du métabolisme de la dopamine tels que *DDC*, *DBH* et *MAOA* (Czech *et al.* 2012; Wu *et al.* 2009b). Ainsi, de par la régulation de gènes importants du système dopaminergique, *SRY* pourrait contribuer au biais de sexe observé dans les pathologies impliquant la dopamine telles que la maladie de Parkinson, la schizophrénie ou les ADHD (qui présentent respectivement un sex-ratio garçon/fille de 1,5/1; 2,7/1 et 4/1) (Taylor *et al.* 2007; Bao & Swaab 2010). De plus, si l'association de *DDC* et *DBH* avec les ASD reste incertaine (Toma *et al.* 2013; Lauritsen *et al.* 2002; Yu *et al.* 2016; Yrigollen *et al.* 2008; Anderson *et al.* 2008), de nombreuses études impliquent *MAOA* dans la DI (Brunner *et al.* 1993) et les ASD (Tassone *et al.* 2011; Cohen *et al.* 2003; Cohen *et al.* 2011; Piton *et al.* 2014; Verma *et al.* 2014) amenant ainsi l'hypothèse d'un potentiel effet indirect de *SRY* via la régulation de *MAOA* dans l'étiologie de ces pathologies.

## ii. Autres gènes candidats

Le peu de gènes du chromosome Y codant pour des protéines sont majoritairement exprimés dans le système reproducteur mâle en ayant notamment un rôle primordial dans la spermatogénèse. Néanmoins, certains sont également exprimés dans le cerveau et sont des gènes candidats à la DI ainsi qu'aux ASD.

C'est le cas de *NLGN4Y*, homologue de *NLGN4X* codant pour une protéine d'adhésion synaptique dont des mutations ont été identifiées dans des familles affectées par la DI et/ou les ASD (Jamain *et al.* 2003; Laumonnier *et al.* 2004; Talebizadeh *et al.* 2006). Une seule mutation faux-sens dans *NLGN4Y* a été retrouvée chez un patient autiste et chez son père ayant des difficultés d'apprentissage (mais seulement chez un patient sur les 335 criblés) (Yan *et al.* 2008). Il semble ainsi que, si elles participent à l'étiologie de l'autisme, les mutations dans les gènes codant pour les protéines de la famille des NLGN (dont font partie les NLGN4X/Y) sont extrêmement rares. Cependant, une étude d'expression sur du sang de patients XYY montre que le niveau d'expression de *NLGN4Y* (qui est alors présent en deux copies) est corrélé avec les comportements autistiques présentés par ces patients, impliquant *NLGN4Y* dans l'augmentation du risque d'autisme chez les hommes XYY (Ross *et al.* 2015).



D'autres gènes du chromosome Y méritent également d'être cités. C'est le cas de *PCDH11Y* qui est exprimé dans le cerveau et qui code pour une autre protéine d'adhésion cellulaire impliquée dans la formation des synapses (Grant, 2003). Une étude d'association par approche gène candidat s'est cependant montrée non fructueuse, ne trouvant aucune différence entre les individus autistes et contrôles (Durand *et al.* 2006). Les gènes *ZFY* et *TSPY* quant à eux sont également exprimés dans le cerveau (Mayer *et al.* 1998). Tout comme son homologue *TSPX*, *TSPY* code pour une protéine qui interagit avec CASK, une protéine d'échafaudage synaptique (Kido *et al.* 2011) et dont des mutations ont été identifiées chez des patients souffrant d'ASD (Piton *et al.* 2011; Sanders *et al.* 2012). Cependant, l'implication de ces gènes du chromosome Y dans la DI et les ASD n'est que très peu probable car, à l'heure actuelle, aucun lien n'a été trouvé dans la littérature.

L'une des explications à l'excès de garçons observé dans la DI et les ASD implique l'existence de facteurs de susceptibilité spécifiques des hommes ce qui les rendrait plus sensibles à ces pathologies. De par sa présence uniquement chez les hommes, le chromosome Y et les gènes qui lui sont spécifiques sont des candidats logiques. Néanmoins, il existe très peu de preuves jusqu'à présent pour connaître réellement le poids de cette théorie du chromosome Y dans l'étiologie du biais de sexe présent au sein de la DI et des ASD.

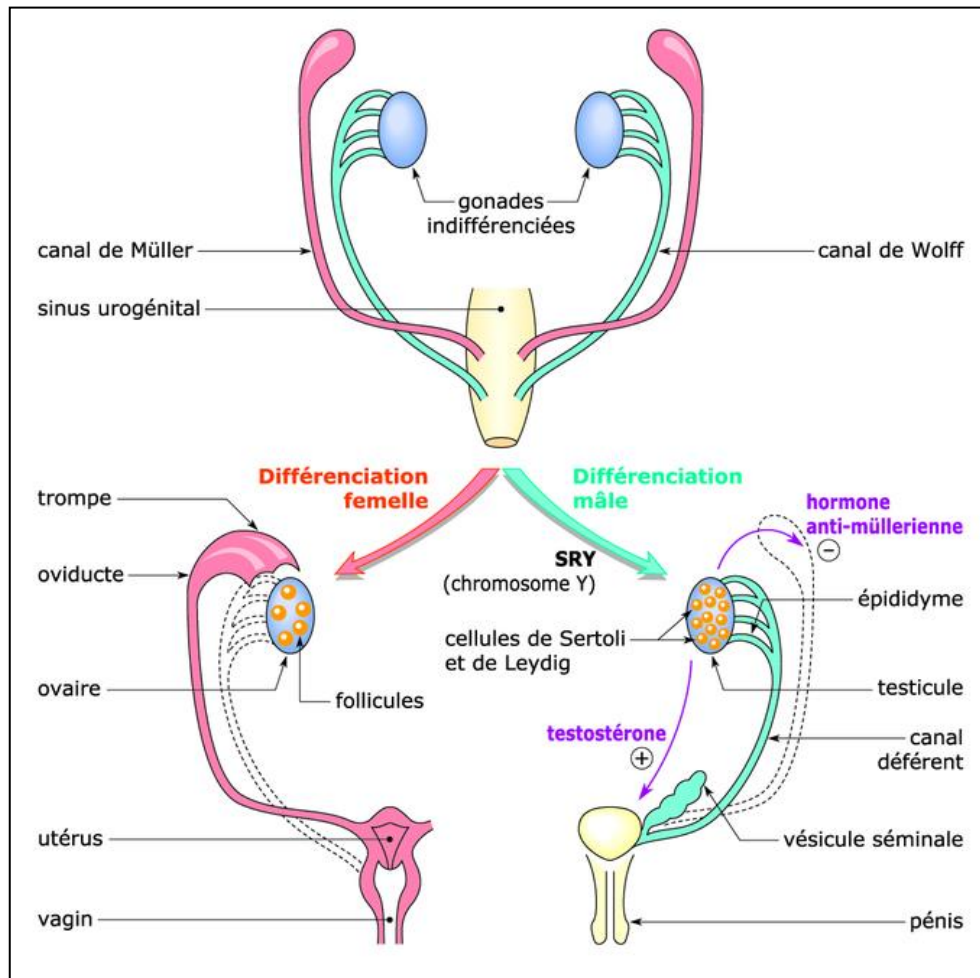
## **2. L'implication des hormones masculines**

La moindre perturbation de l'environnement fœtal pendant la grossesse peut avoir des conséquences dramatiques sur le développement du fœtus et ce même plusieurs années après la naissance. Alors que le biais de sexe observé dans les maladies à apparition tardive reflète des interactions complexes et de longue durée entre gènes et environnement (notamment le contexte hormonal), la situation peut être considérée comme moins complexe quand il s'agit des maladies à apparition précoce, comme c'est le cas pour les NDD. En effet, de par leur apparition pendant l'enfance, les effets des hormones sexuelles pubertaires peuvent être négligés dans le cas des NDD. Des rappels sur la détermination sexuelle ainsi que les mécanismes endocrinologiques impliquant notamment la production d'hormones (androgènes, hormone anti-müllérienne (AMH) et inhibine B) par les gonades masculines pendant la vie fœtale et néonatale et leurs liens avec l'autisme sont discutés ici.



### i. La détermination sexuelle

Chez l'Homme, l'expression du facteur de transcription codé par le gène *SRY* est nécessaire et suffisante pour initier la différenciation sexuelle mâle en dirigeant le développement des gonades indifférenciées de l'embryon en testicules (notamment en différenciant les cellules précurseurs en cellules de Sertoli) et ce entre la 6<sup>ème</sup> et 7<sup>ème</sup> semaine de gestation. Ce mécanisme implique l'induction de l'expression de *SOX9* par *SRY* menant à une cascade d'évènements incluant également l'expression du gène *FGF9*. Les cellules de Sertoli produisent alors l'hormone anti-müllérienne (AMH) qui permet la régression apoptotique des canaux de Müller et également la différenciation des cellules de Leydig autour de la 7<sup>ème</sup> semaine de gestation. Chez le fœtus fille, l'absence de *SRY* mène à la différenciation par défaut des gonades en ovaires et la non production d'AMH permet le maintien des canaux de Müller qui donneront le tractus génital femelle (trompes utérines et utérus). Chez le fœtus garçon, les cellules de Leydig nouvellement différenciées synthétisent alors l'hormone stéroïdienne masculine principale (la testostérone) qui permet le développement des canaux de Wolff en structures génitales internes mâles (épididyme et canal déférent; voir schéma récapitulatif **Figure 4**). Ainsi, le sexe gonadique mâle (présence de testicules) est déterminé par le sexe chromosomique (présence du chromosome Y, plus particulièrement de *SRY*) *via* la production d'hormones par les testicules en cours de maturation. Néanmoins, ces hormones testiculaires auraient un rôle bien plus vaste que la simple formation des organes génitaux mâles pendant le développement et sont également importantes plus tardivement dans la vie (comme il en sera notamment discuté pour les androgènes dans une partie ultérieure).



**Figure 4.** Schéma récapitulatif de la différenciation sexuelle chez l'Homme.

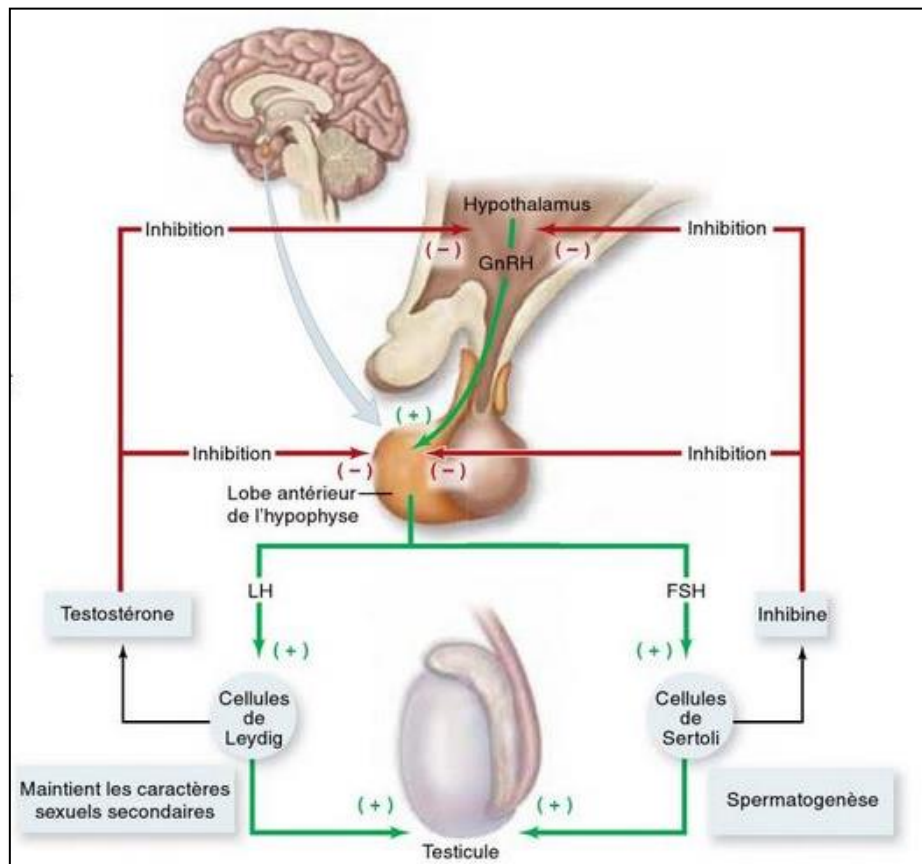
A partir d'une structure indifférenciée, l'expression du gène *SRY* chez les mâles entraîne : 1) la production de l'hormone anti-müllérienne (par les cellules de Sertoli) provoquant la disparition des canaux de Müller et, 2) la production de testostérone (par les cellules de Leydig) entraînant le développement des voies génitales mâles à partir des canaux de Wolff. Chez les femelles, l'évolution est spontanée, due à l'absence d'hormones mâles : les canaux de Müller persistent et ceux de Wolff disparaissent formant le tractus génital féminin (Anon, 2016).

## ii. Production des hormones testiculaires et rétrocontrôles

Le niveau d'expression des hormones testiculaires varie au cours de la vie d'un homme et, à l'âge adulte, leur production est soumise à une régulation fine faisant intervenir l'axe hypothalamo-hypophysaire siégeant à la base du cerveau (schéma récapitulatif **Figure 5**).

L'hypothalamus sécrète l'hormone de libération des gonadotrophines hypophysaires (GnRH) qui stimule la production de l'hormone folliculo-stimulante (FSH) et de l'hormone lutéinisante (LH) par l'adénohypophyse. Contrairement à chez la femme où les gonadotrophines sont produites de façon cyclique pour préparer périodiquement le système reproducteur à la fécondation et à la grossesse, la production chez l'homme est relativement constante (même si légèrement pulsatile car liée au cycle diurne-nocturne). La FSH exerce une action trophique sur les cellules de Sertoli et agit avec les androgènes pour maintenir la spermatogénèse. Elle stimule également la production d'inhibines par ces mêmes cellules qui exercent une rétro-inhibition sur la sécrétion de FSH par l'adénohypophyse. La LH quant à elle a une action trophique sur les cellules de Leydig et stimule la sécrétion de testostérone qui à son tour exerce une rétro-inhibition sur la sécrétion de LH. La testostérone régule donc la production de LH mais n'a aucun effet sur la FSH, ce sont les inhibines qui sont à l'origine du contrôle de la production de FSH.

Chez l'Homme, il existe deux inhibines (A et B), mais seule la B aurait ce rôle régulateur chez l'adulte. Outre son rôle dans le rétrocontrôle négatif de la sécrétion de FSH par l'hypophyse antérieure, elle favorise la production de testostérone par les cellules de Leydig et ralentit l'activité aromatasique permettant la transformation de la testostérone en œstradiol. Elle est ainsi très importante pour la spermatogénèse et le niveau plasmatique d'inhibine B évalue assez bien la fonction testiculaire, et en particulier, le fonctionnement des cellules de Sertoli (sa concentration est réduite dans bon nombre d'infertilités, bien plus que les concentrations de FSH ou de LH).

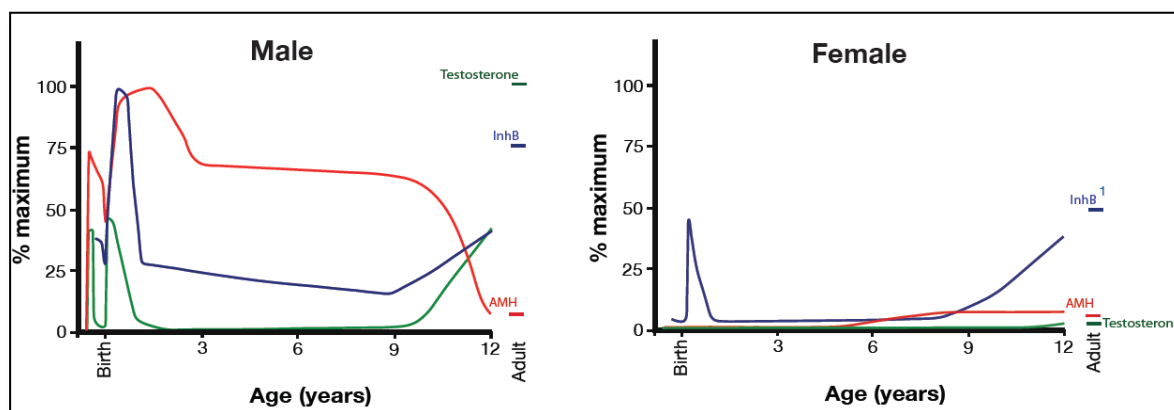


**Figure 5. Contrôle de la production d'hormones par les testicules.** L'hypothalamus sécrète la GnRH qui stimule la sécrétion de LH et de FSH par l'hypophyse antérieure. La LH stimule la sécrétion de testostérone par les cellules de Leydig. La testostérone est impliquée dans le développement et le maintien des caractères sexuels secondaires et stimule la spermatogénèse. La FSH stimule les cellules de Sertoli des tubules séminifères, ce qui facilite la spermatogénèse. La FSH stimule aussi la sécrétion d'inhibine par les cellules de Sertoli. La testostérone et l'inhibine, à leur tour, inhibent par un rétrocontrôle négatif les sécrétions de LH et de FSH (Johnson *et al.* 2011).

### iii. Liens entre l'AMH, l'inhibine B et les ASD

Comme vu précédemment, l'AMH joue un rôle très important pendant le développement en initiant la différenciation gonadique en testicules alors que l'inhibine B est importante à partir de l'adolescence pour la spermatogénèse (de Kretser *et al.* 2004) et le contrôle de l'activité testiculaire par la rétro-inhibition de l'hypophyse antérieure. Cependant, l'inhibine B est également produite sous forme d'un pic autour de la naissance et ce bien avant le début de la spermatogénèse, posant la question de son rôle à cette période.

Pendant le développement, l'AMH et l'inhibine B sont produites à un niveau bien supérieur de celui de l'âge adulte, l'AMH continuant même d'être sécrétée de façon très importante pendant les dix premières années de vie (**Figure 6**). Seulement, très peu d'études rapportent un lien entre ces hormones sertoliennes et le développement du cerveau.



**Figure 6.** Evolution du taux de production des différentes hormones masculines depuis le développement *in utero* jusqu'à l'âge adulte. InhB : inhibine B, AMH : hormone anti-müllérienne (Pankhurst & McLennan, 2012).

D'après des études réalisées chez la souris, l'AMH serait impliquée dans le dimorphisme sexuel de certaines zones du cerveau en contrôlant notamment le nombre de neurones et serait requise pour le développement social ainsi que l'identité sexuelle (Wang, Protheroe, *et al.* 2009; Wittmann & McLennan 2013a; Wittmann & McLennan 2013b). Une étude mesurant le niveau d'AMH et d'inhibine B chez des garçons souffrant d'ASD a révélé que le taux de ces hormones n'est pas altéré chez ces enfants (Pankhurst & McLennan 2012). Cependant, le niveau d'AMH est négativement corrélé avec la sévérité de leurs traits autistiques (évalué par le score ADI-R) et le taux d'inhibine B est, à l'inverse de l'AMH,

positivement corrélé avec leur score ADI-R (et ce de façon bien plus significative que l'est l'AMH). Il est à noter que les comportements répétés et stéréotypés (symptômes présentant pourtant un biais de sexe) (Rubenstein *et al.* 2015) ne sont pas influencés par ces deux hormones. A la vue de ces résultats, il semble que l'AMH et l'inhibine B ne soient pas à elles seules responsables de l'apparition des traits autistiques chez ces enfants (car leur taux n'est pas anormal) mais sont néanmoins de possibles modulateurs de leur phénotype dans un contexte génétique et environnemental particulier.

#### **iv. L'hormone testiculaire majeure : la testostérone**

La testostérone est produite par les cellules de Leydig très tôt pendant le développement du fœtus mâle, absente pendant l'enfance et de nouveau synthétisée à partir de l'adolescence. Elle joue de multiples rôles aussi bien pendant le développement qu'à l'âge adulte et ce non seulement sur le système reproducteur mâle mais également sur d'autres organes, dont le cerveau. De nombreuses études ont également montré qu'il existait une association entre la testostérone et les ASD. L'ensemble de ces points, de son action moléculaire à ses liens avec l'autisme, seront détaillés dans le chapitre suivant.

De nombreuses hypothèses tentent ainsi d'expliquer le biais de sexe observé dans la DI et les ASD. Même si elles n'ont pas toutes le même poids à la vue des données expérimentales, elles ne sont pas exclusives et pourraient toutes participer à la mise en place de cet excès de garçons dans ces pathologies. Toutes ces hypothèses s'accordent pour dire que les cas d'ASD chez les femmes sont le produit d'une charge génétique importante alors que les cas masculins nécessitent une plus faible charge génétique fortement influencée par une exposition environnementale. Tout ceci a été illustré par les différents exemples de facteurs protecteurs présents chez les femmes (pour la plupart génétiques et/ou épigénétiques : pénétrance incomplète, XCI, empreinte) ainsi que l'existence de facteurs de susceptibilité chez les hommes (ceux ayant le plus de poids étant environnementaux : hormones).



## **Chapitre 2. Les androgènes**

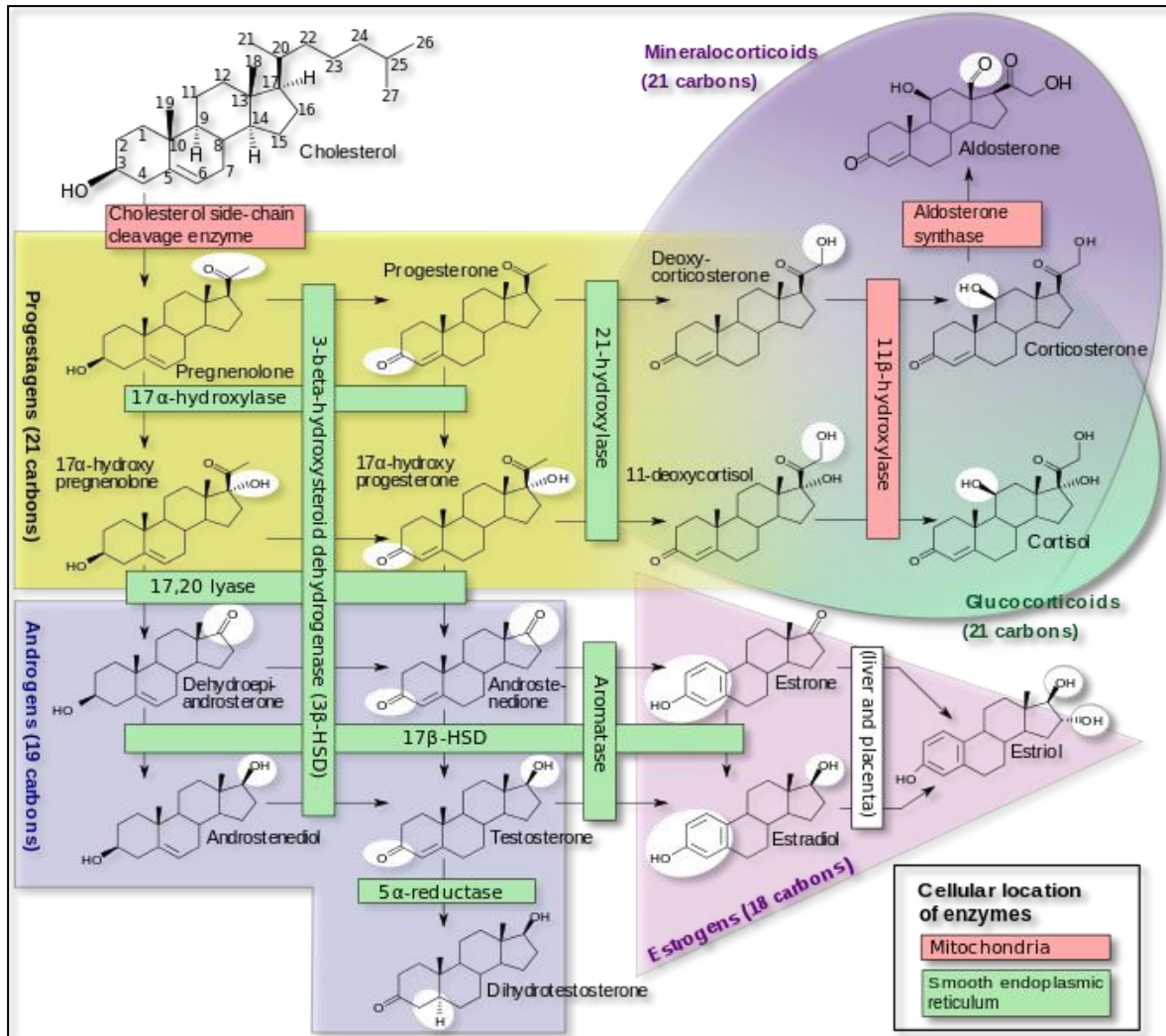
Pour revue voir (Tostain *et al.* 2004).

### **I. Biosynthèse, production, transport et métabolisme**

#### **1. Biosynthèse de la testostérone**

La testostérone est une hormone stéroïdienne de 19 carbones faisant partie du groupe des androgènes (voir **Figure 7**). Chez l'homme, elle est synthétisée à 95% dans les cellules de Leydig à partir du cholestérol (Schulze, 1984). Celui-ci est transformé en prégnénolone (dans la mitochondrie) qui est ensuite hydroxylée (par la 17- $\alpha$ -hydroxylase) puis clivée (par la 17,20-lyase) en déhydroépiandrostérone (DHEA) qui peut alors être convertie en testostérone (dans le réticulum endoplasmique). Le cortex surrénalien synthétise les 5% restants de testostérone à partir de l'androstènedione mais également d'autres hormones stéroïdiennes, les glucocorticoïdes et les minéralocorticoïdes grâce à l'expression des enzymes 21 $\alpha$ - et 11 $\beta$ -hydroxylases. Les cellules de Leydig quant à elles sont dépourvues de ces enzymes ce qui leur permet de synthétiser uniquement les androgènes. Il est à noter que la testostérone est également produite par les femmes, l'unique source étant le cortex surrénalien. Chez les femmes, la testostérone fabriquée par le cortex surrénalien sera transformée en œstrogènes grâce à l'action de l'enzyme aromatasase dans la graisse corporelle. Etant donné que la production de testostérone par les femmes est négligeable (10 fois plus faible que chez l'homme) et qu'elle exerce son action *via* sa conversion en œstrogènes, la suite de ce chapitre concernera principalement la testostérone chez les hommes.



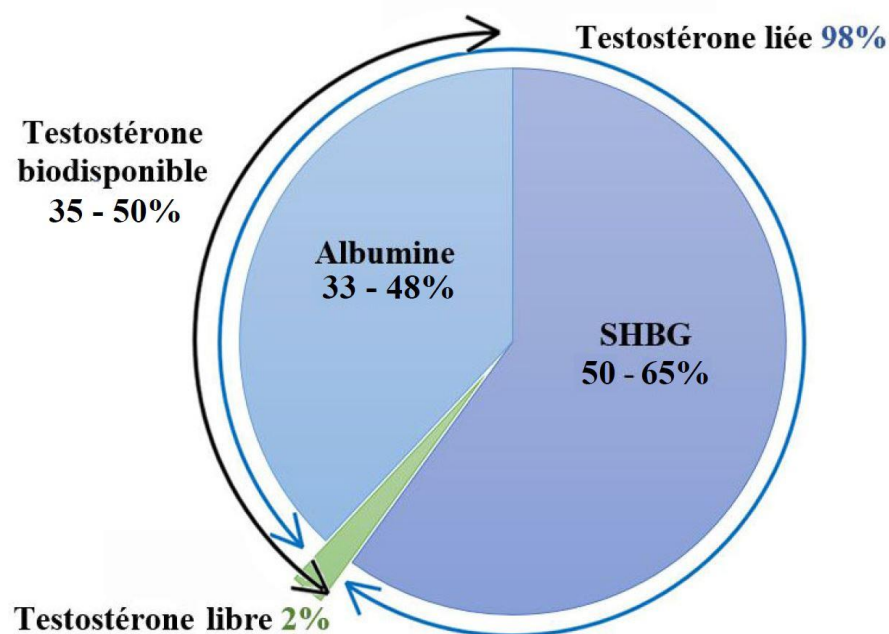


**Figure 7. Biosynthèse des stéroïdes.**

## 2. Transport et métabolisme

La production de testostérone par un homme adulte sain est estimée entre 5 et 7,5mg par jour (Vierhapper *et al.* 2000). Le contenu en testostérone du testicule de l'homme adulte est seulement de 50 $\mu$ g, ce qui montre que la quasi-totalité de la testostérone produite est libérée dans la circulation (Bagatell & Bremner 2003). Elle diffuse librement à travers la membrane des cellules de Leydig et sa concentration s'équilibre rapidement dans les différents compartiments testiculaires puis, la plus grande partie de la testostérone contenue dans le liquide interstitiel diffuse alors dans le sang veineux. Dans le plasma, 98% de la testostérone est liée aux protéines (l'albumine et les  $\beta$ -globulines liant les hormones stéroïdiennes ou SHBG), les 2% restants représentant la testostérone libre.

La testostérone liée à la SHBG, à l'inverse de la fraction liée à l'albumine, se dissocie difficilement et n'est pas disponible pour l'utilisation tissulaire. La testostérone considérée comme biologiquement active car disponible pour être assimilée par les tissus représente ainsi la somme de la testostérone libre et de la testostérone liée à l'albumine, soit près de 50% de la testostérone totale (**Figure 8**). Ces protéines de liaison servent de réserve de testostérone (qui serait sinon rapidement métabolisée par le foie) et permettent également d'amortir les variations du taux de testostérone afin de garder ce taux constant (Nieschlag *et al.* 2012). Il faut savoir que la SHBG peut également lier ses métabolites de la testostérone de façon plus ou moins forte (Dunn *et al.* 1981).



**Figure 8.** Répartition de la testostérone plasmatique.

Chez l'homme, une fois libérée dans le plasma, une petite partie des androgènes est aromatisée en œstrogènes (principalement dans le tissu graisseux mais aussi en quantité moins importante dans la peau, le rein, l'os et le cerveau) (Winters & Dalkin 2004). Respectivement 80% et 95% de l'œstradiol et de l'œstrone plasmatique retrouvée chez l'homme adulte en sont issus, le reste provenant directement des testicules (principalement produites par les cellules de Sertoli) (Dorrington & Armstrong 1975). Le niveau plasmatique de l'œstradiol est de 73 à 184pmol/L pour une production totale d'environ 184nmol/jour.

De plus, la testostérone est convertie dans certaines cellules cibles en dihydrotestostérone (DHT ou androstanolone) par les 5- $\alpha$ -réductases (au nombre de trois). La réductase de type 3 est la plus exprimée dans la plupart des tissus, notamment la prostate, les autres tissus génitaux et le cerveau (Yamana *et al.* 2010). La concentration plasmatique de la DHT est d'environ 10% de celle de la testostérone et est très difficilement détectable (Wudy *et al.* 1999). La DHT ayant un effet plus puissant que la testostérone (affinité pour leur récepteur augmentée), la transformation de celle-ci en DHT est donc une façon d'amplifier l'action de la testostérone dans les tissus cibles.

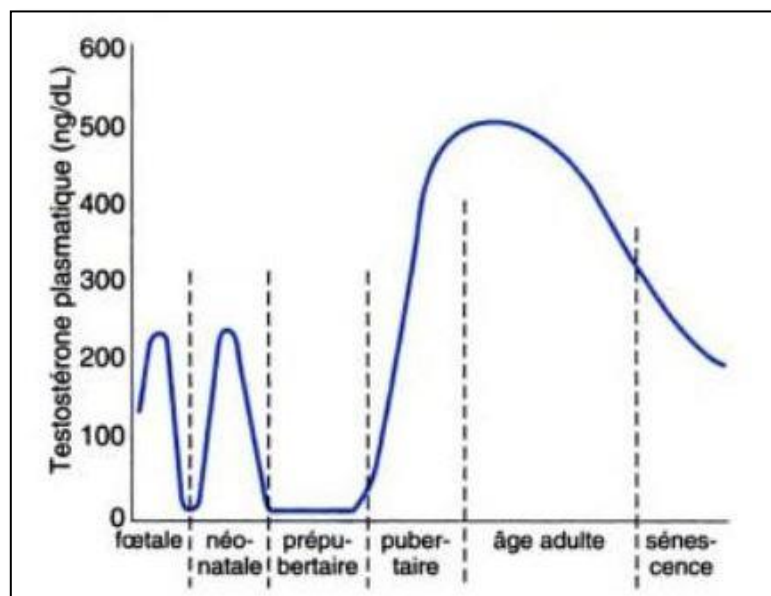
En résumé, la testostérone peut avoir des effets directs (en agissant sur son récepteur, le récepteur aux androgènes, AR) mais également indirects par le biais des métabolites actifs qui augmentent et diversifient ses effets biologiques. Elle joue alors le rôle de pro-hormone et peut être métabolisée en DHT qui activera elle aussi l'AR, ou en œstrogènes qui activeront les récepteurs aux œstrogènes (ESR1 et ESR2). L'action globale de la testostérone reflète ainsi la réponse intégrée de chaque tissu à la testostérone, la DHT et aux œstrogènes.

Enfin, il faut noter que la testostérone peut finalement être métabolisée (notamment par le foie) en 17-cétostéroïdes (surtout en androstérone). Ces 17-cétostéroïdes sont des androgènes faibles dont l'affinité pour l'AR ne dépasse pas 20% de celle de la testostérone. Ces composés sont ensuite éliminés avec la testostérone dans l'urine et la bile. De cet équilibre entre système de synthèse et systèmes de transformation et de dégradation dépend la stabilité du taux de testostérone. Cet ensemble métabolique est très actif puisque la demi-vie de la testostérone est de 12 minutes, ce qui implique une fabrication permanente par le testicule.

### **3. Production et taux de testostérone au cours de la vie**

Les androgènes sont produits par les gonades (testicules et ovaires) ainsi que le cortex des glandes surrénales. La testostérone est le représentant le plus connu des androgènes. Nous avons vu que, pendant le développement du fœtus mâle, la testostérone est produite par les cellules de Leydig nouvellement différenciées sous forme de deux vagues principales : la première pendant le second trimestre de la grossesse mène à la différenciation sexuelle mâle (comme évoqué à la fin du chapitre précédent) alors que la seconde a lieu dans les trois mois après la naissance (le rôle de ce pic de production est encore mal connu).

La première vague est due à une augmentation très importante du nombre et de l'activité des cellules de Leydig par l'hormone chorionique gonadotrope humaine maternelle (hCG) (Müller & Skakkebaek 1992). La seconde vague quant à elle est due à l'action des gonadotrophines produites par l'hypophyse antérieure (notamment la LH), amenant le taux de testostérone proche de celui observé durant la puberté. Après la période néonatale, les cellules de Leydig régressent et entrent en repos, les testicules restant inactives et la production de testostérone proche de zéro jusqu'à la période de maturation finale du système reproducteur, l'adolescence. C'est alors de nouveau l'augmentation du taux de LH qui entraîne une nouvelle phase de prolifération et de différenciation des cellules de Leydig amenant leur nombre à celui de l'âge adulte (Saez, 1994). Enfin, la diminution de la testostérone liée à l'âge (notamment causée par la diminution de la sécrétion de LH) entraîne une atrophie progressive puis la disparition des cellules de Leydig accompagnée du déclin des capacités de stéroïdogénèse (Paniagua *et al.* 1991). Le profil de production de la testostérone au cours de la vie d'un homme est illustré par la **Figure 9**.



**Figure 9.** Niveau de testostérone plasmatique à différentes périodes de la vie chez les sujets mâles. On observe deux pics de production précoce de testostérone, l'un *in utero* et le second à la naissance. La testostérone est ensuite produite à un taux maximum à partir de l'adolescence puis son taux diminue chez les hommes sénescents (Ganong, 2005).

Le taux de sécrétion de la testostérone chez un adulte mâle normal est de 14 à 31 $\mu$ mol/jour. Même si le dosage de la fraction libre est accessible *via* la salive (Ostatníková *et al.* 2002), très peu d'études ont été réalisées et les données disponibles dans la littérature concernent principalement le dosage de testostérone totale (libre et liée). Ainsi, le niveau plasmatique de testostérone totale est d'environ 10 à 35nmol/L chez l'homme adulte contre 0,5 à 3nmol/L chez la femme. Plus tôt dans la vie, les deux vagues de production de la testostérone sont retrouvées dans le plasma chez les fœti (dosée dans le liquide amniotique) et les nouveau-nés mâles (dosée dans le sang de cordon) mais absentes chez les femelles. La DHT plasmatique quant à elle est quasiment indétectable avec les techniques utilisées (1nmol/L pour les femmes adultes contre 1,7nmol/L pour les hommes), avant tout puisqu'elle est majoritairement convertie au sein des cellules cibles. Les niveaux de testostérone dans le cerveau proviennent de mesures *post-mortem* d'adultes sénescents réalisées lors d'études sur le rôle protecteur des hormones dans le vieillissement du cerveau et ne sont donc pas connus pour les fœti, enfants et adolescents (voir **Table 2**).

Période	Plasma		Cerveau
	Homme	Femme	Homme
<b>Fœtale</b> (8 - 24 semaines)	Pic à ~ 10 nmol/L	< 0,5 nmol/L	nd.
<b>Néonatale</b> (0 - 6 mois)	Pic entre 2,5 - 8 nmol/L	< 0,5 nmol/L	nd.
<b>Enfance</b> (1 - 10 ans)	< 0,7 nmol/L	< 0,5 nmol/L	nd.
<b>Puberté</b> (10 - 20 ans)	1,7 - 17,5 nmol/L	0,5 - 3 nmol/L	nd.
<b>Age adulte</b> (20 - 60 ans)	10 - 35 nmol/L	0,5 - 3 nmol/L	~ 2 ng/g
<b>Sénescence</b> (> 75 ans)	0,4 - 9 nmol/L	0,3 - 1,7 nmol/L	0,25 ng/g

**Table 2. Dosages de la testostérone plasmatique et cérébrale à différentes périodes de la vie chez les sujets mâles et femelles.**

Synthèse de (Wudy *et al.* 1995; Wudy *et al.* 1999; Rosario *et al.* 2004) (nd. : non disponible).

#### 4. Cas des neurostéroïdes

Il a été suggéré depuis les années 1980 que le cerveau pouvait également être considéré comme un organe stéroïdogénique (pour revue, lire (Compagnone & Mellon 2000)). En effet, la plupart des enzymes présentes dans les organes stéroïdogéniques plus classiques (gonades et glandes surrénales) ont été trouvées dans le cerveau aussi bien par mesure de leur niveau d'ARNm, de protéines que par leur activité enzymatique. Le cerveau peut ainsi synthétiser *de novo* des stéroïdes, appelés neurostéroïdes pour les différencier des stéroïdes produits par des organes stéroïdogéniques classiques. Les neurostéroïdes incluent également les stéroïdes métabolisés par le cerveau à partir de précurseurs circulants. De plus, le cerveau possède des enzymes modificatrices de stéroïdes qui peuvent les convertir en une variété de composés neuroactifs.

Les neurostéroïdes principalement retrouvés dans le cerveau sont la prégnénone, la progestérone et la DHEA. Cependant, les stéroïdes plasmatiques libres passent très facilement la barrière hémato-encéphalique et peuvent également être retrouvés dans le cerveau. Il est alors difficile de clairement distinguer les stéroïdes produits localement de ceux qui y ont été transportés depuis la circulation sanguine.

L'expression des enzymes stéroïdogéniques et donc la production de ces neurostéroïdes est régulée au cours de la vie. Cette régulation n'est pas seulement "région spécifique" mais également dépendante du type cellulaire rendant leur action beaucoup plus complexe. Ils exercent ainsi plusieurs actions biologiques pendant l'embryogénèse mais également chez l'adulte. Ces actions peuvent passer par une régulation génomique *via* les récepteurs hormonaux nucléaires mais également par une régulation non-génomique par l'intermédiaire de voies de signalisation intracellulaires. Les neurostéroïdes sont ainsi impliqués dans la myélinisation, la croissance des neurites (importante pour l'organisation du cerveau en développement) et exercent un effet neurotrophique et neuroprotecteur (qui sont impliqués dans les neuropathies chez les sujets d'âge avancé). De plus, l'activité de plusieurs récepteurs de neurotransmetteurs (notamment les récepteurs GABA de type A ou GABA<sub>A</sub> et les récepteurs glutamate de type N-méthyl-D-aspartique ou NMDA) peuvent être modulées respectivement par la prégnénone et la DHEA. Enfin, les neurostéroïdes possèdent des effets observables plus largement au niveau comportemental en ayant des propriétés anxiolytiques, antidépresseurs, anesthésiques, anticonvulsivants ou en influençant les phases du sommeil et les capacités mnésiques.

## II. Le récepteur aux androgènes (AR)

Les stéroïdes libres comme la testostérone peuvent facilement diffuser à travers la membrane plasmique des cellules. Elle exerce ensuite son action cellulaire *via* son récepteur spécifique, le récepteur aux androgènes (AR).

### 1. Généralités

#### i. Classification

L'AR est un facteur de transcription dépendant d'un ligand (ici les androgènes comme la testostérone ou la DHT) appartenant à la famille des récepteurs aux hormones stéroïdiennes également appelée NR3 (pour Nuclear Receptor Subfamily 3). Cette famille comprend différents groupes :

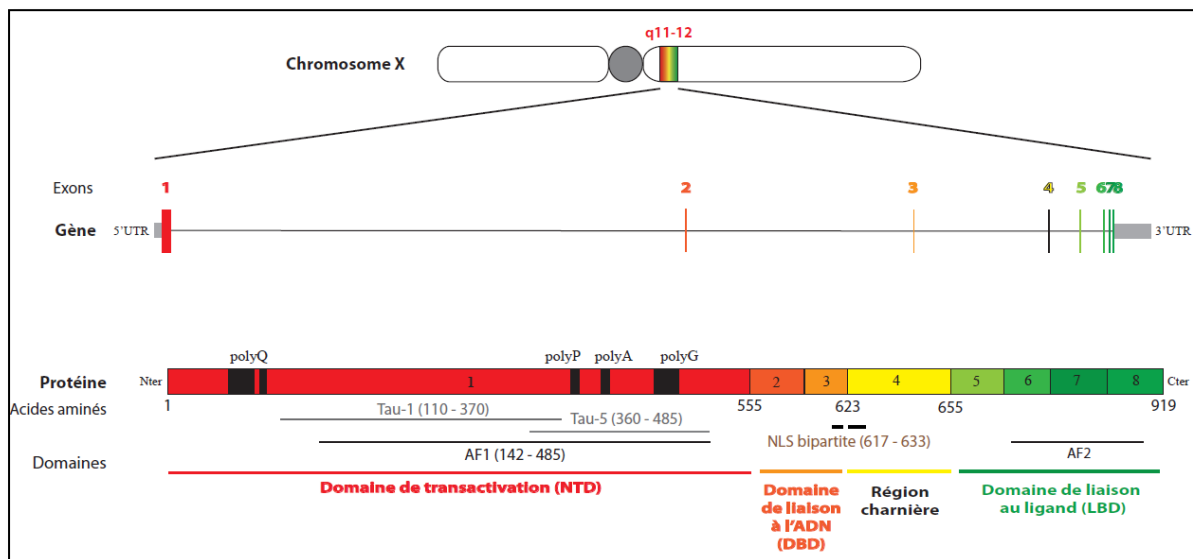
- **Groupe A : Les récepteurs aux œstrogènes (ER ou ESR)** comprenant :
  - Le récepteur  $\alpha$  (ER $\alpha$  ou ESR1 ou NR3A1) et
  - Le récepteur  $\beta$  (ER $\beta$  ou ESR2 ou NR3A2).
- **Groupe B : Les récepteurs liés aux récepteurs aux œstrogènes (ERR)** comprenant les récepteurs  $\alpha$  (ERR $\alpha$  ou NR3B1),  $\beta$  (ERR $\beta$  ou NR3B2) et  $\gamma$  (ERR $\gamma$  ou NR3B3).
- **Groupe C : Les récepteurs aux cétostéroïdes** regroupant :
  - Le récepteur aux glucocorticoïdes (GR ou NR3C1),
  - Le récepteur aux minéralocorticoïdes (MR ou NR3C2),
  - Le récepteur à la progestérone (PGR ou NR3C3) et
  - Le récepteur aux androgènes (AR ou NR3C4).

Le gène du récepteur aux androgènes est localisé sur le chromosome X, plus précisément au locus Xq11-Xq12 (Lubahn, Joseph, Sullivan, *et al.* 1988; Brown *et al.* 1989). Le transcrit qui en résulte possède 2757 nucléotides comprenant 8 exons et des introns variant de 0,7 à 2,6 kb. La protéine résultante est composée de 919 acides aminés pour une masse moléculaire de 110 kDa (van Laar *et al.* 1989).



## ii. Organisation de l'AR

Des analyses comparatives de l'organisation structurale des récepteurs NR3 ont montré une forte homologie, ces récepteurs comprenant trois domaines fonctionnels principaux : le domaine de transactivation NH2 terminal (NTD, résidus 1 à 554) suivi du domaine de liaison à l'ADN (DBD, résidus 555 à 622) et du domaine de liaison du ligand COOH terminal (LBD, résidus 665 à 919) connecté au DBD par une région charnière (résidus 623 à 664; **Figure 10**). La structure de l'AR a été décrite pour la première fois par (Brinkmann *et al.* 1989). Jusqu'à maintenant, aucune structure tridimensionnelle de l'AR complet n'a été résolue. Il existe néanmoins des structures partielles, notamment du DBD et du LBD, qui ont permis d'obtenir des informations cruciales sur son fonctionnement.



**Figure 10. Structure du récepteur aux androgènes humain.** Le gène codant pour le récepteur aux androgènes se situe sur le chromosome X, en région q11-12. Il est constitué de huit exons qui codent pour une protéine de 919 acides aminés d'un poids moléculaire de 110 kDa. Le domaine de transactivation (NTD) comprend deux unités activatrices de la transcription (Tau-1 et Tau-5) ainsi que des répétitions de glutamines (polyQ), proline (polyP), alanines (polyA) et glycines (polyG). Le domaine de liaison à l'ADN est formé de deux "doigts de zinc", la région charnière comprend un signal de localisation nucléaire bipartite (NLS) et le domaine de liaison au ligand contient un domaine d'activation AF2.



### a. Le domaine de transactivation NH2 terminal (NTD)

Cet important domaine de l'AR codé par l'exon 1 représente plus de la moitié de la protéine. Sa structure modulaire peut former une interface sur laquelle d'autres protéines, notamment des facteurs nucléaires, peuvent se fixer : il intervient ainsi de façon prédominante dans la régulation de la transcription. Même si la totalité du NTD est requis pour une action complète du récepteur (Simental *et al.* 1991), le domaine d'activation AF1 possède à lui seul un pouvoir activateur de la transcription particulièrement marqué. Plus précisément, l'AF1 contient deux unités d'activation de la transcription distinctes, Tau-1 (qui contient un motif de type FxxLF (FQNLF)) et Tau-5 (qui contient un motif WHTLF). Tau-5 est constitutivement active alors que Tau-1 nécessite la liaison du ligand sur le LBD (Doesburg *et al.* 1997).

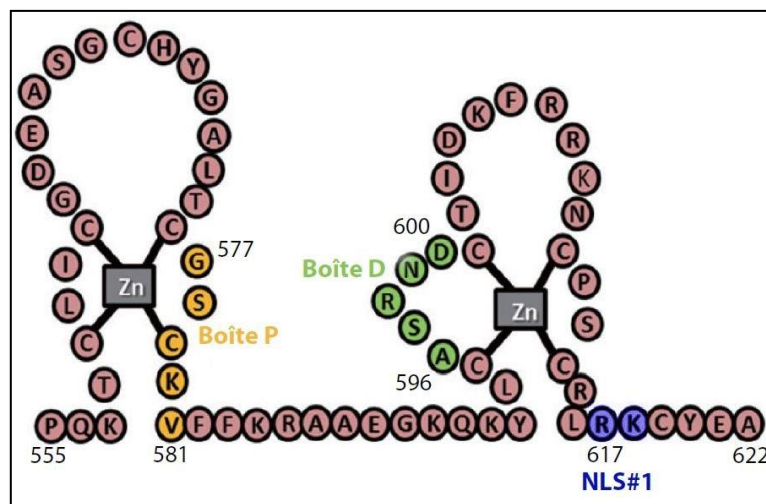
Ce domaine présente également les répétitions polymorphiques de l'AR (une répétition de glutamines (polyQ codée par (CAG)<sub>n</sub>CAA) (Zhang *et al.* 1994) et une répétition de glycines (polyG codée par (GGN)<sub>n</sub>) ainsi que des répétitions d'alanines (polyA) et de prolines (polyP). Dans la population générale, la taille du polyQ varie de 9 à 38 répétitions (avec des variations interethniques) (Choong & Wilson 1998; Edwards *et al.* 1992; Hsing *et al.* 2000) alors que le polyG est compris entre 16 et 27 résidus. La longueur des polyP et polyQ influence la fonction activatrice de la transcription. En effet, la longueur des répétitions est inversement proportionnelle à l'activation de transcription du gène cible (Beilin *et al.* 2000), probablement en jouant sur l'affinité différentielle des protéines co-activatrices de l'AR (Hsiao *et al.* 1999). Le polymorphisme du polyQ de l'AR est également associé à certaines maladies.

Enfin, le NTD pourrait réguler la localisation subcellulaire de l'AR en promouvant la localisation cytoplasmique grâce à une région de 200 acides aminés (acides aminés 50 à 250 du NTD) (Dar *et al.* 2014).

### b. Le domaine de liaison à l'ADN (DBD)

Le DBD est le domaine central de l'AR permettant la fixation du récepteur sur ses séquences-cibles de l'ADN présentes au sein des promoteurs et enhancers des gènes régulés par les androgènes. Il est codé par les exons 2 et 3 et est extrêmement conservé entre les différents récepteurs aux hormones stéroïdiennes (Freedman *et al.* 1988).

Ce domaine est composé de deux doigts de zinc c'est-à-dire deux anses protéiques maintenues à leur base par quatre résidus cystéine liés à un ion zinc (**Figure 11**). Le premier doigt permet la reconnaissance de la séquence de l'ADN cible, les trois acides aminés positionnés à sa base (boite proximale, P) étant particulièrement critiques car entrant directement en contact avec le grand sillon de l'ADN (Berg, 1989; Truss & Beato, 1993). Le second doigt possède une boite D (distale) très basique, favorise la dimérisation de l'AR et stabilise l'interaction ADN-récepteur. Les détails de l'interaction AR-ADN ainsi que les séquences reconnues seront vus dans une partie ultérieure.



**Figure 11. Structure du domaine de liaison à l'ADN de l'AR.**

Le domaine est organisé autour de deux doigts de zinc ( $Zn^{2+}$ ). La boite P (proximale) permet la reconnaissance et la fixation aux séquences d'ADN cibles et la boite D (distale) favorise la dimérisation de l'AR et stabilise l'interaction ADN-récepteur. Adaptée de (Tan *et al.* 2015)

### c. Le domaine de liaison du ligand COOH terminal (LBD)

Ce domaine carboxyl-terminal est codé par les exons 5 à 8 du gène *AR* et est très conservé au sein de la famille des récepteurs aux stéroïdes. Sa fonction principale est de lier les ligands spécifiques du récepteur (les androgènes) au sein d'une poche. La très haute affinité spécifique de liaison avec les androgènes est favorisée par les protéines chaperonnes de choc thermique HSP70 et HSP90 qui maintiennent l'AR dans une conformation inactive, favorable à la liaison avec le ligand (Fang *et al.* 1996; Housley *et al.* 1990). Certains résidus cachés dans la poche du LBD rendent celle-ci flexible, permettant de lier

tout un panel d'androgènes naturels (testostérone, DHT) mais aussi des agonistes et antagonistes synthétiques (par exemple le R1881 et le bicalutamide respectivement). Les molécules les plus affines sont celles qui établissent le plus d'interactions avec le LBD : c'est ainsi que la DHT est plus affine pour l'AR que la testostérone (Pereira de Jésus-Tran *et al.* 2006). La nature des acides aminés qui composent cette poche est spécifique et non conservée entre les différents récepteurs stéroïdiens ce qui leur confère leur spécificité et la sélectivité de fixation du ligand.

De plus, le LBD intervient dans d'autres fonctions du récepteur comme la dimérisation (Wong *et al.* 1993) et la régulation transcriptionnelle (en empêchant l'activation constitutive du récepteur en l'absence d'hormone) (Jenster *et al.* 1991; Simental *et al.* 1991). En effet, la liaison avec un androgène active une zone spécifique du LBD, dénommée AF2, qui, tout comme l'AF1 présent dans le NTD, permet la liaison de co-activateurs de l'AR au niveau de motifs LxxLL et permet ainsi la modulation de son activité transcriptionnelle (Moilanen *et al.* 1997; He *et al.* 1999).

#### d. La région charnière

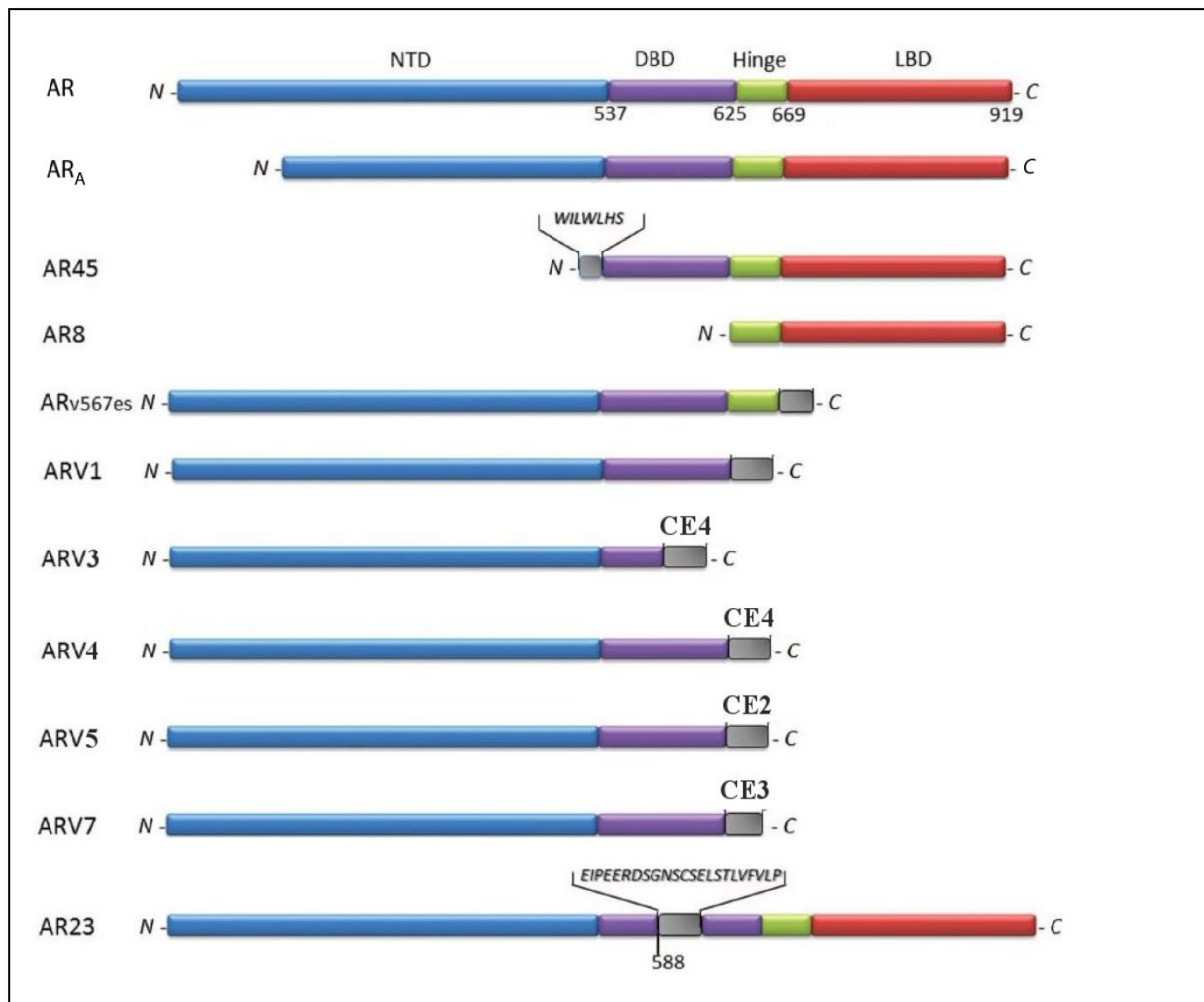
Située entre le DBD et le LBD, la région charnière est un "spacer" flexible codé par l'exon 4 du gène *AR*. Cette région est impliquée dans les changements conformationnels entraînés par la liaison de l'AR avec son ligand (Kuil & Mulder 1994). Elle possède également un signal de localisation nucléaire (NLS) bipartite à la jonction avec le DBD (acides aminés 617 à 633) (Zhou *et al.* 1994) permettant l'importation nucléaire de l'AR lié à son ligand. De plus, ce domaine serait aussi responsable d'interactions avec des protéines de modulation, essentiellement inhibitrices, de l'activité transcriptionnelle du récepteur (Loy *et al.* 2003; Poukka *et al.* 1999).

#### iii. Les isoformes de l'AR

La forme la plus fréquente de l'AR est une protéine de 110 à 114 kDa (van Laar *et al.* 1989) composée de 910 à 919 acides aminés. Ces variations de longueur viennent du nombre variable des répétitions de triplets dans l'exon 1 du gène (plus particulièrement les polyQ et polyG) (Lubahn, Joseph, Sullivan, *et al.* 1988; Lubahn, Joseph, Sar, *et al.* 1988). En effet, si les domaines de liaison à l'ADN et au ligand

ont une très forte homologie entre les récepteurs stéroïdiens, il n'en est pas de même pour le domaine NH2 terminal. Il faut également noter l'existence d'une isoforme minoritaire nommée A. Il en résulte une protéine de 87 kDa dépourvue de sa partie N-terminale dont la fonction, même si encore mal connue, semble être la même que celle de l'isoforme majoritaire de 110 kDa (Wilson & McPhaul 1994; Gao & McPhaul 1998). Néanmoins, l'origine de cette isoforme est soumise à débat puisqu'elle serait issue, soit d'une initiation de traduction interne (Met188) (Wilson & McPhaul, 1994), soit d'une protéolyse *in vitro* pendant l'extraction des cellules (Gregory *et al.* 2001) ce qui impacte sur la véracité de son rôle *in vivo*.

A l'heure actuelle, pas moins de neuf variants additionnels ont été décrits dans la littérature (**Figure 12**, pour revue, lire (Pelekanou *et al.* 2013)). Ces isoformes sont des variants d'épissage, conséquences de sauts d'exons ou d'épissage d'exons cryptiques (CE). Elles ont été décrites et impliquées dans le cancer de la prostate (notamment ceux à progression avancée échappant aux traitements anti-androgéniques) (Zhang *et al.* 2011) mais elles sont également exprimées dans d'autres tissus sains. Certaines ne possèdent pas le domaine NTD et ont un rôle inhibiteur sur l'isoforme majoritaire de l'AR (comme AR45 qui forme des hétérodimères AR-AR45) (Ahrens-Fath *et al.* 2005). D'autres au contraire ont un LBD et/ou la région charnière remplacé(s) par un nouveau domaine C-terminal et sont constitutivement actives (c'est le cas d'ARv567es, ARV1, V3, V4, V5 et V7). Un autre variant d'épissage est l'AR23 qui possède une insertion de 23 acides aminés dans le DBD lui enlevant toute fonction nucléaire. Enfin, le dernier variant d'épissage décrit est l'AR8 qui ne présente ni NTD ni DBD et est principalement retrouvé à la membrane plasmique où il a un rôle dans la réponse aux facteurs de croissance.



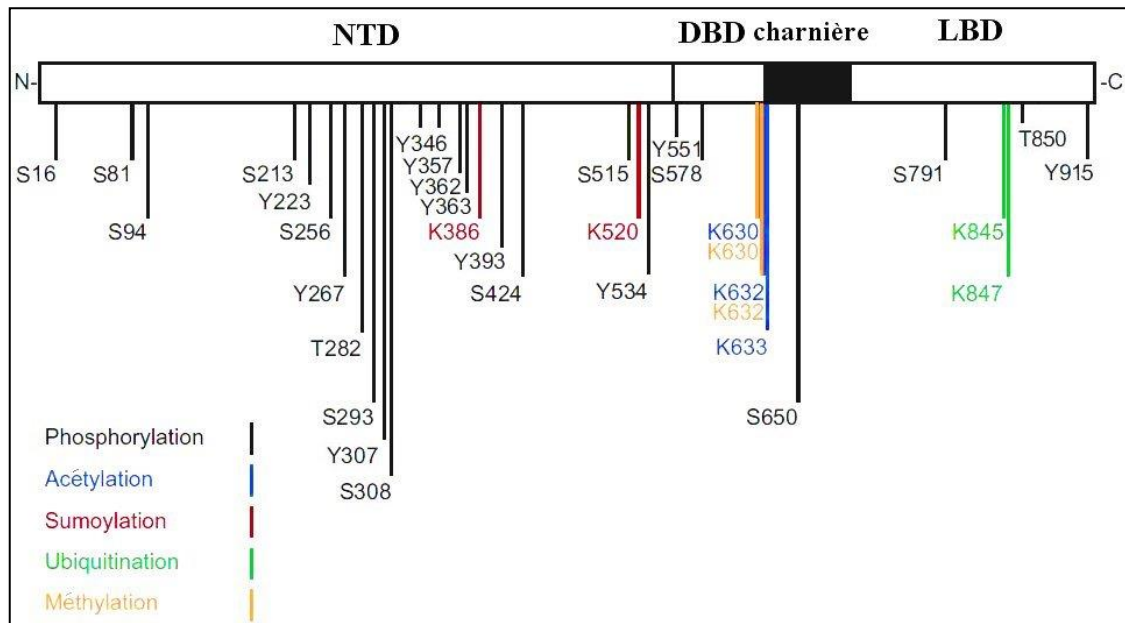
**Figure 12.** Structure de l'AR et de ses principales isoformes.

Adaptée de (Pelekanou *et al.* 2013; van der Steen *et al.* 2013).

#### iv. Modifications post-traductionnelles

Après leur traduction, la plus grande partie des protéines subissent différents types de modifications covalentes. L'AR peut ainsi être phosphorylé, acétylé, méthylé, ubiquitiné et SUMOylé (**Figure 13**, pour revue voir (Gioeli & Paschal 2012; van der Steen *et al.* 2013)). Ces modifications post-traductionnelles modifient notamment la stabilité, la localisation cellulaire, la liaison à l'ADN, l'activité transcriptionnelle ainsi que le recrutement des co-régulateurs de l'AR. La modification majoritaire de l'AR est la phosphorylation dont plus de 75% des sites sont retrouvés dans le NTD indiquant leur rôle dans la régulation de son activité transcriptionnelle (Zhou *et al.* 1995; Kuiper *et al.* 1993). La plupart des études ont été réalisées dans le cadre du cancer de la prostate et sur l'isoforme majoritaire de l'AR

et des études approfondies sur les autres isoformes et variants seraient nécessaires pour comprendre plus globalement la régulation de ces différents mécanismes. De plus ceux-ci sont extrêmement complexes à décrypter puisqu'il semble que les modifications post-traductionnelles de l'AR peuvent interagir fonctionnellement entre elles (Coffey & Robson 2012).



**Figure 13. Modifications post-traductionnelles de l'AR.**

Sont représentés les domaines de l'AR avec les modifications post-traductionnelles connues. Les phosphorylations sont indiquées en noir, les acétylations en bleu, les SUMOylations en rouge, les ubiquitinations en vert et les méthylation en orange.

D'après (Gioeli & Paschal 2012; van der Steen *et al.* 2013)

## 2. Pathologies associées à l'AR

L'AR est l'acteur central de la signalisation androgénique puisqu'il est le médiateur qui établit la connexion fonctionnelle entre le signal hormonal et la réponse cellulaire. Ainsi, des anomalies de sa structure ou encore des dérégulations de son mode de fonctionnement sont associées à diverses pathologies telles que l'insensibilité aux androgènes, le syndrome de Kennedy ou encore le cancer de la prostate. L'ensemble des mutations du gène *AR* sont référencées dans la "Androgen Receptor Gene Mutations Database World Wide Web Server" de l'Institut Lady Davis (<http://androgendb.mcgill.ca>) (Gottlieb *et al.* 2012).

### **i. L'insensibilité aux androgènes (AIS)**

Le syndrome d'insensibilité aux androgènes (AIS) est une pathologie dans laquelle les individus mâles atteints synthétisent normalement des androgènes mais l'AR ne transmet pas le signal cellulaire en résultant. De ce fait, même s'ils sont génétiquement des hommes (caryotype 46,XY), ils développent des caractères sexuels secondaires féminins et sont stériles. Ces individus sont élevés comme des femmes et le diagnostic de la maladie est posé à la puberté à cause de l'absence de menstruations. Il existe différents degrés d'atteinte : le syndrome d'insensibilité totale aux androgènes (phénotype génital externe féminin), partielle (phénotype génital externe ambigu) ou modérée (phénotype génital externe masculin). En effet, les mutations causales dans le gène *AR* touchent la plupart du temps le DBD ou le LBD et entraînent un spectre de conséquences sur l'activité de l'AR, allant d'une faible diminution à l'abolition totale de son activité. Quatre types de mutations ont été retrouvées à ce jour :

- 1) des mutations ponctuelles causant des substitutions ou des codons stop prématurés,
- 2) des insertions ou délétions de nucléotides causant un décalage du cadre de lecture,
- 3) des délétions totales ou partielles du gène et 4) des mutations introniques qui affectent l'épissage de l'ARN *AR* (McPhaul, 2002).

Ces individus, notamment ceux avec une complète insensibilité aux androgènes, possèdent un chromosome Y, des testicules qui produisent un niveau normal d'androgènes mais n'ont pas d'AR fonctionnel. Ils offrent ainsi l'opportunité d'étudier l'importance relative des androgènes, des œstrogènes et des chromosomes sexuels dans le dimorphisme sexuel observé dans le développement du cerveau car leur contribution respective est toujours soumise à débat.

### **ii. Le syndrome de Kennedy**

La maladie de Kennedy, ou atrophie musculaire bulbo-spinale, est une maladie rare du motoneurone. Son mode de transmission est récessif et lié au chromosome X. Elle se manifeste chez les hommes entre 30 et 60 ans par des tremblements et crampes musculaires évoluant lentement vers une faiblesse musculaire des membres et de la face associée à des dérégulations endocrinologiques (baisse de la fertilité) (Kennedy *et al.* 1968).

D'un point de vue moléculaire, ce syndrome est dû à une expansion des triplets CAG codant pour le polyQ présent dans le NTD de l'AR. Chez les individus sains, la répétition est de 9 à 38 résidus glutamines (Choong & Wilson 1998) mais, chez les individus souffrant de la maladie de Kennedy, leur nombre excède les 40 (La Spada *et al.* 1991). Il semblerait que la taille du polyQ soit inversement corrélée à l'âge d'apparition de la faiblesse musculaire chez les patients (Igarashi *et al.* 1992). Enfin, l'instabilité somatique de cette expansion reste limitée (Biancalana *et al.* 1992) et le mécanisme par lequel elle augmente reste inconnu (Zhang *et al.* 1994).

Pour ce qui est de la physiopathologie de cette maladie, les conséquences de l'expansion du polyQ chez les malades peut être double. Tout d'abord, comme abordé précédemment, la taille de ce polyQ modifie le repliement de l'AR ce qui influe sur son activité transcriptionnelle, la diminuant d'environ 50% pour une expansion de 40 à 50 résidus (Mhatre *et al.* 1993) en modifiant notamment l'affinité de fixation avec ses co-régulateurs (Hsiao *et al.* 1999; Cowan *et al.* 2003). D'autre part, cet AR dont la conformation est anormale est résistant à la protéolyse. Des fragments de l'AR s'accumulent dans les cellules, formant des inclusions cytotoxiques pour les motoneurones périphériques (Walcott & Merry, 2002; Cowan *et al.* 2003). Ceux-ci dégénèrent puis meurent menant peu à peu à l'atrophie musculaire. Cette cytotoxicité est spécifique à la fois de l'expansion polyQ et des motoneurones ce qui implique qu'aucune autre mutation de l'AR (comme celles retrouvées dans l'AIS) ne cause de dégénérescence des motoneurones.

Ainsi, l'allongement du polyQ de l'AR est clairement la cause de la maladie de Kennedy, entraînant à la fois une perte de fonction de l'AR (diminution de son activité transcriptionnelle responsable notamment de la baisse de fertilité) et un gain de fonction spécifiquement toxique pour les motoneurones (inclusion de fragments cytotoxiques).

### **iii. Le cancer de la prostate**

Les androgènes et l'AR jouent un rôle important dans le développement, la différenciation et le maintien de l'appareil génital masculin. C'est notamment le cas de la prostate dans laquelle ils sont également impliqués dans son dysfonctionnement, en particulier dans la croissance anarchique des cellules cancéreuses prostatiques. En effet, l'AR est exprimé à différents degrés dans les tumeurs primaires

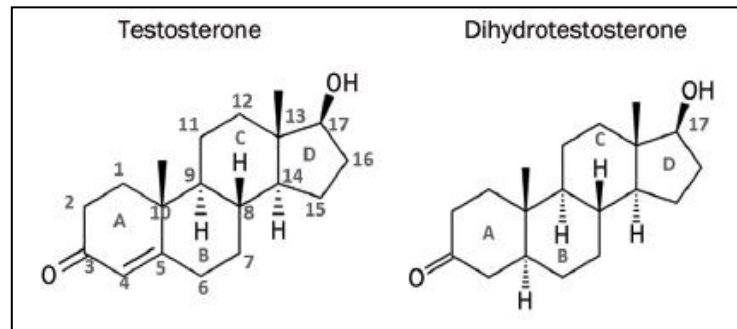


(Ruizeveld de Winter *et al.* 1994; Chodak *et al.* 1992) et des études menées chez l'homme et les modèles animaux révèlent une relation entre le niveau cellulaire de l'AR et la progression de la maladie (Henshall *et al.* 2001; Ricciardelli *et al.* 2005). Cependant, bien que des études sur les modèles animaux suggèrent qu'une élévation de l'expression de l'AR pourrait initier le développement du cancer de la prostate (Stanbrough *et al.* 2001), il n'y a pas de preuve directe de son implication comme cause du cancer de la prostate, mais il pourrait exacerber un cancer préexistant. Des anomalies du gène *AR* sont également retrouvées dans les cellules cancéreuses prostatiques, telles que des amplifications du gène lui-même ou des mutations activatrices de l'AR. Par exemple, la mutation p.T877A dans le LBD est retrouvée dans les LNCaP, le modèle cellulaire du cancer de la prostate le plus utilisé, et confère à l'AR une activation constitutive sans nécessité de ligand spécifique (Veldscholte *et al.* 1992).

De plus, l'une des composantes génétiques du cancer de la prostate réside dans le polymorphisme de l'AR puisqu'il existe une relation inverse entre la taille du polyQ du NTD de l'AR et le risque de développer un cancer de la prostate (Irvine *et al.* 1995; Giovannucci *et al.* 1997; Hsing *et al.* 2000). L'AR est donc un acteur clé du cancer de la prostate chez l'homme et est source de nombreuses investigations en tant que cible thérapeutique potentielle. Cet engouement ne va cesser d'augmenter dans les années à venir car il y a de plus en plus d'études montrant l'implication de l'AR dans d'autres cancers tels que le cancer du sein (Lobaccaro *et al.* 1993), du larynx (Goulioumis *et al.* 2011), du foie (Chen *et al.* 2012) et des testicules (Garolla *et al.* 2005).

### **3. Mécanisme d'action de la testostérone via l'AR**

Comme il l'a déjà été évoqué, ce sont la testostérone et son métabolite la DHT qui assurent l'essentiel de l'action des androgènes chez l'homme et ce *via* leur liaison à l'AR. Seules la testostérone et la DHT libres entrent dans les cellules cibles et ce par simple diffusion au travers de la membrane plasmique. Il faut savoir que la DHT diffère de la testostérone par l'absence d'une double liaison dans le cycle A (**Figure 14**), ce qui lui confère une affinité de liaison pour l'AR deux à trois fois plus élevée que la testostérone et une constante de dissociation cinq fois plus faible (Grino *et al.* 1990; Zhou *et al.* 1995). Le complexe DHT-récepteur se lie également plus facilement et plus stablement à l'ADN.



**Figure 14.** Structures moléculaires de la testostérone et de la dihydrotestostérone.

### i. Activation et translocation de l'AR

Le mécanisme moléculaire de l'action des androgènes *via* l'AR est résumé dans la **Figure 15**. En l'absence de ligand (testostérone ou DHT) fixé sur le LBD, l'AR est inactif et principalement cytoplasmique. Il fait alors partie d'un complexe macromoléculaire comprenant des protéines inhibitrices (les HSP70 et 90) qui le maintiennent dans une conformation inactive incapable de lier l'ADN. En effet, l'AR est replié sur lui-même ce qui empêche toute dimérisation ou interaction avec d'autres protéines co-activatrices ou l'ADN. Cependant, cette conformation est très favorable à la liaison avec le ligand (Fang *et al.* 1996; Housley *et al.* 1990).

Une fois le ligand fixé dans la poche du LBD de l'AR, le complexe AR-HSPs se dissocie. La fixation du ligand rend également l'AR beaucoup plus stable notamment grâce à des interactions intramoléculaires qui se produisent entre le motif Tau-1 du NTD et l'AF2 du LBD (Zhou *et al.* 1995; Langley *et al.* 1995; Doesburg *et al.* 1997; Wilson 2011). Ces interactions intramoléculaires (appelées interactions N/C) entraînent un changement conformationnel allostérique qui amorce la fixation de co-activateurs de l'AR. Ce recrutement a lieu entre les motifs LxxLL des protéines co-régulatrices et la région AF2 du LBD de l'AR (Needham *et al.* 2000). Les sites de dimérisation (vraisemblablement présents dans le LBD) (Nemoto *et al.* 1994) ainsi que le NLS sont également mis à jour ce qui permet la translocation de l'homodimère AR-AR dans le noyau au travers des pores nucléaires.

L'import nucléaire de l'AR est dépendant de l'importine  $\alpha$  qui le lie dans le cytoplasme puis le libère rapidement après l'entrée dans le noyau (Kawata *et al.* 2008; Ni *et al.* 2013). Cette translocation débute en quelques minutes et semble être complète en 30 min (Jenster *et al.* 1991). Il est à noter qu'une étude réalisée en 2008, mais jamais reproduite depuis, suggère la présence de deux NLSs additionnels localisés dans le NTD et le LBD (Kawata *et al.* 2008). Après avoir joué son rôle régulateur (abordé dans la suite de cette partie) et/ou lorsque les androgènes ne sont plus présents dans la cellule, l'AR est exporté dans le cytoplasme et n'est pas systématiquement dégradé. En effet, il peut subir un cycle d'import et d'export à travers la membrane nucléaire plusieurs fois après une nouvelle stimulation par les androgènes (Tyagi *et al.* 2000).

## ii. Reconnaissance des éléments de réponses aux androgènes (ARE) par l'AR

Après sa translocation dans le noyau, l'homodimère AR-AR se fixe *via* les deux doigts de zinc de son DBD sur des séquences d'ADN propres à l'AR, les ARE (pour Androgen Response Element). Les ARE consistent en deux demi-sites hexamériques (5'-TGTTCT-3') séparés par 3 pb (Ham *et al.* 1988; Roche *et al.* 1992). La dimérisation de l'AR est coopérative puisque la liaison de seulement un AR sur un demi-site est beaucoup plus faible. Ces demi-sites répétés peuvent être plus ou moins dégénérées, directs (DR3) ou inversés (IR3). Les ARE de type IR3 (le principal étant 5'-GTTACAnnnTGTTCT-3') sont des éléments classiques ressemblant beaucoup aux éléments de réponses des récepteurs aux glucocorticoïdes et à la progestérone (GRE/PRE) et sont donc non spécifiques car également reconnus par ces deux récepteurs. Les ARE de type DR3 sont quant à eux sélectifs de l'AR.

Les règles concernant la spécificité ou non d'un ARE vis à vis de l'AR ne sont cependant pas toujours très claires. En effet, certains ARE classiques de type IR3 sont spécifiques de l'AR et certains de type DR3 sont non spécifiques, contrairement à ce qui a été évoqué précédemment (Denayer *et al.* 2010). La **Table 3** résume les différentes classes d'ARE qui ont été mis en évidence expérimentalement.

<b><u>Non spécifiques</u></b>			
<b>Haute affinité</b>	GTTACA aac TGTTCT AGTACT tga TGTTCT TGTACA gga TGTTCT AGCACT tgc TGTTCT GAAACA gcc TGTTCT	<b>Faible affinité</b>	ATAGCA tct TGTTCT TGAAGT tcc TGTTCT ATCCTA tgt TGTTCT AGAACA aaa TGTACA
<b><u>Spécifiques de l'AR</u></b>			
<b>Haute affinité</b>	GGTTCT tgg AGTACT TGGTCA ggc AGTTCT GGCTCT ttc AGTTCT	<b>Faible affinité</b>	AGCAGG ctg TGTCCC

**Table 3. Classification des ARE revue dans (Claessens *et al.* 2001).**

Parce que le DBD est extrêmement conservé au sein de la famille des récepteurs stéroïdiens (notamment les trois acides aminés de la base du doigt de zinc permettant la reconnaissance de l'ADN), la question de la spécificité de chacun de ces récepteurs pour des éléments de réponses qui leur sont propres reste sujette à débat. Ainsi, les autres récepteurs ne reconnaissent pas les ARE sélectifs qui sont en fait des éléments de réponse non consensus (de type DR3, la plupart du temps dégénérés, par exemple 5'-GGTTCTnnnAGTACT-3') (Claessens *et al.* 1996; Verrijdt *et al.* 2000; Schoenmakers *et al.* 2000). D'après les données structurales, il semblerait que l'AR contiennent une interface additionnelle qui stabilise le complexe AR-AR-ARE, la stabilité de la dimérisation des autres récepteurs, notamment du GR, n'étant pas assez forte pour permettre l'interaction avec ces ARE sélectifs (Shaffer *et al.* 2004). De plus, la fixation du ligand sur l'AR entraîne des interactions N/C très fortes permettant le recrutement de co-facteurs qui sont indispensables à la reconnaissance de ces ARE sélectifs (Claessens *et al.* 2008).

### **iii. Régulation de la transcription (action génomique)**

Une fois fixé sur les ARE (classiques ou sélectifs) situés dans les promoteurs et/ou enhancers des gènes androgéno-dépendants, l'AR va réguler la transcription de ceux-ci par le recrutement de nombreux co-facteurs de type co-activateurs ou co-répresseurs qui modulent notamment l'état de la chromatine. L'homodimère AR-AR rassemble également d'autres

molécules additionnelles comme des facteurs de transcription et l'ARN polymérase II. On compte à ce jour plus de 200 protéines pouvant interagir avec l'AR, beaucoup d'entre-elles interagissant également avec d'autres récepteurs stéroïdiens. Les gènes androgéno-dépendants possédant des ARE peuvent soit être activés ou réprimés par la fixation de l'AR. Pour ces gènes directement régulés, la modification de la quantité d'ARNm puis de protéine produite est rapide (environ 30 min, c'est la réponse primaire). Il existe également une réponse secondaire aux androgènes prenant place dans les 12 à 48h après l'exposition aux androgènes. Dans ce cas, c'est une régulation indirecte de l'expression de ces gènes *via* les gènes cibles primaires régulés par l'AR.

L'exemple le plus connu des gènes androgéno-dépendants, car le plus étudié, est le gène *PSA* qui est un biomarqueur de l'évolution du cancer de la prostate. Ainsi, il a été identifié deux ARE dans le promoteur de ce gène (ARE I en -170 (Riegman *et al.* 1991) et ARE II en -394 (Cleutjens *et al.* 1996)) et une région enhancer en -4200 contenant plusieurs ARE (ARE III) (Cleutjens *et al.* 1997). Depuis, de nombreuses études transcriptomiques (par SAGE pour Serial Analyses of Gene Expression, puce d'expression ou séquençage d'ARNm (RNA-seq)) ont été réalisées, majoritairement dans les cellules modèles du cancer de la prostate ce qui a permis de dresser une liste quasi exhaustive des gènes régulés par les androgènes dans ce type cellulaire (Waghray *et al.* 2001; Xu *et al.* 2001; Velasco *et al.* 2004; Ngan *et al.* 2009; Vaarala *et al.* 2012; Wang *et al.* 2014; Munkley *et al.* 2015; Sowalsky *et al.* 2015). Dans les papiers les plus récents, ces études transcriptomiques sont couplées à des expériences d'immunoprécipitation de la chromatine liée à l'AR couplée à des puces d'expression (ChIP-chip) ou à du séquençage haut débit (ChIP-seq) afin de déterminer précisément où se situent les ARE des gènes régulés par les androgènes (Bolton *et al.* 2007; Lin *et al.* 2009; Takayama *et al.* 2011). L'ensemble de ces techniques (RNA-seq et ChIP-seq) génèrent une grande quantité de données et peuvent être très informatives, notamment pour décrypter les mécanismes de l'action génomique de l'AR.

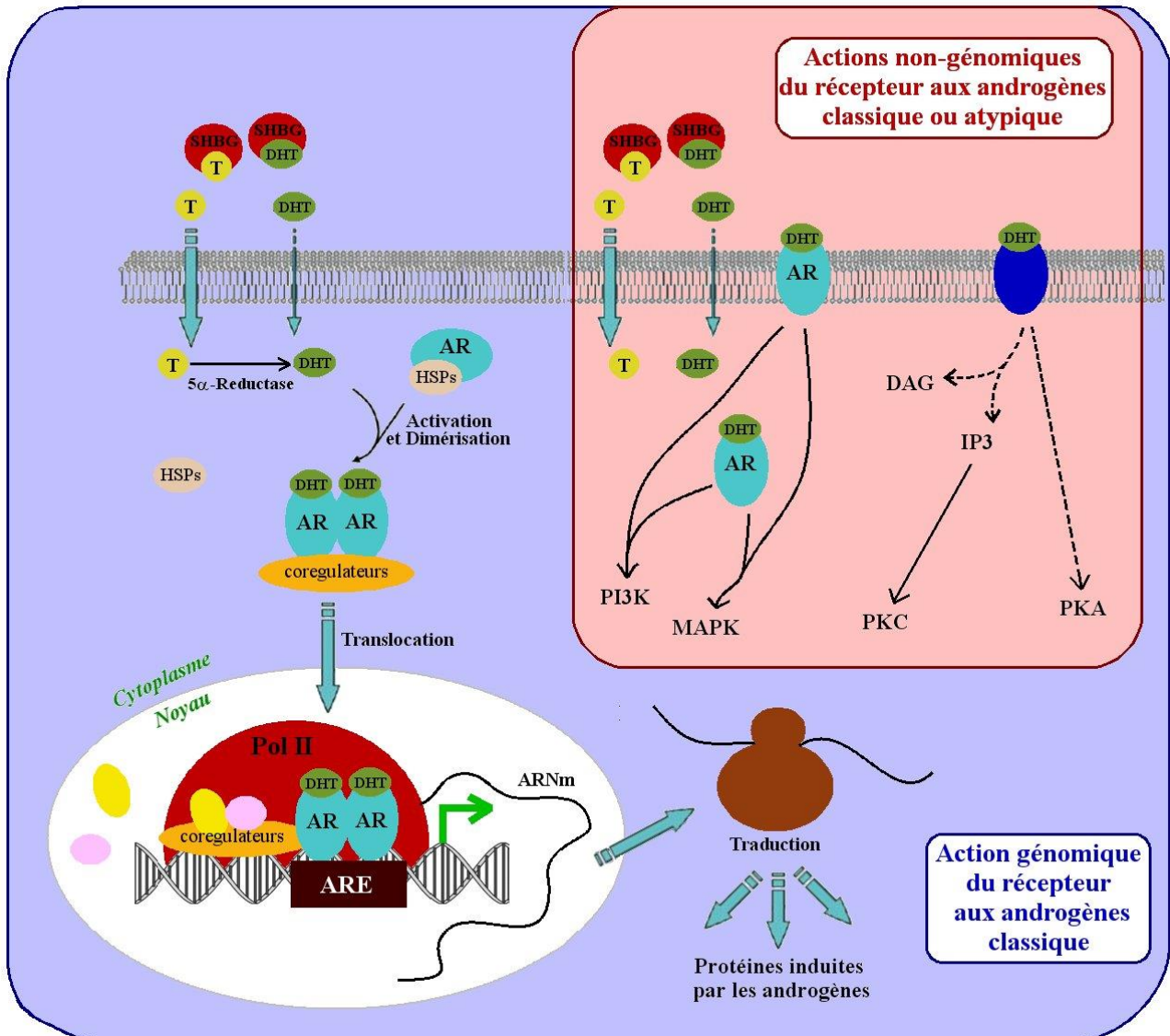
#### **iv. Actions non-génomiques**

L'action génomique de l'AR contrôle la modulation de l'expression des gènes et la synthèse *de novo* de protéines mais un temps de latence est nécessaire entre l'activation de l'AR et la réalisation effective du produit (30 min au minimum). Cette voie génomique contrôle donc des fonctions cellulaires à moyen et long terme.

Il existe ainsi une voie alternative d'activation cellulaire non-génomique de l'AR qui permet des actions à court terme (quelques minutes, voire secondes). Elle permet une adaptation cellulaire rapide à l'environnement en activant ou réprimant des molécules déjà présentes dans la cellule par phosphorylation/déphosphorylation, méthylation/déméthylation, acétylation/déacétylation, etc.

Ce mode d'action a été décrit pour la première fois dans les années 1970 pour les œstrogènes (Pietras & Szego, 1977). Ensuite, différentes équipes ont confirmé qu'il existait des actions des androgènes indépendantes de l'AR nucléaire et qui agissaient sur des récepteurs associés à la membrane plasmique (tels que l'AR lui-même ou bien des canaux ioniques ou des récepteurs couplés à des protéines G) (Guo *et al.* 2002). Cette liaison entraîne ensuite la mobilisation rapide de seconds messagers tels que le calcium, l'inositol triphosphate (IP3), le diacylglycérol (DAG) et la modulation de voies de signalisation intracellulaires par l'intermédiaire des protéines kinases A ou C (PKA ou PKC), la cascade des MAPK (Mitogen-Activated Protein Kinases) ou de la PI3K (PhosphoInositide 3-Kinase) (Simoncini & Genazzani 2003; Falkenstein & Wehling 2000; Losel *et al.* 2003). Si les preuves de l'existence de voies non-génomiques initiées par les androgènes ne sont plus à faire, les mécanismes d'action sont encore mal connus. De plus, le débat concernant un AR membranaire reste ouvert.

Il est à noter que le niveau d'expression du récepteur, sa capacité à être actif transcriptionnellement ainsi que la concentration en hormone sont cruciaux pour déterminer le type de réponse cellulaire (Castoria *et al.* 2003). Ainsi, des organes connus pour ne pas être ciblés par les actions génomiques classiques des androgènes peuvent cependant être vivement régulés par le mode d'action non-génomique (Simoncini & Genazzani 2003). De plus, le mode d'action non-génomique peut aussi avoir pour cibles des facteurs de transcription et leurs partenaires transcriptionnels et ainsi influencer le profil du transcriptome de la cellule. Il semble donc qu'il existe au sein des cellules une collaboration et des interconnexions complexes entre les effets génomiques et non-génomiques des androgènes (**Figure 15**).



**Figure 15.** Les différents modes d'action du récepteur aux androgènes.

Les effets les plus connus de l'AR sont les effets génomiques directs impliquant la régulation de la transcription de gènes cibles *via* la fixation de l'AR et de ses co-régulateurs sur les ARE. En plus de cette voie, l'AR peut initier des effets rapides non-génomiques par l'activation de voies alternatives. L'AR cytosolique peut en effet activer des voies de signalisation telles que celles de la PI3K ou des MAPK. De plus, les androgènes peuvent induire la mobilisation de calcium intracellulaire ou activer les PKA et PKC *via* des récepteurs membranaires différents ou non de l'AR classique.



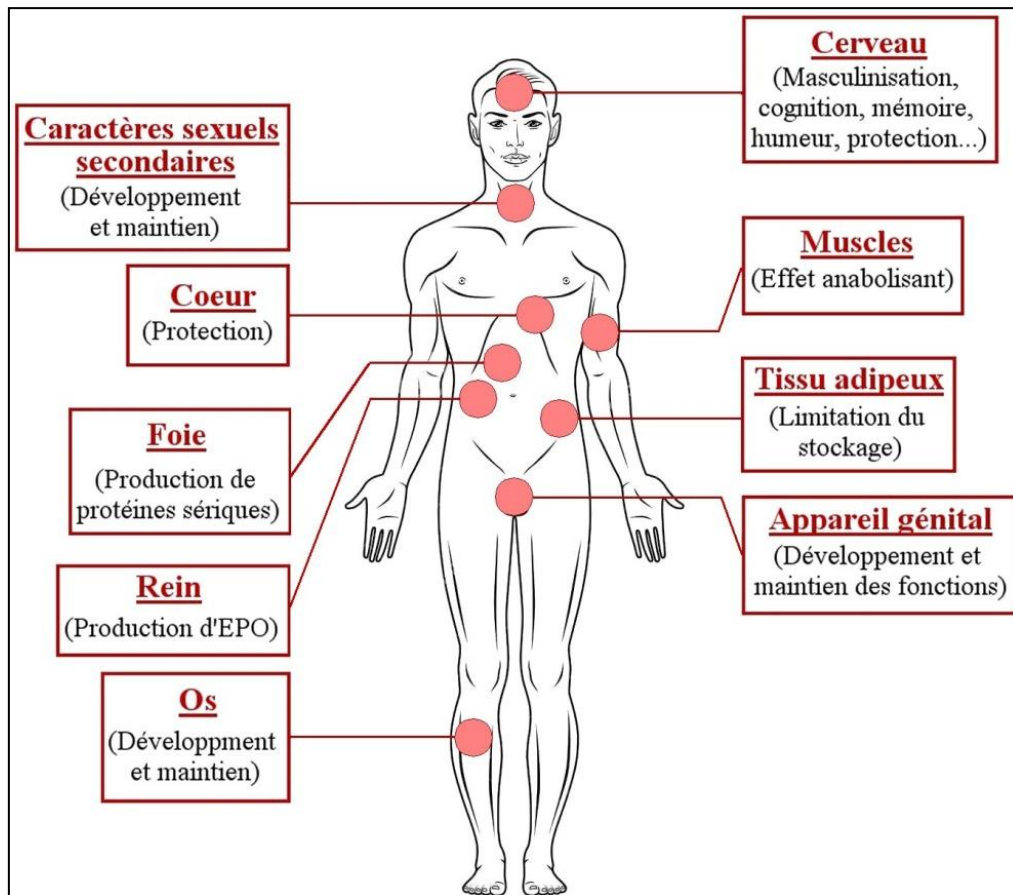
## v. Rôle des co-régulateurs

Comme nous allons l'aborder dans la partie suivante, les androgènes ont des effets différents sur leurs organes cibles. Ceci peut être expliqué en partie par le fait que chaque type cellulaire répond différemment au stimulus des androgènes, même si cette réponse est médiée par le même récepteur intracellulaire. Pour cela, l'AR n'agit pas seul mais est sans cesse en interaction avec des co-régulateurs qui sont essentiels à ses propriétés fonctionnelles. Ceux-ci peuvent être exprimés de façon "tissu spécifique", de manière transitoire ou inductible. Etant donné qu'il existe à ce jour plus de 200 protéines capables d'interagir avec l'AR, il ne s'agit pas ici de rentrer dans les détails. Pour résumer, il existe deux types de co-régulateurs : les co-activateurs qui exercent une action positive sur la transactivation de l'AR et les co-répresseurs qui, au contraire, ont une action négative sur son activité transcriptionnelle. Les mécanismes de régulation de ces co-régulateurs sont multiples. Certains jouent un rôle sur la stabilité du récepteur, sa disponibilité ou sa distribution intracellulaire, son affinité pour le ligand ou sa translocation nucléaire ce qui se traduit au final par une modulation de son activité. D'autres régulent le recrutement de la machinerie transcriptionnelle, le remodelage de la chromatine ou l'affinité de l'AR pour l'ADN au niveau de ses gènes cibles. Les co-régulateurs de l'AR modulent ainsi la réponse de la cellule ou de l'organe aux androgènes

## 4. Effets pléiotropiques des androgènes

Chez l'homme, l'AR est largement distribué dans l'organisme et l'effet des androgènes est pléiotropique comme représenté sur la **Figure 16**. Le but ici n'est pas de faire une description exhaustive du rôle des androgènes dans l'ensemble de ces organes mais simplement d'aborder leurs effets les plus marquants. L'AR est ainsi retrouvé à la fois dans les organes sexuels mais également dans des tissus non-reproducteurs comme les muscles, le cœur, les os, le cerveau, les glandes surrénales, la peau, le foie, les reins ou le tissu adipeux faisant supposer un effet pléiotropique des androgènes chez l'homme (Kimura *et al.* 1993). La testostérone peut agir directement sur l'AR (c'est le cas dans les tissus dépourvus de 5- $\alpha$ -réductase comme le muscle) mais elle peut aussi avoir des effets indirects par le biais de la DHT (comme c'est le cas dans les organes génitaux externes pendant le développement). La raison pour laquelle la testostérone doit être convertie en DHT afin d'agir dans certains tissus mais pas dans d'autres reste une énigme.





**Figure 16.** Principaux effets physiologiques des androgènes sur le développement et le fonctionnement de l'organisme humain mâle.

### i. Caractères sexuels primaires et secondaires

Les concentrations les plus élevées de l'AR sont présentes dans les organes sexuels (prostate, testicules) ce qui est en accord avec le rôle primordial des androgènes dans le développement du phénotype mâle pendant l'embryogénèse, la maturation sexuelle à la puberté et le maintien de la fonction reproductive mâle à l'âge adulte (Isaacs 1984). Il est à noter que la testostérone et son métabolite la DHT ne participent pas équitablement à ces différentes fonctions. Ainsi, la testostérone interviendrait principalement lors de la différenciation sexuelle (maintien et maturation des canaux de Wolff) et la spermatogénèse alors que la DHT est responsable de l'apparition et du maintien des caractères sexuels secondaires (organes génitaux externes, pilosité, voix, acné, musculature) comme il l'a été observé chez des individus déficients pour la 5- $\alpha$ -réductase (Imperato-McGinley & Zhu 2002).

## **ii. Rôles dans le tissu musculaire**

L'AR est exprimé dans le muscle cardiaque dans lequel il aurait un effet protecteur (Wang, Wang, *et al.* 2009) mais également dans les muscles lisses et squelettiques dans lesquels a un effet anabolisant (Baumgartner *et al.* 1999; van den Beld *et al.* 2000). De plus, les androgènes (et plus particulièrement la testostérone) régulent de multiples voies cellulaires permettant de contrôler la masse, la force musculaire et la résistance à la fatigue. Par exemple, ils promeuvent la croissance musculaire en maintenant les myoblastes au stade prolifératif et donc, en retardant leur différenciation (MacLean *et al.* 2008). Les androgènes, *via* l'AR, sont donc indispensables à la physiologie musculaire.

## **iii. Rôles dans le tissu osseux**

Les androgènes sont aussi largement connus pour leurs effets bénéfiques majeurs sur le développement et la maintien du squelette (Orwoll & Klein 1995). Cependant, le rôle des androgènes sur la physiologie osseuse est multiple et complexe puisqu'une partie de leurs effets bénéfiques passerait par la conversion de la testostérone en œstrogènes (Yanase *et al.* 2003) même si le rôle de l'AR reste primordial (Yeh *et al.* 2002; Matsumoto *et al.* 2008). De plus, la testostérone et la DHT ont la capacité d'induire la prolifération des précurseurs d'ostéoblastes, de favoriser la différenciation d'ostéoblastes matures mais aussi d'inhiber leur apoptose (Manolagas *et al.* 2002; Kousteni *et al.* 2003). Ainsi les androgènes auraient un rôle protecteur sur le remodelage osseux.

## **iv. Rôles dans les systèmes nerveux central et périphérique**

L'AR est également exprimé dans le cerveau où il contribue à différents aspects de son développement ainsi qu'au maintien de sa bonne fonction à l'âge adulte. Les androgènes sont impliqués dans l'organisation et la programmation des circuits neuronaux pendant des périodes critiques du développement et participent à la masculinisation du cerveau chez l'homme. Les rôles multiples des androgènes au sein du cerveau ainsi que leurs liens avec des pathologies qui l'affectent telles que la DI et les ASD seront évoqués en détail dans la partie suivante.

### **III. Rôles des androgènes dans le cerveau et liens avec la DI et les ASD**

Comme c'est le cas de nombreux organes, le cerveau est également influencé par les androgènes et ce dès le développement fœtal. L'expression de l'AR, le rôle des androgènes dans le développement du cerveau ainsi que dans le maintien de ses fonctions à l'âge adulte sera abordé dans cette dernière partie. De plus, plusieurs observations suggèrent que les androgènes produits par le fœtus mâle *in utero* seraient impliqués dans les NDD telles que la DI ou les ASD. Nous exposerons donc également les différentes études évoquant l'implication des androgènes fœtaux dans la physiopathologie de l'autisme.

#### **1. Expression de l'AR et rôles des androgènes dans le cerveau**

##### **i. Les comportements reproducteurs et non-reproducteurs**

Comme abordé précédemment, l'AR est exprimé dans de nombreux organes et joue un rôle crucial dans leur développement et fonctionnement. En ce qui concerne le cerveau, l'AR n'est pas retrouvé de façon homogène dans toutes les zones mais dans des régions bien spécifiques. Des études faites chez le rongeur, le primate et l'homme adulte montrent que l'AR est présent dans l'hypothalamus et l'amygdale (Simerly *et al.* 1990; Pomerantz & Sholl 1987; Sholl & Kim 1990; Choate *et al.* 1998). Ces régions du cerveau sont impliquées dans les comportements reproducteurs en lien avec la sexualité, l'agression et l'émotion (tels que la libido, l'intérêt sexuel ou l'excitation) qui sont considérés comme des caractères sexuels secondaires et sont donc connus pour être influencés par les androgènes (car ils apparaissent chez les hommes à la puberté).

L'AR est également présent dans des régions liées à des fonctions non-reproductrices telles que, l'hippocampe (Bezdicikova *et al.* 2007) (impliqué notamment dans la mémoire) ou encore le tronc cérébral, la moelle épinière et le cervelet (Sar & Stumpf 1977; Pomerantz & Sholl 1987) (zones impliquées dans les fonctions motrice et sensorielle) ou encore le cortex où la densité de cellules exprimant l'AR est plus importante que dans les zones liées aux comportements reproducteurs (DonCarlos *et al.* 2003; Puy *et al.* 1995). Les androgènes et notamment la testostérone ont ainsi été impliqués dans la mémoire (notamment spatiale) (Moffat *et al.* 2002), la cognition (Tan & Pu, 2003; Azad *et al.* 2003; Manson, 2008) et l'état d'humeur (en particulier la dépression) (Wang *et al.* 1996).

Chez les rongeurs et les primates, on retrouve cette répartition de l'expression de l'AR dans les régions non-reproductrices dès le stade embryonnaire. Malheureusement, très peu d'études ont été faites sur le cerveau humain et toutes concernent des cerveaux adultes d'où le manque d'informations sur l'expression de l'AR dans le cerveau humain en développement.

## ii. Implication dans le dimorphisme sexuel

D'autre part, les androgènes seraient responsables de plusieurs différences morphologiques entre les cerveaux masculins et féminins. Cependant, les résultats des études divergent en fonction des organismes et des régions étudiés. En effet, pendant le développement chez les rongeurs, il semble qu'il y ai un pic d'expression de l'AR dans certaines régions de l'hypothalamus (notamment l'aire pré-optique) uniquement chez les foeti mâles associé à une perte neuronale pendant la première semaine de vie (McAbee & DonCarlos, 1998; McAbee & DonCarlos, 1999). Les motoneurons de certains noyaux de la moelle épinière sont également concernés par ce dimorphisme, les androgènes et l'AR étant les acteurs indispensables de la masculinisation de cette structure (Goldstein & Sengelaub, 1992). De plus, toujours chez le rongeur, plusieurs études révèlent qu'une exposition précoce du cerveau à la testostérone entraîne un dimorphisme sexuel en régulant le processus de mort cellulaire (Nordeen *et al.* 1985; Dodson & Gorski, 1993; Forger, 2009).

Chez les primates, aucune différence n'est observée au niveau de l'hypothalamus (Pomerantz & Sholl 1987) mais les foeti mâles présentent une latéralité d'expression de l'AR au niveau des cortex frontaux (droit > gauche) et temporaux (gauche > droit) qui est absente chez les femelles. Cette différence de distribution corticale de l'AR pourrait ainsi affecter la différenciation de l'activité corticale qui est sexuellement dimorphique (Sholl & Kim 1990). Les androgènes, *via* l'AR, semblent ainsi jouer un rôle majeur dans l'établissement du dimorphisme sexuel du système nerveux central (Simerly *et al.* 1990), aussi bien d'un point de vue anatomique que fonctionnel (notamment comportemental comme il l'a été abordé précédemment). Néanmoins, dans le cas de l'hypothalamus et des autres zones où l'expression de l'AR est identique chez les deux sexes, les androgènes pourraient influencer le développement du cerveau de façon non dimorphique.

### iii. Localisation cellulaire/subcellulaire de l'AR et conséquences fonctionnelles

D'un point de vue localisation cellulaire, les principales études réalisées chez le rat ont montré que l'AR est quasi exclusivement retrouvé dans les neurones (Simerly *et al.* 1990). Cependant, il semble que des sous-populations d'astrocytes matures l'expriment également mais de façon région et âge dépendante (dans le cortex à l'âge de 10 jours, dans l'hypothalamus à l'âge adulte) (DonCarlos *et al.* 2006; Lorenz *et al.* 2005). Les neurones, astrocytes et oligodendrocytes immatures sont quant à eux AR négatifs. Au sein des neurones matures, l'AR est nucléaire mais également présent dans les axones et les dendrites ce qui implique une fonction à la fois génomique et non-génomique des androgènes dans ce type cellulaire (DonCarlos *et al.* 2003).

Des études ont montré que les androgènes améliorent la survie neuronale et il est maintenant clairement accepté qu'ils possèdent une activité neuroprotectrice (Nordeen *et al.* 1985; Hammond *et al.* 2001; Nguyen *et al.* 2007; Nguyen *et al.* 2010). Les hypothèses suggèrent que les androgènes pourraient agir sur le stress oxydatif (Chisu *et al.* 2006; Ahlbom *et al.* 2001), la modulation du calcium intracellulaire (Zup *et al.* 2014) et sur des protéines impliquées dans l'apoptose *via* l'action non-génomique des voies de signalisation MAPK (Nguyen *et al.* 2005) et PI3K (Bing *et al.* 2014). D'autre part, ils possèdent un effet trophique qui est notamment visualisable dans les motoneurones dont la régénération axonale peut être accélérée par les androgènes après une blessure *in vitro* (Brooks *et al.* 1998) et *in vivo* (Fargo *et al.* 2008; Kujawa *et al.* 1991; Kujawa *et al.* 1993). Ceci est en adéquation avec leur rôle dans la croissance des neurites. En effet, il a été montré chez les rongeurs et l'agneau que les androgènes peuvent moduler l'arborisation des neurones ainsi que la densité synaptique dans certaines régions du système nerveux, telles que le système limbique (Reddy *et al.* 2015), les motoneurones de la moelle épinière (Marron *et al.* 2005) et l'hippocampe (Leranth *et al.* 2003; Ramsden *et al.* 2003; MacLusky *et al.* 2006; Hatanaka *et al.* 2009) et ce *in vitro* et *in vivo*. Cet effet est une réponse relativement rapide de la cellule au stimulus androgénique (effet observé au bout de 2h) et passerait visiblement par une action non-génomique (modulation du calcium et des voies de signalisation MAPK et PKC) (Hatanaka *et al.* 2009). De plus, les astrocytes (qui expriment également l'AR) participeraient également à la ramification des neurites comme il l'a été montré dans l'hypothalamus de rat (Mong *et al.* 1999).

#### iv. Interaction avec la voie des œstrogènes

Il est important de noter que l'action de la testostérone est la conséquence de l'activation de l'AR (par la testostérone ou son métabolite, la DHT) combinée à celle des ESR (par sa conversion en œstrogènes). C'est notamment vrai pour le cerveau puisque l'enzyme aromatase convertissant la testostérone en œstrogènes y est exprimée. Les œstrogènes participeraient ainsi à la masculinisation du cerveau chez les rongeurs (Naftolin *et al.* 1975; Wu *et al.* 2009a). D'autres études sur des souris n'exprimant pas l'AR montrent néanmoins qu'il est indispensable à la masculinisation périnatale du cerveau et qu'il permettrait l'expression des comportements masculins typiques tels que les comportements sexuels ou l'agressivité (respectivement étudiés par une analyse comportementale d'un mâle mis en présence d'une femelle réceptive ou d'un autre mâle « intrus » non agressif) (Raskin *et al.* 2009; Sato *et al.* 2004).

Chez les primates et notamment chez l'Homme, le rôle des androgènes dans la masculinisation du cerveau est bien plus important et les œstrogènes semblent y apporter seulement une faible contribution (Goy *et al.* 1988; Goy & Deputte, 1996; Zuloaga *et al.* 2008). En effet, l'affinité des œstrogènes de la mère pour l' $\alpha$ -fœtoprotéine est bien plus faible chez l'Homme que chez les rongeurs, les laissant agir sur les fœti. Ainsi, si les œstrogènes avaient un effet masculinisant chez l'Homme, ceux d'origine maternelle masculiniserait indifféremment les fœti XX et XY pendant la grossesse.

D'autre part, les phénotypes observés chez les individus ayant différentes anomalies dans la stéroïdogénèse ou des récepteurs associés (AR et ESR) permettent d'affirmer l'implication des androgènes dans la masculinisation du cerveau humain. Ainsi, les rares hommes ayant une mutation dans le gène de l'aromatase ou dans le gène de l'un des ESR ont un phénotype masculin normal (Wilson, 2001). De plus, chez les femmes XX portant une mutation dans le gène de l'aromatase, la déficience en cette enzyme a des conséquences doubles : l'excès d'androgènes et l'insuffisance en œstrogènes entraînent une virilisation excessive de ces femmes (Belgorosky *et al.* 2009). Une virilisation plus ou moins complète est également observée pour les individus XX souffrant d'hyperplasie congénitale des glandes surrénales (CAH) qui ont, dès le début de leur développement, une augmentation de la production de testostérone surrénalienne (Speiser & White, 2003). De façon plus flagrante, les individus XY n'ayant plus d'AR fonctionnel (AIS abordé précédemment) ont une structure cérébrale, un comportement et un profil cognitif « féminin » (Hamann *et al.* 2014;

Hines *et al.* 2003). Il en est de même pour les hommes XY souffrant d'un déficit de production de la DHT (Collaer & Hines, 1995). *In fine*, tout ceci montre bien que la présence du chromosome Y ainsi que des œstrogènes convertis à partir des androgènes ne sont pas suffisants à la masculinisation du cerveau (surtout le cerveau humain) : les androgènes sont indispensables.

Le rôle de l'AR dans le cerveau est donc très complexe et multiple et la conversion de la testostérone en DHT et en œstrogènes au sein du cerveau complique considérablement l'étude du rôle direct de l'AR dans le développement et le fonctionnement de cet organe. De plus, la différence de mécanismes impliqués dans la masculinisation du cerveau entre les rongeurs (où les ESR sont indispensables) et les primates (où l'AR est indispensable) montre à quel point il est important d'étudier les cibles de l'AR dans le contexte de neurones humains.

## **2. Implication des androgènes prénataux dans la DI et les ASD**

### **i. Indices venant des individus souffrant d'anomalies de production des androgènes**

Pendant leur développement *in utero*, certains individus sont exposés à un niveau hormonal anormal influant ainsi sur leur différenciation sexuelle. C'est par exemple le cas des individus XX souffrant de CAH abordé précédemment. Ceux-ci produisent un taux anormalement élevé d'androgènes à cause d'une déficience enzymatique ce qui entraîne une ambiguïté des organes sexuels à la naissance suivie d'une progressive virilisation (Speiser & White, 2003). Ces femmes ont un profil cognitif et comportemental très masculin avec une tendance à l'agressivité (Berenbaum & Snyder, 1995; Berenbaum & Resnick, 1997) et ce à cause du taux élevé de testostérone prénatale (et non postnatale) (Berenbaum *et al.* 2000). Aucune étude n'a directement examiné si les ASD sont plus fréquents chez les femmes souffrant de CAH. Cependant, il a été montré qu'elles possèdent un plus haut "quotient autistique" (AQ) comparées à leurs sœurs non atteintes (en particulier pour les compétences sociales) (Knickmeyer, Baron-Cohen, Fane, *et al.* 2006), sont moins à même de créer des relations sociales (Helleday *et al.* 1993) et présentent plus de difficultés d'apprentissage et de langage (Plante *et al.* 1996; Inozemtseva *et al.* 2008; Nass & Baker, 1991). Ainsi, ces individus peuvent aider à mieux comprendre l'implication de la testostérone prénatale dans l'autisme mais d'autres études sur des échantillons de grande taille sont clairement nécessaires afin de déterminer s'il y a une augmentation de l'incidence des ASD et de la DI chez les personnes souffrant de CAH.



Un autre modèle serait complémentaire aux individus XX souffrant de CAH : ce sont les personnes XY souffrant d'AIS. Ces dernières, lorsqu'elles sont atteintes de la forme complète, ne possèdent pas d'AR fonctionnel mais produisent un niveau normal de testostérone. Dans ce cas, la testostérone est donc utilisée par les ESR, après avoir été convertie en œstrogènes. Cependant, de par la faible incidence de cette condition (entre 1:20000 et 1:99000 naissances vivantes de garçons d'après Orphanet), aucune étude sur les traits autistiques ou l'incidence des ASD chez ces personnes n'a pu être menée.

## ii. Le projet "Testostérone Fœtale" de Cambridge

Les effets des hormones prénatales sur les différences sexuelles chez l'Homme sont difficiles à étudier car il n'est pas possible de manipuler les hormones de fœti humains à des fins expérimentales. Comme abordé précédemment, le rôle de la testostérone fœtale dans le développement humain a été étudié indirectement grâce à des individus exposés à des niveaux anormaux d'hormones *in utero*. D'autre part, il a été trouvé un taux plus élevé de testostérone chez les individus souffrant d'ASD comparés à leurs frères et sœurs non affectés (Hu *et al.* 2009; Tordjman *et al.* 1997) ainsi que chez les femmes souffrant du syndrome d'Asperger (Schwarz *et al.* 2011). Cependant, ces mesures concernent le niveau de testostérone bien après la naissance (enfance et adolescence) et il serait intéressant de connaître quel était leur niveau de testostérone fœtale. De par le faible nombre d'échantillons disponibles, une seule étude a pu être réalisée à l'heure actuelle, révélant une augmentation du taux de testostérone et plus généralement de l'activité stéroïdogénique dans le liquide amniotique de patient souffrant d'autisme (Baron-Cohen *et al.* 2015). De plus, une étude à plus grande échelle a pu mesurer la corrélation entre le niveau de testostérone prénatale mesuré dans le liquide amniotique et des traits relatifs à l'autisme dans des cohortes d'enfants sains (Projet "Testostérone Fœtale" de Cambridge) (Baron-Cohen *et al.* 2006).

La base de ce projet réside dans le dosage de la testostérone du liquide amniotique collecté lors d'aniocentèses réalisées en milieu de grossesse. Une différence mâle/femelle de ce taux de testostérone est observée entre 12 et 18 semaines de gestation, tout comme c'est le cas pour la testostérone fœtale plasmatique. La mesure effectuée durant l'aniocentèse reflète donc fidèlement le taux de testostérone fœtale plasmatique (Finegan *et al.* 1989; Nagamani *et al.* 1979; Robinson *et al.* 1977). L'équipe de S. Baron-Cohen a ainsi effectué une étude longitudinale sur 500 enfants nés en bonne santé



et pour lesquels ce type de dosage prénatal avait été réalisé. Différents comportements ont été évalués à différents âges, tous en lien et pertinents pour l'étude de l'autisme. Leurs résultats ont montré que le taux de testostérone prénatale était inversement proportionnel aux comportements sociaux tels que les contacts visuels (évalués à 12 mois) (Lutchmaya *et al.* 2002), la richesse du vocabulaire (évaluée entre 18 et 24 mois) (Lutchmaya *et al.* 2001), les relations sociales (évaluées à 48 mois) (Knickmeyer *et al.* 2005) et l'empathie (évaluée entre 48 et 96 mois) (Chapman *et al.* 2006; Knickmeyer, Baron-Cohen, Raggatt, *et al.* 2006). A l'inverse, il est proportionnel aux intérêts restreints (Knickmeyer *et al.* 2005) et à la systématisation (Auyeung *et al.* 2006). Plus généralement et pour résumer, le taux de testostérone fœtale serait ainsi corrélé avec les traits autistiques (Auyeung *et al.* 2010) et influencerait le développement social et verbal.

### iii. La théorie du "cerveau masculin extrême"

Les études comportementales menées lors du projet "Testostérone Fœtale" de Cambridge ont également révélé que les comportements observés chez les personnes autistes étaient une exagération des traits de personnalité "masculins". L'exposition prénatale des hommes à la testostérone pourrait faire partie des mécanismes biologiques impliqués dans la mise en place de ces traits de personnalité "masculins". Ainsi, une trop forte exposition prénatale à la testostérone pourrait entraîner une exagération de ces derniers et serait, de ce fait, un facteur de risque à l'autisme. Ceci est la base de la théorie du "cerveau masculin extrême" (Extreme Male Brain theory ou EMB) qui a été proposée par le S. Baron-Cohen (Baron-Cohen 2002). D'un point de vue comportemental, les hommes ont une tendance à la systématisation (c'est-à-dire analyser les variables d'un système pour en tirer les règles qui en gouvernent le comportement) alors que les femmes ont de meilleures capacités verbales et plus d'empathie (à savoir identifier les sentiments des autres et y répondre de façon appropriée). Son équipe a mis au point des tests permettant d'établir un score de systématisation (Systematizing Quotient ou SQ) (Baron-Cohen *et al.* 2003) et d'empathie (Empathy Quotient ou EQ) (Baron-Cohen & Wheelwright, 2004). Ces deux quotients sont intimement liés à l'AQ (Autism-spectrum Quotient) qui est un outil utilisé pour quantifier de façon rapide l'ampleur des traits autistiques présentés par un individu (Baron-Cohen *et al.* 2001; Wheelwright *et al.* 2006). D'après l'ensemble de ces tests, les personnes autistes auraient tendance à plus systématiser ( $SQ_{\text{autiste}} > SQ_{\text{homme}} > SQ_{\text{femme}}$ )

et à présenter moins d'empathie ( $EQ_{\text{autiste}} < EQ_{\text{homme}} < EQ_{\text{femme}}$ ) ce qui leur a permis d'avancer l'hypothèse de l'EMB. Ainsi, les dimorphismes sexuels dans l'anatomie et le fonctionnement cérébral (qui sont, entre autre, les conséquences de l'action des androgènes sur le cerveau) pourraient contribuer à une plus forte susceptibilité des hommes aux facteurs génétiques et environnementaux menant à la différence d'incidence pour les NDD.

#### **iv. Implication des stéroïdes, du gène de l'AR et de ses co-régulateurs dans la DI et les ASD**

Outre le lien entre la testostérone (notamment prénatale) et l'autisme, différentes études ont montré plus largement un lien entre la stéroïdogénèse et l'autisme. Ainsi, plusieurs enzymes impliquées dans la biosynthèse des stéroïdes sont différenciellement exprimées chez les patients autistes (quand comparés à des contrôles) (Hu *et al.* 2009; Schwarz *et al.* 2011). De plus, une étude d'association par approche gène candidat a mis en évidence une association entre des SNP dans des gènes du métabolisme et du transport des stéroïdes et les traits autistiques (en particulier chez les patients souffrant du syndrome d'Asperger) (Chakrabarti *et al.* 2009). Une autre étude d'association concerne directement le gène de l'AR et révèle une plus forte prévalence d'un polyQ court (codé par l'expansion CAG de l'exon 1 du gène) chez les femmes souffrant d'autisme. Une contraction de cette expansion CAG a également été retrouvée chez trois frères souffrant de DI (Kooy *et al.* 1999). En parallèle, une délétion du gène *AR* étendue incluant plusieurs marqueurs proximaux et distaux a été mise en évidence chez deux patients souffrant d'AIS accompagnée de DI (Davies *et al.* 1997). Une duplication de la région Xq12-q13.3 comprenant notamment l'*AR* a également été identifiée dans une grande famille présentant des troubles du développement et des traits autistiques (Kaya *et al.* 2012). Cependant, les données suggèrent plutôt l'implication d'un gène proche de celui de l'AR plutôt que du gène *AR* lui-même (très certainement *OPHNI* suspecté d'être impliqué dans la XLID (Billuart *et al.*, 1998) et les ASD (Piton *et al.* 2011)). De plus, aucune mutation dans le gène *AR* ni de variants rares n'ont été trouvés, plaidant contre la possibilité que les mutations dans le gène *AR* soient communes chez les individus souffrant de DI et d'ASD (Henningsson *et al.* 2009).

Si le gène de l'AR ne semble pas directement associé à l'autisme, plusieurs de ses interacteurs et co-régulateurs l'ont cependant été. Une étude *in silico* comparative personnelle a été réalisée à l'aide de listes de gènes publiées dans la littérature (308 gènes mutés chez des patients avec DI et/ou ASD (Neale *et al.* 2012) et 146 interacteurs de l'AR (Heemers & Tindall 2007)). Cette analyse a révélé un enrichissement significatif d'interacteurs de l'AR parmi les gènes impliqués dans des cas de DI et/ou d'ASD (6 gènes concernés, loi hypergéométrique  $p < 0,01$ ). Par la suite, une étude de la bibliographie récente des 308 gènes publiés par Neale *et al.* a permis d'ajouter 11 autres gènes ayant un lien avec l'AR (gène cible ou interacteur). Les résultats de cette analyse bibliographique sont résumés dans la **Table 4**. Ainsi, il semblerait que de nombreux interacteurs de l'AR aient un rôle primordial dans le développement du cerveau et qu'une mutation dans l'un des 17 cités ci-dessous entraîne de la DI (isolée ou syndromique) accompagnée ou non de traits autistiques. Ceci révèle un potentiel autre mécanisme de régulation du développement du cerveau par l'AR et donne un indice supplémentaire sur son implication dans le dimorphisme sexuel ainsi que dans la plus forte susceptibilité des garçons face aux NDD.

<u>Gène</u>	<u>Syndrome(s) / Phénotype(s) associé(s)</u>	<u>Liens avec la DI et/ou les ASD</u>	<u>Interacteur de l'AR</u>
<b>ATRX</b>	DI syndromique	(Villard <i>et al.</i> 1996)	(Bagheri-Fam <i>et al.</i> 2011)
<b>CBP</b>	DI syndromique avec ou sans autisme (Syndrome Rubinstein-Taybi 1)	(Petrij <i>et al.</i> 1995)	(Aarnisalo <i>et al.</i> 1998)
<b>CHD8</b>	Autisme	(O'Roak <i>et al.</i> 2012)	(Menon <i>et al.</i> 2010)
<b>EP300</b>	DI syndromique (Syndrome Rubinstein-Taybi 2)	(Roelfsema <i>et al.</i> 2005)	(Boorjian <i>et al.</i> 2009)
<b>FLNA</b>	DI syndromique (Syndrome FG)	(Unger <i>et al.</i> 2007)	(Loy <i>et al.</i> 2003)
<b>FOXP1</b>	DI syndromique avec autisme (Rett syndrome (congenital))	(Ariani <i>et al.</i> 2008)	(Obendorf <i>et al.</i> 2007)
<b>FOXP1</b>	DI sévère associée à des troubles du langage avec ou sans autisme	(Hamdan <i>et al.</i> 2010)	(Takayama <i>et al.</i> 2008)
<b>HDAC4</b>	DI syndromique	(Williams <i>et al.</i> 2010)	(Yang <i>et al.</i> 2011)
<b>KLF8</b>	DI non syndromique	(Lossi <i>et al.</i> 2002)	(He <i>et al.</i> 2013)
<b>MEF2C</b>	DI syndromique	(Le Meur <i>et al.</i> 2010; Zweier <i>et al.</i> 2010)	(Wyce <i>et al.</i> 2010)
<b>MLL2</b>	DI syndromique (Kabuki syndrome 1)	(Ng <i>et al.</i> 2010)	(Grasso <i>et al.</i> 2012)
<b>NSD1</b>	DI syndromique (Sotos syndrome)	(Kurotaki <i>et al.</i> 2002)	(Sampson <i>et al.</i> 2001)
<b>PAX6</b>	DI syndromique ou non avec ou sans autisme	(Davis <i>et al.</i> 2008; Malandrini <i>et al.</i> 2001)	(Shyr <i>et al.</i> 2010)
<b>PQBP1</b>	DI syndromique (Renpenning syndrome)	(Kalscheuer <i>et al.</i> 2003)	(Waragai <i>et al.</i> 1999)
<b>PTEN</b>	DI syndromique avec ou sans autisme (Cowden syndrome)	(Hanssen & Fryns 1995; Butler <i>et al.</i> 2005)	(Lin <i>et al.</i> 2004)
<b>TCF4</b>	DI syndromique (Pitt-Hopkins syndrome)	(Zweier <i>et al.</i> 2007)	(Singh <i>et al.</i> 2006)
<b>UBE3A</b>	DI syndromique avec ou sans autisme (Angelman syndrome)	(Kishino <i>et al.</i> 1997)	(Khan <i>et al.</i> 2006)

**Table 4. Co-régulateurs de l'AR impliqués dans des cas de DI et/ou d'ASD.**



# Objectifs de la thèse

---



Les NDD dont font partie la DI et les ASD sont des pathologies à étiologie extrêmement complexe, faisant intervenir des facteurs génétiques et environnementaux. D'un point de vue clinique, ces pathologies ont également la particularité de présenter un excès de garçons plus ou moins conséquent (sex-ratio homme/femme de 1,3 à 1,4 pour la DI et de 4 en moyenne pour les ASD). A travers l'étude bibliographique, nous avons vu que la génétique (notamment les mutations de gènes du chromosome X) et l'environnement (comme le contexte hormonal du fœtus mâle) peuvent expliquer en partie ce biais de sexe.

Ainsi, dans une première partie, l'objectif de ma thèse a été d'étudier l'effet de quelques mutations causales ou potentiellement causales présentes dans des gènes du chromosome X chez des garçons atteints de DI. Ces mutations ont été identifiées par notre équipe par séquençage ciblé de 217 gènes impliqués dans la DI et/ou les ASD chez 106 patients souffrant de DI (Redin *et al.* 2014). J'ai ainsi analysé les conséquences fonctionnelles de mutations pathogènes (délétion héritée de l'exon 17 de *FMRI*, mutation *de novo* c.990+1G>A et héritée c.420-8A>G affectant son épissage) et potentiellement pathogènes (variation faux-sens héritée c.731C>T, p.Pro244Leu dans *PQBPI* et variation faux-sens c.1789C>T, p.Arg597Trp dans *NLGN3*). Une revue de l'ensemble des variations intragéniques de *FMRI* publiées à ce jour a également été réalisée afin de regrouper les évidences et d'examiner leur pathogénicité. Concernant *PQBPI* et *NLGN3*, d'autres variations faux-sens (publiées ou non) ont été étudiées en plus de celles que nous avons identifiées afin d'avoir une vision globale des conséquences fonctionnelles des variations faux-sens dans ces gènes.

Dans une seconde partie, ma thèse s'est plus particulièrement focalisée sur l'une des hypothèses environnementales pouvant expliquer la plus forte vulnérabilité des garçons aux NDD, à savoir, les conséquences cellulaires et moléculaires de l'exposition du cerveau humain en développement aux androgènes. La plupart des études portant sur le rôle des androgènes ont été réalisées sur des modèles murins ou bien sur des cellules humaines de cancer de la prostate. Il en est de même pour les analyses pangénomiques étudiant l'effet des androgènes sur la régulation de la transcription ou identifiant les gènes cibles de l'AR. Très peu de données sont disponibles concernant la régulation de l'expression génique de l'AR dans le système nerveux humain. En effet, seulement quelques gènes ont été montrés comme étant surexprimés dans les neurones ou les lignées cellulaires de neuroblastomes humains, tels que *CRH* (Bao *et al.* 2006), *TUBB2* et *TUBB3* (Butler *et al.* 2001).



Par conséquent, beaucoup de travail reste à faire afin de caractériser la réponse génomique aux androgènes dans les cellules neuronales humaines. Pour combler en partie ce manque, le but de mon travail a été de déterminer quels sont les gènes régulés par les androgènes dans des précurseurs neuronaux humains dérivés de cellules souches embryonnaires humaines et de voir s'ils pouvaient être impliqués dans : 1) le développement du cerveau et/ou la synaptogénèse et 2) des cas de DI et/ou d'ASD. L'objectif ultime est de mieux comprendre le mécanisme d'action de ces hormones masculines dans le cerveau et de mettre en évidence leur rôle dans l'excès de garçons observé dans les NDD.

# Résultats

---



## Chapitre 1.

# Mutations causales dans des gènes du chromosome X

Comme nous l'avons vu dans l'étude bibliographique de ce travail, la DI est une maladie génétiquement hétérogène avec plusieurs centaines de gènes identifiés à l'heure d'aujourd'hui. Après les stratégies actuelles de diagnostic que sont l'array-CGH, le dépistage de l'expansion CGG dans le 5'UTR du gène *FMRI* et le séquençage de quelques gènes suspectés lors de syndromes cliniquement reconnaissables, plus de 60% des patients souffrant de DI et/ou d'ASD n'ont pas de diagnostic moléculaire. Ainsi, au sein de notre laboratoire, nous avons développé une approche de diagnostic par séquençage haut débit ciblé permettant la recherche de variations dans les régions codantes de gènes impliqués dans la DI et/ou les ASD. Dans un premier temps, 106 patients ont été séquencés avec un design comprenant 217 gènes, incluant 99 gènes du chromosome X. Depuis, 834 patients additionnels ont été séquencés pour 275 à 451 gènes impliqués dans la DI et/ou les ASD (en fonction du design utilisé).

Les mutations causales totalement pénétrantes dans des gènes du chromosome X chez les garçons sont l'une des hypothèses avancées dans la littérature pour expliquer le biais de sexe observé dans la DI et les ASD. En complément de mon projet de thèse principal (qui sera abordé dans le 2<sup>ème</sup> chapitre de ces résultats), j'ai ainsi pu réaliser une revue des variations identifiées dans les gènes du chromosome X dans notre cohorte. J'ai également analysé plus spécifiquement les conséquences fonctionnelles de certaines d'entre-elles :

- une délétion héritée de l'exon 17 de *FMRI* chez trois frères, une mutation faux-sens c.990+1G>A *de novo*, une mutation faux-sens héritée c.420-8A>G altérant l'épissage de *FMRI* chez deux jeunes garçons ainsi qu'une revue des variations déjà publiées dans le gène *FMRI*.
- une variation faux-sens héritée c.731C>T, p.Pro244Leu dans *PQBPI* chez deux frères.
- une variation faux-sens héritée c.1789C>T, p.Arg597Trp dans *NLGN3* chez un jeune garçon et ses deux cousins germains maternels ainsi qu'une autre variation identifiée par un collaborateur (c.1540C>T, p.Pro514Ser, données non publiées).

## Matériels et méthodes

### Cohorte de patients

L'ADN de 940 patients souffrant de DI et/ou d'ASD suivis par des généticiens de plus d'une vingtaine d'hôpitaux publics français ont été analysés par séquençage ciblé. L'ensemble des analyses sont le fruit d'un travail commun entre le laboratoire de diagnostic des Hôpitaux Universitaires de Strasbourg et le laboratoire de recherche de l'Institut de Génétique et de Biologie Moléculaire et Cellulaire. Tous les patients ont été préalablement testés pour la mutation récurrente du syndrome de l'X Fragile (expansion CGG > 200), pour la présence de CNV pathogéniques (array-CGH) et en moyenne deux autres gènes ont été investigués par séquençage Sanger. Les données cliniques comprenant notamment la sévérité de la DI ont été évaluées de façon standardisée par les généticiens. Un consentement éclairé expliquant les tests génétiques réalisés a été obtenu des représentants légaux de l'ensemble des patients.

### Sélection des gènes et design de la capture ciblée

Le premier design comprenant 217 gènes (dont 99 gènes de XLID et 118 gènes autosomiques) est décrit en détail dans l'**Annexe 1** (Redin *et al.* 2014) et a été réalisé sur 106 patients. Par la suite, la liste des gènes inclus dans le design a été remise à jour en fonction des données récentes de la littérature allant dans le sens ou à l'encontre de leur implication dans la DI et/ou les ASD. Les gènes ajoutés au design sont principalement des gènes autosomiques récessifs impliqués dans des formes métaboliques de DI ou d'épilepsie ainsi que des gènes nouvellement associés aux NDD. Concernant les gènes de XLID, seuls les quelques nouveaux gènes mis en évidence dans le papier de Hu *et al.* ont été ajoutés (Hu *et al.* 2016). Pour les 834 patients additionnels, 275 ou 451 gènes ont ainsi été séquencés

### Séquençage, analyse et validation des variants

Les bibliothèques de séquençage ont été préparées à partir d'ADN extrait de salive ou de sang périphérique puis une capture SureSelect des gènes d'intérêt a été réalisée. Un séquençage paired-end a été réalisé sur un séquenceur Illumina HiSeq 2000 ou 2500 en multiplexant jusqu'à 16 échantillons par ligne de séquençage. Les méthodes exhaustives concernant l'alignement des séquences, l'annotation des variants ainsi que leur classement sont disponibles dans l'**Annexe 1** (Redin *et al.* 2014). Les variants prédits comme étant

potentiellement pathogènes et correspondant au mode de transmission attendu sont testés de façon plus approfondie (ségrégation au sein de la famille par séquençage Sanger). Les variations sont considérées comme très certainement causales lorsqu'elles co-ségrègent avec la pathologie au sein de la famille et qu'elles sont soit tronquantes, soit des faux-sens déjà décrits de façon convaincante dans la littérature. Les variations sont considérées comme potentiellement causales lorsqu'elles co-ségrègent avec la pathologie au sein de la famille et sont prédites dommageables. C'est notamment le cas des variations faux-sens qui, à l'inverse des variations tronquantes, mènent à la production d'une protéine entière mais différant d'un acide aminé avec la forme sauvage. Des analyses fonctionnelles complémentaires sont ainsi nécessaires pour écarter toute ambiguïté concernant leur pathogénicité.

#### Matériel biologique utilisé et résumé des études fonctionnelles réalisées

Deux situations se sont présentées lorsque nous avons fait les analyses fonctionnelles sur les variants potentiellement causaux dont il est question ici. Pour l'analyse des variations suspectées d'affecter l'épissage (exclusion d'un exon ou rétention d'une partie d'un intron du gène *FMRI*), des RT-PCR ont été réalisées à partir d'ARNm extrait de sang total des individus avec des primers dessinés de part et d'autres de l'anomalie à détecter. En ce qui concerne l'analyse de la délétion intragénique (*FMRI*) ou des variations faux-sens (*PQBPI* et *NLGN3*) suspectées d'affecter la taille, l'expression, la stabilité ou la localisation subcellulaire de la protéine résultante, des études ont été réalisées sur des cellules. La taille, l'expression et la stabilité de la protéine sont déterminées par western blot alors que la localisation est visualisée par immunofluorescence (deux techniques utilisant des anticorps spécifiques de la protéine à étudier). Les cellules utilisées sont des cellules provenant directement du patient (cellules lymphoblastoïdes isolées de sang total puis immortalisées par le virus Epstein-Barr ou fibroblastes provenant de biopsies de peau). Si le matériel biologique du patient n'est pas disponible (ce qui était le cas pour les variants dans *NLGN3*), nous avons alors recours à une surexpression de la protéine mutée dans des lignées de cellules établies et communément utilisées pour ce type d'étude (les COS7 ou HEK293 par exemple). Dans ce cas, il est nécessaire de faire des expériences de mutagenèse sur des plasmides d'expression contenant l'ADNc du gène étudié afin d'y introduire les variations à tester puis de les transfecter dans les COS7 ou les HEK293.

## **Partie 1.**

# **Identification de mutations causales dans *FMR1***

### **I. Le syndrome de l'X Fragile**

Le syndrome de l'X Fragile (FXS) est la forme de DI monogénique la plus fréquente et pourrait également expliquer de 2 à 5% des cas d'ASD (Boyle & Kaufmann, 2010; Kelleher & Bear, 2008; Reddy, 2005). Il est caractérisé par une DI légère à sévère qui peut être associée à des troubles du comportement et à des signes physiques caractéristiques (tels qu'une dysmorphie faciale inconstante ou une macro-orchidie après la puberté) (Hagerman *et al.* 1991). Sa prévalence est estimée à 1:3600 chez les hommes et 1:8000 chez les femmes (Cornish *et al.* 2008). Son mode de transmission est lié au chromosome X avec une expressivité variable ainsi qu'une pénétrance incomplète chez les femmes. Ainsi, les symptômes sont beaucoup moins sévères chez celles-ci, avec par exemple 21 à 50% des garçons FXS qui présentent des traits autistiques contre 1 à 3% des femmes (Moss & Howlin, 2009).

Le FXS est généralement dû à l'inhibition de la transcription du gène *FMR1* (localisé en Xq27.3) ce qui entraîne une absence de synthèse de la protéine résultante, FMRP. L'absence ou la diminution de l'expression de FMRP, qui est une protéine liant et régulant la traduction des ARNm, est responsable d'une dérégulation des voies de signalisation dendritiques ainsi que d'une réduction de la modulation et de la plasticité synaptique ce qui serait à l'origine des phénotypes neurocomportementaux observés (Bassell & Warren, 2008).

Dans 99% des cas, la sous-expression de *FMR1* est due à l'hyperméthylation d'une expansion de la répétition de triplets CGG dans sa région 5' non traduite (> 200 répétitions CGG) (Jacquemont *et al.* 2011; Oberlé *et al.* 1991; Verkerk *et al.* 1991). Cette expansion provient d'allèles prémutés instables (de 55 à 200 répétitions CGG) transmises par les mères porteuses et qui peuvent s'allonger pendant les méioses maternelles. Ces prémutations sont également associées à des syndromes distincts du FXS tels que l'insuffisance ovarienne précoce (FXPOI) chez 25% des femmes porteuses (Bodega *et al.* 2006) ou des tremblements et/ou ataxie (FXTAS) chez les hommes âgés (Biancalana *et al.* 2005).

Les caractéristiques physiques du FXS pouvant être discrètes, voire absentes, le tableau clinique ne permet pas de porter le diagnostic et celui-ci repose donc sur le dépistage moléculaire de tout patient présentant une DI ou des ASD inexplicables (par évaluation de la taille de l'expansion CGG par PCR ou Southern Blot).

Même s'il est clair que l'expansion de la répétition CGG de *FMRI* est la cause la plus fréquente du FXS, d'autres types de mutations ont également été rapportées chez des patients souffrant de FXS (Luo *et al.* 2015). Des délétions du gène entier (Luo *et al.* 2014), de la région 5' non traduite (Hammond *et al.* 1997) ou de l'exon 1 (Viveiros *et al.* 2015) ont ainsi été trouvés. De plus, quelques mutations faux-sens intragéniques entraînant l'expression d'une protéine anormale dont la fonction est dérégulée voire abolie ont également été mises en évidence (De Boulle *et al.* 1993; Lugenbeel *et al.* 1995; Wang *et al.* 1997; Grønskov *et al.* 2011; Myrick *et al.* 2014). Cependant, celles-ci restent rares étant donné le faible nombre de séquençage du gène *FMRI* effectué chez des patients souffrant de DI et/ou d'ASD et sont donc sûrement sous-évaluées (Collins *et al.* 2010; Handt *et al.* 2014).

## **II. Résultats, discussion et perspectives**

**Publication** : Intragenic *FMRI* disease-causing variants: a significant mutational mechanism leading to Fragile-X Syndrome

**Quartier A\***, Poquet H\*, Gilbert-Dussardier B, Rossi M, Casteleyn AS, Des Portes V, *et al.*, en révision chez Eur J Hum Genet.

**But de l'étude** : Rapporter cinq nouveaux patients souffrant d'un syndrome de l'X Fragile atypique causé par des mutations intragéniques du gène *FMRI*. Les trois nouvelles mutations identifiées (une délétion du dernier exon de *FMRI* héritée chez trois frères, une mutation *de novo* c.990+1G>A et une mutation héritée c.420-8A>G chez deux jeunes garçons) ont été analysés d'un point de vue fonctionnel afin de déterminer leur potentielle pathogénicité (effet sur l'expression de l'ARNm *FMRI* ainsi que sur la protéine FMRP pour la première mutation, effet sur l'épissage de *FMRI* pour les deux dernières). Une revue de la littérature concernant l'ensemble des variations déjà identifiées dans le gène *FMRI* a également été réalisée afin d'évaluer le degré de leur pathogénicité.

**Contribution** : Majeure. Réalisation des expérimentations pour la partie fonctionnelle, traitement, analyse et interprétation des données, écriture du papier avec les Dre Piton et Pre Faivre.





## Title

### **Intragenic *FMRI* disease-causing variants: a significant mutational mechanism leading to Fragile-X Syndrome**

**Angélique Quartier**,<sup>1#</sup> Hélène Poquet,<sup>2,3</sup> Brigitte Gilbert-Dussardier,<sup>4</sup> Massimiliano Rossi<sup>5</sup>, Anne-Sophie Casteleyn<sup>6</sup>, Vincent Des Portes<sup>7</sup>, Claire Feger,<sup>8</sup> Elsa Nourisson<sup>8</sup>, Paul Kuentz,<sup>2,3</sup> Claire Redin,<sup>1</sup> Julien Thevenon,<sup>2,3</sup> Anne-Laure Mosca-Boidron,<sup>9</sup> Patrick Callier,<sup>9</sup> Jean Muller,<sup>8,10</sup> Gaetan Lesca<sup>5</sup>, Frédéric Huet,<sup>3</sup> Véronique Geoffroy,<sup>10</sup> Salima El Chehadeh,<sup>3</sup> Matthieu Jung,<sup>1</sup> Benoit Trojak,<sup>11</sup> Stéphanie Le Gras,<sup>1</sup> Daphné Lehalle,<sup>2</sup> Bernard Jost,<sup>1</sup> Stéphanie Maury,<sup>11</sup> Alice Masurel,<sup>3</sup> Patrick Edery<sup>5</sup>, Christel Thauvin-Robinet,<sup>2,3</sup> Bénédicte Gérard,<sup>8</sup> Jean-Louis Mandel,<sup>1,8</sup> Laurence Faivre,<sup>3,4##</sup> Amélie Piton,<sup>1,8##</sup>

<sup>1</sup> Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR-7104, Inserm U964, Université de Strasbourg, Collège de France, Illkirch, France,

<sup>2</sup> EA 4271 GAD « Génétique des Anomalies du Développement » & FHU TRANSLAD, CHU Dijon, Université de Bourgogne, France

<sup>3</sup> Centre de Génétique et Centre de Référence Maladies Rares 'Anomalies du Développement, CHU Dijon, France

<sup>4</sup> Service de Génétique et Centre de Référence Anomalies du Développement-Ouest, CHU de Poitiers; EA3808, Université de Poitiers

<sup>5</sup> Centre de Référence Anomalies du Développement, Service de génétique, Hospices Civils de Lyon et INSERM U1028, CNRS UMR5292, GENDEV Team, Centre de Recherche en Neurosciences de Lyon, France.

<sup>6</sup> Service de neuropédiatrie, CHU Angoulême, France

<sup>7</sup> Reference Center for Fragile X and Rare Intellectual Disabilities.

<sup>8</sup> Laboratoire de diagnostic génétique, Institut de Génétique Médicale d'Alsace, Hôpitaux Universitaires de Strasbourg, France

<sup>9</sup> Laboratoire de génétique moléculaire et de Cytogénétique, Plateau Technique de Biologie, CHU Dijon, France

<sup>10</sup> Laboratoire de Génétique Médicale, Institut de Génétique Médicale d'Alsace, INSERM U1112, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg, Strasbourg, France.

<sup>11</sup> Service de Psychiatrie Générale et Addictologie, CHU, Dijon, France

# These authors should be considered as equally contributors.

\* Corresponding author Amélie Piton and Laurence Faivre

The authors declare no conflict of interest.

Address for correspondence:

Amélie Piton, PhD

Laboratoire "Mécanismes génétiques des maladies neurodéveloppementales", IGBMC, 1 Rue Laurent Fries, 67400 Illkirch-Graffenstaden, France

Tel: +33369551652

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

Laurence Faivre, MD-PhD,

Centre de Génétique et Centre de Référence Anomalies du développement et Syndromes Malformatifs, Hôpital d'Enfants, 14 rue Gaffarel, CHU Dijon, France.

Tel: +33380295313 Fax: +33380293266

E-mail: [laurence.faivre@chu-dijon.fr](mailto:laurence.faivre@chu-dijon.fr)

## Running Title

Intragenic *FMRI* changes in Fragile-X Syndrome

## Abstract

Fragile X Syndrome (FXS) is a frequent genetic form of Intellectual Disability (ID). The main recurrent mutagenic mechanism causing FXS is the expansion of a CGG repeat sequence in the 5'-UTR of the *FMRI* gene, therefore routinely tested in ID patients. We report here three *FMRI* intragenic pathogenic variants not affecting this sequence, identified using High Throughput Sequencing (HTS): a previously reported hemizygous deletion encompassing the last exon of *FMRI*, too small to be detected by array-CGH and inducing decreased expression of a truncated form of FMRP protein, in three brothers with ID (family 1) and two splice variants in boys with sporadic ID: a *de novo* variant c.990+1G>A (family 2) and a maternally-inherited c.420-8A>G variant (family 3). After clinical reevaluation, the five patients presented features consistent with FXS (mean Hagerman's scores = 15). We conducted a systematic review of all rare non-synonymous variants previously reported in *FMRI* in ID patients and show that six of them are convincing pathogenic variants. This study suggests that intragenic *FMRI* variants, although much less frequent than CGG expansions, are a significant mutational mechanism leading to FXS and demonstrates the interest of HTS approaches to detect them in ID patients with a negative standard work-up.

## Key words

*FMRI*, Fragile-X Syndrome, punctual variants, next-generation sequencing, intellectual disability

## Introduction

Fragile X Syndrome (FXS) is the most common known form of inherited Intellectual Disability (ID) and its neurodevelopmental phenotype often overlaps Autism Spectrum Disorder (ASD) <sup>1,2</sup>. FXS occurs in both genders with an estimated prevalence around 1/5000 in males and 1/8000 in females due to variable expressivity and incomplete penetrance. The high prevalence, the inconstant morphological phenotype especially in young children (long habitus, a long narrow face, prominent forehead, large prominent ears, hyperextensible finger joints, single palmar crease and macroorchidism at puberty) and the behavioral features, which are not always discriminative (hyperactivity and/or short attention span, tactile defensiveness, hand flapping and/or hand-biting, poor eye contact, perseverative speech), <sup>3</sup> justifies systematic FXS genetic testing in all patients with unexplained ID or ASD <sup>4</sup>.

FXS is caused by the absence of Fragile X Mental Retardation Protein (FMRP), an intracellular RNA binding protein. This absence dysregulates dendritic signaling pathways and affects brain hippocampal synaptic plasticity <sup>5,6</sup> through incompletely understood mechanisms. The main recurrent mutagenic mechanism leading to the absence of FMRP is CGG repeat expansion in the 5'-untranslated region (5'UTR) of *FMR1* <sup>7</sup>, inducing abnormal methylation of this region followed by transcriptional silencing. This repeat expansion is believed to explain at least 99% of FXS cases <sup>8</sup>. Around 1-2% of all children with developmental delay are found to have FXS caused by CGG -expansion<sup>4</sup>. However, others pathogenic *FMR1* variants have also been reported, including large gene deletions, 5'UTR or exon 1 deletions associated or not with a full *FMR1* CGG expansion in the mother <sup>9-15</sup>, as well as a very small number of punctual variants in the *FMR1* coding regions. Direct sequencing of *FMR1* was not frequently carried out, and only a few groups

have screened for variants in *FMRI* coding regions in clinically relevant cohorts presenting<sup>16,17</sup> or not<sup>18-20</sup> alterations of FMRP at the protein level. Two other studies were conducted on unspecific ID patients<sup>21,22</sup>. This lack of *FMRI* sequence screening has led to a marked deficit in the description of small exonic deletions or punctual variants.

In this study, we report the clinical and molecular descriptions of three *FMRI* variants identified by targeted High-Throughput Sequencing (HTS) of several hundred ID-genes in 940 patients with unexplained ID and negative results following routine FXS screening: a maternally-inherited deletion of the last exon of *FMRI* previously reported in a family with presumed X-linked ID<sup>23</sup> and two novel intronic substitutions, one *de novo* and one maternally-inherited, affecting splicing in two sporadic male patients with ID. We also reviewed the other *FMRI* point variations described in the literature in patients with ID and concluded that six of them are convincingly pathogenic or likely pathogenic. Together with our results, there are now nine intragenic pathogenic variants in *FMRI* outside the 5'UTR causing FXS.

## Patients and methods

### *Patients*

Patients III:1, III:2 and III:3 from pedigree 1 are three male siblings aged 47, 48 and 50 years. All display ID and were negative for routine FXS testing and array-CGH analysis, despite a suspected X-linked inheritance pattern (**Figure 1A**). They were part of the series of 100 patients with marfanoid habitus (MH) and an associated ID-phenotype published in 2013<sup>24</sup>. Proband III:1 was included in a cohort of 106 ID patients who underwent targeted HTS of 217 ID-genes and for whom we previously reported the sequencing results<sup>23</sup>. Patients from pedigree 2 (**Figure 3A**) and from pedigree 3

(**Figure 3B**) presented undiagnosed ID and underwent routine FXS testing and array-CGH analysis. They were both included in second series of 834 additional ID patients for whom we sequenced coding regions of 275-451 ID-genes. This study was approved by the local Ethics Committee of the Strasbourg University Hospital (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale -CCPPRB). For all patients, written informed consent for genetic testing was obtained from their legal representative. For all the patients presented in this paper, families also gave consent for the publication of photographs.

*Targeted high-throughput sequencing, bioinformatic pipeline and variant ranking*

DNA was prepared as described previously<sup>25</sup>. It was enriched in coding sequences of 275 to 451 genes known to cause ID, including *FMRI*, by target custom capture (SureSelect, Agilent). Libraries were prepared separately for each patient. Parental DNA was pooled untagged by 14-16 (mothers and fathers separately) and each pool was treated as one sample for the library preparation. The enriched libraries were tagged and pooled by 32 in one lane of an Illumina sequencer (HiSeq2500) for a 100bp paired-end run. Read mapping and variant calling were performed following standard procedures, and variants were annotated and filtered using VaRank *via* Alamut HT (Interactive Biosoftware)<sup>23,26</sup>. Candidate variants harboring a frequency compatible with the incidence of ID were selected<sup>27</sup>. We retrieved candidate variants reported in EVS (<http://evs.gs.washington.edu/EVS/>) or ExAC (<http://exac.broadinstitute.org/>) databases in no more than one male for variants in X-linked genes (as we cannot exclude the possibility, although unlikely, that a particular carrier from the general population might have mild ID). Pools of untagged parental DNA (mothers and fathers separately) were analyzed to check for the presence of each candidate variation. Remaining variants, predicted as

potentially pathogenic and matching the mode of inheritance associated with the affected gene, were further tested for validation. The potential effect of variants on splicing was analyzed using BDGP Splice Site Prediction by Neural Network (NnsplICE) <sup>28</sup>, MaxEntScan <sup>29</sup>, Human Splicing Finder <sup>30</sup> and GeneSplicer <sup>31</sup> *via* Alamut HT. Putative heterozygous/homozygous/hemizygous CNVs were highlighted using the previously described method based on a depth-of-coverage comparison between the index sample and eight other random samples from the same sequencing lane <sup>25</sup>. For the X-chromosome, coverage was normalized according to the patient's gender. Confirmation of *FMR1* punctual variants and analysis of the consequences of *FMR1* variants at the RNA and protein levels are detailed in **Supplementary methods**. Once confirmed, the three variants were submitted to the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>).

#### *Reverse clinical phenotyping*

All patients were reevaluated using the Hagerman checklist <sup>3</sup>. Patients from family 1 underwent clinical and behavioural evaluations after obtaining the molecular diagnosis, in order to determine if they were *a posteriori* compatible with the FXS phenotype. The index patient III:1 had a standardized assessment by a multidisciplinary team including a clinical geneticist, a psychiatrist and a neuropsychologist, in the presence of his referral caregiver, in order to provide a neurodevelopmental, psychocognitive, emotional and adaptive picture of his functioning. The clinical examination comprised a neurological evaluation, the examination of malformative or dysmorphic features (face, trunk, extremities and whole body, **Figure 1**). The interview concerned sleeping, feeding and digestive features. HistoFry data were collected using the Achenbach System of Empirically Based Assessment for Adult, ASEBA. The Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) was used. It included verbal and non-verbal tasks and four



index scores: the Verbal Comprehension Index (VCI), the Perceptual Reasoning Index (PRI), the Working Memory Index (WMI), and the Processing Speed Index (PSI). The Full Scale Intelligence Quotient (FSIQ) was determined, as were specific index skills, significant gaps between subtests, and more generally strengths and weaknesses of the intellectual cognitive profile. Patient II:2 was also assessed for social cognitions using the Social Perception domain of the Neuropsychological Assessment-Second Edition (NEPSY-II), with an analysis of the recognition of facial emotions. Finally, the Vineland Adaptive Behavior Scales-Second Edition (Vineland™-II, VABS) was used in a semi-structured interview with the referral caregiver, to determine ages of development in each area<sup>32</sup>. It assesses socialization (interpersonal relationships, games and leisure, coping), communication (receptive, expressive, and written), autonomy in daily living activities (personal, domestic, community) and motor skills (gross, fine). Sibling patients III:1 and 3 were phenotyped in their specialized residential institution in the presence of the care and educational team by a psychiatrist and a clinical research assistant trained in administering the Vineland Adaptive Behavior Scales-Second Edition (Vineland™-II, VABS).

#### *Literature and database review of FMRI variations*

We compiled the rare non-synonymous variations previously identified in the different studies which screened for *FMRI* coding sequences in patients with ID over the last twenty years. These studies can't be classified in three groups. The first group concerned studies conducted on a cohort of ID patients with a phenotype evocative of FXS but negative after CGG expansion testing<sup>16-20</sup>. *FMRI* coding regions were screened for punctual variations (using heteroduplex analysis, single strand conformation analysis, genomic or cDNA Sanger sequencing) either directly or only in patients in whom the FMRP protein was found to be absent (by Western Blot). The second category of studies

analyzed *FMRI* coding sequences in male ID patients, with no specific clinical orientation, using high resolution melting, Sanger or HTS)<sup>21-23</sup>. The third one included large-scale studies in which *FMRI* was sequenced together with other genes in unspecific ID (large panels, X-exome, exome, genome)<sup>33-41</sup>. For each non-synonymous variation, we compiled genetic and functional evidence of its pathogenicity and its frequency in general population database such as Exome Variant Server. The clinical neurodevelopmental and morphological data available for patients were also compiled and classified as highly, moderately, poorly or not evocative of FXS. HTS studies were also considered: 1,228 ID individuals from large ID panels including *FMRI*<sup>39-41</sup>, 627 patients who underwent X-exome sequencing<sup>35,36,38</sup> and 1,325 patients analyzed by WES<sup>33,34,37,42</sup>.

## Results

### *FMRI* exon 17 deletion previously identified in three brothers with ID leads to a truncated *FMRP* protein (family 1)

In family 1 (**Figure 1A**), deletion of the last exon of *FMRI* (exon 17) was previously detected by targeted HTS of 217 genes performed on patient III:1<sup>23</sup>. No reads were obtained for this exon in the patient whilst a high depth of coverage was observed for the other samples of the same sequencing batch (depth of coverage of 210 +/- 92 reads for males, **Table S1**). An array-CGH previously performed on the patient was reanalyzed *a posteriori* and confirmed this result, revealing two consecutive probes (below the detection threshold of three) deleted in this region (**Figure 2A**). The deletion of exon 17 was confirmed by PCR using primers located at either side of the exon revealing no amplification in the three affected brothers (III:1, III:2 and III:3) (**Figure 2B**). Genomic qPCR analysis revealed that this deletion was present at the heterozygous state in their

unaffected mother and was not found in their two maternal aunts, leading to a reassuring genetic counseling for the other family members (**Figure 2C**). RT-qPCR showed that *FMRI* RNA was expressed at a normal level in blood extract and lymphoblastoid cell lines from patient III:3 (**Figure 2D**), suggesting it escapes nonsense-mediated mRNA decay. We investigated the consequences of the loss of exon 17 and its acceptor splice site on *FMRI* mRNA. We observed retention of the beginning of intron 16 in *FMRI* mRNA from patient III:3 while this was not observed in mRNA from an unrelated individual 4 (**Figure 2E**). This insertion of intron 16 in the mRNA, r.1737\_1738ins1737+1\_?, leads to a frameshift resulting in a premature stop codon at position 589 instead of 633: p.Ile580fs\*9. Western blot analysis confirmed the existence of a truncated form of FMRP protein in patient's cells with a ~10% reduction in size (~75kDa vs. ~85 kDa) consistent with the loss of 44 amino acids (**Figure 2F**). The expression of this truncated form in patients was lower than the expression of normal FMRP protein observed in controls (~50% lower levels of truncated FMRP in lymphoblastoid cells from patients III:1 and III:3) suggesting reduced stability of the truncated protein.

The three affected brothers of this family experienced a diagnostic odyssey lasting many years. Patient III:1, aged 50, is a protected adult under tutorship, living in a specialized institution with patient III:3. He is the most severely affected of the siblings. He has a narrow, elongated face with a prominent forehead (**Figure 1**), macroorchidia and mild gynecomastia. When examined, he was non-verbal pronouncing only syllables and a few simple words. He suffered from currently controlled epilepsy. He presents behavioral problems (long-standing food selectivity, opposition attitudes, one episode of aggressiveness towards others and himself). Patient III:2, aged 48, presents moderate ID and is a protected adult under tutorship, living and working in a specialized institute.

He has a long, thin habitus in spite of standard measurements, and an elongated face with thick lower lip (**Figure 1**). He was described as fearful, avoided eye contact and had an anxious bearing. He did not express himself spontaneously and answered questions briefly in a few words. He had developed coping skills for daily life situations. Patient III:3, aged 47, was a protected adult under tutorship. He had a long, thin habitus in spite of standard measurements, a narrow, elongated face and prominent ears and forehead. He had macroorchidia. He was non-verbal. He did not present any aggressiveness towards others or self-injurious behavior. He developed social relationships despite some non-pervasive rituals and stereotypies. Complete neuropsychological evaluations were performed to assess the severity of impairments and their impact on intellectual functioning and adaptive behavior (Vineland, WAIS-IV, NEPSY-II, supplementary information and **Table 1**). Hagerman scores for the three brothers were calculated. Two of the three brothers had a score above 16 (**Table 2**).

*Identification of two novel intronic variants affecting FMR1 splicing (families 2 and 3)*

In the second family (**Figure 3A**), targeted HTS of 451 ID-genes revealed a substitution c.990+1 G>A, predicting to disrupt the splice donor site of exon 10, in one boy presenting sporadic ID. This substitution was not present in maternal blood DNA suggesting a *de novo* inheritance. Analysis of RNA from the patient's blood demonstrated that this variant led to an aberrant splicing event, skipping exon 10 in its entirety. The loss of this 110pb exon (r.881\_990del) leads to a frameshift after amino acid 294 and to the appearance of a premature stop codon at position 305: p.Lys295Asnfs\*11. The resulting truncated protein lacks essential domains such as the nuclear export signal and the RGG domain responsible for the interaction of FMRP with target mRNAs. The proband, aged 10 years, was the first child of healthy parents, with no other case of ID in the family. A

younger brother had normal development. The pregnancy and delivery were uneventful. His birth length (50cm), weight (3350g) and head circumference (34.5cm) were normal. His motor skills were delayed as he could not walk before the age of 24 months. At age 3 years, he had no language and normal hearing. At that time, he began to show features of repetitive behavior as well as hand flapping, and avoided physical contact. Physical examination revealed hypotonia with valgus feet and knees. He had a high forehead, a coarse face, a wide mouth with an everted lower lip. His ears were of normal size, but abnormally shaped (**Figure 3A**). He had clinodactyly of the 5<sup>th</sup> fingers, square feet with short toes and a shawl scrotum. At age 10 years, his height was at +1 SD, weight at -0.5 SD, and head circumference at the mean. His behavior was marked by autistic traits, making physical examination difficult. He did not have any neuropsychological evaluation. He needed to be educated in a specialized school. His shawl shaped scrotum could have suggested an Aarskog syndrome, but the lack of other major signs as facial dysmorphism and growth retardation excluded this diagnosis. Metabolic analyses, electroencephalogram and cerebral MRI were normal, as well as CGG-repeat test of the *FMRI* gene, *DMPK* (type 1 dystrophic myotonia) and *ARX* genes, and array-CGH. No pathogenic variant was found in *FGDI*. His Hagerman score was calculated as high (15/26) considering his young age (no testicular phenotype observable) and the absence of familial history (consistent with *de novo* status of the variant), but he presented no attention deficit disorder (**Table 2**).

In the third family, targeted HTS of 451 ID-genes revealed an intronic substitution c.420-8A>G (**Figure 3B**). This substitution was present in maternal blood. This variation was predicted to decrease the use of the normal splice acceptor site according to three prediction programs (score decrease for SpliceSite Finder: -100%, Genesplicer: -100% and MaxEnt: -52%, no change for NnsplICE and Human Splicing Finder). It also created a

cryptic intronic acceptor site with high prediction scores of being used (SpliceSite Finder: 83.8 [0-100], MaxEnt: 8.1 [0/16], NnsplICE: 1 [0-1], GenesplICE: 5 [0-15] and Human Splicing Finder: 84 [0-100]). Analysis of mRNA in the patient's blood demonstrated that this variant led to at least one aberrant splicing event (**Figure 3C**) with the use of this cryptic intronic acceptor site, leading to the retention of 7 nucleotides of intron 5 (r.419\_420ins420-7\_420-1). This leads to a frameshift after amino acid 140 with the appearance of a premature stop codon three amino acids downstream, p.Met140Ilefs\*3. The resulting truncated protein lacks essential domains such as the nuclear export signal and the RGG domain responsible for the interaction of FMRP with target mRNA. This proband, aged 20, was the second child of unrelated healthy parents. The older sister does not present any cognitive or health problems. The pregnancy was characterized by maternal depression without exposure to teratogens; the delivery was uneventful. His birth measurements and APGAR scores were normal (length: 52cm, weight: 3310g, head circumference: 33.5cm, APGAR 10/10). He presented a strabismus which was corrected by surgery at the age of 7 years. He had axial hypotonia in his first year and walked at the age of 23 months. Language delay was noticed without hearing impairment. He presented stereotypies and features of repetitive behavior such as rocking and hand flapping, and a high level of change-related anxiety. He went to a specialized school. At age 20.5 years, he had a normal height, weight and head circumference (176.6cm, 59.2kg and 57.5cm). A long face and mild clinodactyly of the 3rd and 5th fingers were noted. Karyotype, array-CGH, and the *FMR1* CGG expansion test were negative. Standard metabolic tests did not reveal any anomaly and investigations for *ARX* recurrent 24bp duplication or 22q11 microdeletion were returned negative. No X-inactivation bias was identified in the patient's mother. Brain MRI, performed at 21 months, did not reveal any significant abnormality: a mildly short splenium was noted. His Hagerman score was 14/26 (**Table 2**).

*Other FMR1 point variations reported in individuals with ID and in the general population*

Whereas it is clear that CGG-repeat expansion in the *FMR1* promoter is by far the most frequent cause of FXS, other types of variations, particularly *FMR1* deletions, have also been reported in FXS<sup>9</sup>. Deletions of the entire gene or of its first exon<sup>10-12</sup>, associated or not with full *FMR1* CGG expansion in the mother<sup>13-15</sup>, have been found by conventional CGG expansion testing or by microarray analysis. Since 1992, screening for variations in the *FMR1* coding sequence, performed in Fragile-X-evocative or nonspecific ID patients without expanded CGG repeats and subsequent methylation<sup>16-20,22,23</sup>, has led to the identification of several rare intragenic non-synonymous variations within the *FMR1* locus (**Table 3 and Table S2**). We have also compiled the rare non-synonymous variants identified in *FMR1* during large-scale studies (large panel, X-exome or whole exome sequencing) performed on ID patients by other teams: only missense variants but no truncating variants were identified in the 3,180 individuals tested<sup>33-42</sup>. We reanalyzed all these variants in *FMR1* in the light of new public sequencing data (Exome Variant Server and the ExAC project, supposed to be free or at least not enriched in severe developmental disorders), prediction tools and functional evidence. Altogether, besides the disease-causing truncating variants we detailed in this paper, six previously reported variants could also be reclassified as disease-causing (pathogenic-class 5 or likely pathogenic-class 4 variants, summarized in **Table 3** and in **Figure 4**). Half of them are truncating changes: (p.Thr125Leufs\*35, c.52-1\_52delinsTA and p.Ser27\*) and half of them are missense changes (p.(Phe126Ser), p.(Gly266Glu) and p.Ile304Asn), not reported in EVS nor ExAC, occurring *de novo* and/or and having a functional impact *in vitro* (neurons) or *in vivo* (knock-in mice)<sup>16-19</sup>. For the other missense variants reported, some are predicted to be damaging (SIFT, Polyphen2), but neither the amino acid changes by themselves nor the

phenotypes of patients make it possible to draw any clear conclusions about their pathogenicity and we thus considered them “variants of uncertain significance” (VUS)<sup>11,20-22</sup>. Some missense variants were even classified as “likely benign”. One of them, p.Asp438Asn, predicted to affect the nuclear localization signal of *FMRI*, was identified by targeted HTS of 217 ID genes in two brothers with no morphological symptoms of FXS but with FXS neurodevelopmental phenotype: further familial segregation analysis showed that one unaffected half-brother also carries this variation too thus suggesting that it is a benign variant<sup>23</sup>. Besides these non-synonymous variations, an intronic variant, c.990+14C>T, first described as altering the splicing of exon 10 was reported twenty years ago in three unrelated ID males<sup>20</sup>, but this result was controversial since this variation was also identified in healthy control individuals<sup>43</sup>. Data from ExAC confirmed the polymorphic status of this variation, present at the hemizygous state in more than 4,000 males of this general population.

In this ExAC general population, three variations predicted to affect splice sites are reported at the hemizygous state: a recurrent disruption of the exon 12 splice acceptor site present in 44 hemizygous males, and two unique substitutions of either the exon 12 splice donor site or the exon 16 splice acceptor site. The skipping of these exons is thought to lead to in-frame deletions in FMRP protein sequence (21 amino acids lost for exon 12 and 8 amino acids for exon 16). Moreover, exon 12 is found alternatively spliced and included in only some of the different isoforms. This suggests that exon 12 and its encoded amino acids (a.a. 376-396 in NM\_002024.5) might be dispensable. A total of 29 missense variants affecting the longest isoform NM\_002024.5 of *FMRI* predicted to have benign or damaging effects, are also reported at the hemizygous state in 236 males of the ExAC population (**Table S3**). Therefore, data from this general population underline that one



should interpret non-synonymous variations identified in the *FMRI* gene with caution, even if they affect conserved amino acids and are predicted as damaging, and especially in patients with unspecific ID not clinically evocative of FXS.

## Discussion

We report here the cases of three brothers with ID of presumed X-linked inheritance and a phenotype compatible with FXS and two sporadic male patients from other families, all negative for routine FXS testing. In the first family, deletion of the last exon of *FMRI* (exon 17) detected by targeted HTS of 217 genes performed on index patient<sup>23</sup> was confirmed in the two affected brothers by PCR. This deletion led to a decreased expression level of a truncated form of FMRP. These are the first cases of a single exon deletion not comprising the 5'UTR or exon 1 of the *FMRI* gene, and therefore the first deletion to escape both expansion testing and CNV search by array-CGH. This illustrates the power of HTS in identifying small Copy Number Variants (CNV). In the second and third families, variants that altered splicing and led to a frameshift and probably an early truncated protein were identified. They were *de novo* and maternally-inherited, respectively.

FXS is almost exclusively diagnosed by molecular testing for CGG repeat expansion. Indeed, as FXS is not always clinically obvious, especially for prepubertal children, the CGG-repeat test, which is cheap and fast, is the standard of care for any child not meeting developmental milestones and is one of the most frequently ordered genetic tests. Among all children with developmental delay, the actual diagnostic yield of FXS testing is only 1–2%<sup>4</sup>. CGG-repeat expansion in the *FMRI* promoter is the most frequent cause of FXS, but deletions of the entire *FMRI* gene<sup>10-12</sup> or of the 5'UTR have also been

reported in FXS <sup>9</sup>. Rare intragenic non-synonymous variations have also been identified within the *FMRI* locus <sup>16-20,22,23,35,41,42</sup> but after reanalysis in the light of new public sequencing data, prediction tools and functional evidence, we showed that only six could be reclassified as convincing.

Altogether, in the cohort of 940 French individuals with nonspecific ID and a negative *FMRI* CGG expansion test, three intragenic pathogenic *FMRI* variants were identified, thus accounting for a non-negligible proportion of ID cases (0.3%). However, if we consider the other HTS studies performed on non-specific ID-patients, the mutation yield is lower: only one convincing disease-causing missense variant p.(Phe126Ser) in 3,180 individuals with ID (plus four variants of unknown significance). Therefore, except for recurrent CGG expansion, the *FMRI* gene is infrequently mutated in ID as compared to other genes of the same or a smaller size, such as *RPS6KA3* or *MECP2* <sup>44</sup>.

The six patients reported in the literature carrying convincing pathogenic variants in *FMRI* mainly come from series of patients with clinical features evocative of FXS. They all present a long face and large ears, and except for the prepubertal patient, macroorchidism. We describe here five other patients with unexplained ID who were negative for Fragile-X testing of the CGG expansion carrying *FMRI* changes. Fine reverse phenotyping of the three adult siblings of the first family carrying the exon 17 deletion revealed the dysmorphism and behavioral phenotype of the FXS spectrum. They present various dysmorphic features that have been reported in FXS, although more typical in patient III:2 than in his siblings, and two of them had macroorchidism. Although non-specific, their cognitive evaluations were also compatible with FXS. Two of the brothers obtained high scores for the FXS evaluation established by Hagerman *et al.* (17/26 and 18/26 for individual III:1 and III:3 respectively), corresponding to what is observed in

patients carrying the CGG repeat expansion (mean score of 17.3)<sup>3</sup>. The third brother had a lower score (14/26), as his dysmorphic features were less pronounced, but his score was still above the mean score obtained for non-FXS ID patients<sup>3</sup>. Reverse phenotyping was more difficult in the young boy from the second family and the young adult of the third family, since the dysmorphic features and macroorchidism tend to appear or increase with age, but they had some neuropsychological features, especially autistic traits, consistent with FXS. In contrast, they both presented without hyperactivity and without macroorchidism and had moderately high Hagerman scores (15/26 and 14/26 respectively). It was particularly difficult to suspect FXS in the patient of the second family as, due to the *de novo* occurrence of this variant, there was no history of ID in his family. Together, considering the five patients described here and the six from the literature carrying pathogenic variants in *FMR1*, we found no obvious clinical differences between patients with intragenic changes compared and those with classic CGG expansions. The two boys of the second and third families present a milder phenotype than do the three affected brothers of the first family, even though they carry variants leading to more prematurely truncated FMRP protein, lacking several essential functional domains. This study confirmed that other types of *FMR1* changes, besides classic CGG expansions, which could not be diagnosed through standard tests, may cause an FXS phenotype.

In conclusion, this report enlarges the molecular description of FXS, with the description of five individuals carrying intragenic pathogenic variants of *FMR1*. In total, nine pathogenic non-synonymous changes (loss-of-function and missense variants) are now described in *FMR1* in eleven individuals with a compatible Fragile-X phenotype. We demonstrate the interest of HTS in identifying *FMR1* intragenic deletions and punctual variants in ID patients with a negative standard work-up. The increasing use of HTS in

CGG-expansion negative patients with ID will lead to the identification of all types of variants (SNVs and CNVs) affecting *FMRI* coding sequences and determine the true frequency of *FXS* in ID. However, classical CGG expansion will still need to be tested separately, as current HTS methods fail to detect them. Making a diagnosis of *FXS* is of particular importance at two levels. First, genetic counseling for patients' families is essential due to the high risk of recurrence of this dominant X-linked form of ID. Second, numerous targeted therapeutics for *FXS* are being developed<sup>45,46</sup>, and some *FXS* patients are already benefiting from some of these treatments<sup>47,48</sup>.

### **Acknowledgements**

The authors thank the families for their participation to the study. The authors also thank the French Ministry of Health (PHRC national 2008) as well as the Regional Council of Burgundy, the Fondation Jerome Lejeune, the Agence de la Biomédecine, the Agence Nationale de la Recherche (ANR-10-IDEX-0002-02) for their financial support. They thank Ricardos Tabet and people from the IGBMC sequencing platform for their technical and bioinformatic support. The authors also thank the Centre National de Génotypage (Jean-François Deleuze, Robert Olasso, Anne Boland and the technicians and bioinformaticians) for their participation in library preparation and DNA sequencing. We are very grateful to Philip Bastable for language editing.

### **Conflict of Interest Statement**

The authors declare no conflict of interest

## References

1. Boyle L, Kaufmann WE: The behavioral phenotype of FMR1 mutations. *Am J Med Genet C Semin Med Genet* 2010; **154C**: 469-476.
2. Tranfaglia MR: The psychiatric presentation of fragile x: evolution of the diagnosis and treatment of the psychiatric comorbidities of fragile X syndrome. *Dev Neurosci* 2011; **33**: 337-348.
3. Hagerman RJ, Amiri K, Cronister A: Fragile X checklist. *Am J Med Genet* 1991; **38**: 283-287.
4. Rauch A, Hoyer J, Guth S *et al*: Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A* 2006; **140**: 2063-2074.
5. Bassell GJ, Warren ST: Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron* 2008; **60**: 201-214.
6. Comery TA, Harris JB, Willems PJ *et al*: Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. *Proc Natl Acad Sci U S A* 1997; **94**: 5401-5404.
7. Oberle I, Vincent A, Abbadi N *et al*: New polymorphism and a new chromosome breakpoint establish the physical and genetic mapping of DXS369 in the DXS98-FRAXA interval. *Am J Med Genet* 1991; **38**: 336-342.
8. Jacquemont S, Birnbaum S, Redler S, Steinbach P, Biancalana V: Clinical utility gene card for: fragile X mental retardation syndrome, fragile X-associated tremor/ataxia syndrome and fragile X-associated primary ovarian insufficiency. *Eur J Hum Genet* 2011; **19**.
9. Wells RD: Mutation spectra in fragile X syndrome induced by deletions of CCG\*CCG repeats. *J Biol Chem* 2009; **284**: 7407-7411.
10. Coffee B, Ikeda M, Budimirovic DB, Hjelm LN, Kaufmann WE, Warren ST: Mosaic FMR1 deletion causes fragile X syndrome and can lead to molecular misdiagnosis: a case report and review of the literature. *Am J Med Genet A* 2008; **146A**: 1358-1367.
11. Luo S, Huang W, Xia Q, Du Q, Wu L, Duan R: Mutational analyses of the FMR1 gene in Chinese pediatric population of fragile x suspects: low tolerance for point mutation. *J Child Neurol* 2015; **30**: 803-806.
12. Vengoechea J, Parikh AS, Zhang S, Tassone F: De novo microduplication of the FMR1 gene in a patient with developmental delay, epilepsy and hyperactivity. *Eur J Hum Genet* 2012; **20**: 1197-1200.
13. Hammond LS, Macias MM, Tarleton JC, Shashidhar Pai G: Fragile X syndrome and deletions in FMR1: new case and review of the literature. *Am J Med Genet* 1997; **72**: 430-434.
14. Luo S, Huang W, Xia Q *et al*: Cryptic FMR1 mosaic deletion in a phenotypically normal mother of a boy with fragile X syndrome: case report. *BMC Med Genet* 2014; **15**: 125.
15. Viveiros MT, Santos MD, Dos Santos JM *et al*: Screening for fragile X syndrome in males from specialized institutions in the northeast region of Brazil. *Genet Mol Res* 2015; **14**: 6897-6905.

16. Gronskov K, Brondum-Nielsen K, Dedic A, Hjalgrim H: A nonsense mutation in FMR1 causing fragile X syndrome. *Eur J Hum Genet* 2011; **19**: 489-491.
17. Lugenbeel KA, Peier AM, Carson NL, Chudley AE, Nelson DL: Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. *Nat Genet* 1995; **10**: 483-485.
18. De Boulle K, Verkerk AJ, Reyniers E *et al*: A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat Genet* 1993; **3**: 31-35.
19. Myrick LK, Nakamoto-Kinoshita M, Lindor NM, Kirmani S, Cheng X, Warren ST: Fragile X syndrome due to a missense mutation. *Eur J Hum Genet* 2014; **22**: 1185-1189.
20. Wang YC, Lin ML, Lin SJ, Li YC, Li SY: Novel point mutation within intron 10 of FMR-1 gene causing fragile X syndrome. *Hum Mutat* 1997; **10**: 393-399.
21. Collins SC, Bray SM, Suhl JA *et al*: Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. *Am J Med Genet A* 2010; **152A**: 2512-2520.
22. Handt M, Epplen A, Hoffjan S, Mese K, Epplen JT, Dekomien G: Point mutation frequency in the FMR1 gene as revealed by fragile X syndrome screening. *Mol Cell Probes* 2014; **28**: 279-283.
23. Redin C, Gerard B, Lauer J *et al*: Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *J Med Genet* 2014; **51**: 724-736.
24. Callier P, Aral B, Hanna N *et al*: Systematic molecular and cytogenetic screening of 100 patients with marfanoid syndromes and intellectual disability. *Clin Genet* 2013; **84**: 507-521.
25. Redin C, Le Gras S, Mhamdi O *et al*: Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alstrom syndromes. *J Med Genet* 2012; **49**: 502-512.
26. Geoffroy V, Pizot C, Redin C *et al*: VaRank: a simple and powerful tool for ranking genetic variants. *PeerJ* 2015; **3**: e796.
27. Piton A, Redin C, Mandel JL: XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am J Hum Genet* 2013; **93**: 368-383.
28. Reese MG, Eeckman FH, Kulp D, Haussler D: Improved splice site detection in Genie. *J Comput Biol* 1997; **4**: 311-323.
29. Eng L, Coutinho G, Nahas S *et al*: Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Hum Mutat* 2004; **23**: 67-76.
30. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C: Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: e67.

31. Pertea M, Lin X, Salzberg SL: GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001; **29**: 1185-1190.
32. Sparrow SS, Cicchetti DV: The behavior inventory for rating development (BIRD): assessments of reliability and factorial validity. *Appl Res Ment Retard* 1984; **5**: 219-231.
33. de Ligt J, Willemsen MH, van Bon BW *et al*: Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 2012; **367**: 1921-1929.
34. Hamdan FF, Srour M, Capo-Chichi JM *et al*: De novo mutations in moderate or severe intellectual disability. *PLoS Genet* 2014; **10**: e1004772.
35. Hu H, Haas SA, Chelly J *et al*: X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Mol Psychiatry* 2016; **21**: 133-148.
36. Philips AK, Siren A, Avela K *et al*: X-exome sequencing in Finnish families with intellectual disability--four novel mutations and two novel syndromic phenotypes. *Orphanet J Rare Dis* 2014; **9**: 49.
37. Rauch A, Wieczorek D, Graf E *et al*: Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 2012; **380**: 1674-1682.
38. Tarpey PS, Smith R, Pleasance E *et al*: A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet* 2009; **41**: 535-543.
39. Martinez F, Caro-Llopis A, Rosello M *et al*: High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. *J Med Genet* 2016.
40. Tzschach A, Grasshoff U, Beck-Woedl S *et al*: Next-generation sequencing in X-linked intellectual disability. *Eur J Hum Genet* 2015; **23**: 1513-1518.
41. Grozeva D, Carss K, Spasic-Boskovic O *et al*: Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Hum Mutat* 2015; **36**: 1197-1204.
42. Wright CF, Fitzgerald TW, Jones WD *et al*: Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* 2015; **385**: 1305-1314.
43. Vincent JB, Gurling HM: Point mutation in intron 10 of FMR1 is unlikely to be a cause of fragile X syndrome. *Hum Mutat* 1998; **12**: 431-432.
44. Chelly J, Mandel JL: Monogenic causes of X-linked mental retardation. *Nat Rev Genet* 2001; **2**: 669-680.
45. Hagerman RJ, Des-Portes V, Gasparini F, Jacquemont S, Gomez-Mancilla B: Translating molecular advances in fragile X syndrome into therapy: a review. *J Clin Psychiatry* 2014; **75**: e294-307.
46. Jacquemont S, Berry-Kravis E, Hagerman R *et al*: The challenges of clinical trials in fragile X syndrome. *Psychopharmacology (Berl)* 2014; **231**: 1237-1250.
47. Greiss Hess L, Fitzpatrick SE, Nguyen DV *et al*: A Randomized, Double-Blind, Placebo-Controlled Trial of Low-Dose Sertraline in Young Children With Fragile X Syndrome. *J Dev Behav Pediatr* 2016; **37**: 619-628.
48. Leigh MJ, Nguyen DV, Mu Y *et al*: A randomized double-blind, placebo-controlled trial of minocycline in children and adolescents with fragile x syndrome. *J Dev Behav Pediatr* 2013; **34**: 147-155.

### Legends to figures

**Figure 1: Pedigree (A) and photographs of the first family of three affected brothers with ID** (Patient III:1 B, E, F, patient III:2 C, G, H, patient III:3 D, I, J).

**Figure 2: Identification of a deletion of exon 17 of *FMRI* leading to a truncated FMRP protein in the three brothers (first family)**

(A) Location of the deleted array-CGH probes (B) PCR amplification of *FMRI* exon 14 and exon 17 (Neg: experiment negative control; Ctrl-1M: unrelated male individual 1) (C) relative quantification of genomic DNA by qPCR amplification of exon 17 (Ctrl-2F: unrelated female individual 2) (D) Relative quantification of *FMRI* mRNA by qPCR analysis on blood cDNA and on mRNA extracted from lymphoblastoid cell lines (n=6) (Ctrl-3M and 4M: unrelated males individuals 3 and 4) (E) PCR amplification of blood cDNA using a reverse primer located at the beginning of intron 16 (F) Western blot analysis of FMRP protein expression on lymphoblastoid cell lines from FXS, patients III:1 and III:3 and unrelated male control individual 3. Normal size of FMRP protein is indicated by the dotted line. Semi-quantification of FMRP protein level (normalized to GAPDH) shows a reduction in FMRP expression in patients' cell lines compared to controls (n=5). (Student's t test: ns: non-significant; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001).

**Figure 3: Identification of two novel *FMRI* splice variants in two boys with sporadic ID (second and third families).**

(A) Pedigree and photograph of the patient from the second family carrying a *de novo* c.990+1G>A variant which affects the donor splice site of exon 10 and leads to the skipping of the entire exon in *FMRI* mRNA. (B) Pedigree and photograph of the patient

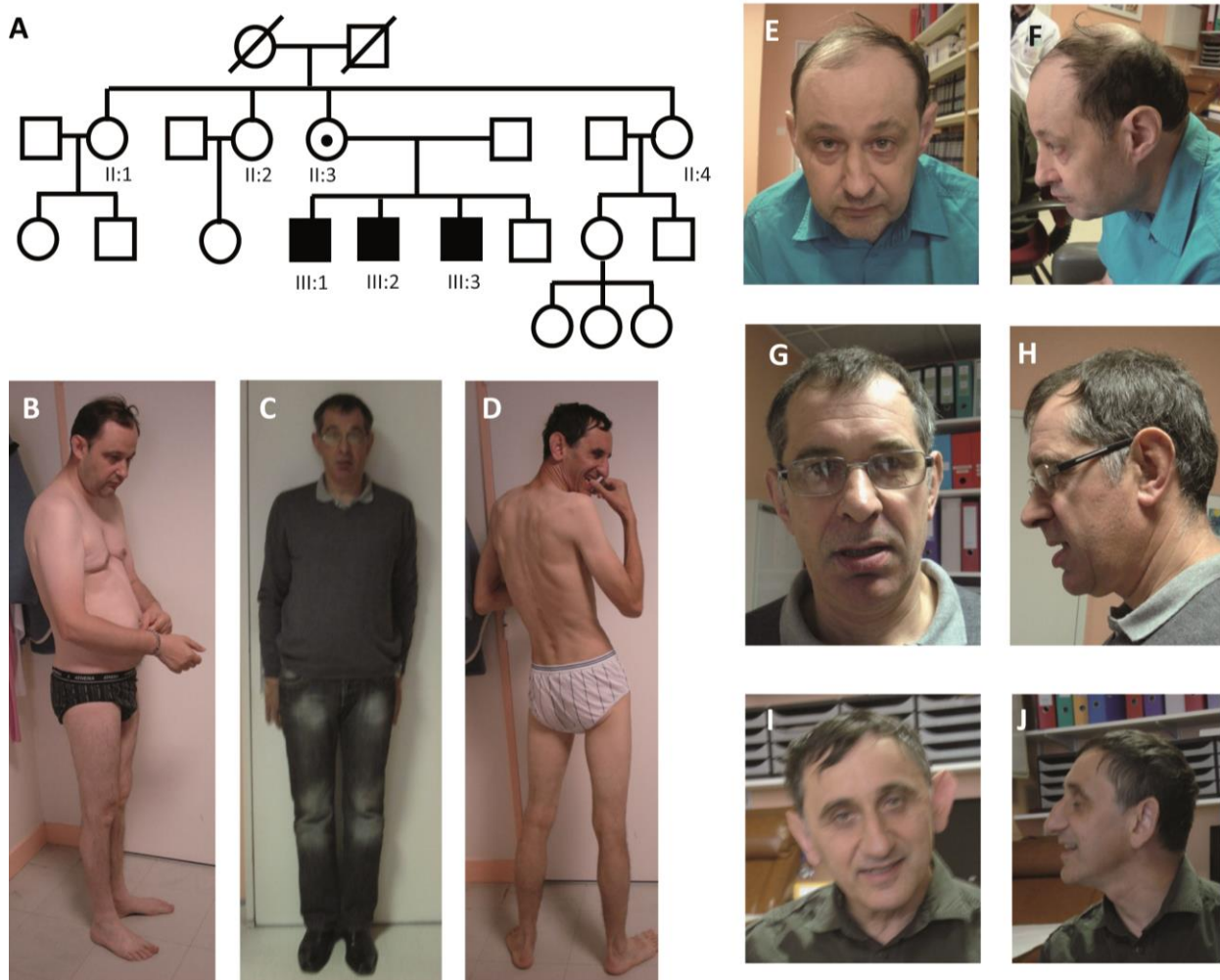


from the third family carrying a c.420-8A>G variant altering the splice acceptor site of exon 6 and leading to the use of alternative intronic cryptic sites.

**Figure 4: Representation of the FMRP protein with the different *FMRI* pathogenic variants and variants of unknown significance identified in ID patients.**

NLS: nuclear localization signal, KH: K homology domain; NES: nuclear export signal;  
RGG: glycine-arginine rich domain.

**Figure 1**



**Figure 2**

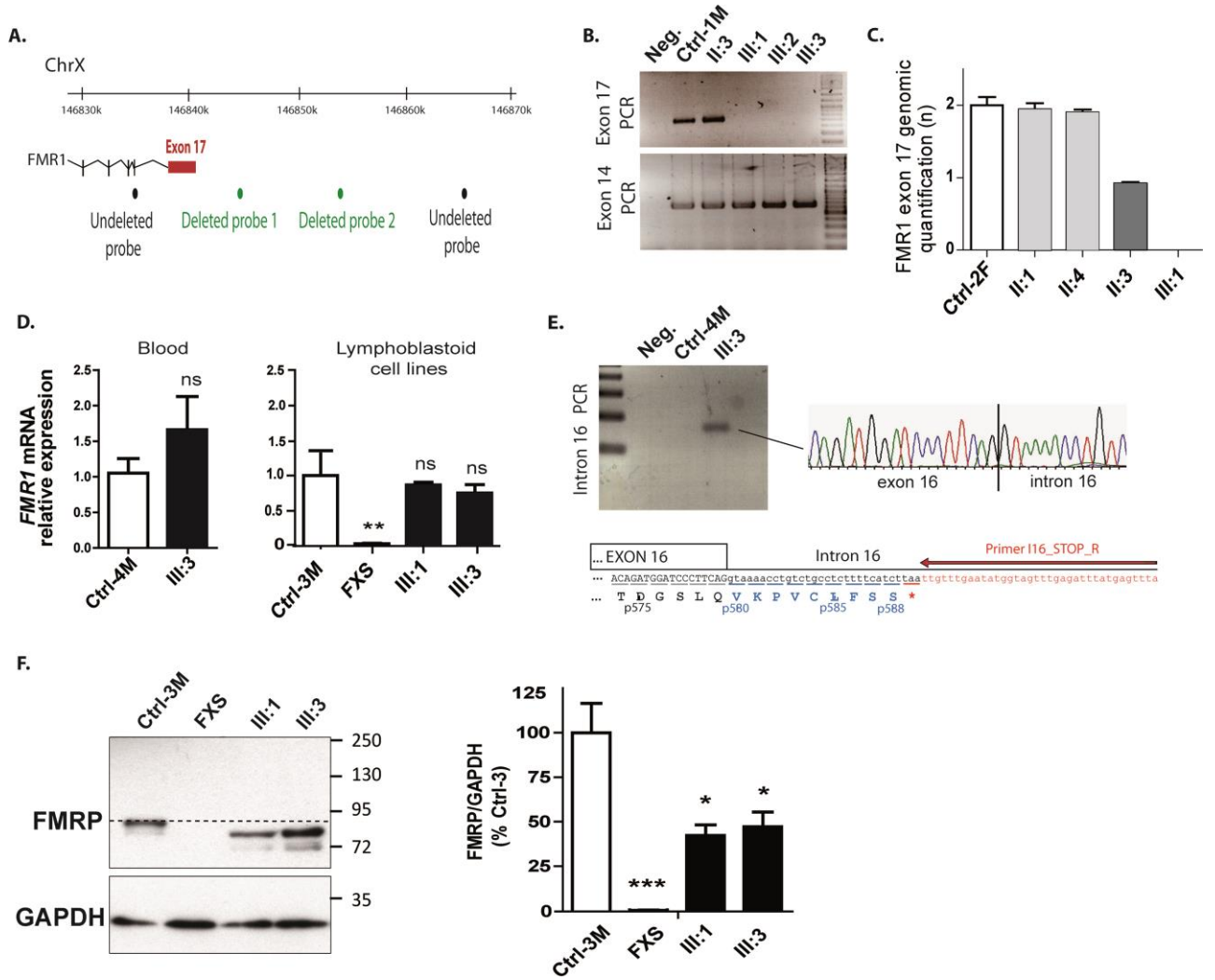


Figure 3

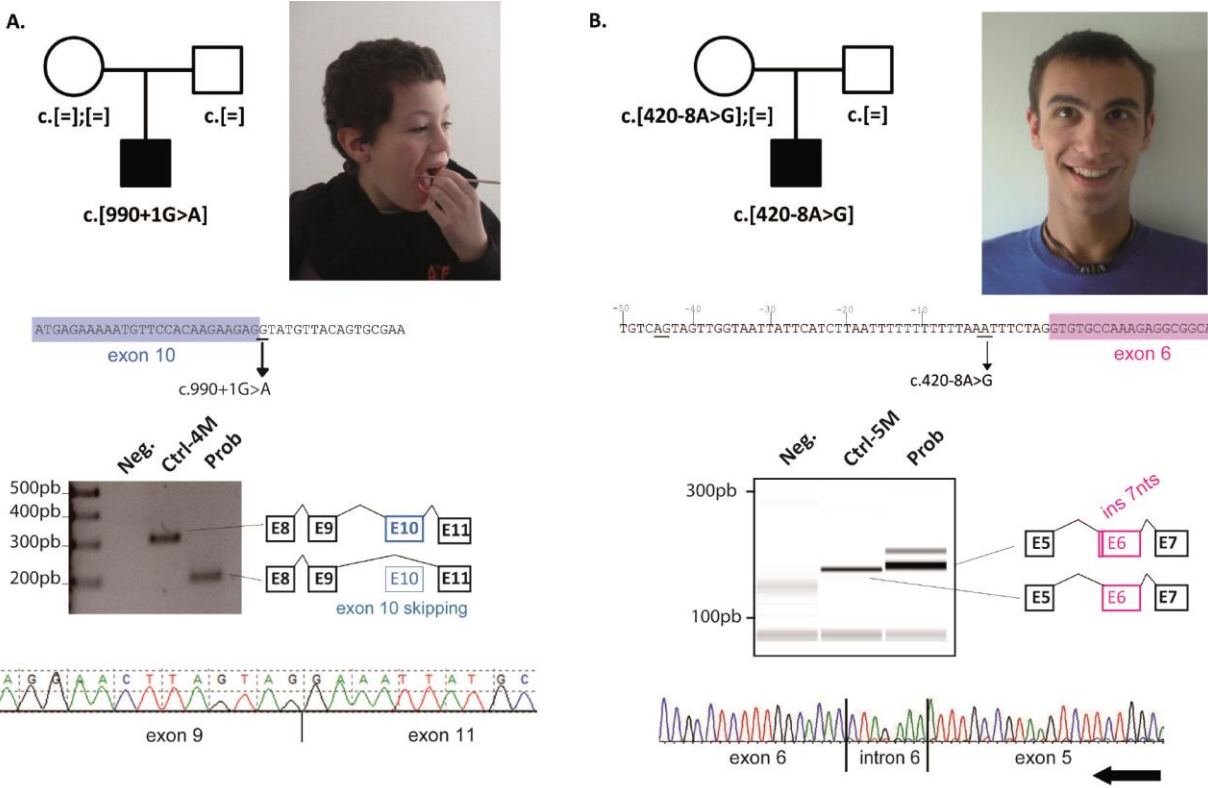
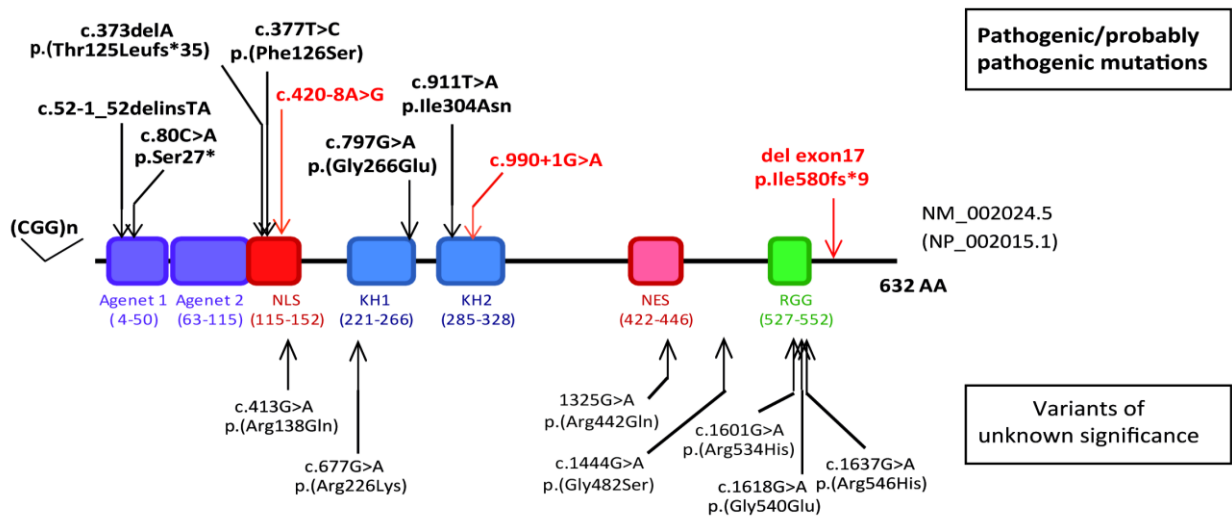


Figure 4



**Table 1. Results of Vineland scales in the four domains, communication, autonomy, socialization and motricity in the three siblings of family 1.**

	<b>III:1</b>	<b>III:2</b>	<b>III : 3</b>
<b>Communication</b>	<b>1 year 6 months</b>	<b>3 years 4 months</b>	<b>1 year 6 months</b>
Receptive	1 year 4 months	3 years 11 months	1 year 6 months
Expressive	1 year 8 months	3 years 3 months	1 year 7 months
Written	1 year 6 months	<3 years	1 year 6 months
<b>Autonomy</b>	<b>3 years 7 months</b>	<b>6 years 2 months</b>	<b>4 years 6 months</b>
Personal	3 years 7 months	7 years 8 months	4 years 6 months
Familial	4 years 1 month	6 years 5 months	4 years 9 months
Social	3 years 1 month	5 years 6 months	4 years 1 month
<b>Socialization</b>	<b>1 year 6 months</b>	<b>3 years 1 month</b>	<b>3 years 3 months</b>
Interpersonal relationships	1 year 11 months	2 years 4 months	2 years 11 months
Games and leisure	<1 year	1 year 5 months	8 years
Adaptation	4 years 9 months	1 year 4 months	7 years 5 months
<b>Motricity</b>	<b>2 years</b>	<b>5 years 3 months</b>	<b>2 years 7 months</b>
Global	1 year 10 months	>5 years 11 months	2 years 2 months
Fine	1 year 4 months	5 years 2 months	3 years

**Table 2. Hagerman scores for the five patients (according to Hagerman *et al.*, 1991)**

	<i>Patient III:1</i>	<i>Patient III:2</i>	<i>Patient III:3</i>	<i>Proband</i>	<i>Proband</i>
	<i>Family 1</i>	<i>Family 1</i>	<i>Family 1</i>	<i>Family 2</i>	<i>Family 3</i>
Intellectual deficiency	2	2	2	2	2
Hyperactivity	1	1	2	0	0
Short attention span	2	2	2	0	2
Tactilely defensive	2	1	1	2	1
Hand flapping	0	0	0	2	2
Hand biting	0	0	0	2	2
Poor eye contact	2	2	2	2	2
Perseverative speech	2	1	2	0	2
Hyperextensible MP joints	1	1	1	2	1
Large or prominent ears	1	1	2	1	0
Large testicles	2	1	2	0	0
Simian crease or sydney line	0	0	0	2	0
Family history of ID	2	2	2	0	0
<b>Total score</b>	<b>17</b>	<b>14</b>	<b>18</b>	<b>15</b>	<b>14</b>

0: not present; 1: borderline or present in the past; 2: definitely present

**Table 3. Review of the intragenic *FMRI* variants previously identified in ID patients**

Variation	Exon affected	Reference	Pathogenicity	Inheritance	FXS phenotype
c.52-1_52delinsTA; p.?	Exon 2	17	Pathogenic	Maternally inherited	+++
c.80C>A ; p.Ser27*	Exon 2	16	Pathogenic	Maternally inherited	++/+++
c.373delA ; p.Thr125Leufs*35	Exon 5	17	Pathogenic	<i>de novo</i>	++
c.420-8A>G ; p.Met140Ilefs*3	Exon 6	<i>this report</i>	Pathogenic	Maternally inherited	++
c.990+1G>A; p.Lys295Asnfs*11	Exon 10	<i>this report</i>	Pathogenic	<i>de novo</i>	++
g.(?_147030199)_(147046357_?) del; p.Ile580fs*9	Exon 17	<sup>23</sup> , <i>this report</i>	Pathogenic	Maternally inherited	+++
c.377T>C ; p.(Phe126Ser)	Exon 5	42	Likely Pathogenic	<i>De novo</i>	NA
c.797G>A; p.(Gly266Glu)	Exon 8	19	Likely Pathogenic	NA	+++
c.911T>A; p.Ile304Asn	Exon 10	18	Likely Pathogenic	<i>de novo</i>	++
c.413G>A; p.(Arg138Gln)	Exon 5	21	Unknown Significance	NA	-
c.677G>A; p.(Arg226Lys)	Exon 8	41	Unknown Significance	NA	NA
c.1325G>A; p.(Arg442Gln)	Exon 14	42	Unknown Significance	Maternally inherited	NA
c.1444G>A; p.(Gly482Ser)	Exon 14	22	Unknown Significance	NA	NA
c.1601G>A; p.(Arg534His) (2x)	Exon 15	22	Unknown Significance	Maternally inherited Maternally inherited	NA +
c.1618G>A; p.(Gly540Glu)	Exon 15	35	Unknown Significance	NA	NA
c.1637G>A; p.(Arg546His)	Exon 15	20	Unknown Significance	NA	NA

FXS: Fragile-X syndrome; Phenotype highly (+++), moderately (+), poorly (+) or not (-) evocative of a FXS.  
N.A.: clinical data not available. Nomenclatures c. are indicated in NM\_002024.5 isoform

## **Abbreviations**

Fragile X Syndrome (FXS), Fragile X Mental Retardation Protein (FMRP), Intellectual Disability (ID), Autism Spectrum Disorder (ASD), High Throughput sequencing (HTS), Full Scale Intelligence Quotient (FSIQ), Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), Processing Speed Index (PSI), Copy Number variant (CNV)

## Supplementary information

Patient III:1, aged 50, was a protected adult under tutorship, and lived in a specialized institute with one of his brothers (patient III:3). He was the most severely affected of siblings. He was non-verbal, pronounced syllables and few simple words in echolalia. He succeeded to make himself understood within the institution, in spite of important socio adaptive and communication difficulties in community. He suffered from currently controlled epilepsy which has been in the past twice complicated by status epilepticus. The caregiver mentioned one episode of tantrum crisis with destructive behavioral disorder, aggressiveness towards others and self-injurious which necessitated admission in a psychiatric hospital. His treatment combined antipsychotic, anxiolytic and sedative drugs. Long-standing food selectivity was reported, now exacerbated by opposition attitudes. These behaviors seemed to be part of a withdrawal movement with hypersomnia and clinophilia, in a context of suspected endogenous mood vulnerability. Positioned on the sidelines of peer group, he might be both rigid and permeable to environment. He was apathetic during testing. Beyond asthenia, repetitive yawning seemed to correspond to behavioral stereotypies. He expressed little facial emotion and had poor eye contact. Wechsler scales could not be carried due to severity of ID. Vineland scales highlighted a deep deficit of global and adaptive functioning despite of psychosocial rehabilitation support. Autonomy domain was the most developed subdomain. Communication and socialization were collapsed (**Table S2**). Patient III:2, aged 48, was described as fearful and in withdrawal towards peers, vulnerable to surrounding excitement, moody and quick tempered. He avoided eye contact and had anxious bearing. He did not express himself spontaneously, answered questions briefly in a few words, and was sensitive to encouragement and reinsurance. He had developed coping skills for daily life situations. The neuropsychological standardized assessment allowed us to determine a Full Scale Intelligence Quotient (FSIQ) of 44 on the WAIS-IV. The four index scores were



significantly lower than standard level (Verbal Comprehension Index (VCI) = 61, Perceptual Reasoning Index (PRI) = 56, Working Memory Index (WMI) = 50, Processing Speed Index (PSI) = 50) and the cognitive profile was marked by significant dissociation between VCI and WMI. Thus, reasoning subtests (similarities, matrix, and visual puzzles) were more successful than visual attention and working memory items (impaired selection of relevant visual data, disrupted anterograde and retrograde digit span respectively 2 and 0). VABS-II highlighted a global adaptive behaviour delay in the four tested areas (communication, autonomy, socialization and motor skills), but with heterogeneous results. Indeed, autonomy in every day's life was the most developed domain thanks to effective strategies. Communication and socialization were less developed (**Table S2**). He acquired most of gross and fine assessed motor skills. Integration of both Wechsler and Vineland scores enabled the diagnosis of moderate ID by combining intellectual functioning and global adaptive profiles. This phenotype was narrowed through targeted supplementary testing focused on academic learning, graphics, social cognitions and psycho affective development. A large deficit of social recognition of facial emotions was observed on Social Perception domain of NEPSY-II. Patient III:3, aged 47, was a protected adult under tutorship. He was non-verbal and pronounced unintelligible series of syllables and words in echolalia. He did not present any aggressiveness towards others nor self-injurious disorder. He adopted satisfactory institutional life peer group adjustment and integration, investing occupational and mediation activities, maintaining special relationships, despite of some non-pervasive rituals and stereotypies. His psychotropic treatment combined selective serotonin reuptake inhibitor anti-depressant, antipsychotic, anxiolytic and sedative drugs. Sleeping and feeding could be moderately disrupted during anxio-depressive phases. He also suffered from endogenous mood vulnerability but disinhibition remained accessible to framework. During the interview, worried and introverted behavior was replaced by a

pleasant contact and a smiling face. Wechsler scales could not be carried due to severity of ID. Vineland scales highlighted a deep deficit of global and adaptive functioning despite of psychosocial rehabilitation support. Like his brothers, autonomy was the highly developed domain (**Table S2**). Socialization scores were better than communication scores.

## **Supplementary methods**

### *Array-CGH*

Before targeted NGS, an array-CGH analysis had been performed for patient III:1. A custom-designed array was designed with approximately 240,000 oligonucleotides manufactured by Diagnogene™, (Division Imaxio, Biopôle Saint-Beauzire, France) for the purpose of a research project aiming at deciphering the molecular and cytogenetic bases of marfanoid habitus and ID<sup>1</sup>. This array contained oligonucleotides selected from the Agilent online library (earray; <https://earray.chem.agilent.com/earray>) with increased probe density within 7 selected genes. The entire genome was covered with an average resolution of 20kb (NCBI, hg18). The procedures of array-CGH were performed according to the Agilent instructions with minor modifications<sup>1</sup>. Slides were scanned using the Agilent G2565 Microarray scanner and images quantified using Agilent features extraction (v9.0) and a graphical overview was obtained using Genomic Workbench software (v5.0). Mapping data were analyzed on the human genome sequence using Ensembl ([www.ensembl.org](http://www.ensembl.org); hg19). CNV were assessed using the Database of Genomic Variants (<http://projects.tcag.ca/variation/>). CNV were considered starting from 3 or more deviating probes.

*Sanger sequencing validation*

The genomic and mRNA references used for *FMRI* in this paper are NG\_007529.2 and NM\_002024.5 respectively. c.990+1G>A and c.420-8A>G variants were confirmed by PCR amplification (FMR1\_E6F 5'-CCTGTAGTGTAGAGTGGTCCCTGAAT-3' and FMR1\_E6R 5'-ACAAGCAGTGCTAGGGAAAGTGAC-3'; FMR1\_E10F 5'-AGTAGG GCTGTGCTTACTGCTTTG-3' and FMR1\_E10R 5'-TGCCATTAACACTCGACCCA GTGA-3') followed by Sanger sequencing in the two patients and their mothers. Pedigree concordance was checked using highly polymorphic microsatellite markers (PowerPlex 16 HS System, Promega, Madison, WI, USA) or by analysing parental pools for other rare variations.

*FMRI genomic amplification*

PCR with specific primers to amplify *FMRI* exon 14 (5'CCCCAAAATAGCATATCTGTGCCTTAGC3' and 5'ACCCACTGTGAATTGGAA ACCATGT3') and exon 17 (5'TGGATGCAGTGAGATGACCAGTGT3' and 5'GGTGAATGATCACCCAATGCAGGAA3') were performed on blood DNA for the three brothers, the mother and an unrelated male individual 1 (Ctrl-1M). Quantitative PCR using exon 17 specific primers (5'-TGCGCACGGGTAAAGATCGTAA-3' and 5'-AGGGTACTCCATTCACGAGTGGTT-3') were performed on patient III:2, the mother II:1 and the two maternal aunts, as well as on an unrelated female control individual 2 (Ctrl-2F). The relative number of copy of *FMRI* in the different individuals was calculated using the 2-( $\Delta\Delta C_t$ ) method with beta globin (primer F: 5'-ACACAACACTGTGTTCACTAGC-3' and primer R: 5'-CAACTTCATCCACGTTCCACC-3') as a reference gene.

### *Cell culture*

Lymphoblastoid cells lines were established by Epstein-Barr virus (EBV) immortalization of blood lymphocytes obtained from patients III:1 and III:3, an unrelated male control individual 3 (Ctrl-3M), and a Fragile-X patient (FXS) with classical expansion (>200 repetitions). They were maintained in suspension cultures in RPMI 1640 supplemented with 0 to 10mM HEPES, 10 to 20% fetal bovine serum, 0 to 2mM Glutamine, 1% antimycotic and Gentamicin (40µg/mL).

### *FMRI mRNA extraction, analysis and quantification*

RNA was extracted from blood (PAXgene Blood RNA System, Preanalytix, Hombrechtikon, Switzerland) for the patient III:3 of the first family, the probands of the second and the third family, and an unrelated male individual 4 (Ctrl-4M) according to manufacturer's instructions. RNA was also extracted from lymphoblastoid cell lines obtained from patients III:1 and III:3, Ctrl-3M and the FXS patient carrying the classical CGG expansion using TRI Reagent® protocol (Molecular Research Center, Cincinnati, OH, USA). For cDNA synthesis 500ng of total RNA were reverse transcribed (RT) using random hexamers and SuperScript™ II Reverse Transcriptase according to manufacturer's recommendation. To study the potential retention of intron 16 in *FMRI* mRNA due to the deletion of exon 17 identified in the first family, we performed PCR on blood cDNA from patient III:3 and Ctrl-4M using following primers (FMR1\_RT\_E16F 5'-AGGAGGCTTCA AAGGAAACGAC-3' and FMR1\_I16\_STOP\_R 5'-TAAACTCATAAATCTCAAACCTA CCATATTCAAACAA-3'). To check the effect on splicing of the c.990+1G>A and c.420-8A>G variants identified in families 2 and 3, PCR with specific primers (FMR1\_RT\_E8F 5'-GTACCTGGGGTCACTGCTATTGAT-3' and FMR1\_RT\_E11R 5'-GGGGCATTAGGTCCAACCCTT-3'; FMR1\_RT\_E5F 5'-GATCAAGCTGGATGTGCCAGAA-3' and

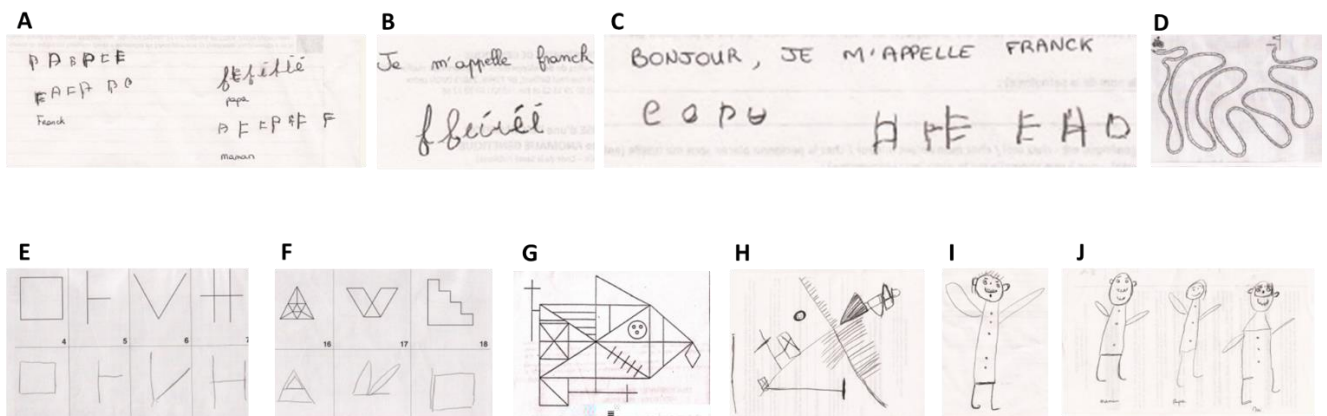
FMR1\_RT\_E7R 5'-ATGTGCTCGCTTTGAGGTGACTTC-3') was also performed on blood cDNA from the probands and unrelated male individuals (Ctrl-4M, Ctrl-5M). All the PCR products obtained were migrated on agarose gel or on capillary DNA analyzer (Caliper LabChip GX, PerkinElmer) were sequenced (at GATC, Germany or on capillary sequencer ABI3500, Applied Biosystem). In order to quantify *FMRI* mRNA in the brothers of the first family, quantitative RT-PCR were performed on LightCycler 480 II using the QuantiTect SYBR Green PCR Master Mix (Qiagen, Netherlands). All qPCR reactions were performed in triplicate. Primers used were: FMR1\_RT\_E10F 5'-CAGGAGATTGTGGACAAGT CAGGA-3', FMR1\_RT\_E11R 5'-GGGGCATTAGGTCCAACCCTT-3', GAPDH\_RT\_F 5'-GAAGGTGAAGGTCGGAGTC-3' and GAPDH\_RT\_R 5'-TTGATGGCAACAATAT CCACTT-3'. Reaction specificity was controlled by post-amplification melting curve analysis. The relative expression of *FMRI* vs. *GAPDH* was calculated using the  $2^{-(\Delta\Delta Ct)}$  method. A parametric Student's t-test was performed in order to compare the *FMRI* relative expression.

#### *Western blot analysis*

For Western Blot analysis, lymphoblast cells were lysed in RIPA buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.25% sodium deoxycholate, 1% NP-40 and protease inhibitors). A total of 50µg of protein lysate were separated on 8% SDS-PAGE and transferred to polyvinylidene fluoride membrane. Membranes were blocked in Tween 0.1%-TBS + 5% defatted milk followed by incubation with the monoclonal anti-FMRP antibody FM-1C3<sup>2</sup> at 1:5000 (overnight) or a with monoclonal anti-GAPDH antibody (Sigma-Aldrich, MO, USA) at 1:10000 (less than 1h). Incubation with appropriate secondary HRP-labelled antibody was followed by detection with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, MA, USA). The density of each protein band of interest was measured and normalized to GAPDH by the ImageJ software. A parametric Student's t-test was performed in order to compare the normalized FMRP expression.

**Figure S1: Drawings of patient III:2**

(A) When asking for writing his first name, “mummy” and “daddy”, he drew a series of letters which did not correspond with the requested words (B, C): In a position to copy a sentence, the patient wrote a series of letters which did not correspond with the model neither in terms of form nor of number of digits (D) Concerning graphics, he could regulate enough his gesture to draw a line within narrow and winding path (E, F) Simple geometric shapes were successfully reproduced but complex figures as stairs placed him in difficulty (G, H) Rey figure test was failed since the patient could not respect visuo-spatial relationships between components but did what he could to draw some details of the shape (I) The drawing of a man equaled an age development of 6 years and 9 months. The main body parts were represented (J) The drawing of the family only contained three individuals, he and his parents. His brothers with whom he therefore rarely interacted were not included. He made body map errors, drew sometimes 4 eyes and 2 mouths for the same face or 2 feet for the same leg. There was no gender or generation differentiation.



**Table S1: Number of reads obtained by DNA sequencing of individual III:1 and other individual form the same serie**

**Table S2: Non-synonymous rare variants identified in ID patients by direct screening of *FMR1* coding sequence or by large-scale HTS approaches**

ADHD: attention deficit with hyperactivity disorder; ASD: autism spectrum disorder; ID: Intellectual disability; FXS: Fragile-X syndrome; NDD: neurodevelopmental disorder; HA: heteroduplex analysis; SSCP: Single Strand Conformation Polymorphism; HRM: high resolution melting; HTS: high throughput sequencing; WES: whole exome sequencing; X-ES: X-exome sequencing; Phenotype highly (+++), moderately (+), poorly (+) or not (-) evocative of a FXS. N.A.: clinical data not available;

**Table S3: Non-synonymous variants present in the ExAC general population**

**References:**

- 1.Callier P, Aral B, Hanna N *et al*: Systematic molecular and cytogenetic screening of 100 patients with marfanoid syndromes and intellectual disability. *Clin Genet* 2013; **84**: 507-521.
- 2.Devys D, Lutz Y, Rouyer N, Bellocq JP, Mandel JL: The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat Genet* 1993; **4**: 335-340.

**Table S1.**

FMR1 Ex	chr	Ind1 = III.2	Ind																	Mean						
			M								F								Males		Females					
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Ind	Mean	Ind	Mean	
FMR1 Ex1	chrX	146993830	48	59	33	54	71	25	66	102	101	127	83	96	77	140	58	61	156	142	83	63,8	20,5	130,0	22,5	
FMR1 Ex2	chrX	147003367	147003584	158	203	145	169	151	145	418	700	693	742	511	549	458	824	330	385	953	880	580	337,0	168,5	810,3	105,1
FMR1 Ex3	chrX	147006971	147007236	174	204	156	199	182	153	340	657	666	741	516	550	491	828	313	383	947	878	544	335,9	157,8	799,3	117,7
FMR1 Ex4	chrX	147009754	147009995	152	190	148	203	151	166	328	613	606	671	458	560	415	751	292	387	875	770	512	317,5	147,8	722,3	104,1
FMR1 Ex5	chrX	147010093	147010406	175	204	143	213	173	158	378	693	702	780	489	559	492	816	309	393	969	872	554	338,8	159,2	814,9	105,2
FMR1 Ex6-7	chrX	147011453	147011790	172	224	162	215	179	161	370	714	722	782	549	553	447	836	282	423	1004	906	561	343,8	159,3	838,6	112,7
FMR1 Ex8-9	chrX	147013919	147014304	183	233	163	213	175	207	443	851	898	975	656	695	594	992	373	505	1176	1114	698	412,9	212,7	1017,1	124,3
FMR1 Ex10	chrX	147017932	147018221	165	194	154	192	173	186	358	662	693	719	497	577	459	829	306	429	970	829	562	340,6	160,4	790,1	114,9
FMR1 Ex11	chrX	147018844	147019301	193	240	166	228	186	206	447	860	872	966	644	691	571	1027	414	508	1216	1084	669	414,2	202,5	1015,6	135,3
FMR1 Ex12	chrX	147019527	147019768	138	189	131	179	147	141	310	550	595	658	425	486	381	769	265	358	830	771	449	288,4	130,1	706,3	111,2
FMR1 Ex13	chrX	147022004	147022269	189	210	162	211	168	169	392	728	758	820	524	604	483	894	373	460	1069	929	585	361,8	170,7	875,3	125,5
FMR1 Ex14	chrX	147024627	147024868	146	198	111	170	151	116	278	540	541	593	391	492	362	679	231	330	757	679	430	271,7	128,4	638,3	87,5
FMR1 Ex15	chrX	147026249	147026737	174	224	152	220	197	183	408	774	799	839	593	639	516	964	343	441	1132	1078	647	380,3	186,4	952,0	150,8
FMR1 Ex16	chrX	147026961	147027226	147	179	141	167	144	158	317	600	617	688	437	511	433	750	267	326	889	831	512	299,3	144,6	743,7	116,0
FMR1 Ex17	chrX	147030174	147030391	0	146	115	136	129	84	210	389	398	446	321	335	274	490	169	259	581	495	337	209,6	92,1	470,6	71,5



**Table S2. Part 1/2**

Variation	Type	Exon affected	Diagnostic approach	Number of patients	Screening approach	Reference	Number of hemizygote males in ExAC/EVS	Functional validation	Pathogenicity
c.52-1_52delinsTA (p.?)	Truncating	Exon 2	Clinical suggest. FXS	-	Western Blot + <i>FMR1</i> screening by HA	Lugenbeel, <i>et al.</i> , 1995	0	-	Pathogenic
c.80C>A (p.Ser27*)	Truncating	Exon 2	Clinical suggest. FXS	-	Western Blot + <i>FMR1</i> screening by sequencing	Gronskov <i>et al.</i> , 2011	0	-	Pathogenic
c.373delA (p.Thr125Leufs*35)	Truncating	Exon 5	Clinical suggest. FXS	-	Western Blot + <i>FMR1</i> screening by HA	Lugenbeel, <i>et al.</i> , 1995	0	-	Pathogenic
c.420-8A>G (p.Met140Ilefs*3)	Truncating	Exon 6	ID individuals	824	HTS (large panel)	This report	0	-	Pathogenic
c.990+1G>A (p.Lys295Asnfs*11)	Truncating	Exon 10	ID individuals	824	HTS (large panel)	This report	0	-	Pathogenic
Deletion	Truncating	Exon 17	ID individuals	106	HTS (large panel)	Redin <i>et al.</i> , 2014	0	-	Pathogenic
c.377T>C (p.Phe126Ser)	Missense	Exon 5	NDD individuals	1133	HTS (WES)	Wright <i>et al.</i> , 2014	0	Proba. Damaging (Polyphen2) Deleterious (SIFT) Dist. de Grantham: 155	Likely Pathogenic
c.797G>A (p.Gly266Glu)	Missense	Exon 8	Clinical suggest. FXS	-	<i>FMR1</i> mutation screening (DNA sequencing)	Myrick <i>et al.</i> , 2014	0	Proba. Damaging (Polyphen2) Deleterious (SIFT) Dist. de Grantham: 98	Likely Pathogenic
c.911T>A (p.Ile304Asn)	Missense	Exon 10	Clinical suggest. FXS	-	<i>FMR1</i> mutation screening cDNA sequencing)	De Bouille <i>et al.</i> , 1993	0	Proba. Damaging (Polyphen2) Deleterious (SIFT) Dist. de Grantham: 149	Likely Pathogenic

Variation	Type	Exon affected	Diagnostic approach	Number of patients	Screening approach	Reference	Number of hemizygote males in ExAC/EVS	Functional validation	Pathogenicity
c.413G>A (p.Arg138Gln)	Missense	Exon 5	male ID individuals	963	FMR1 mutation screening (DNA sequencing)	Collins <i>et al.</i> , 2010	ExAC: 4 rs20016341 3	No effect on AMPA internalization, or binding to polyribosomes in Fmr1 KO neuron. Affects synaptic function of fmr1 drosophila. No rescue of presynaptic action potential (AP) broadening in Fmr1 KO mice. Disruption of interaction with the large-conductance calcium-activated potassium (BK) channels (Myrick <i>et al.</i> , 2015; Alpatov <i>et al.</i> , 2015)	Unknown Significance
c.677G>A (p.Arg226Lys)	Missense	Exon 8	ID individuals	986	HTS (large panel)	Grozeva <i>et al.</i> , 2015	0	Benign (Polyphen2) Deleterious (SIFT) Dist. de Grantham: 24	Unknown Significance
c.767A>G (p.Asp256Gly)	Missense	Exon 8	ID patients	986	HTS (large panel)	Grozeva <i>et al.</i> , 2015	0	Benign (Polyphen2) Deleterious (SIFT) Dist. de Grantham: 94	Unknown Significance
c.1325G>A (p.Arg442Gln)	Missense	Exon 14	NDD individuals	1133	HTS (WES)	Wright <i>et al.</i> , 2014	0	Proba. Damaging (Polyphen2) Deleterious (SIFT) Dist. de Grantham: 43	Unknown Significance
c.1444G>A (p.Gly482Ser)	Missense	Exon 14	male ID individuals	508	FMR1 mutation screening (HRM)	Handt <i>et al.</i> , 2014	0	Poss. Damaging (Polyphen2) Tolerated (SIFT) Dist. de Grantham: 56	Unknown Significance
c.1618G>A (p.Gly540Gln)	Missense	Exon 15	X-linked families	405	HTS (X-ES)	Hu <i>et al.</i> , 2016	0	Poss. Damaging (Polyphen2) Deleterious (SIFT) Dist. de Grantham: 98	Unknown Significance

Variation	Type	Exon affected	Diagnosis approach	Number of patients	Screening approach	Reference	Number of hemizygotic males in ExAC/EVS	Functional validation	Pathogenicity
c.1637G>>A (p..Arg546His)	Missense	Exon 15	Clinical suggest. FXS	-	FMR1 mutation screening (SSCP)	Wang <i>et al.</i> , 1997	0	Poss. Damaging (Polyphen2) Deleterious (SIFT) Dist. de Grantham: 29	Unknown Significance
c.1231G>>A (p..Ala411Thr)	Missense	Exon 13	ID patients	986	HTS (large panel)	Grozeva <i>et al.</i> , 2015	ExAC: 2	Benign (Polyphen2) Tolerated (SIFT) Dist. de Grantham: 58	Unlikely pathogenic
c.1312G>>A (p..Asp438Asn)	Missense	Exon 14	ID patients	106	HTS (large panel)	Redin <i>et al.</i> , 2014	0	Proba. Damaging (Polyphen2) Deleterious (SIFT) Dist. de Grantham: 23	Unlikely pathogenic
c.1601G>>A (p..Arg534His) (2x)	Missense	Exon 15	ID male patients	508 males	FMR1 mutation screening (HRM)	Handt <i>et al.</i> , 2014	0 rs78189334 0	Benign (Polyphen2) Tolerated (SIFT) Dist. de Grantham: 29	Unlikely pathogenic

**Table S2, Part 2/2**

<i>Variation</i>	<i>Inheritance</i>	<i>Morphological phenotype</i>	<i>Neurodevelopmental and behavioral phenotype</i>	<i>FXS phenotype</i>
c.52-1_52delinsTA (p.?)	Maternally inherited	Adult Male: Prominent wide forehead, long, thin face, mild facial asymmetry, mild prognathism, wide prominent ears, high arched palate, macroorchidism, soft velvety skin, and increased mobility of the hand joints. Mother mildly retarded	ID and ASD, cooperative, anxious, eye avoidance. Toe walking, hand flapping, obsession with neatness, soft and rapid speech with classical cluttering	+++
c.80C>A (p.Ser27*)	Maternally inherited	Adult Male: Classical fragile X phenotype, with an elongated face, high, broad forehead, low-set large ears, prognathia, enlarged testes, hypotonia, hypermobility, with hyperextensible joints. Mother with mild-to-moderate ID	No active language except for a few repeated words used out of context, autistic features with little eye contact, perseveration and limited interests. Epilepsy from age 4.	++/+++
c.373delA (p.Thr125Leufs*35)	<i>de novo</i>	Pre-pubertal Male: wide prominent ears, prominent head, long face with coarse features, epicanthal folds, and hypermobile joints. Normal genitalia	Mild ID, reduced balance, poor motor skills and infantile speech with perseveration of topic and 'cluttering' of phrases, poor eye contact and increased activity	++
c.420-8A>G (p.Met140Ilefs*3)	<i>de novo</i>	<i>see supplementary text</i>	<i>see supplementary text</i>	++
c.990+1G>A (p.Lys295Asnfs*11)	<i>de novo</i>	<i>see supplementary text</i>	<i>see supplementary text</i>	++
<b>Deletion</b>	Maternally inherited	<i>see supplementary text</i>	<i>see supplementary text</i>	+++
c.377I>C (p.Phe126Ser)	<i>de novo</i>	NA	NA	NA

<i>Variation</i>	<i>Inheritance</i>	<i>Morphological phenotype</i>	<i>Neurodevelopmental and behavioral phenotype</i>	<i>FXS phenotype</i>
<b>c.797C&gt;A</b> (p.Gly266Gln)	NA	Teenager (16 years): Normal measurements, large ears, tall forehead, long face, large ears, hypermobility of the elbows and small joints of the hands, flat feet, and macroorchidism.	Severe ID (general abilities all around the 3- to 4-year level at age 16), walking achieved at 24 months, achievement of urinary continence at age 15, ASD, disruptive outbursts at adolescence with some aggression, poor attention span and hyperactivity, with a diagnosis of ADHD that responded favorably to treatment with methylphenidate.	+++
<b>c.911T&gt;A</b> (p.Ile304Asn)	<i>de novo</i>	Adult Male: Height 90 <sup>th</sup> p., Weight 50 <sup>th</sup> p., OFC 50 <sup>th</sup> , 98 <sup>th</sup> p. Asymmetric elongated face, prominent supraorbital ridges, sunken eyes, mandibular prognathism, coarse features, large ears, thick lower lip, brachydactyly, flat feet, valgus deformity of the knees, macroorchidism	Sitting at 20 months, standing without support at 4 years, walking alone at 5 years. Profound ID (IQ below 20), institutionalized, primitive contact with surroundings, no verbal communication except some sounds. Focal seizures successfully treated. Spastic paraparesis of lower extremities	++
<b>c.413G&gt;A</b> (p.Arg138Gln)	NA	No dysmorphic features commonly associated with FXS	Global developmental delay, ID, and intractable seizures but no other behavioral, neurological features commonly associated with FXS	-
<b>c.677G&gt;A</b> (p.Arg226Lys)	NA	NA	NA	NA
<b>c.767A&gt;G</b> (p.Asp256Gly)	Unknown	NA	NA	NA

<i>Variation</i>	<i>Inheritance</i>	<i>Morphological phenotype</i>	<i>Neurodevelopmental and behavioral phenotype</i>	<i>FXS phenotype</i>
c.1325G>A (p..Arg442Gln)	Maternally inherited	NA	NA	NA
c.1444G>A (p..Gly482Ser)	Unknown	Young Male (4 years): Sparse clinical information. Tall stature (height on the 98th centile).	Psychomotor delay, language impairment and low attention span	NA
c.1618G>A (p..Gly540Glu)	Unknown	NA	NA	NA
c.1637G>A (p..Arg546His)	Unknown	NA	NA	NA
c.1231G>A (p..Ala411Thr)	Unknown	NA	NA	NA
c. 1312G>A (p..Asp438Asn)	Maternally inherited, present in the unaffected maternal half-brother	8 y-o and 9 y-o brothers: No dysmorphic features commonly associated with FXS	ID, autistic features, behavioral and sleep disorders, and behavioral instability	+
c.1601G>A (p..Arg534His) (2x)	Maternally inherited, present in the affected sister	Pre-pubertal Male (11 years) and his sister (13 years): head circumferences, heights and weights in the normal percentiles	Developmental delay, special school.	NA
	Maternally inherited	Pre-pubertal Male (11 years): macrocephalic, and his height and weight were close to the 97th percentile. Urinary incontinence Mother: no phenotype	Developmental delay, learning disability, short attention span and agitation outbursts.	+



**Table S3.**

Chr	Position	RSID	Reference	Alternate	Transcript Consequence	Consequence	Exon	Filter	Annotation	Polyphen2 (HumVar)	SIFT	Grantham	Allele Count	Allele Number	Number of Homozygotes	Number of Hemizygotes	Allele Frequency
X	146993710	.	G	C	c.13G>C	p.Val5Leu	exon 1	PASS	missense	Benign	Tolerated	32	1	15366	0	1	0.00006508
X	147007110	.	G	C	c.157G>C	p.Val53Leu	exon 3	PASS	missense	Benign	Tolerated	32	4	87599	0	1	0.00004566
X	147010288	.	A	G	c.382A>G	p.Lys128Glu	exon 5	PASS	missense	Proba. Damaging	Deleterious	56	1	87625	0	1	0.00001141
X	147010319	.	G	A	c.413G>A	p.Arg138Gln	exon 5	PASS	missense	Benign	Deleterious	43	9	87473	0	4	0.0001029
X	147011478	rs782278808	A	G	c.431A>G	p.Gln144Gly	exon 6	PASS	missense	Poss. Damaging	Deleterious	98	1	74930	0	1	0.00001335
X	147011480	rs29281	G	T	c.433G>T	p.Alal45Ser	exon 6	PASS	missense	Benign	Tolerated	99	827	75780	30	166	0.01091
X	147011552	.	G	A	c.505G>A	p.Val169Ile	exon 6	PASS	missense	Benign	Tolerated	29	1	86284	0	1	0.00001159
X	147011655	.	T	A	c.522T>A	p.Asn174Lys	exon 7	PASS	missense	Benign	Tolerated	94	21	87487	0	15	0.0002400
X	147011711	.	G	A	c.578G>A	p.Arg193His	exon 7	PASS	missense	Proba. Damaging	Deleterious	29	1	87633	0	1	0.00001141
X	147014050	.	A	G	c.737A>G	p.Lys246Arg	exon 8	PASS	missense	Poss. Damaging	Tolerated	26	12	87685	0	1	0.0001369
X	147014101	.	A	G	c.788A>G	p.His263Arg	exon 8	PASS	missense	Benign	Tolerated	29	2	87591	0	1	0.00002283
X	147014220	rs139029212	A	G	c.818A>G	p.Lys273Arg	exon 9	PASS	missense	Benign	Tolerated	26	27	87258	0	10	0.0003094
X	147014260	.	A	G	c.858A>G	p.Ile286Met	exon 9	PASS	missense	Proba. Damaging	Deleterious	10	1	86659	0	1	0.00001154
X	147019016	.	A	G	c.1022A>G	p.Asn341Ser	exon 11	PASS	missense	Benign	Tolerated	46	1	87226	0	1	0.00001146
X	147019028	.	T	C	c.1034T>C	p.Val345Ala	exon 11	PASS	missense	Benign	Tolerated	64	8	87374	0	3	0.00009156
X	147019058	rs781853998	A	G	c.1064A>G	p.His355Arg	exon 11	PASS	missense	Benign	Tolerated	29	1	87596	0	1	0.00001142
X	147019084	.	C	T	c.1090C>T	p.His364Tyr	exon 11	PASS	missense	Benign	Tolerated	83	1	87675	0	1	0.00001141
X	147019617	.	G	A	c.1126-1G>A	splice acceptor	exon 12	PASS	splice acceptor	-	-	-	97	87765	2	44	0.001105
X	147019660	.	C	T	c.1168C>T	p.Pro390Ser	exon 12	PASS	missense	Benign	Deleterious	98	1	87763	0	1	0.00001139
X	147019672	.	G	C	c.1180G>C	p.Ala394Pro	exon 12	PASS	missense	Benign	Deleterious	27	1	87759	0	1	0.00001139
X	147019675	.	T	C	c.1183T>C	p.Tyr395Arg	exon 12	PASS	missense	Benign	Deleterious	101	1	87758	0	1	0.00001139
X	147019681	.	G	C	c.1188-1G>C	splice donor	exon 12	PASS	splice donor	-	-	-	1	87754	0	1	0.00001140
X	147022137	.	G	A	c.1231G>A	p.Ala411Thr	exon 13	PASS	missense	Benign	Tolerated	58	7	87697	0	2	0.00007982
X	147026428	.	C	A	c.1511C>A	p.Ser504Tyr	exon 15	PASS	missense	Proba. Damaging	Deleterious	144	3	75269	0	1	0.00004095

Chr	Position	RSID	Reference	Alternate	Transcript Consequence	Consequence	Exon	Filter	Annotation	Popphen2 (HumVar)	SIFT	Grantham	Allele Count	Allele Number	Number of Homozygotes	Number of Hemizygotes	Allele Frequency
X	147026496	.	C	T	c.1579C>T	p.Arg27Cys	exon 15	PASS	missense	Proba. Damaging	Deleterious	180	1	60980	0	1	0.00001640
X	147026514	.	C	T	c.1597C>T	p.Arg33Tyr	exon 15	PASS	missense	Proba. Damaging	Deleterious	101	1	50289	0	1	0.00001989
X	147027052	.	A	G	c.1655-2A>G		exon 16	PASS	splice acceptor	-	-	-	1	79040	0	1	0.00001265
X	147027059	.	G	A	c.1660G>A	p.Asp54Asn	exon 16	PASS	missense	Poss. Damaging	Tolerated	23	4	80299	0	3	0.00004981
X	147027062	.	G	A	c.1663G>A	p.Asp55Asn	exon 16	PASS	missense	Benign	Tolerated	23	2	80647	0	1	0.00002480
X	147027080	.	A	C	c.1681A>C	p.Asn561His	exon 16	PASS	missense	Poss. Damaging	Tolerated	68	1	81484	0	1	0.00001227
X	147030216	.	G	A	c.1751G>A	p.Cys384Tyr	exon 17	PASS	missense	Benign	Tolerated	194	9	87382	0	6	0.0001030
X	147030264	.	G	T	c.1799G>T	p.Ser600Ile	exon 17	PASS	missense	Poss. Damaging	Tolerated	142	1	87684	0	1	0.00001140
X	147030285	.	C	T	c.1820C>T	p.Thr607Met	exon 17	PASS	missense	Poss. Damaging	Deleterious	81	6	87696	0	5	0.00006842
X	147030297	.	G	T	c.1833G>T	p.Arg611Leu	exon 17	PASS	missense	Proba. Damaging	Deleterious	102	1	87713	0	1	0.00001140





## **Partie 2. Identification d'une mutation potentiellement causale dans *PQBPI***

### **I. Le syndrome de Renpenning**

Le syndrome de Renpenning (RENS) est un groupe de XLID syndromiques caractérisées par une DI moyenne à sévère associée à une microcéphalie, une maigreur ainsi qu'une petite taille (en comparaison des autres membres de la famille) (Stevenson *et al.* 2005). De par son mode de transmission lié au chromosome X, ces manifestations cliniques sont uniquement présentes chez les hommes, les femmes porteuses étant asymptomatiques. Le RENS possède des variantes phénotypiques qui incluent le syndrome de Golabi-Ito-Hall, d'Hamel et de Sutherland-Haan. Ces syndromes ont chacun été décrits dans seulement une famille.

Le RENS est causée par une mutation du gène *PQBPI*. Ce gène est exprimé dans plusieurs régions du cerveau (notamment dans le cervelet et l'hippocampe) et la protéine résultante est particulièrement présente dans les noyaux des neurones (Waragai *et al.* 1999). Cette protéine contient différents domaines dont un domaine WW (responsable de l'interaction avec l'ARN polymérase et donc impliqué dans l'activation de la transcription) (Komuro *et al.* 1999b), un domaine riche en acides aminés polaires (interagissant avec les polyQ des protéines) (Waragai *et al.* 1999) et un domaine Cter (contenant une région impliquée dans l'interaction avec la protéine du spliceosome U5-15kD) (Takahashi *et al.* 2010). *PQBPI* serait ainsi impliquée dans le métabolisme des ARNm en agissant à la fois sur la transcription, l'épissage mais également leur transport jusqu'à la synapse. Tout ceci est crucial pour le développement et le maintien du réseau neuronal et explique en quoi des altérations de *PQBPI* peuvent influencer les fonctions cognitives (Kunde *et al.* 2011).

D'un point de vue génétique, les mutations identifiées chez des patients souffrant du RENS concernent majoritairement la répétition de 6 AG présente dans l'exon 5 de *PQBPI*. Parmi 29 familles affectées par une XLID phénotypiquement cohérente avec le RENS, cinq présentaient une mutation dans cette répétition (AG)<sub>6</sub> à savoir, soit une insertion d'un AG, soit une délétion de 2AG (respectivement c.463-464dupAG et c.459-462delAGAG) (Kalscheuer *et al.* 2003). Néanmoins, il existe une forte variabilité phénotypique concernant les mutations affectant la répétition (AG)<sub>6</sub> puisque : 1) l'insertion d'un AG (c.463-464dupAG) a également été associée au syndrome de Sutherland-Haan

(décrit chez sept hommes d'une seule grande famille) et 2) la délétion d'un AG (c.461-462delAG) a été associée au syndrome d'Hamel qui est la variante la plus sévère du RENS (rapporté dans une famille dont deux frères et deux oncles maternels souffraient d'une DI sévère, trois d'entre eux étant décédés dans la petite enfance) (Hamel *et al.* 1994). Des microdélétions dans l'exon 5 (c.575-576delAG et c.547-569del(23pb)) (Lenski *et al.* 2004; Cossée *et al.* 2006), des insertions d'1pb dans l'exon 6 (c.641insC et c.631insA) (Lenski *et al.* 2004; Rejeb *et al.* 2011), une délétion d'1pb dans l'exon 7 (c.718delT) (Hu *et al.* 2016) ainsi qu'une mutation non-sens dans l'exon 6 (c.586C>T) (Germanaud *et al.* 2011) ont été identifiées chez des patients atteints du RENS. Le point commun de toutes ces variations pathogènes est qu'elles entraînent systématiquement l'expression d'une protéine tronquée ne possédant plus le domaine Cter ce qui altère sa fonction.

A l'heure actuelle, une seule mutation faux-sens pathogène dans *PQBP1* a été identifiée dans l'exon 5 (c.194A>G, p.Tyr65Cys), plus particulièrement dans le domaine WW réduisant ainsi l'efficacité d'interaction de PQBP1 avec ses partenaires. De ce fait, cette mutation est considérée comme pathogène car entraînant une altération de la fonction de PQBP1 au niveau de la régulation de la transcription et de l'épissage. Cette mutation a été mise en évidence dans seulement une famille avec trois individus affectés du syndrome de Golabi-Ito-Hall, une autre variante du RENS (Lubs *et al.* 2006; Sudol *et al.* 2012). Très récemment, une autre variation faux-sens a été décrite dans une famille avec XLID (c.727C>T, p.Arg243Trp) (Hu *et al.* 2016). La pathogénicité de cette seconde variation faux-sens reste néanmoins à être déterminée.

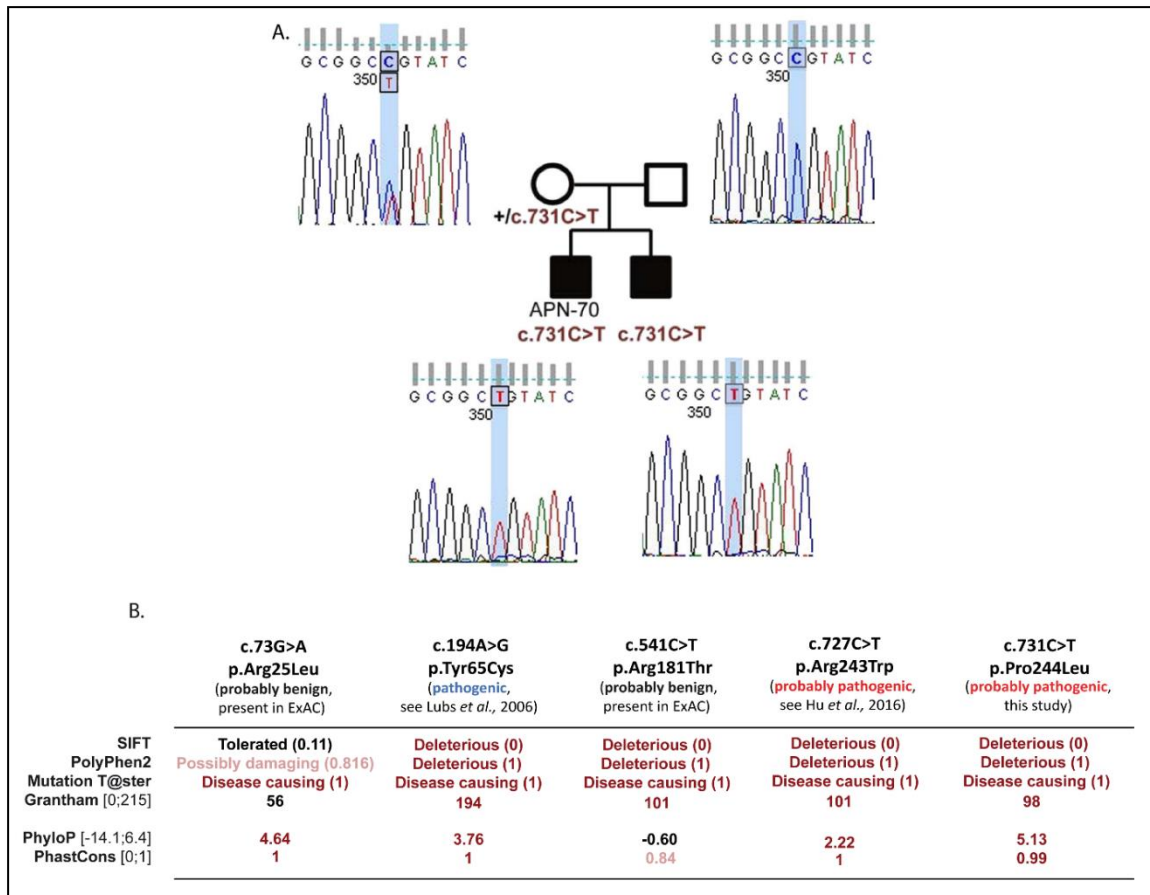
## II. Résultats

### 1. Identification d'une mutation faux-sens potentiellement causale dans *PQBP1* chez deux frères souffrant de DI et d'ASD

Le cas index (APN-70) est un garçon né en 1988 souffrant de DI modérée accompagnée d'une faible autonomie (langage, lecture et écriture pauvre de niveau CP), de troubles de la communication, de l'élocution et des interactions sociales et de difficultés d'apprentissage. Les comportements autistiques avec des mouvements stéréotypés ont commencé à l'âge de 3 ans. Son frère présente les mêmes signes cliniques (**Figure 17**).

Les tests génétiques préliminaires menés chez APN-70 sont revenus négatifs pour l'analyse de l'expansion CGG du gène *FMR1* mais l'array-CGH a cependant révélé une duplication 16q22.2 héritée du père non atteint et donc certainement bénigne. Après analyse par séquençage ciblé de 217 gènes, une variation faux-sens c.731C>T dans le gène *PQBP1* a été détectée chez le patient APN-70. Cette variation est présente à l'état hémizygotique chez les deux frères atteints, héritée de leur mère asymptomatique et absente chez le père (**Figure 17A**).

Au niveau protéique, la variation faux-sens c.731C>T a pour conséquence une substitution d'une proline en position 244 par une leucine (p.Pro244Leu). Cet acide aminé est très conservé au cours de l'évolution et ce changement est prédit délétère par les logiciels de prédiction (SIFT, PolyPhen2, Mutation T@ster et Grantham). Les deux autres variations faux-sens décrites dans la littérature (c.194A>G, p.Tyr65Cys et c.727C>T, p.Arg243Trp) sont également prédites délétères (**Figure 17B**) (Lubs *et al.* 2006; Hu *et al.* 2016). Cependant, à part pour la substitution p.Tyr65Cys pour laquelle une altération de la fonction de PQBP1 a été montrée, l'effet pathogène des changements p.Arg243Trp et p.Pro244Leu n'a pas été démontré. Ceci nous a amené à étudier de façon plus approfondie l'effet de ces deux variations faux-sens sur la fonction de PQBP1 afin d'en déterminer la pathogénicité. Nous étudierons également en parallèle la mutation pathogène faux-sens c.194A>G, p.Tyr65Cys ainsi qu'une mutation tronquante identifiée dans deux familles regroupant six patients souffrant de RENS (c.459-462delAGAG, p.Arg153SerfsX41) (Kalscheuer *et al.* 2003). Comme contrôle, nous avons choisi deux variations faux-sens présentes chez deux garçons de la population générale (Exome Aggregation Consortium, ExAC, base de données regroupant les données de séquençage de plus de 60000 individus) (c.73G>A, p.Arg25Leu et c.541C>T, p.Arg181Thr) qui sont prédites moins délétères ou touchent un nucléotide moins conservé (**Figure 17B**).



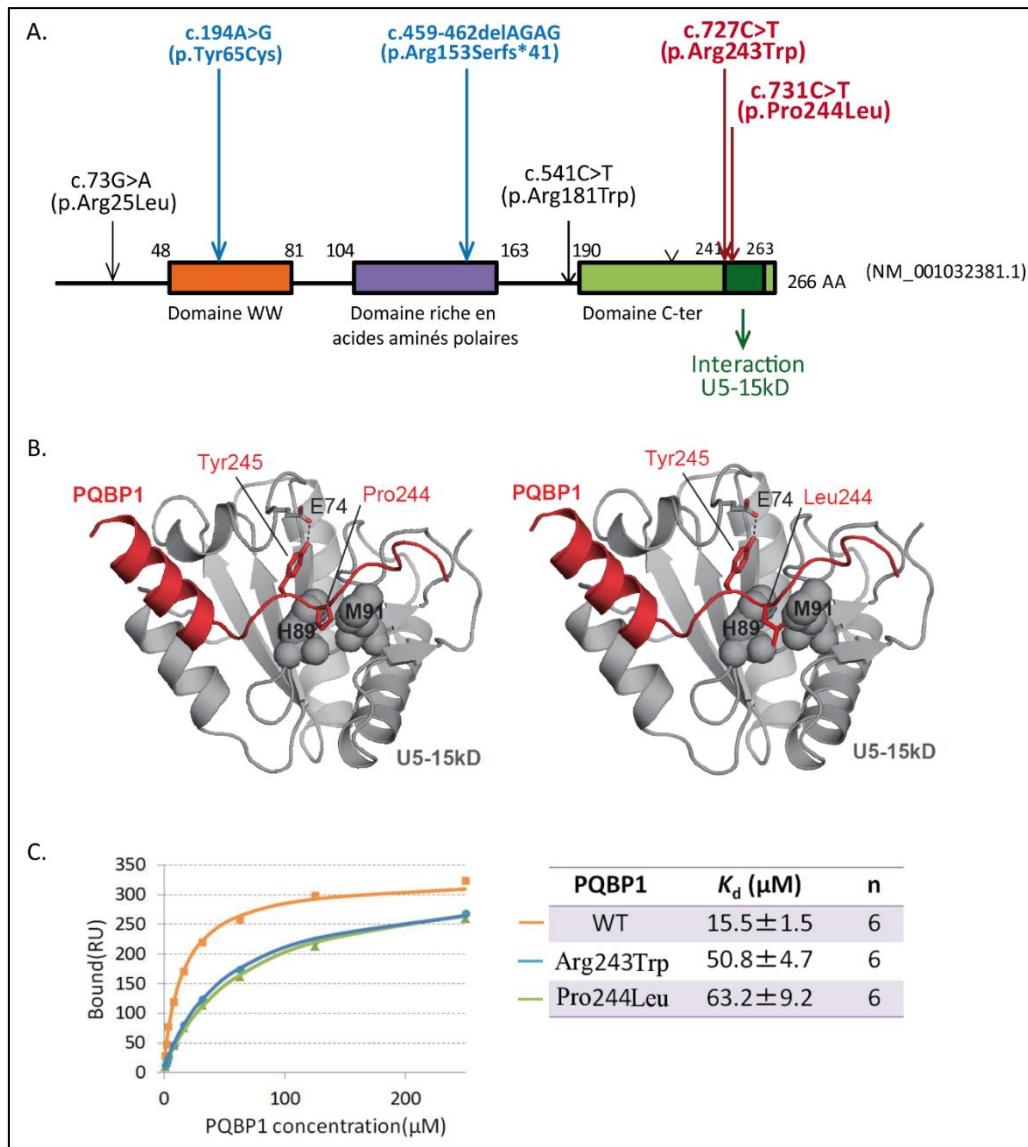
**Figure 17. Identification d'une mutation faux-sens potentiellement causale dans *PQBP1*.**

**A.** Pedigree du patient APN-70 montrant les chromatogrammes du séquençage Sanger pour chaque individu : la variation c.731C>T est présente chez les deux frères atteints, à l'état hétérozygote chez la mère et est absente chez le père. **B.** Scores de prédiction de la seule mutation faux-sens démontrée comme pathogène (c.194A>G) (Lubs *et al.* 2006), des deux variations faux-sens étudiée ici pour lesquelles l'effet pathogène reste à prouver (c.727C>T et c.731C>T) (Hu *et al.* 2016) et des deux variations faux-sens contrôles probablement bénignes car retrouvées chez deux garçons dans ExAC (c.73G>A et c.541C>T).

## 2. Etudes fonctionnelles

PQBP1 est une protéine capable de réguler l'épissage de multiples ARNm (Wang *et al.* 2013). Elle interagit avec de nombreux facteurs d'épissage tels que WBP11 (*via* le domaine WW) (Komuro *et al.* 1999a) et U5-15kD (*via* le domaine Cter) (Takahashi *et al.* 2010) et fait également partie du spliceosome précoce (Makarova *et al.* 2004).

Il a été montré qu'U5-15kD lie PQBP1 sur son motif YxxPxxVL. La variation c.731C>T, p.Pro244Leu identifiée par notre équipe ainsi que la variation c.727C>T, p.Arg243Trp précédemment rapportée (Hu *et al.* 2016) sont localisées dans le domaine Cter et plus particulièrement dans la région permettant l'interaction avec ce facteur du spliceosome U5-15kD (**Figure 18A**). Néanmoins, elles sont situées juste en amont de ce motif et ne font pas partie des acides aminés critiques identifiés (Tyr245, Pro248, Val251 et Leu252). En effet, il a été montré que la mutation de l'acide aminé Tyr245 en Asp245 déstabilise l'interaction avec U5-15kD de par la perte des contacts hydrophobes, entraînant ainsi une perte de fonction de PQBP1 *in vitro* (Mizuguchi *et al.* 2014). Etant donné leur proximité, nous avons voulu tester si la substitution Pro244Leu ainsi que la substitution Arg243Trp identifiée par Hu *et al.* induisaient le même phénotype (en collaboration avec M. Mineyuki et H. Okasawa, Tokyo, Japon). Une modélisation *in silico* a révélé que les variations p.Arg243Leu et p.Pro244Leu ne sont pas prédites pour affecter la liaison hydrogène formée entre l'acide aminé 245 de PQBP1 et l'acide aminé 74 de la protéine U5-15kD (**Figure 18B**). Cependant, ces deux variations faux-sens diminuent de façon modérée l'interaction avec la protéine U5-15kD (**Figure 18C**), suggérant qu'elles pourraient avoir des conséquences sur la régulation de l'épissage. Nous avons ainsi testé l'effet des variations dans PQBP1 sur l'épissage d'ARN pré-messagers cibles pour lesquels un épissage alternatif a été mis en évidence chez des neurones corticaux d'embryons de souris (Wang *et al.* 2013). Nous nous sommes concentré sur deux d'entre eux : *BCL2L1* pouvant donner un transcrit court, Bcl-xS codant pour un facteur pro-apoptotique et un transcrit long, Bcl-xL codant pour un facteur anti-apoptotique et *NRCAM* (Moore *et al.* 2010; Wang *et al.* 2013). Des RT-PCR semi-quantitatives ont été réalisées sur des ARN extraits de lignées lymphoblastoïdes de trois contrôles, de deux patients porteurs de mutations pathogènes dans *PQBP1* (c.194A>G et c.459-462delAGAG) et de deux patients porteurs des variations faux-sens étudiées ici (c.727C>T et c.731C>T). L'isoforme Bcl-xS de *BCL2L1* et le gène *NRCAM* sont très peu exprimés dans ce type cellulaire et sont donc difficilement visibles sur un gel d'agarose classique. Cependant, les bandes sont visualisables et quantifiables au Bioanalyzer. La quantification relative des différentes isoformes de *NRCAM* ne révèle néanmoins aucune différence d'épissage chez les patients porteurs de variations dans *PQBP1* comparés aux contrôles (et ce à la fois pour les mutations pathogènes et les variations faux-sens testées ici). Le ratio Bcl-xL/S des isoformes du gène *BCL2L1* semble en revanche significativement diminué chez les patients porteurs de variations dans *PQBP1*, indépendamment du type de variation (**Figure 19A** et **B**).

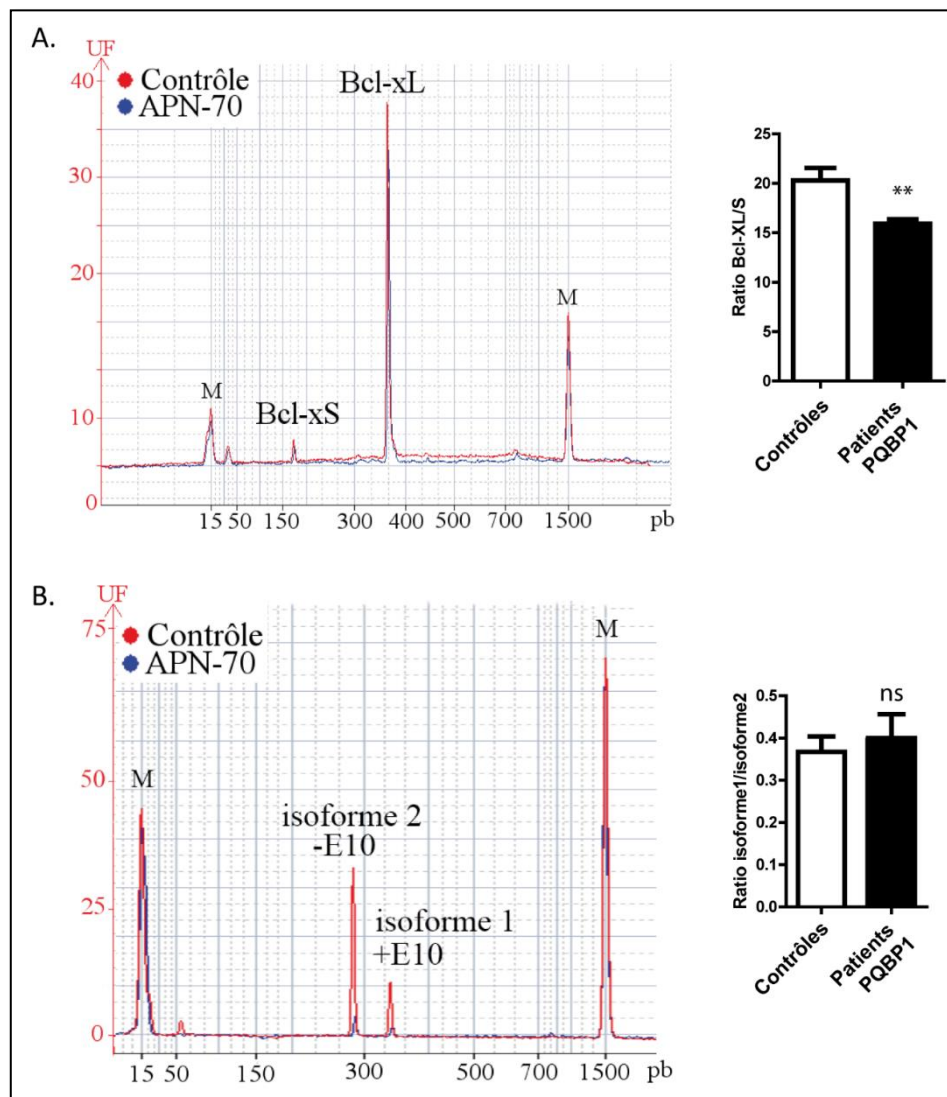


**Figure 18.** Etude de l'effet de variations dans PQBP1 sur l'interaction avec U5-15kD.

**A.** Schéma représentant les différents domaines de la protéine PQBP1 ainsi que la localisation des variations étudiées. En rouge sont représentées les variations potentiellement pathogènes testées ici, en bleu les différentes mutations pathogènes publiées dans la littérature chez des patients souffrant de RENS ou de XLID et en noir les variations vraisemblablement bénignes car retrouvées chez deux hommes dans ExAC. Les variations c.727C>T, p.Arg243Trp (Hu *et al.* 2016) et c.731C>T, Pro244Leu identifiée dans par notre équipe sont situées dans le domaine d'interaction avec la protéine du spliceosome U5-15kD (Takahashi *et al.* 2010).

**B.** Modélisation *in situ* de l'interaction entre U5-15kD (en gris) et PQBP1 (en rouge). A gauche est représenté l'interaction avec PQBP1 sauvage (Pro244) et à droite l'interaction avec PQBP1 muté (Leu244).

**C.** Test de liaison entre PQBP1 sauvage ou muté (Arg243Trp ou Pro244Leu) et U5-15kD. Les courbes obtenues ont permis le calcul du  $K_d$  pour chacune des protéines PQBP1.



**Figure 19.** Etude de l'effet de variations dans PQBP1 sur l'épissage de deux transcrits cibles.

A gauche sont représentés deux exemples d'électrophérogrammes obtenus par Bioanalyzer pour un contrôle et le patient APN-70 permettant la quantification des différentes isoformes de *BCL2L1* (A) et *NRCAM* (B). M : marqueur, UF : Unité de fluorescence. A droite sont représentés les ratios des différentes isoformes de *BCL2L1* (A) et de *NRCAM* (B) pour l'ensemble des individus contrôles (n = 3) et des patients porteurs de variations dans PQBP1 (n = 5) (moyennes et écarts moyens, t-test de Student ns : non significatif ; \*\* p < 0,01).



### III. Discussion et perspectives

Le phénotype des deux frères porteurs de la variation c.731C>T rapportés dans la présente étude n'est pas tout à fait comparable avec le RENS de par l'absence d'anomalies physiques (pas de maigreur ni de petite taille caractéristiques du RENS). Néanmoins, la dysmorphie est compatible avec le RENS, le périmètre crânien étant inférieur à la moyenne (entre -1,5 et -2SD) mais ne passant pas en dessous du seuil de microcéphalie (-3SD). Si la mutation c.194A>G, p.Tyr65Cys a des conséquences certaines sur l'interaction de PQBP1 avec WBP11, un facteur d'épissage (Tapia *et al.* 2010), celles des variations c.727C>T et c.731C>T, touchant le domaine d'interaction avec U5-15kD n'altèrent pas la liaison avec cette protéine mais la diminue de façon modérée (**Figure 18B** et **C**). Les études *in vitro* préliminaires concernant l'épissage de *NRCAM* dans des cellules lymphoblastoïdes de contrôles et de patients porteurs de différentes variations dans *PQBP1* n'ont pas permis de mettre en évidence d'altération chez les patients comparés aux contrôles (**Figure 19B**). Néanmoins, les résultats préliminaires révèlent une légère altération de la régulation de l'épissage de *BCL2L1* chez les patients porteurs de variations dans *PQBP1* (**Figure 19A**). En effet, *PQBP1* est un régulateur de l'épissage de *BCL2L1* en favorisant l'expression de Bcl-xL (ratio Bcl-xL/S > 1) ce qui a été confirmé par des expériences de sous-expression de *PQBP1* (par siRNA) dans lesquelles les formes Bcl-xL et xS étaient identiquement exprimées (ratio Bcl-xL/S = 1) (Wang *et al.* 2013). Dans notre cas, la diminution du ratio Bcl-xL/S dans les cellules de patient *PQBP1* fait passer ce ratio de 20 à 16 ce qui correspond à une diminution de la régulation de l'épissage de *BCL2L1* par *PQBP1* mais pas à une abolition totale de celle-ci. Le rôle de *PQBP1* dans l'épissage est crucial puisque son efficacité d'épissage est diminuée de 80% dans les lymphocytes de patients souffrant du syndrome de Golabi-Ito-Hall (mutation perte de fonction c.194A>G, p.Tyr65Cys) (Tapia *et al.* 2010). Ainsi, en plus de *NRCAM* et *BCL2L1* testés ici, beaucoup d'autres transcrits voient leur épissage régulé par *PQBP1* (Wang *et al.* 2013; Iwasaki & Thomsen, 2014). Afin d'être plus exhaustif et d'étudier le profil global d'épissage des ARN pré-messagers, des RNA-seq sont actuellement en cours de réalisation sur les lignées lymphoblastoïdes des trois contrôles, de deux patients porteurs de mutations pathogènes dans *PQBP1* (c.194A>G et c.459-462delAGAG) et de deux patients porteurs des variations faux-sens étudiées ici (c.727C>T et c.731C>T) (**Figure 18A**). Le séquençage paired-end 2x100bp couplé à une analyse bioinformatique adaptée permettra de mettre en évidence l'ensemble des ARN pré-messagers cibles de *PQBP1* et ceux dont l'épissage est altéré lorsque *PQBP1* est muté.

De plus, PQBP1 est également impliqué dans la régulation de la transcription (interaction avec l'ARN polymérase ainsi que des facteurs de transcription) (Komuro *et al.* 1999b) et des données préliminaires laissent à penser que l'expression globale de *BCL2L1* et *NRCAM* est augmentée dans les lignées lymphoblastoïdes de patients porteurs de variations dans PQBP1 comparés aux contrôles. Ces RNA-seq permettront donc de confirmer ces données et de révéler d'autres différences d'expression génique entre contrôles et patients.

D'autre part, il a été montré que la diminution de l'expression de PQBP1 dans des cultures primaires de neurones de souris réduit la croissance des dendrites notamment en altérant l'épissage alternatif d'ARNm impliqués dans le développement des neurites (Wang *et al.* 2013). En complément de l'étude moléculaire par RNA-seq, une étude cellulaire pourra être envisagée afin d'évaluer un potentiel impact des variations dans *PQBP1* sur la morphologie des cellules neuronales et en particulier sur la croissance des neurites. Pour cela, différentes stratégies pourront être réalisées en parallèle, dans des lignées de neuroblastomes humains (BE(2)C ou SH-SY5Y) et/ou dans des précurseurs neuronaux humains (hNSCs) : 1) Surexpression de PQBP1 sauvage ou muté ; 2) Sous-expression de PQBP1 à l'aide d'ARN interférents (siRNA) afin de mimer une perte de fonction et 3) Création d'hNSCs exprimant stablement PQBP1 muté par la technologie de modification du génome CRISPR/Cas9.

Pour compléter, les expérimentations actuellement en cours visent à étudier l'effet fonctionnel de l'ensemble des variations précédemment citées (**Figure 18A**) *in vivo*, chez le poisson-zèbre. Même si la variabilité phénotypique inter et intrafamiliale est très forte dans le RENS, des caractéristiques constantes sont retrouvées telles que la DI, le retard de croissance, la microcéphalie et la maigreur (Germanaud *et al.* 2011). Notre choix s'est porté sur l'étude de la microcéphalie car celle-ci peut être facilement reproduite et rapidement quantifiée chez le poisson-zèbre, comme cela a déjà été montré avec d'autres gènes (Golzio *et al.* 2012). Ce phénotype sera quantifié après inactivation de PQBP1 endogène (par injection de morpholino) puis expérience de sauvetage par surexpression de *PQBP1* humain sauvage ou muté (par injection des ARNm correspondants dans des œufs de poissons-zèbres sauvages, en collaboration avec C. Golzio).

*In fine*, l'ensemble des résultats des analyses réalisées *in vitro* et *in vivo* permettront de décrypter les conséquences fonctionnelles de différents types de variations perte de fonction dans *PQBPI*. Ces données permettront également de déterminer si les deux variations faux-sens (c.727C>T, p.Arg243Trp et c.731C>T, p.Pro244Leu) affectent la fonction de *PQBPI*. De plus, nous pouvons envisager de faire des corrélations génotype/phénotype pour certains endophénotypes du RENS (notamment la microcéphalie et le retard de croissance) en fonction du type de perte de fonction (variations faux-sens ou tronquantes).

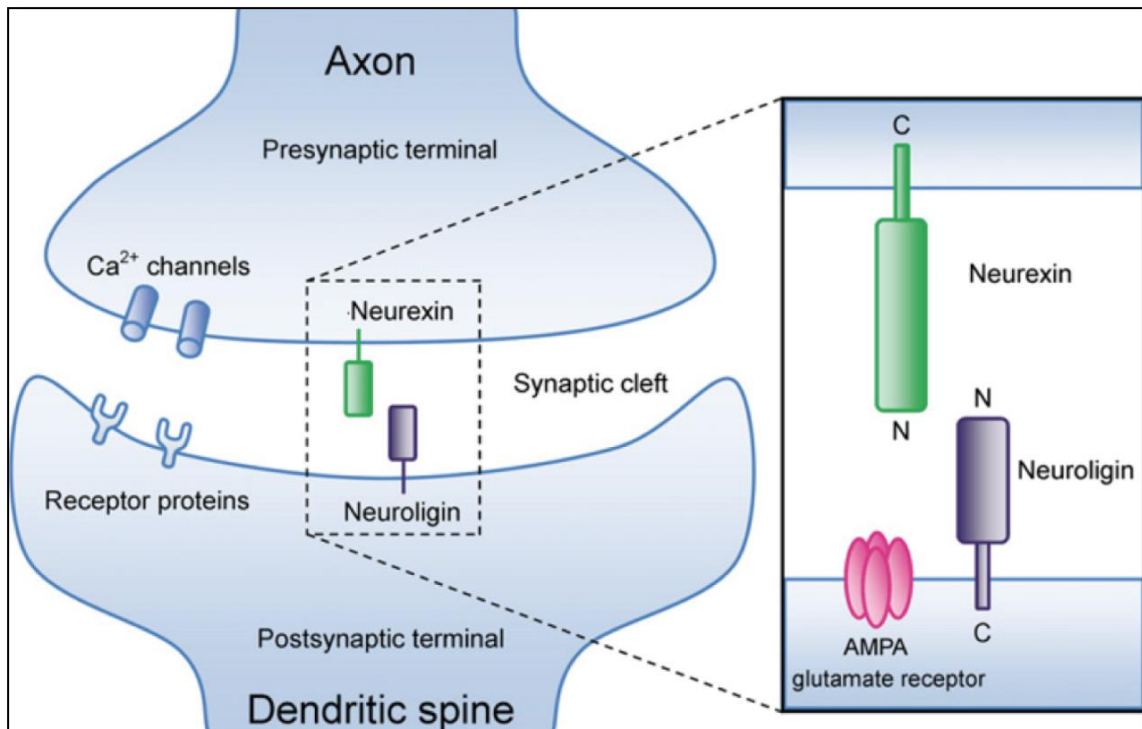
## **Partie 3. Identification d'une mutation potentiellement causale dans *NLGN3***

Dans les articles originaux, les variations identifiées dans *NLGN3* sont annotées sur des isoformes différentes. Par souci de clarté, l'ensemble des annotations ont été homogénéisées et font référence à l'isoforme NM\_018977.3 qui correspond au variant 2 de la *NLGN3* auquel il manque un exon alternatif dans la partie 5' codante du gène mais ne décalant pas le cadre de lecture. Cette isoforme 2 est ainsi plus courte de 20 acides aminés comparée à l'isoforme 1 mais possède une fonction identique.

### **I. *NLGN3*, un gène d'autisme ?**

Les neuroligines (NLGN) sont une famille de protéines d'adhésion cellulaire post-synaptiques. Elles lient les neurexines (NRXN), une famille de récepteurs membranaires pré-synaptiques dont elles sont les ligands endogènes. Ces deux familles de protéines transmembranaires se lient l'une à l'autre *via* leurs domaines extracellulaires pour promouvoir les connections synaptiques entre les neurones (**Figure 20**).

Il existe quatre neuroligines, dont les NLGN3 et NLGN4 qui sont codées par des gènes du chromosome X. La NLGN4X possède un homologue, la NLGN4Y, codée par un gène du chromosome Y. Les différentes NLGN sont très conservées au sein des vertébrés, et sont toutes enrichies dans les terminaisons post-synaptiques. La NLGN1 est principalement retrouvée dans les synapses excitatrices, la NLGN2 dans les synapses inhibitrices et la NLGN3 dans les deux types de synapses (Chubykin *et al.* 2007; Graf *et al.* 2004).



**Figure 20.** Structure du complexe trans-synaptique neurexine-neuroligine. Est donné ici l'exemple d'une synapse excitatrice. Les neurexines sont localisées sur les terminaisons pré-synaptiques et les neuroligines sur les épines post-synaptiques. L'encadré montre l'interaction entre les domaines extracellulaires des neurexines et des neuroligines à la synapse (Reichelt *et al.* 2012).

L'ensemble des NLGN médient le signal au travers de la synapse mais elles ne sont cependant pas essentielles à la cohésion synaptique. Il a été montré que les NLGN améliorent la formation des synapses *in vitro* (Scheiffele *et al.* 2000) mais qu'elles ne sont pas requises pour la génération des synapses *in vivo* (Varoqueaux *et al.* 2006) ce qui implique qu'elles n'entrent pas en jeu dans l'établissement des synapses mais qu'elles contribuent à leur bon fonctionnement (notamment la modulation des circuits neuronaux en réponse à une activité synaptique) (Chubykin *et al.* 2007). En effet, les NLGN possèdent un domaine intracellulaire interagissant avec des protéines de la densité post-synaptique. Leur dysfonctionnement altère ainsi les propriétés fonctionnelles de la synapse telles que la transmission synaptique.

A la vue de leur rôle crucial dans le fonctionnement synaptique, les NRXN et NLGN sont de bons candidats pour les NDD. L'expression de l'ARN *NLGN3* a été trouvée significativement diminuée dans les lignées lymphoblastoïdes de patients souffrant d'ASD (Yasuda *et al.* 2011). De plus, de nombreuses mutations ont été mises en évidence dans les gènes *NRXN1* et aussi mais surtout *NLGN4X* chez des patients souffrant d'autisme. Cependant, très peu de mutations ont été trouvées dans le gène *NLGN3*. En effet, la toute première mutation publiée dans les NDD a été celle de l'équipe de T. Bourgeron (Jamain *et al.* 2003). Ils ont mis en évidence une mutation faux-sens (c.1351C>T, p.Arg451Cys) chez deux frères l'ayant hérité de leur mère hétérozygote qui présente un très fort biais d'inactivation du X porteur de la mutation (92:8) (Gong *et al.* 2008). L'un des frères présente le syndrome d'Asperger et le second souffre d'autisme accompagné de DI. Il a été montré par la suite que cette mutation altère le transport de la protéine résultante qui est retenue dans le réticulum endoplasmique. La faible fraction de protéine atteignant la membrane plasmique possède une affinité de liaison diminuée pour la NRXN1 suggérant que la mutation p.Arg451Cys est une mutation perte de fonction (Comoletti *et al.* 2004; Chih *et al.* 2004; De Jaco *et al.* 2006). Cette mutation a également été introduite chez la souris et ces dernières présentent une altération des interactions sociales, une amélioration des capacités d'apprentissage (mémoire spatiale) et une augmentation de la transmission synaptique inhibitrice mais aucun effet sur les synapses excitatrices. L'impact sur les synapses inhibitrices n'étant pas retrouvé chez les souris *Nlgn3* KO, il semblerait que la mutation p.Arg451Cys se comporte également comme une mutation gain de fonction (Tabuchi *et al.* 2007). Des études supplémentaires sont néanmoins nécessaires pour définitivement comprendre l'effet pathogène de la mutation p.Arg451Cys.

A la suite de la découverte de cette première mutation causale, de nombreuses études de séquençage du gène *NLGN3* ont été menées sur des cohortes d'individus souffrant d'ASD (avec ou sans DI). Ces études concernaient des cohortes d'origine géographique très diverse et la majorité d'entre-elles se sont révélées négatives (Talebizadeh *et al.* 2004; Vincent *et al.* 2004; Ylisaukko-oja *et al.* 2005; Gauthier *et al.* 2005; Blasi *et al.* 2006; Piton *et al.* 2011; Wermter *et al.* 2008; Avdjieva-Tzavella *et al.* 2012; Liu *et al.* 2013; Volaki *et al.* 2013). Quelques études ont mis en évidence des variations dans *NLGN3* mais celles-ci étaient soit des variants synonymes, soit des variants prédits bénins par les logiciels de prédictions (Steinberg *et al.* 2012; Yanagi *et al.* 2012; Mikhailov *et al.* 2014; Xu *et al.* 2014).

Seule une autre variation faux-sens potentiellement pathogène a été identifiée chez un patient souffrant d'ASD accompagnés de DI, d'hyperactivité et de traits dysmorphiques (c.1022T>C, p.Val341Ala) mais aucune validation fonctionnelle n'a été réalisée à ce jour (Yu *et al.* 2013). Les mutations dans *NLGN3* représentent donc très probablement une cause très rare d'autisme. De plus, dans ExAC, la plupart des variations retrouvées sont synonymes et il n'y a aucune variation non-sens. Concernant les variations faux-sens, elles sont sous-représentées (95 observées contre 204,2 attendues, Z-score 3,74) ce qui implique que ce gène ne tolère pas ce type de variations et est soumis à une forte contrainte.

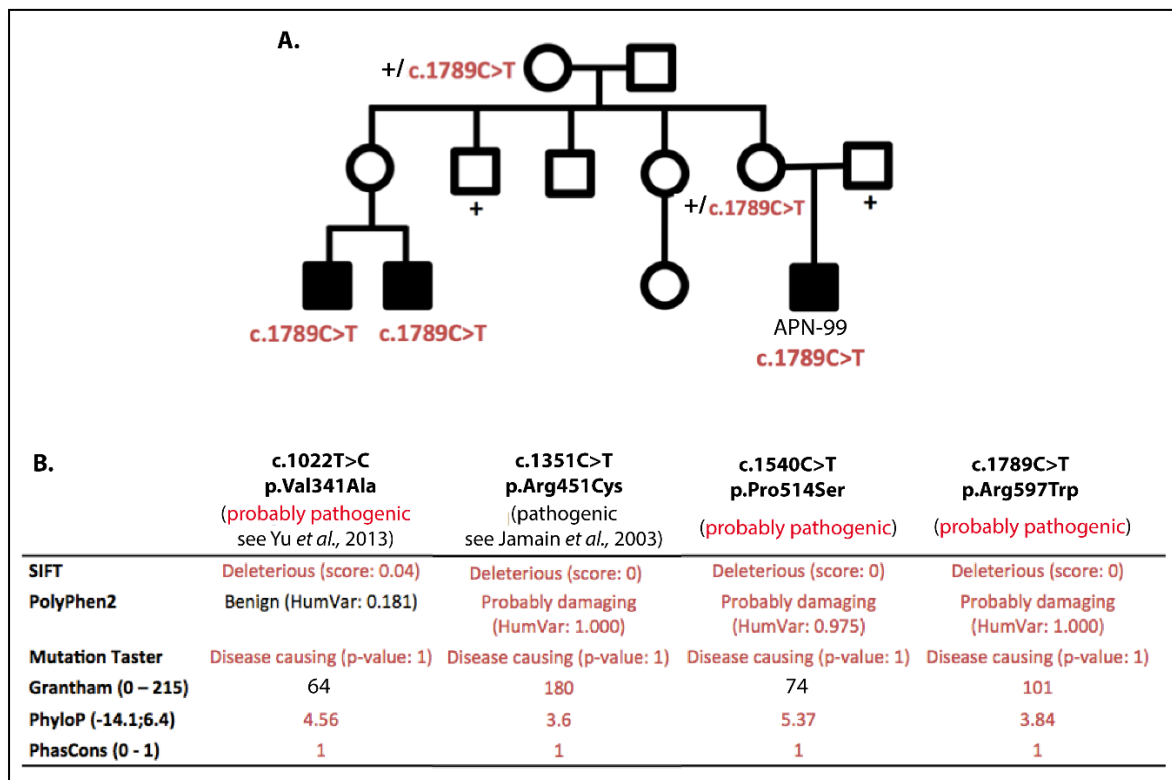
## II. Résultats

### 1. **Identification d'une mutation faux-sens potentiellement causale dans *NLGN3* chez trois jeunes garçons souffrant de DI et d'ASD**

Le cas index (APN-99) est un garçon né en 2007 souffrant de DI sévère avec absence de langage accompagnée d'autisme et de troubles du sommeil. Ses deux cousins germains maternels son également atteints (**Figure 21**). Les tests génétiques préliminaires sont revenus négatifs pour l'analyse de l'expansion CGG du gène *FMRI* mais l'array-CGH a cependant révélé une délétion 5q15 héritée du père non atteint et donc certainement bénigne. Après analyse par séquençage ciblé de 217 gènes, une variation faux-sens c.1789C>T dans le gène *NLGN3* a été détectée chez le patient APN-99. Cette variation est présente à l'état hémizygotique chez APN-99 et héritée de sa mère asymptomatique (où la variation est présente à l'état hétérozygotique). De plus, l'analyse de co-ségrégation révèle que la variation est également présente à l'état hémizygotique chez ses deux cousins maternels souffrant d'ASD et absente chez le père et l'oncle maternel non atteints (**Figure 21A**).

En parallèle, un de nos collaborateurs australiens a mis en évidence une autre variation faux-sens chez un patient souffrant d'ASD avec DI et problèmes de langage (c.1540C>T, p.Pro514Ser, communication personnelle de J. Gecz). Des analyses *via* des logiciels de prédiction révèlent que l'ensemble de quatre variations faux-sens publiées ou décrites à ce jour affectent des nucléotides et acides aminés très conservés et qu'elles sont toutes délétères et donc potentiellement causales (**Figure 21B**). Celles-ci entraînent des substitutions d'acides aminés dans le domaine extracellulaire de la *NLGN3*, domaine d'interaction avec les NRXN. Ainsi, nous avons décidé d'étudier de façon plus approfondie

l'effet de la seule mutation pathogène validée à ce jour (c.1351C>T, p.Arg451Cys) (Jamain *et al.* 2003), de la variation faux-sens identifiée dans (Yu *et al.* 2013) (c.1022T>C, p.Val341Ala) et des deux variations faux-sens présentées ici (c.1540C>T, p.Pro514Ser et c.1789C>T, p.Arg597Trp) sur la protéine NLGN3 afin de déterminer la pathogénicité des variations nouvellement décrites. Comme contrôle négatif, nous avons choisi d'également étudier un polymorphisme prédit bénin présent à proximité (c.1894A>G, p.Thr632Ala présent chez 71 hommes dans ExAC et également chez un homme et quatre femmes d'une cohorte d'individus contrôles séquencée dans (Blasi *et al.* 2006).



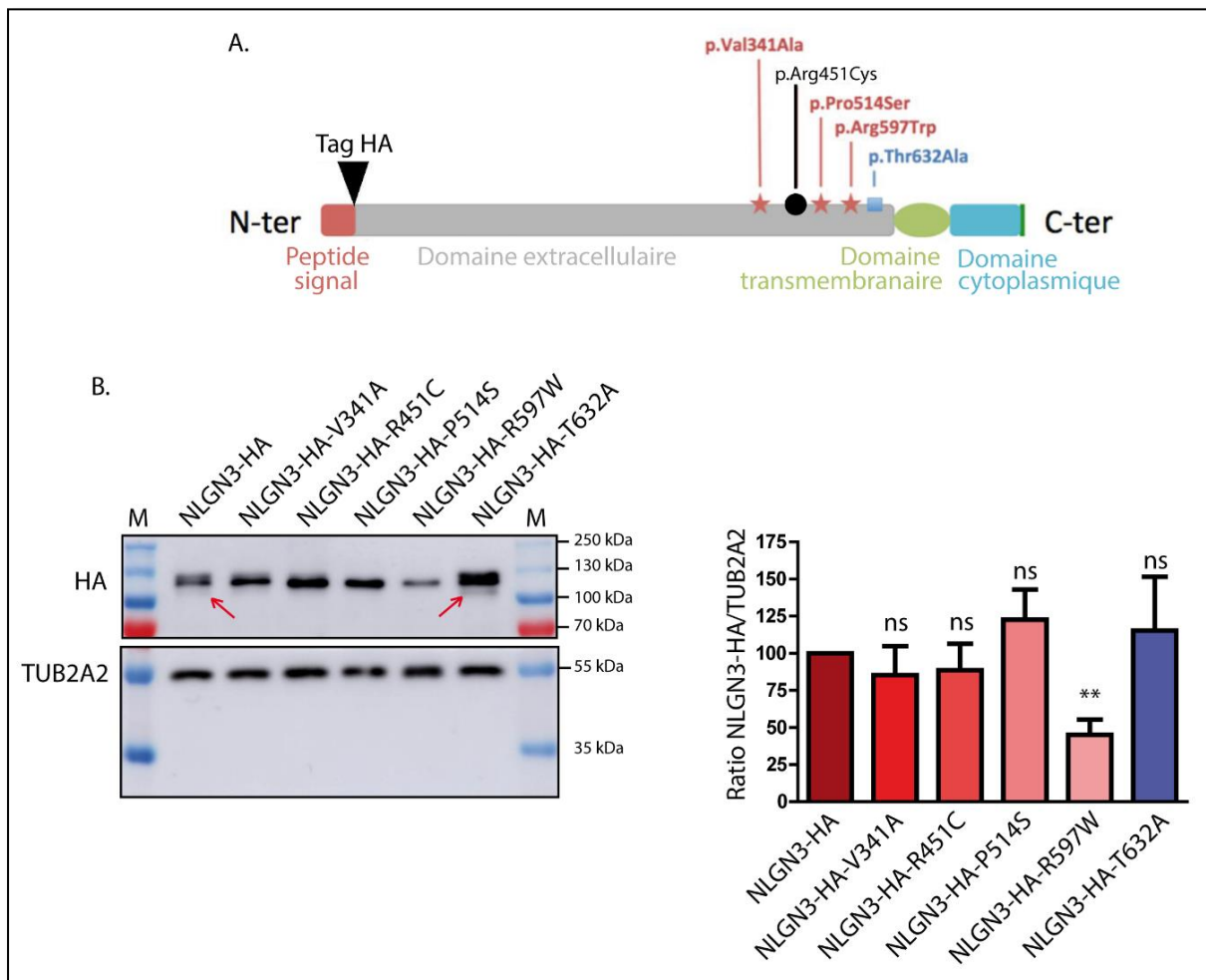
**Figure 21. Identification d'une mutation faux-sens potentiellement causale dans NLGN3.**

**A.** Pedigree du patient APN-99 : la variation est présente à l'état hémizygote chez le patient et ses deux cousins germains maternels également atteints, à l'état hétérozygote chez sa mère et sa grand-mère non atteintes et est absente chez son père et ses deux oncles maternels également non atteints. **B.** Scores de prédiction comparant les trois variations faux-sens potentiellement pathogènes (Yu *et al.* 2013) ainsi que de la seule mutation faux-sens pathogène validée dans la littérature (Jamain *et al.* 2003; Yu *et al.* 2013). L'ensemble de ces variations présente une très forte conservation au niveau nucléotidique ainsi qu'un effet prédit délétère sur la protéine.



## 2. Etudes fonctionnelles

Le but de ces études fonctionnelles est de déterminer si les différents variants affectent l'expression, la stabilité et la localisation de la protéine NLGN3 pour valider leur caractère pathogène. Pour cela, nous avons utilisé un plasmide d'expression pCDNA3.1 contenant l'ADNc de NLGN3 (NM\_018977.3). Le tag HA a ensuite été inséré entre le peptide signal (permettant l'adressage de la protéine à la membrane) et le domaine extracellulaire et les différentes variations précédemment citées ont été introduites par mutagenèse dirigée (**Figure 22A**). Les NLGN3 sauvages ou mutées ont été surexprimées dans des cellules facilement transfectables telles que les COS7 (cellules de rein simiennes). Un western blot réalisé sur ces cellules avec l'anticorps anti-HA révèle que la NLGN3-HA est exprimée dans les COS7 et ce quelque soit la variation insérée. La seule variation entraînant une baisse d'expression significative est la p.Arg597Trp identifiée chez le patient APN-99. De plus, il semble qu'une bande additionnelle plus légère soit présente pour la NLGN3 sauvage ainsi que pour la variation bénigne p.Thr632Ala (**Figure 22B**).



**Figure 22.** Effets de variations dans *NLGN3* sur la stabilité et l'expression de la protéine.

**A.** Représentation des domaines protéiques de la NLGN3 et localisation des variations étudiées. Le tag HA a été insérée entre le peptide signal et le domaine extracellulaire. Les étoiles rouges indiquent les variations potentiellement pathogènes, le cercle noir la seule mutation pathogène validée (Jamain *et al.* 2003) et le carré bleu le polymorphisme prédit bénin. **B.** Western blot montrant la surexpression de la NLGN3-HA sauvage ou porteuse des différentes variations dans les COS7. Les flèches rouges indiquent une bande plus légère absente des NLGN3 porteuses des variants potentiellement pathogènes. M : marqueur. Le graphique représente la quantification de l'expression de la NLGN3-HA normalisée par la TUBA2A. Test de Student ns : non significatif, \*\*  $p < 0.01$ ,  $n = 4$ .

### III. Discussion et perspectives

Les données préliminaires présentées ici ne permettent pas pour l'instant de conclure quant à la pathogénicité des différents variants testés. Néanmoins, la variation c.1789C>T, p.Arg597Trp entraîne une baisse d'expression significative de la NLGN3, laissant suggérer une perte de fonction partielle. Concernant les autres variations potentiellement pathogènes (c.1022T>C, p.Val341Ala ; c.1351C>T, p.Arg451Cys et c.1540C>T, p.Pro514Ser), l'expression de la protéine résultante ne semble pas affectée. Cependant, il semble que le nombre de bandes soit différent, une bande plus légère n'étant présente que pour la NLGN3 sauvage ainsi que pour la variation bénigne c.1894A>G, p.Thr632Ala. De nouveaux western blots sont nécessaires afin de confirmer ces observations et ils seraient intéressant de mieux caractériser l'expression des différentes isoformes de NLGN3 en fonction des variations insérées. Ces différentes bandes pourraient également représenter différentes modifications post-traductionnelles de la NLGN3 et les variations potentiellement pathogènes testées pourraient influencer sur celles-ci. Il serait intéressant d'étudier plus en détails des modifications si l'étude des isoformes s'avère infructueuse. En parallèle, des expériences complémentaires d'immunofluorescence sont en cours de réalisation afin d'étudier la localisation subcellulaire de la NLGN3 car sa mauvaise localisation pourrait nous indiquer une éventuelle modification de sa fonction.

Une étude récente effectuée sur des glioblastomes dérivés de patients a révélé que la NLGN3 membranaire des neurones actifs pourrait être clivée. La forme soluble résultante (correspondant au domaine extracellulaire) ainsi sécrétée pourrait alors agir sur la prolifération des cellules gliales en activant la voie PI3K (Venkatesh *et al.* 2015). Ce mécanisme avait déjà été décrit pour la NLGN1 chez les rongeurs et serait impliqué dans la maturation des épines dendritiques et la transmission synaptique (Peixoto *et al.* 2012; Suzuki *et al.* 2012). Les différents variants identifiés dans *NLGN3* sont dans le domaine extracellulaire sécrété. Il serait ainsi intéressant de tester si le clivage de NLGN3 est affecté chez les différents mutants en quantifiant par exemple la proportion de la NLGN3 soluble sécrétée.

*In fine*, l'ensemble des résultats des analyses réalisées permettront de décrypter les conséquences fonctionnelles de différentes variations identifiées dans *NLGN3* et de conclure quant à leur pathogénicité. Ce gène n'étant que très peu muté chez les patients souffrant d'ASD et/ou de DI, l'établissement de la pathogénicité de ces variations nouvellement identifiées permettra de confirmer ses implications dans ces pathologies.

## Partie 4.

### Discussion du chapitre 1.

Parmi les patients souffrant de DI isolée ou syndromique, une cause génétique a été identifiée dans moins de 50% des cas lors du diagnostic standard comprenant le caryotype, l'array-CGH (10 à 15%) et le test de l'expansion CGG de *FMRI* (1 à 2%). Ce chiffre est même bien inférieur en ce qui concerne le diagnostic des patients atteints d'ASD, condition cliniquement et génétiquement extrêmement hétérogène. Afin de remédier à ce manque, un test diagnostique utilisant le séquençage ciblé de gènes impliqués dans la DI et/ou les ASD a été mis en place par notre équipe et a été transféré en routine au laboratoire de diagnostic des Hôpitaux Universitaires de Strasbourg. Ainsi, parmi les 940 patients souffrant de DI isolée ou syndromique (dont environ un tiers présentant également des ASD) séquencés, 625 ont été analysés à l'heure actuelle et une mutation causale ou potentiellement causale a été identifiée dans 26,1% des cas (certaines étant encore en cours de validation). Les mutations suivant un mode de transmission autosomique dominante sont les plus nombreuses (18,4% des cas), suivies de celles liées au chromosome X dans 6,9% des cas et finalement les autosomiques récessives dans 1% des cas. De façon intéressante, les femmes présentent plus de mutations pathogéniques que les hommes (34,8% *versus* 23,3%,  $p = 0,0052$ ), en particulier des mutations autosomiques dominantes (27,4% *versus* 16,3%,  $p = 0,0014$ ). Des données similaires ont été retrouvées pour les mutations *de novo* dans une cohorte de 996 patients souffrant de DI analysés par séquençage ciblé (Grozeva *et al.* 2015) ainsi que dans une cohorte de 4293 patients souffrant de DI (et 3287 patients additionnels déjà publiés) analysés par séquençage d'exome (McRae *et al.* 2016). Une étude récente de CNV chez des patients souffrant d'ASD a révélé la même tendance puisque les femmes atteintes présentent plus de CNV autosomiques délétères (Jacquemont *et al.* 2014) suggérant qu'elles aient besoin d'une plus forte "charge" génétique pour développer de telles pathologies.

Au total 45 mutations causales ou potentiellement causales ont été identifiées dans des gènes du chromosome X au sein de notre cohorte, dont 36 chez des hommes (voir détails dans la **Table 5**). Dans ce premier chapitre, il a été donné l'exemple de cinq d'entre-elles identifiées chez des hommes au sein de trois gènes du chromosome X (impliqués dans des DI syndromiques : *FMRI* et *PQBPI* ou dans des cas d'autisme : *NLGN3*).

Le FXS (causé par une anomalie dans *FMRI*) a une prévalence de l'ordre de 1:3600 hommes et représente de 2 à 5% des cas d'autisme. Si les mutations ponctuelles dans *FMRI* ont certes une fréquence bien plus faible que la mutation récurrente (expansion CGG > 200), elles ne doivent pas être négligées. En effet, au sein de notre cohorte de 940 individus souffrant de DI associée ou non à des ASD et négatifs pour l'expansion CGG de *FMRI*, trois mutations intragéniques pathogènes ont été identifiées dans *FMRI* ce qui représente 0,32% des cas. Ainsi, proposer, en plus du test moléculaire classique évaluant la taille de l'expansion CGG de *FMRI*, un séquençage complet de ce gène à tout patient présentant une DI ou des ASD non spécifiques permettrait de mettre en évidence des anomalies affectant *FMRI* et donc de déterminer leur fréquence réelle dans la DI et les ASD. Le FXS, syndrome monogénique lié au chromosome X, pourrait ainsi représenter une plus grande part des cas de DI et d'ASD. De plus, le diagnostic du FXS est particulièrement important à deux autres niveaux. Tout d'abord, le conseil génétique pour les familles de patients est essentiel en raison du risque élevé de récurrence de cette XLID dominante. Ensuite, de nombreuses thérapies ciblées pour le FXS sont en cours de développement et certains patients atteints de FXS bénéficient déjà de certains de ces traitements.

En ce qui concerne les mutations dans *PQBPI*, elles ont été associées au RENS depuis 2003. La première famille identifiée comprenait 20 hommes affectés (Renpenning *et al.* 1962; Lenski *et al.* 2004). Dans une revue de 2005, Stevenson *et al.* regroupent les cas publiés ce qui correspondait au total à 52 patients appartenant à dix familles (Stevenson *et al.* 2005). Depuis, 26 hommes atteints dans dix familles supplémentaires ont été rapportés (les délétions dans l'hexamère AG de l'exon 5 étant les mutations les plus récurrentes) (Fichera *et al.* 2002; Fichera *et al.* 2005; Cossée *et al.* 2006; Martínez-Garay *et al.* 2007; Rejeb *et al.* 2011; Germanaud *et al.* 2011). Le nombre de patients identifiés reste ainsi faible et ce malgré des caractéristiques cliniques désormais reconnaissables (mises en évidence notamment grâce à des revues abordant précisément les tableaux cliniques) (Kleefstra *et al.* 2004; Germanaud *et al.* 2011). Le RENS semble donc être un désordre très rare dont la prévalence reste inconnue.

Les mutations causales dans *NLGN3* sont quant à elles encore plus rares. Seule une mutation pathogène et une variation de pathogénicité inconnue ont été décrites dans littérature et deux variations additionnelles sont présentées ici. La toute première mutation décrite (c.1351C>T, p.Arg451Cys) (Jamain *et al.* 2003) a été validée fonctionnellement

même si les études sont contradictoires concernant son effet perte ou gain de fonction (Comoletti *et al.* 2004; Chih *et al.* 2004; De Jaco *et al.* 2006; Tabuchi *et al.* 2007). Des études fonctionnelles sont encore nécessaires pour déterminer la pathogénicité des trois autres variations faux-sens. Néanmoins, même si ces dernières s'avèrent pathogènes, les mutations dans *NLGN3* restent extrêmement rares chez les patients souffrant d'autisme.

Pour résumer, il existe plus de 100 gènes du chromosome X impliqués dans la DI et les ASD. Trois d'entre eux ont été abordés ici, l'un causant un syndrome fréquent (*FMR1/FXS*), un second causant un syndrome beaucoup plus rare (*PQBPI/RENS*) et le troisième dont seulement une mutation validée a été identifiée chez des patients atteints d'autisme (*NLGN3*). L'étude des Prs JL. Mandel et J. Chelly avait mis en évidence des mutations dans un gène de XLID, *ARX*, chez 6,6% des hommes souffrant de XLID et dans 0,13% des cas sporadiques de DI (Mandel & Chelly, 2004). Dans notre cohorte de 940 patients comprenant des DI isolée et syndromique, le séquençage d'une centaine de gènes de XLID a révélé la présence d'une variation causale ou potentiellement causale dans un gène du chromosome X chez 7,6% des hommes, mais aussi, chez 5,3% des femmes. Les mutations causales dans des gènes du chromosome X ne sont ainsi pas présentes chez 30 à 40% des hommes atteints comme il aurait fallu que ce soit le cas pour expliquer totalement le biais de sexe de 1,3 à 1,4/1 observé dans la DI. De plus, l'absence de différence significative entre le nombre de mutations liées au chromosome X identifiées chez les hommes et chez les femmes au sein de notre cohorte confirme que les mutations causales de gènes de XLID ne peuvent pas expliquer la totalité de l'excès de garçon observé dans la DI et encore moins dans les ASD qui présentent un biais de sexe encore plus marqué. De ce fait, il est clair que d'autres facteurs interviennent (facteurs protecteurs chez les femmes ou sensibilisants chez les hommes; génétique et/ou environnementaux).

Sexe	Numéro d'identification	Gène	Isoforme	Coordonnées génomiques de la variation (g.)	Conséquence au niveau de l'ARN (c.)	Conséquence au niveau protéique (p.)	Transmission	Statut	Classe
F	B00H4KZ	CNKSR2	NM_014927.3	g.21544984G>A	c.958-1G>A	(p.?)	de novo	hétérozygote	3+
F	B00H4N7	DDX3X	NM_001356.4	g.41203076G>A	c.765+1G>A	(p.?)	de novo	hétérozygote	5
F	MMPN-288	DDX3X	NM_001356.4	g.41196720T>C	c.103+2T>C	(p.?)	de novo	hétérozygote	5
F	B00H4IP	DDX3X	NM_001356.4	g.41204681_41204685dup	c.1195_1199dup	p.(Tyr400*)	de novo	hétérozygote	5
F	APN-215	IL1RAPL1	NM_014271.3	g.29417419_29417420del	c.697_698del	p.(Val233Tyrfs*6)	de novo	hétérozygote	4
F	APN-142	MECP2	NM_004992.3	g.153296777G>A	c.502C>T	p.(Arg168*)	de novo	hétérozygote	5
F	APN-130	MECP2	NM_004992.3	g.153296363G>A	c.916C>T	p.Arg306Cys	de novo	hétérozygote	5
F	APN-164	NAA10	NM_003491.3	g.153197863G>A	c.247C>T	p.(Arg83Cys)	de novo	hétérozygote	5
F	APN-228	NAA10	NM_003491.3	g.153197863G>A	c.247C>T	p.(Arg83Cys)	de novo	hétérozygote	4
H	APN-138	ATRX	NM_000489.3	g.76972632G>A	c.109C>T	p.(Arg37*)	maternelle	hémizygotique	5
H	MMPN117	ATRX	NM_000489.4	g.76972632G>A	c.109C>T	p.(Arg37*)	maternelle	hémizygotique	5
H	APN-232	CASK	NM_003688.3	g.41530723C>G	c.490G>C	p.(Gly164Arg)	maternelle	hémizygotique	4
H	MMPN-148	CASK	NM_003688.3	g.41524622C>T	c.616G>A	p.(Gly206Ser)	maternelle	hémizygotique	4
H	SGT160051	CNKSR2	NM_014927.3	g.21619399G>A	c.1977-1G>A	(p.?)	maternelle	hémizygotique	5
H	APN-137	CUL4B	NM_003588.3	g.119681009_119681010del	c.811_812del	p.(Gln271Aspfs*11)	maternelle	hémizygotique	5
H	APN-26	FMR1	NM_002024.5	g.(?_147030202)_ (147046357_?)del	-	-	maternelle	hémizygotique	5
H	MMPN-202	FMR1	NM_002024.5	g.147018133G>A	c.990+1G>A	p.Lys295Asnfs*11	de novo	hémizygotique	4
H	B00H4P4	FMR1	NM_002024.5	g.147011459A>G	c.420-8A>G	p.Met140Ilefs*3	maternelle	hémizygotique	4
H	-	GRIA3	NM_000828.4	g.122561802G>A	c.1888G>A	p.(Gly630Arg)	maternelle	hémizygotique	5
H	MMPN-204	GRIA3	NM_000828.4	g.122538739G>A	c.1474G>A	p.(Gly492Ser)	maternelle	hémizygotique	3+

Sexe	Numéro d'identification	Gène	Isoforme	Coordonnées génomiques de la variation (g.)	Conséquence au niveau de l'ARN (c.)	Conséquence au niveau protéique (p.)	Transmission	Statut	Classe
H	APN-113	HCFC1	NM_005334.2	g.153230153G>A	c.218C>T	p.(Ala73Val)	maternelle	hémizygote	5
H	APN-82	IL1RAPL1	NM_014271.3	g.29935696_29935705del	c.894_903del	p.(Trp299Thrfs*18)	maternelle	hémizygote	5
H	APN-68	IQSEC2	NM_001111125.1	g.53268395G>A	c.3097C>T	p.(Gln1033*)	de novo	hémizygote	5
H	APN-34	KDM5C	NM_004187.3	g.53228250C>G	c.2152G>C	p.(Ala718Pro)	de novo	hémizygote	5
H	APN-135	KDM5C	NM_004187.3	g.53240784dup	c.1297delinsTG	p.(Glu433*)	maternelle	hémizygote	5
H	B00H4KG	KDM5C	NM_004187.3	g.53227806_53227807del	c.2381_2382del	p.Leu794Glnfs*6	maternelle	hémizygote	5
H	MMPN-67	KIAA2022	NM_001008537.2	g.73961580G>A	c.2812C>T	p.Leu938Phe	maternelle	hémizygote	3+
H	APN-162	KIAA2022	NM_001008537.2	g.73961667del	c.2725del	p.(Ala909Profs*13)	maternelle	hémizygote	5
H	APN-171	L1CAM	NM_000425.3	g.153129873T>G	c.3226A>C	p.(Thr1076Pro)	maternelle	hémizygote	4
H	APN-16	MAOA	NM_000240.3	g.43590942_43590943delinsTT	c.797_798delinsTT	p.Cys266Phe	maternelle	hémizygote	4
H	APN-3	MECP2	NM_004992.3	Réarrangement complexe	-	-	maternelle	hémizygote	5
H	B00H4HR	MECP2	NM_004992.3	g.153296860G>A	c.419C>T	p.(Ala140Val)	maternelle	hémizygote	4
H	APN-99	NLGN3	NM_018977.3	g.70389249C>T	c.1849C>T	p.(Arg617Trp)	maternelle	hémizygote	3++
H	MMPN-124	OPHN1	NM_002547	dup exons 4-5	-	-	?	hémizygote	5
H	B00H4TT	PAK3	NM_002578.3	g.110391067G>T	c.379G>T	p.(Asp127Tyr)	maternelle	hémizygote	3++
H	APN-105	PHF8	NM_015107.2	g.54028583C>G	c.1141+5G>C	(p.?)	maternelle	hémizygote	5
H	B00H4TK	PLP1	NM_001128834.2	g.103043401T>G	c.658T>G	p.Cys220Gly	maternelle	hémizygote	4
H	APN-70	PQBP1	NM_005710.2	g.48760294C>T	c.731C>T	p.Pro244Leu	maternelle	hémizygote	3++
H	APN-204	PTCHD1	NM_173495	del gène entier	-	-	maternelle	hémizygote	5
H	APN-110	SLC16A2	NM_006517.3	g.73749067T>C	c.1412T>C	p.(Leu471Pro)	maternelle	hémizygote	5
H	B00H4MP	SLC6A8	NM_005629.3	g.152959835_152959836del	c.1429_1430del	p.(Ser477Glyfs*24)	maternelle	hémizygote	5



Sexe	Numéro d'identification	Gène	Isoforme	Coordonnées génomiques de la variation (g.)	Conséquence au niveau de l'ARN (c.)	Conséquence au niveau protéique (p.)	Transmission	Statut	Classe
H	APN-43	SLC9A6	NM_001042537.1	g.135080258_135080262del	c.526-9_526-5del	p.(Val177_Arg202del)	maternelle	hémizygote	5
H	MMPN-15	SYN1	NM_006950.3	g.47464406G>A	c.745C>T	p.(Gln249*)	maternelle	hémizygote	4
H	B00H4KH	USP9X	NM_001039590.2	g.41029493G>A	c.2877+5G>A	(p.?)	maternelle	hémizygote	3++
H	B00H4L7	ZDHC9, RAB39B	NM_001008222.2, NM_171998.3	g.128957700G>A, g.154490388G>C	c.442C>T, c.342C>G	p.(Arg148Trp), p.(Tyr114*)	maternelle	hémizygote	4

**Table 5. Mutations causales et potentiellement causales identifiées au sein de notre cohorte de patients souffrant de DI isolée ou syndromique.**

Sont colorés en rouge pâle les variants étudiés dans ce premier chapitre. La classe des variations réfère aux catégories de pathogénicité recommandées par l'ACMG (Richards *et al.* 2015) : 5 : pathogène, 4 : potentiellement pathogène, 3+ et 3++ : signification inconnue avec quelques évidences suggérant leur pathogénicité. F : femme, H : homme.

*The end of this manuscript will be written in English to enable a better assessment  
by an English speaker examiner.*



## **Chapter 2.**

# **Effects of the androgens on human neural precursors**

One of the possible explanations for the male excess observed in the neurodevelopmental disorders (NDD) such as intellectual deficiency (ID) and autism spectrum disorder (ASD) is the role of prenatal androgens during male brain development. It is known that an early androgens exposure during a critic *in utero* period affects the brain development and this could explain why males are more susceptible to develop such kind of pathologies. This second part of my work mainly focused on this androgens' theory and tried to assess the effects of androgens on human neuronal precursors.

## **I. Preliminary results**

### **1. Choice of the cellular model**

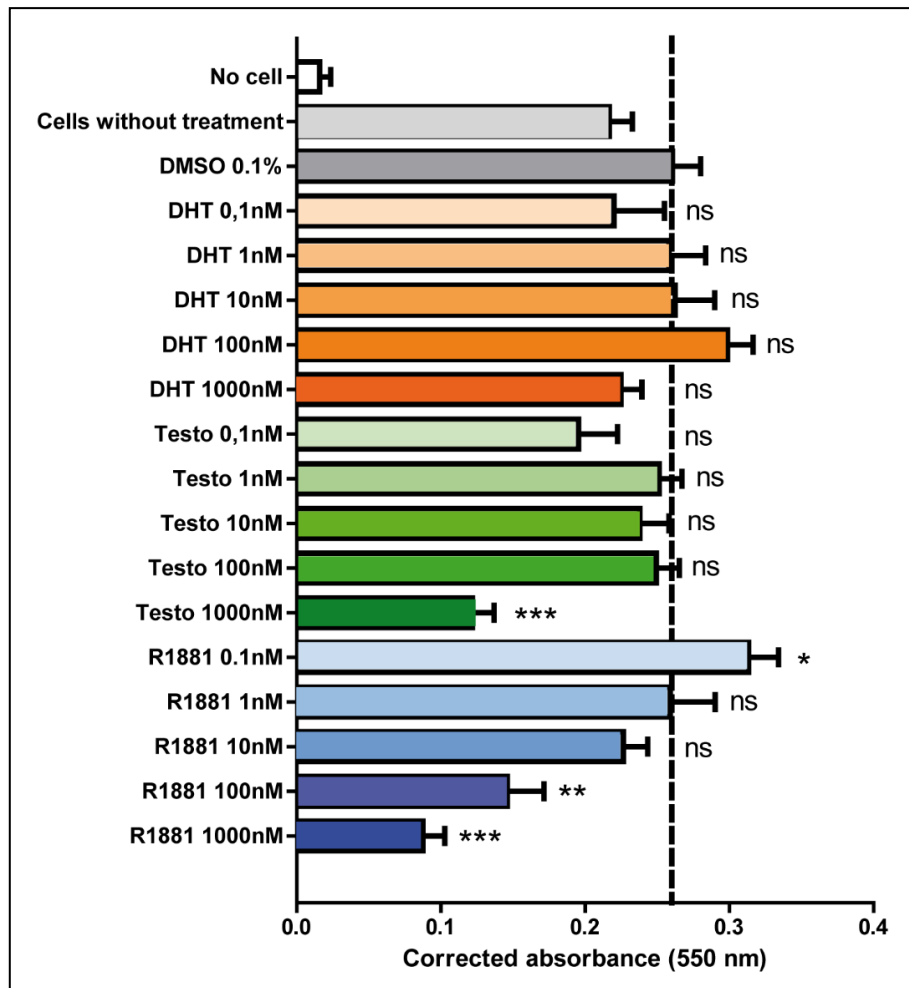
Most of the genome-wide studies analyzing the effects of androgens on the transcription regulation (RNA-seq) or identifying the AR target genes in human (ChIP-chip or ChIP-seq) have been performed on prostate cancer cell lines (mainly LNCaP). Because the AR transcription regulation is, as many other transcription factors, cell specific, the use of human neuronal cells is essential to decipher the role of androgens during brain development. For the present study, we decided to use human neural stem cells (hNSCs) derived from human embryonic stem cells (hESCs).

hNSCs are a stable population of self-renewing cells whose differentiation protocole was published by (Boissart *et al.*, 2013). Briefly, they can be easily maintained in a stable state by cultivating them in a specific medium containing growth factors (EGF, FGF-2 and BDNF) and neuronal differentiation can be induced by removing these growth factors. After differentiation, more than 90% of the cells are neurons, the other 10% being astrocytes and oligodendrocytes. The resulting neurons are mainly cortical neurons of the superficial layers. According to all of this, these cells are a useful model to study brain-related diseases.

The hNSC line we used is derived from male hESC line SA001 (derived from a surplus embryo from *in vitro* fertilization performed by Cellartis AB, Göteborg, Sweden). The hNSCs karyotype was normal (46,XY) and the genotyping by DNA chip (Infinium from Illumina) did not reveal any loss of heterozygosity (studies done by the team of Dre Benchoua who trained us and gave us the hNSC line). We looked at the *AR* gene and we did not find any mutations which could affect its function and the polyQ track was in the normal range (21 CAG). The full hNSCs characterization (quality control, AR expression at mRNA and protein level, AR isoforms and AR stabilization and translocation after androgens treatment) is described in the paper below (Quartier *et al.*, under review in Mol. Psychiatry).

## 2. Choice of androgens doses and cytotoxicity

Concerning the androgens doses, the physiological concentration range in men is from 10 (embryonic period) up to 30nM (during adulthood) for testosterone (and seemingly less for DHT) (Wudy *et al.*, 1999, 1995). In order to be sure that these concentrations are tolerated by hNSCs and derived-neurons a MTT test was performed with several androgens either natural (testosterone and DHT) or synthetic (R1881) with concentrations from 100-fold less to 100-fold more than the physiological concentration. The MTT assay (for 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) is a colorimetric assay for assessing cell metabolic activity and by extension, cell survival. As shown in **Figure 23**, DHT is not cytotoxic, testosterone starts to be cytotoxic at 1 $\mu$ M whereas R1881 leads to a significant cell death at 100nM and 1 $\mu$ M. Therefore the 10 and 100nM concentrations will be used for the natural androgens while the 1 and 10nM concentrations will be used for the R1881 (known to be as efficient as natural androgens with a 10-fold less concentration).



**Figure 23. Androgens cytotoxicity.** MTT assay showing the absorbance at 550nm revealing cell survival for each treatment (test performed in triplicate for a final  $14 < n < 35$ , Student's t-test: ns: non-significant, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).

### 3. Microarray on neurons differentiated from hNSCs

As a first trial, a microarray was performed on neurons differentiated from hNSCs during 17 days. After 17 days of differentiation and 24 hours androgens treatment (with the three above-cited androgens at the two determined concentrations), only few genes were found differentially expressed compared to the basal condition (DMSO; **Table 6**). Moreover, these significantly regulated genes were mainly known to be expressed by astrocytes (such as *AQP4*, *S100A10*...). These results led us to the conclusion that the cellular population after this hNSCs differentiation was enriched in astrocytes and/or, that the astrocytes were the main cell type expressing *AR*. Seeing that, we decided to pursue

our project on hNSCs because: 1) they represent a stable and homogenous cellular population and do not require further differentiation steps avoiding differentiation bias; 2) they correspond to an earlier development stage that is more relevant for our study concerning brain development and 3) they express four times more *AR* at the mRNA level.

Treatment	Significantly regulated genes	Overlap
DHT 10nM	145	64
DHT 100nM	156	
Testo 10nM	7	0
Testo 100nM	6	
R1881 1nM	43	4
R1881 10nM	12	

**Table 6.** Microarray results summary performed on neurons differentiated 17 days from hNSCs. Significantly regulated genes were selected with the following parameters: Student's test  $p < 0.01$ , fold change  $> |1.3|$  and intensity  $> 5$  (24 hours treatments,  $n = 3$ ).

## II. Results

**Paper:** Genes and mechanisms regulated by androgens as possible contributors to the male excess observed in autism

**Quartier A\***, Redin C, Chatrousse L, Haumesser N, Maglott-Roth A, Brino L, Jost B, Keime C, Le Gras S, Benchoua A, Mandel JL, Piton A, under review in Mol Psychiatry.

Aim of the study: Deciphering the role of androgens in human neural stem cells at the molecular (by RNA-seq and ChIP-seq) and cellular level (with proliferation, neuronal survival and neurite outgrowth experiments). This study could shed light on the role of androgens during brain development and their potential implication in the male excess observed in some neurodevelopmental disorders (such as intellectual deficiency or autism).

### Results:

- Human neural stem cells (hNSCs) express the androgen receptor (AR) and respond to androgens treatment (AR stabilization and relocation from the cytoplasm to the nucleus).
- Androgens lead to low but significant changes in the expression of 77 genes, the top-genes being *RGCC*, *RNF144B*, *NRCAM*, *TRIM22*, *FAM107A*, *LAMA2* and *IGFBP5*. These regulations are AR-dependent.
- 11,610 genomic sequences are found enriched in AR-bound chromatin. AR-bound sequences were identified in one third of the androgens-regulated genes.
- The androgens-regulated genes and the AR-bound genomic sequences are enriched in genes involved in ID or ASD such as *NRCAM* and *FAM107A*.
- Androgens promote hNSCs proliferation through AR and presumably *via FAM107A* upregulation.
- Androgens prevent neuronal death during differentiation in stressful conditions.

Contribution: Major. Performed most of the experiments, analyzed and interpreted the processed data, wrote the paper along with Dre Piton.

*NB:* According to their size, only the first lanes of the supplementary tables are enclosed to have an overview of their content.





1 **Genes and mechanisms regulated by androgens as possible contributors to**  
2 **the male excess observed in autism**

3 **Angélique Quartier**<sup>1,2</sup>, Claire Redin<sup>1,2</sup>, Laure Chatrousse<sup>3</sup>, Nicolas Haumesser<sup>1</sup>,  
4 Anne Maglott-Roth<sup>4</sup>, Laurent Brino<sup>4</sup>, Céline Keime<sup>5</sup>, Stéphanie Le Gras<sup>5</sup>, Alexandra Benchoua<sup>3</sup>,  
5 Jean-Louis Mandel<sup>1,2,6@\*</sup>, Amélie Piton<sup>1,6@\*</sup>

6

7 <sup>1</sup> Department of Translational Medicine and Neurogenetics, IGBMC, Illkirch, France;

8 <sup>2</sup> Chair of Human Genetics, Collège de France, Paris, France;

9 <sup>3</sup> CECS, I-STEM, AFM, Corbeil-Essonnes, France

10 <sup>4</sup> High-throughput screening platform, IGBMC, Illkirch, France;

11 <sup>5</sup> Microarrays and deep sequencing platform, IGBMC, Illkirch, France;

12 <sup>6</sup> Laboratory of Genetic Diagnostic, Hôpitaux Universitaires de Strasbourg, France

13

14 The authors declare no conflict of interest.

15 \* should be considered at co-last authors

16

17 Address for correspondence:

18 Jean-Louis Mandel, MD, PhD

19 Laboratoire "Mécanismes génétiques des maladies neurodéveloppementales",

20 IGBMC

21 1, rue Laurent Fries,

22 67400, Illkirch, France

23 Tel: +33388653210

24 E-mail: jlmandel@igbmc.fr

25

26 Amélie Piton, PhD

27 Laboratoire "Mécanismes génétiques des maladies neurodéveloppementales",

28 IGBMC

29 1, rue Laurent Fries,

30 67400, Illkirch, France

31 Tel: +33369551652

32 E-mail: piton@igbmc.fr

33 **Abstract**

34 Autism spectrum disorders (ASD) and intellectual disability (ID) are common  
35 neurodevelopmental disorders (ND) with clinical and genetic overlaps and that affects more  
36 males. This gender bias, very strong for ASD and notable for ID, remains unexplained. One  
37 hypothesis is that prenatal exposure to androgens during male brain development may favor an  
38 increased susceptibility of males to ND. We decided to use human neural precursors (hNSCs)  
39 differentiated from embryonic stem cells to investigate the role of androgens during brain  
40 development. We show that hNSCs are a suitable model to study androgens actions. We  
41 observed that androgens increase hNSCs proliferation and protected them against stress-induced  
42 cell death. RNA-sequencing revealed that androgens treatment led to subtle but robust changes  
43 in the expression of less than hundred genes. Androgen receptor (AR) knock-down in hNSCs  
44 abolished these regulations, while they were replicated in neuroblastoma cells SH-SY5Y  
45 overexpressing AR. A ChIP-sequencing experiment allowed the identification of AR-target  
46 genomic regions in one third of the androgens-regulated genes and in numerous other genes.  
47 We observed that androgens-target genes are enriched in genes involved in ASD or ID. Among  
48 these, *NRCAM* encodes a neural cell adhesion protein and *FAM107A*, encoding a stress-  
49 responding protein that mediated at least partially the effects of androgens on cell proliferation.  
50 The regulation of these genes and pathways by androgens during brain development might  
51 therefore participate to the increased susceptibility of male brain, exposed to other genetic and  
52 environmental factors, to develop ND. This approach might also be relevant for other sex-biased  
53 ND.

54

## 55 **Introduction**

56 Autism spectrum disorders (ASD) are characterized by deficits in social communication  
57 and restricted interests and/or stereotyped repetitive behaviors, often associated with  
58 impairment in language development and intellectual disability (ID). ASD affect 1.2% of the  
59 population <sup>1</sup>. Environmental and genetic factors play a role in the etiology of autism, as revealed  
60 by twin and familial studies <sup>2</sup>. Numerous studies have been led in order to identify common  
61 variants involved in the susceptibility to ASD, under the hypothesis of a complex multigenetic  
62 phenotype. Despite the lack of replication between these different studies and the difficulty to  
63 link the positive signals identified to a causative variant and to functional effects, there has been  
64 recent support for the hypothesis that common variants with low odds ratios play indeed a role  
65 in autism etiology, confirming that genetic of ASD is at least partially multifactorial <sup>3</sup>. Other  
66 studies suggested that oligogenic inheritance implicating rare variants (single nucleotide  
67 polymorphisms SNPs or copy number variants CNVs) may play a role in autism etiology <sup>4, 5</sup>.  
68 Environmental factors seem also to participate equally to genetic factors <sup>6</sup>. On the other hand,  
69 some forms of ASD are caused by single genetic events: chromosomal anomalies, CNVs or  
70 intragenic and point mutations affecting one single gene such as *SHANK3*, *CHD8*, *POZG* or  
71 *TBR1* among others <sup>7</sup>. The genes involved in monogenic forms of ASD are usually also involved  
72 in ID, but the part of monogenic forms in the genetic architecture of ASD seems to be lower  
73 than for ID.

74 Numerous studies have concluded that ASD show a much higher frequency in males,  
75 about 4:1 and even 6:1 among those without associated ID <sup>8</sup>. Under or misdiagnosis in females  
76 with ASD have been recently discussed and might participate to the phenomenon especially for

77 high-functioning ASD<sup>9</sup>. Up to now there is no clearly established model explaining this sex  
78 bias<sup>10</sup>. Several genes located on the X-chromosome are involved in monogenic syndromic  
79 (*FMRI*, *MECP2*, *MAOA*, etc) or non-syndromic (*NLGN3*, *NLGN4X*, *ILIRAPL1*, etc) forms of  
80 ASD but, their cumulative frequency remains very low and fails to explain this huge male  
81 excess. Moreover, pathogenic mutations in X-linked ASD genes are also observed in girls with  
82 ASD. Several studies which focused on the X-chromosome to identify genetic causes of ASD  
83 did not identify an excess of pathogenic mutations in males on genes located on this  
84 chromosome<sup>11,12</sup>. Surprisingly, an excess in number and size of CNVs was observed in ASD  
85 females compared to ASD males<sup>13</sup> suggesting that the mutation load is higher in affected  
86 females than in affected males. In parallel, it was shown that this mutation load is higher in  
87 males with ASD and low intellectual quotient (IQ) compared to males with ASD without ID<sup>14</sup>.  
88 One might therefore suppose that in individuals with ASD and no ID, who are mainly males,  
89 more complex genetic and environmental risk factors are involved, including those having a  
90 sex-specific impact. The biological origin of this sex-specific inequality could be due to a  
91 “protective effect” in females, as proposed by Robinson *et al.* and Jacquemont *et al.*<sup>13,15</sup>, or a  
92 “predisposing effect” in males, as defended by Baron-Cohen *et al.*<sup>16</sup>. The authors went even  
93 further setting the “extreme male brain” theory suggesting that males even in the general  
94 population present more autistic features than females<sup>16-18</sup>. The biological mechanisms that  
95 might be responsible for this inequality in predisposition could be classified in two categories  
96 that are not mutually exclusive<sup>16,17</sup>: 1) risk/protection factors linked to sex chromosomes X or  
97 Y: role of X-linked genes escaping X-inactivation<sup>19</sup> or submitted to specific parent-of-origin  
98 allelic imprinting<sup>20,21</sup>, influence of total chromatin size<sup>22</sup>, predisposing variants in Y-linked  
99 genes<sup>23-26</sup> Sex determining region Y (SRY) gene<sup>27-29</sup> or 2) risk/protection factors due to sex-

100 specific hormonal impregnation: hormones which are not secreted during neural development  
101 at the same level in males and females could have a role during brain development and therefore  
102 modulate the susceptibility of the brain to develop ASD or other neurodevelopmental disorders  
103 (ND) when combined to other genetic or environmental insults.

104         In support of the hormonal hypothesis, several observations suggest that fetal androgens  
105 (produced early by Leydig cells of the developing testis) are implicated in ASD. Studies have  
106 correlated testosterone levels measured in amniotic fluid, with measures of traits related to  
107 autism in cohorts of normal children followed at various ages (the Cambridge Fetal Testosterone  
108 project)<sup>30</sup>. Elevated testosterone levels were also reported in females with Asperger syndrome  
109 <sup>31</sup> and some studies even showed a differential expression of genes involved in steroids  
110 biosynthesis in ASD patients compared to controls <sup>32,33</sup>. Other arguments suggesting the role of  
111 fetal androgens stem from behavior and cognitive studies in females with congenital adrenal  
112 hyperplasia, a genetic disease leading to overproduction of adrenal androgens during fetal  
113 development, who are more likely to present autistic traits <sup>34</sup>. We could therefore make the  
114 hypothesis that androgens prenatal exposure may act as a sex-specific environmental risk factor  
115 increasing penetrance of other genetic (or environmental) susceptibility factors, contributing to  
116 a global increase susceptibility of males for ASD.

117         There are two major peaks of testosterone production during early life in males, one  
118 during the second trimester of pregnancy and one during the first three postnatal months, periods  
119 of great importance in nervous system development <sup>35</sup>. Testosterone readily passes the blood-  
120 brain barrier, and its receptor, the androgen receptor (AR), is expressed in various brain cells  
121 (particularly in neurons) and brain regions <sup>36</sup>. In rodents, AR is expressed in areas important for  
122 the regulation of emotions, cognition and behavior (limbic system, cortex, CA1 hippocampus,

123 etc). Androgens seem to play notably a major role in establishing anatomical and functional  
124 (especially behavioral) sexual dimorphism in the nervous system<sup>37,38</sup>. Moreover, several studies  
125 showed that early exposure to testosterone or to its more active metabolite 5-alpha-  
126 dihydroxytestosterone (DHT), has a neuroprotective activity<sup>39, 40</sup> and modulates synaptic  
127 density and neurite outgrowth in some brain regions<sup>41,42</sup>.

128         Nevertheless, almost all genome wide data of androgens effect on transcription  
129 regulation and AR targets in humans have been obtained on prostate cancer or muscle cells<sup>43</sup>,  
130<sup>44</sup> and little is known concerning AR target genes in neuronal cells. We therefore decided to  
131 study the effect of androgens treatments in male human pluripotent neural stem cells (hNSCs)  
132 differentiated from embryonic stem cells. We performed RNA-sequencing and identified genes  
133 whose expression is regulated by androgens, and demonstrated that this regulation involves the  
134 AR. We also showed that androgens protect neurons in differentiation from cell death in stress  
135 conditions. We observed an enrichment of genes mutated in monogenic forms of ASD and/or  
136 ID and of genes differentially expressed in ASD patients among these androgens-target genes.  
137 Moreover, one of the main androgens-target genes, *NRCAM*, which encodes a neuronal cell  
138 adhesion molecule regulating neurite outgrowth, was previously described as an autism  
139 associated gene: several SNPs in its promoter or intronic regions appeared significantly  
140 associated with ASD<sup>45,46</sup>.

141

142 **Materials and methods**

143 **Reagents, plasmids, small interfering RNAs and antibodies**

144 Dimethyl sulfoxide (DMSO), 5alpha-dihydrotestosterone (DHT), testosterone, R1881,  
145 dexamethasone, L-685458 and retinoic acid were purchased from Sigma-Aldrich (St Louis,  
146 MO, USA). pE-AR plasmid was kindly provided by Dr. Jocelyn Ceraline (Faculté de Médecine,  
147 Strasbourg, France). The siRNA were purchased from Dharmacon (GE Healthcare, Lafayette,  
148 CO, USA). For western blots and immunostainings, the following specific primary antibodies  
149 were used: polyclonal anti-AR N-20, polyclonal anti-AR C-19, polyclonal anti-GR M-20,  
150 polyclonal anti-RGCC C-14, polyclonal anti-FAM107A N-17 (all from Santa Cruz, Heidelberg,  
151 Germany), monoclonal anti-GAPDH, monoclonal anti-RNF144B (LifeSpan BioSciences,  
152 Seattle, WA, USA), monoclonal anti-NRCAM (R&D Systems, Minneapolis, MN, USA),  
153 monoclonal anti-TRIM22 (Origene, Rockville, MD, USA), PerCP-Cy5.5-conjugated SOX2  
154 monoclonal antibody (BD Biosciences, Franklin Lakes, NJ, USA), monoclonal anti-HuC/D  
155 (Thermo Fisher Scientific, Waltham, MA, USA) and chicken anti-TUJ1 (Merck Millipore  
156 Darmstadt, Germany). The secondary antibodies used were HRP-labelled goat anti-mouse IgG  
157 and HRP or Alexa-488 goat anti-rabbit IgG (Jackson immunoresearch, Baltimore, PA, USA).

158 **Cell lines culture and treatment**

159 The male hESC line SA001 was obtained from Cellartis (Goteborg, Sweden). The work  
160 involving hESC line was supervised by the French Bioethics Agency (Agreement number NOR  
161 AFSB 12002 43S). Human neural stem cells (hNSCs) were derived as described in Boissart *et*  
162 *al.*<sup>47</sup>. Briefly, hNSCs were seeded on poly-ornithine and laminin-coated (Sigma-Aldrich) dishes



163 and maintained in N2B27 medium (DMEM/F12 and Neurobasal medium (1:1) supplemented  
164 with N2, B27, 2-mercaptoethanol (all from Invitrogen, Carlsbad, CA, USA)) containing BDNF  
165 (20ng/mL), FGF-2 (10ng/mL) (both from PeproTech, Rocky Hill, NJ, USA) and EGF ((R&D  
166 Systems; 10ng/mL). For the neuronal differentiation, cells were cultured in N2B27 medium  
167 without growth factors. The medium was changed every four days with systematic addition of  
168 2µg/mL of laminin to avoid detachment and clumping of neurons. Quality control was  
169 performed as described in **Supplementary Information**.

170 LNCaP and SH-SY5Y cells were obtained from the American Type Culture Collection (ATCC,  
171 Manassas, VA, USA). LNCaP were seeded in CellBind® flasks (Corning, NY, USA) in RPMI  
172 1640 medium with 10mM HEPES, 10% fetal calf serum, 1mM sodium pyruvate and gentamicin  
173 (40µg/mL). SH-SY5Y were maintained in RPMI 1640 medium without HEPES, 10% fetal calf  
174 serum and gentamicin (40µg/mL).

175 Cells were placed in depleted medium (phenol-red free medium with charcoal treated serum for  
176 LNCaP and SH-SY5Y; phenol-red free medium with no N2 or B27 for hNSCs) during 48h  
177 hours before androgens treatments.

### 178 **RNA isolation, reverse transcription-PCR and real-time quantitative PCR**

179 RNA samples extracted from hNSCs were tested with RT-PCR and RT-qPCR as described in  
180 **Supplementary Information**, primer list in **Supplementary Table S8**.

181

182 **mRNA-sequencing (RNA-seq)**

183 Three batches (biological replicates) of treated hNSCs were used with technical replicates: four  
184 DMSO *vs.* three DHT 100nM for batch 1, three DMSO *vs.* three DHT 100nM for batch 2 and  
185 three DMSO *vs.* three DHT 100nM for batch 3. After RNA isolation, libraries of template  
186 molecules suitable for high throughput DNA-sequencing were created using “TruSeq™ RNA  
187 Sample Preparation v2 Kit” (Illumina, San Diego, CA, USA; see **Supplementary**  
188 **Information**). DNA libraries were checked for quality and quantified using 2100 Bioanalyzer.  
189 The libraries were loaded in the flow cell at 7.5pM concentration and clusters were generated  
190 in the Cbot and sequenced in the Illumina Hiseq 2500 as single-end 50 base reads. Image  
191 analysis and base calling were performed using RTA 1.18.61 and CASAVA 1.8.2.

192 **Chromatin immunoprecipitation combined with direct sequencing (ChIP-seq)**

193 ChIP protocol was adapted from Massie and Mills *et al.* and Adli and Bernstein *et al.* <sup>43, 48</sup>  
194 (see **Supplementary Information**). Cells were treated with DMSO or androgens during 4h.  
195 AR polyclonal antibodies used for ChIP were AR N-20 and anti-AR 3279.4 (homemade).  
196 ChIP enrichment was tested by real-time PCR (see **Supplementary Information**) and the  
197 remainder was used for single-end library preparation. ChIP-seq libraries were prepared using  
198 NEXTflex ChIP-Seq Kit (#5143-02, Bioo Scientific, Austin, TX, USA) following the  
199 manufacturer's protocol (V12.10) with some modifications detailed in the **Supplementary**  
200 **Information**. Prior to analyses DNA libraries were checked for quality and quantified using a  
201 2100 Bioanalyzer. The libraries were loaded in the flow cell at 8pM concentration and clusters  
202 were generated using the Cbot and sequenced on the Illumina Hiseq2500 as single-end 50 base

203 reads following Illumina's instructions. Image analysis and base calling were performed using  
204 RTA 1.18.61 and bcl2fastq 1.8.2.

### 205 **Bioinformatic analysis**

206 RNA-seq and ChIP-seq analysis were performed as described in **Supplementary Information**.  
207 GEO accession number: GSE86458

### 208 **Pathways analysis and genes lists selection**

209 Significant differential expressed genes (Benjamini corrected p-value <0.05 and at least |1.2|  
210 fold changes) were considered for the Ingenuity Pathways Analysis (IPA): a core analysis  
211 considering direct and indirect relationships was performed on the list of genes obtained.  
212 Functional annotation was also performed using the Database for Annotation, Visualization and  
213 Integrated Discovery (DAVID 6.7, current release) <sup>49</sup>. Biological processes, cellular  
214 components and molecular functions of Gene Ontology Consortium (GO) were used for the  
215 functional annotations.

216 List of genes mutated in at least one ASD/ID patient have been compiled from a combination  
217 of gene lists already published <sup>50-52</sup> (**Supplementary Table S9**). List of genes differentially  
218 expressed in autistic patients data available from microarray performed in blood, brain and  
219 lymphoblastoid cell lines <sup>33,53-57</sup>. List of genes associated (from genome wide association studies  
220 or candidate genes studies) with ASD were retrieved from AutismKb  
221 (<http://autismkb.cbi.pku.edu.cn/>; **Supplementary Table S9**).

222

## 223 **Plasmid and small interfering RNA transfection**

224 To overexpress AR in SH-SY5Y, cells were transfected at about 60% of confluency with the  
225 pE-AR plasmid by using lipofectamine 2000 transfection reagent (Thermo Fisher Scientific).  
226 To knockdown AR expression, hNSCs were seeded at 30 000 cells/cm<sup>2</sup> and transfected with a  
227 pool of 4 AR siRNAs (20nM each) or scramble siRNA by using INTERFERin reverse  
228 transfection protocol (Polyplus-transfection, Illkirch-Graffenstaden, France). After 36h of  
229 incubation, cells were exposed to DMSO or 100nM DHT for 24h in the depleted medium. Cells  
230 were harvested for total RNA or protein extraction.

## 231 **Western Blot and immunofluorescence**

232 Protein analyses by western blot and immunofluorescence were performed using standard  
233 protocols as described in **Supplementary Information**.

## 234 **Proliferation assay**

235 Reverse transfection of hNSCs (20,000 cells/cm<sup>2</sup>) was done in 96-well plates with 20nM siRNA  
236 and INTERFERin in OptiMEM. Untreated wells were used to determine baseline phenotype  
237 and cell viability, negative control siRNA (scramble siRNA) to distinguish sequence-specific  
238 silencing from non-specific affects, positive control siRNA (PLK1 siRNA) to measure  
239 efficiency of siRNA uptake into cells and test siRNA (*AR*, *FAM107A* or *RGCC* siRNA) to  
240 knockdown target genes. The day after, cells were treated with DMSO or DHT 100nM.  
241 Each day a plate was fixed and stained with DAPI and analysis was done to count the number  
242 of nucleus in each condition using CellInsight automated microscope and HCS Studio software  
243 (Thermo Fisher Scientific).

244 **Neurite outgrowth assay**

245 The protocol used was described in Boissard et al., 2013 <sup>47</sup>. Briefly, hNSCs were seeded  
246 automatically in 384-well plates using the Bravo automated liquid Handling Platform (Agilent,  
247 Santa Clara, CA, USA) and differentiated in neurons during 14 days. The compounds were  
248 delivered 6h after plating and replaced when changing the medium at days 4, 7 and 11 using the  
249 same platform. For high-content screening assays, the fluorescent labeling was performed in an  
250 automated manner using the Bravo liquid handling system (anti-TUJ1 labelling and Hoechst  
251 nuclear staining). Fluorescence image acquisitions and analyses were performed using the  
252 Cellomics Arrayscan automated microscope (Thermo Fisher Scientific). To quantify neurite  
253 outgrowth and branching points, the Cellomics 'neuronal profiling' bioapplication was used to  
254 track TUJ1<sup>+</sup> neurites. A minimum of 200 neurons per well were analyzed in at least four wells.

255 **Statistical analysis**

256 Student's t-test was used to determine if androgens treatments were significantly different from  
257 basal DMSO condition (RT-qPCR, proliferation and neurite outgrowth experiments).  
258 The hypergeometric test was used to calculate the statistical significance of genes or peaks lists  
259 overlap. The population of genes expressed in hNSCs (according to RNA-seq DMSO data) was  
260 used as population size to consider the cell-specific gene expression.

261

262 **Results**263 **hNSCs is a relevant model to study androgens effects**

264 Human pluripotent stem cell-derived neural stem cells (hNSCs) represent a  
265 phenotypically stable population of self-renewing cells differentiated from human embryonic  
266 stem cells. They can be maintained for long-term expansion using a cocktail of growth factors  
267 <sup>47</sup>. We have used them to study the effect of androgens on human neural cells. In order to control  
268 the stability of the hNSCs, the quality control of these cells was performed by SOX2/HuC/D  
269 immunostaining, self-renewing undifferentiated hNSCs being characterized by a SOX2<sup>+</sup> and  
270 HuC/D<sup>-</sup> staining (**Supplementary Figure S1**). We checked the androgen receptor (AR)  
271 expression by RT-qPCR on mRNA extracted from hNSCs, hNSCs differentiated during 17 days  
272 into neurons (ND17), neuroblastoma cell line SH-SY5Y and a well-studied prostate cancer cell  
273 line known to very highly express AR, LNCaP (**Figure 1**). We showed that hNSCs express a  
274 significant level of *AR* mRNA, however 20 fold lower than in LNCaP. This expression  
275 decreased threefold after neuronal differentiation (**Figure 1A**). Both protein-coding isoforms  
276 (NM\_00004.3 (large) and NM\_001011645.2 (short), **Figure 1C**) are expressed in LNCaP,  
277 hNSCs and ND17 (**Figure 1B**). Western blots performed with two different antibodies  
278 (one specific for the long isoform and the other reacting with both isoforms) confirmed the  
279 presence of the two coding isoforms, the full length 110kDa and the 45kDa (lacking the N-terminal  
280 protein-interacting domain) in LNCaP and hNSCs (**Figure 1C**).

281 To check if AR translocates to the nucleus after androgens treatment, hNSCs were  
282 exposed to 5-alpha-dihydrotestosterone (DHT, a natural metabolite of testosterone and a  
283 high-affinity ligand of AR) during 4h. Immunostaining with an anti-AR antibody (N-20)

284 revealed that AR was diffuse in the cytosol and the nucleus in the basal state (DMSO treatment,  
285 **Figure 1D, top panel**) and the cytosolic staining was translocated to the nucleus after DHT  
286 100nM treatment (**Figure 1D, bottom panel**). In addition, we observed a strong increase of the  
287 full length AR protein level after androgens treatment, while mRNA level was not affected  
288 (**Figure 1E** and **Supplementary Figure S3**). This fits with the known stabilization of AR  
289 protein when bound to its ligand <sup>58</sup>. As there are many crosstalks between the AR and other  
290 steroid receptors, we assessed their expression and showed that the glucocorticoid receptor  
291 (GR), is the only other steroid receptor expressed in hNSCs. Treatments with androgens do not  
292 affect GR protein level and glucocorticoid treatment does not affect AR protein level  
293 (**Supplementary Figure S2**).

294 Together, the data revealed that AR is well expressed in hNSCs, and responds to  
295 androgens treatment by relocation from the cytoplasm to the nucleus, where it can exert its direct  
296 regulation of gene expression, making these cells a relevant human neural model to study the  
297 effects of androgens on gene expression regulation.

## 298 **Identification of androgens-target genes in hNSCs**

299 To identify the genes regulated by androgens in hNSCs, we performed transcriptomic  
300 analysis (RNA-seq) comparing gene expression in three different series of hNSCs treated with  
301 DMSO (total n = 10) or with DHT 100nM (total n = 9) during 24h. Batch effect was adjusted  
302 as described in **Supplementary Information**. In one of these series, treatments with other  
303 androgens at several concentrations (DHT 10nM, testosterone 10 and 100nM and R1881 1nM)  
304 were also performed. A list of 106 genes which expression was significantly changed after DHT  
305 100nM treatment was established (fold change above |1.2| and p-value adjusted for multiple

306 testing under 0.05, **Supplementary Table S1**). Seventy-seven of them were significantly  
307 expressed in hNSCs (above 100 reads normalized and divided by gene length for each condition)  
308 and comprise 52 up-regulated and 25 down-regulated genes (red dots on the volcano plot,  
309 **Figure 2A**). In comparison, retinoic acid (RA) regulates 2 760 genes (**Supplementary**  
310 **Table S6**). Fold changes in gene expression were also lower for DHT (between -1.83 and 2.16)  
311 than for RA (between -11.49 and 71.22).

312 We performed pathways analyses and observed an enrichment in genes involved in  
313 "nervous system development and function" (IPA, **Supplementary Table S2**) among the  
314 androgens-regulated genes. The most enriched functional category concerned the regulation of the  
315 morphology of nervous system and more precisely regulation of neurite growth and branching.  
316 Gene ontology (GO) term analysis (DAVID, **Supplementary Table S3**) revealed four terms  
317 enriched in our dataset (non-adjusted p-value < 0.005): '*cell adhesion*', '*biological adhesion*',  
318 '*regulation of cell growth*' and '*vitamin binding*'. Interestingly, 47 of the 77 DHT-regulated genes  
319 are also expressed in human prostate (DAVID tissue expression analysis, data not shown).

320 The androgens-target genes identified were ranked according to statistical significance  
321 (adjusted p-value) and reproducibility (regulation by other androgens, testosterone and R1881,  
322 data not shown). The "top-five" target genes (large red dots on **Figure 2A**) regulated by  
323 androgens in hNSCs were: *RGCC* (fold change 2.11, p-value  $5.22E^{-36}$ ), *RNF144B*  
324 (fold change 1.74, p-value  $3.94E^{-19}$ ), *NRCAM* (fold change 1.33, p-value  $6.73E^{-17}$ ), *TRIM22*  
325 (fold change 1.35, p-value  $8.08E^{-10}$ ) and *FAM107A* (fold change 1.49, p-value  $1.15E^{-4}$ ).  
326 These best genes are involved in many different biological processes such as cell cycle/cell  
327 growth (*RGCC* and *FAM107A*), ubiquitination (*RNF144B*), axon growth/ neuron migration  
328 (*NRCAM*) or transcription (*TRIM22*, **Figure 2B**).



## 329 **Regulation of target gene expression by androgens acts through the AR**

330 To confirm the effect of androgens on the expression of these genes, we performed  
331 RT-qPCR and western blot analyses in hNSCs. We confirmed an increase of *RGCC* at the  
332 mRNA and protein level after different androgens treatments (**Figure 2C**). The same results  
333 were obtained for *RNF144B*, *NRCAM*, *TRIM22* and *FAM107A* (**Supplementary Figure S3**).  
334 We then checked if this up-regulation by androgens involved the AR. No androgens regulation  
335 was observed in SH-SY5Y cells that do not express AR (**Figure 1**). When we overexpressed  
336 AR in these cells using an AR expression vector (pE-AR; **Supplementary Figure S4A**),  
337 we observed an up-regulation of mRNA levels of target genes after DHT treatment (**Figure 2D**  
338 and **Supplementary Fig S4B**). In contrast, when AR was efficiently knocked-down in hNSC  
339 by AR siRNA transfection (**Supplementary Figure S5A**), we observed an abolition of the effect  
340 of DHT treatment on gene expression (**Figure 2E** and **Supplementary Figure S5B**). Altogether  
341 these results confirm that DHT acts through AR to regulate gene expression in hNSCs.

## 342 **DHT affects hNSCs proliferation**

343 Several studies performed in rodents suggested that androgens could play a role in cell  
344 proliferation<sup>59,60</sup>. We observed an enrichment in genes involved in the regulation of cell growth  
345 among androgens-target genes (**Supplementary Table S3**) and two of the “top- five” target  
346 genes (*RGCC* and *FAM107A*) are involved in the negative regulation of exit from mitosis and  
347 cell growth<sup>61</sup>. We observed that DHT treatment significantly increased hNSCs proliferation  
348 compared to basal condition (**Figure 3A**). *AR* knock-down by siRNA transfection prevented  
349 this enhancement of proliferation by DHT. Knock-down of *FAM107A* leads to a similar  
350 phenotype, suggesting that enhancement of hNSCs proliferation required the up-regulation of

351 *FAM107A* by DHT though AR binding. Interestingly, the down-regulation of *RGCC* abolished  
352 hNSCs proliferation and caused cell death at day 4. This drastic effect is independent of  
353 androgens treatment suggesting that AR is not involved in this process and prevent us to study  
354 the implication of *RGCC* on DHT-mediated increase of cell proliferation.

### 355 **DHT prevents neuronal death during differentiation in stressful conditions**

356 Several studies have suggested a neuroprotective role of androgens<sup>39, 40</sup>. To study if  
357 androgens may have a neuroprotective effect during neuronal differentiation, hNSCs were  
358 differentiated into neurons during 17 days with either classical medium or medium depleted in  
359 N2/B27 supplements, which represents a stressful environment for cells, then treated with DHT  
360 100nM (**Figure 3B**). At the end of the differentiation (day 17), no difference was observed  
361 between basal condition (DMSO) and DHT treatment with the classical medium: neurons were  
362 adherent and formed a complex network with many neurites (**Figure 3B, top panel**). In contrast,  
363 in depleted medium, hNSCs died during differentiation and no neurons were obtained after 17  
364 days (**Figure 3B, bottom panel**). This phenomenon was prevented by DHT treatment.  
365 To quantify this neuroprotective effect, the number of wells containing living and differentiated  
366 cells was assessed every 3 days (**Figure 3B, Supplementary Figure S6**). At day 14 in depleted  
367 medium, cell survived in only 40% of wells in basal condition while they survived in 90% when  
368 they were treated by DHT. At day 17, no cells survived without DHT treatment while 67% of  
369 wells treated by DHT showed differentiated living neurons. These results showed that DHT  
370 prevented neuronal death induced by stressful conditions during differentiation.

371

## 372 **DHT has no measurable effect on neurite outgrowth**

373           Androgens have been shown to modulate neurite outgrowth, dendritic spine and synapse  
374 formation in different models of rodent neuronal cells<sup>41, 42, 62-65</sup>. To test if androgens modulate  
375 neurite outgrowth in this model, hNSCs were treated during their differentiation with a broad  
376 range of DHT concentrations, with L-685458, a molecule demonstrated to inhibit neurite  
377 outgrowth<sup>47</sup> or with retinoic acid (RA), known to enhance neurite outgrowth. After verifying  
378 that DHT treatment does not affect the proliferation/differentiation balance during hNSCs  
379 differentiation (at days 4, 7 and 14, data not shown), we measured total neurite length, total  
380 neurite number and branching points formed per neuron after 14 days of neuronal  
381 differentiation. Higher concentrations of L-685458 (> 4 $\mu$ M) decrease the length of neurites and  
382 number of branching point per neuron, while almost all L-685458 concentrations tested affect  
383 the number of neurites (**Figure 3C**). High RA concentrations, as expected, increased the three  
384 parameters measured. On the contrary, no significant variations were observed after DHT  
385 treatment. The same results were obtained with testosterone and R1881 treatments  
386 (**Supplementary Figure S7**) suggesting that androgens have no detectable effects on neurite  
387 outgrowth in hNSCs.

## 388 **Identification of AR target sequences**

389           AR is a ligand-induced transcription factor which binds to specific target sequences on  
390 DNA. To identify these AR target sequences in hNSCs, we performed a chromatin  
391 immunoprecipitation of AR followed by DNA-sequencing (ChIP-seq). We first performed a  
392 ChIP-seq with two different anti-AR antibodies after 4h of DHT 100nM treatment on LNCaP  
393 cell lines and confirmed that this allows identification of AR target DNA sequences

394 (**Supplementary Figure S8** and **Supplementary Table S4**). ChIP-seq experiments were then  
395 performed with the optimized protocol in hNSCs after DHT (10 and 100nM) or testosterone  
396 (100nM) treatment. In total, 11 610 genomic sequences, corresponding to 9 192 genes, were  
397 bound by the AR in hNSCs after an androgens treatment. There was less than 1% overlap with  
398 peaks identified in LNCaP (data not shown). Motif analysis revealed that the AR target genomic  
399 sequences are, among other, enriched in SP1, SP2, EGR1 and NRF1 motifs (**Supplementary**  
400 **Figure S9**). The AR and GR/NR3C1 motifs, although not found in the “top-enriched” motifs,  
401 were still slightly enriched (GR/NR3C1: z-score 2.77 pvalue 0.006; AR: zscore 1.95 pvalue 0.05).

402 We identified AR enriched peaks in 36% of the DHT regulated genes (twenty-eight  
403 genes detailed in **Supplementary Table S5**, hypergeometric law  $p = 0.016$ ). Among them, one  
404 peak identified in *FAM107A* corresponds to an uncharacterized GRE element. Three peaks were  
405 identified in *NRCAM*, the most enriched located in the third intron of the gene (chr7:107949253-  
406 107950043), overlapping a putative ARE/GRE sequence 5'-AGAACA<sub>n</sub>nnTGTTCT-3'  
407 (**Figure 4B**). This peak included putative GRE/NR3C1 and FOXA1 binding sequences.  
408 We confirmed the enrichment of this sequence in AR-bound immunoprecipitated chromatin by  
409 ChIP-qPCR (**Figure 4C**).

#### 410 **Androgens-target genes are enriched in genes involved in ASD and/or ID**

411 We compared the list of genes regulated by androgens in hNSCs with genes involved in  
412 ASD/ID (combination of gene lists already published, see methods; **Table 1**). Among the  
413 77 DHT-regulated genes, six are known to be mutated in at least one ASD/ID patient,  
414 corresponding to a borderline significant enrichment ( $p = 0.029$ ). Moreover, almost one quarter  
415 (15 genes,  $p = 0.0012$ ) of the DHT-regulated genes have been found differentially expressed in

416 autistic patients (data available from transcriptomic microarrays performed in blood, brain and  
417 lymphoblastoid cell lines, see methods). A significant overlap was found with the genes mutated  
418 in monogenic forms of ASD/ID genes and genes containing at least one AR-bound genomic  
419 sequence (212 genes,  $p = 0.0026$ ) as well as with those differentially expressed in ASD patients  
420 (490 genes,  $p = 2.23E^{-11}$ ). Genes previously associated with ASD by candidate gene studies or  
421 GWAS were not enriched among androgens-regulated genes.

422

## 423 **Discussion**

424 Most of the studies deciphering androgens effects on transcription regulation in human  
425 have been performed on prostate cancer cell lines. To better understand the role of androgens  
426 during brain development, we analyzed androgens effect on neuronal cells. We chose a human  
427 *in vitro* model of neuronal cells instead of *in vivo* mouse model as observations suggested that  
428 testosterone role in brain masculinizing in rodents mainly acts *via* its conversion into estradiol  
429 by aromatase, which might not be the case in human<sup>66,67</sup>. SH-SY5Y neuroblastoma cells, already  
430 used in the literature to assesses testosterone effects, were not used as surprisingly no AR  
431 expression was observed in five batches of these cells coming from different origins.  
432 Indeed, contradictory results concerning AR expression in this cell line can be found in literature<sup>68,69</sup>.  
433 As the aromatase *CYP19A1* converting testosterone into estrogen and the estrogen receptors  
434 *ESR* are expressed in this cell line, some testosterone effects observed might therefore be  
435 mediated by its conversion into estrogen. The cellular model we decided to use here was already  
436 published as an "amenable model to study psychiatric disorders"<sup>47,70</sup>. The use of hNSCs was  
437 particularly relevant here because: 1) they model fetal brain; 2) they express AR at the mRNA

438 and protein level at higher level than differentiated neurons, which could indicate that androgens  
439 action is more important in these precursor cells than in mature neurons and 3) one can more  
440 readily obtain the cell number necessary for RNA-seq and ChIP-seq experiments.

441         Only discrete changes in a small number of genes were observed in hNSCs after a DHT  
442 treatment, as compared to genes regulated by RA. This is in agreement with the very subtle  
443 differences in neuroanatomy or behavior observed between prepubertal human male and female.  
444 Moreover, the low amplitude observed can still have important functional consequences as it is  
445 clearly established that some genes implicated in ND are dosage sensitive (*e.g.* *DYRK1A*,  
446 *MECP2* or other chromatin modifiers). In this way, androgens could potentiate the effect of a  
447 variation in one of its target genes.

448         We observed an enrichment in genes involved in monogenic forms of ASD or ID and in  
449 genes differentially expressed in ASD patients among the target genes of androgens in neuronal  
450 precursors. In this way, androgens could act as sensitizing factor in male by modulating  
451 expression of genes involved in ND. As an example, *FAM107A* is up-regulated by androgens  
452 treatment in hNSCs and AR binds a sequence located in *FAM107A* second intron. The ENCODE  
453 project found several ARE/GRE consensus sequences in *FAM107A* in mouse liver or human  
454 epithelial cells but they do not overlap with the peak we identified. Previous studies showed a  
455 regulation of *FAM107A*, by the activation of the GR <sup>71</sup>. We showed that *FAM107A*  
456 up-regulation by AR appears required for androgens enhancement of cell proliferation.  
457 In rodents, *Fam107a* is involved in neuronal cell survival <sup>72</sup> and was described as a regulator of  
458 neuronal migration and can impact neurite outgrowth and modulate synaptic efficacy <sup>73</sup>.  
459 Virus-induced hippocampal *Fam107a* up-regulation in mouse improves spatial memory <sup>73</sup> and,

460 in lateral septum, increases sociability<sup>74</sup>. This gene is highly expressed in brain, and particularly  
461 during embryogenesis in rodents<sup>72</sup>. *FAM107A* was found to be differentially expressed in  
462 cerebellum of autistic patients (nine patients *vs.* four controls, FC 1.54<sup>56</sup>), and in other patients  
463 suffering from ND, such as schizophrenia or bipolar disorders<sup>75</sup>.

464 We identified an up-regulation of *NRCAM* expression after androgens treatment in  
465 neuronal precursors, as well as at least three AR-bound sequences in intronic regions of this  
466 gene. Genetic associations between SNPs in the *NRCAM* gene and autism have been  
467 inconsistently reported<sup>45, 46, 76, 77</sup>, and a specific haplotype of eight SNPs located between the  
468 promoter region and intron 2 (rs3763463 to rs2216259) appeared significantly associated to  
469 ASD<sup>45, 46</sup>. AR-bound regions do not directly overlap the associated SNPs described, but other  
470 SNPs located in the promoter or in intronic regions might affect *NRCAM* expression  
471 (**Figure S4A** and **Supplementary Table S7**). *NRCAM* was described as a potential target of the  
472 neuronal specific splicing factor A2BP1, known to be involved in ASD<sup>78</sup>, and differences in  
473 splicing events were observed between ASD patients with low expression of A2BP1 *vs.* control  
474 individuals<sup>57</sup>. These differences in *NRCAM* splicing events between ASD patients and control  
475 individuals were recently replicated by others, involving microexons regulation<sup>79</sup>. *Nrcam* is  
476 expressed in the central nervous system during development and encodes a neural cell adhesion  
477 molecule involved in neuronal migration, axonal guidance, neurite outgrowth, dendritic spine  
478 remodeling, etc (for review<sup>80</sup>). *Nrcam*-null mice showed decrease in sociability for males and  
479 in learning for both sexes<sup>81</sup>. No truncating mutations of *NRCAM* have been described in  
480 monogenic forms of ASD until now. However, numerous deletions encompassing *NRCAM* are  
481 reported in individuals with autism, ID or global developmental delay in the Decipher database.

482

483 One of those, observed in a patient with autism, is an intragenic deletion of the first coding  
484 exons of *NRCAM* (exons 4 to 11). Coe *et al.* also reported another intragenic deletion  
485 encompassing the second and third non-coding exons in one patient with ND (**Figure S4A**)<sup>82</sup>.  
486 *NRCAM* expression is regulated by the Wnt/ $\beta$ -catenin pathway<sup>83</sup> that has been proposed to be  
487 involved in ASD physiopathology<sup>84</sup>. AR has been shown to interact with  $\beta$ -catenin in an  
488 androgens-dependent manner and shuttles it to the nucleus<sup>85</sup>. However, we did not find  
489 co-translocation of AR and  $\beta$ -catenin into the nucleus after androgens treatment in hNSCs  
490 (data not shown) and we observed no significant overlap between  $\beta$ -catenin target genes and  
491 androgens-regulated genes (RNA-seq data). A modest enrichment was found between  $\beta$ -catenin  
492 target genes and hNSCs AR target genes (CHIP-seq data; hypergeometric law  $p = 0.0018$ ).

493 We showed that DHT treatment enhances proliferation of hNSCs. At least two of its  
494 target genes, *RGCC* and *FAM107A*, play a role in cell proliferation. *RGCC* knock-down has a  
495 drastic inhibitory effect on cell growth, effect previously observed in human lung cancer<sup>86</sup>.  
496 In parallel, knock-down of *FAM107A* cancels proliferation increase caused by androgens.  
497 We also demonstrated that DHT has a protective effect preventing cells from death when they  
498 are facing neuronal differentiation in stressful conditions, *i.e* without specific supplementation  
499 (vitamin, insulin, antioxidant molecules, etc). If we observed that RA induces neurite outgrowth  
500 during hNSCs differentiation, we were unable to detect any effect of DHT.  
501 Previous observations showed that regulation of cell death by androgens might play a role in  
502 establishing sexual dysmorphism of the brain in mouse<sup>60</sup>.

503 We demonstrated that androgens regulation of at least the “top-five” of these target genes  
504 acts *via* its receptor, the AR. The six most enriched motifs included binding sites for SP1, EGR1



505 and NRF1, all known to interact with AR<sup>87-89</sup>. We also found a significant enrichment in the  
506 closely-related<sup>90</sup> ARE and GRE motifs among the genomic regions bound by AR after a DHT  
507 treatment and in FOXA1/2 motifs. Those factors are nuclear factors described as AR interactors  
508<sup>91</sup>. More precisely, FOXA1 facilitates the recognition of ARE by AR<sup>92,93</sup>. AR-bound regions  
509 identified in hNSCs were different from those identified in prostate cancer cells, suggesting that  
510 interaction between AR and its proteic partners might vary between tissues and cell types.  
511 An AR-bound sequence was identified in only one third of the DHT-regulated genes found by  
512 transcriptomic studies suggesting that the remaining genes might be regulated by an indirect  
513 action of AR.

514 In conclusion, we used neuronal precursors, a cellular model relevant for studying ND,  
515 to analyze the role of androgens during brain development. We observed that androgens increase  
516 neuronal precursor proliferation and protect cells against stress-induced death during their  
517 differentiation. This study also revealed a subtle regulation of less than hundred genes by  
518 androgens. These androgens-target genes are enriched in genes associated to ASD, such as  
519 *FAM107A* or *NRCAM*. The regulation of these genes and pathways by androgens during brain  
520 development might therefore participate to the increased susceptibility of male brain, exposed  
521 to other genetic and environmental factors affecting the same genes and pathways, to develop  
522 ND. Our approach might also be relevant for other sex-biased ND such as Gilles de la Tourette  
523 or Hirschsprung disease. A better understanding of the mechanisms underlying differential risk  
524 and resilience to disease between both sexes may be of importance to identify therapeutic targets  
525 and pharmacological approaches for these diseases.

526

527 **Acknowledgments**

528 We thank Jocelyn Ceraline for sharing the pE-AR plasmid, Vanessa Ueberschlag-Pitiot and  
529 Daniel Metzger for sharing dexamethasone and anti-AR (C-19 and homemade 3279.4) and anti-  
530 GR (M-20) antibodies, all of them and Fabien Guimiot for scientific discussions, Bernard Jost,  
531 Christelle Thibault-Carpentier, Tao Ye and Constance Vagne from IGBMC sequencing and  
532 bioinformatic platforms for technical and bioinformatic help and IGBMC cell culture service  
533 for their technical support. We also thank the training students that participated to this study:  
534 Grace Gan, Léna Boztepe and Ines El Faroukhi. This study was supported by grant from Agence  
535 Nationale de la Recherche (HARTAGeNe) and the institute benefits from ANR-10-IDEX-0002-  
536 02. Angélique Quartier received a fellowship from the Fondation pour la Recherche Médicale.

537

538 **Conflict of Interest**

539 The authors declare no conflict of interest.

540

541 **Supplementary Information**

542 Supplementary information is available at Molecular Psychiatry's website

543

544 **References**

- 545 1. Elsabbagh M, Divan G, Koh YJ, Kim YS, Kauchali S, Marcin C *et al.* Global prevalence of autism  
546 and other pervasive developmental disorders. *Autism Res* 2012; **5**(3): 160-179.  
547
- 548 2. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E *et al.* Autism as a strongly  
549 genetic disorder: evidence from a British twin study. *Psychol Med* 1995; **25**(1): 63-77.  
550
- 551 3. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G *et al.* Individual common variants exert  
552 weak effects on the risk for autism spectrum disorderspi. *Hum Mol Genet* 2012; **21**(21): 4781-  
553 4792.  
554
- 555 4. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB *et al.* Most genetic risk for autism  
556 resides with common variation. *Nat Genet* 2014; **46**(8): 881-885.  
557
- 558 5. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders.  
559 *Hum Mol Genet* 2010; **19**(R2): R176-187.  
560
- 561 6. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial  
562 risk of autism. *JAMA* 2014; **311**(17): 1770-1777.  
563
- 564 7. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism  
565 disease mechanisms through genetics. *Nat Med* 2016; **22**(4): 345-361.  
566
- 567 8. Newschaffer CJ, Croen LA, Daniels J, Giarelli E, Grether JK, Levy SE *et al.* The epidemiology of  
568 autism spectrum disorders. *Annu Rev Public Health* 2007; **28**: 235-258.  
569
- 570 9. Dworzynski K, Ronald A, Bolton P, Happe F. How different are girls and boys above and below  
571 the diagnostic threshold for autism spectrum disorders? *J Am Acad Child Adolesc Psychiatry*  
572 2012; **51**(8): 788-797.  
573
- 574 10. Lai MC, Lombardo MV, Auyeung B, Chakrabarti B, Baron-Cohen S. Sex/gender differences and  
575 autism: setting the scene for future research. *J Am Acad Child Adolesc Psychiatry* 2015; **54**(1):  
576 11-24.  
577
- 578 11. Piton A, Gauthier J, Hamdan FF, Lafreniere RG, Yang Y, Henrion E *et al.* Systematic resequencing  
579 of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol*  
580 *Psychiatry* 2011; **16**(8): 867-880.  
581
- 582 12. Nava C, Lamari F, Heron D, Mignot C, Rastetter A, Keren B *et al.* Analysis of the chromosome X  
583 exome in patients with autism spectrum disorders identified novel candidate genes, including  
584 TMLHE. *Transl Psychiatry* 2012; **2**: e179.  
585
- 586 13. Jacquemont S, Coe BP, Hersch M, Duyzend MH, Krumm N, Bergmann S *et al.* A higher  
587 mutational burden in females supports a "female protective model" in neurodevelopmental  
588 disorders. *Am J Hum Genet* 2014; **94**(3): 415-425.  
589

- 590 14. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de  
591 novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**(7526): 216-221.  
592
- 593 15. Robinson EB, Koenen KC, McCormick MC, Munir K, Hallett V, Happe F *et al.* A multivariate twin  
594 study of autistic traits in 12-year-olds: testing the fractionable autism triad hypothesis. *Behav*  
595 *Genet* 2012; **42**(2): 245-255.  
596
- 597 16. Baron-Cohen S, Lombardo MV, Auyeung B, Ashwin E, Chakrabarti B, Knickmeyer R. Why are  
598 autism spectrum conditions more prevalent in males? *PLoS Biol* 2011; **9**(6): e1001081.  
599
- 600 17. Schaafsma SM, Pfaff DW. Etiologies underlying sex differences in Autism Spectrum Disorders.  
601 *Front Neuroendocrinol* 2014; **35**(3): 255-271.  
602
- 603 18. Baron-Cohen S. The extreme male brain theory of autism. *Trends Cogn Sci* 2002; **6**(6): 248-254.  
604
- 605 19. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene  
606 expression in females. *Nature* 2005; **434**(7031): 400-404.  
607
- 608 20. Skuse DH. Imprinting, the X-chromosome, and the male brain: explaining sex differences in the  
609 liability to autism. *Pediatr Res* 2000; **47**(1): 9-16.  
610
- 611 21. Crespi B. Turner syndrome and the evolution of human sexual dimorphism. *Evol Appl* 2008;  
612 **1**(3): 449-461.  
613
- 614 22. Wijchers PJ, Yandim C, Panousopoulou E, Ahmad M, Harker N, Saveliev A *et al.* Sexual  
615 dimorphism in mammalian autosomal gene regulation is determined not only by Sry but by sex  
616 chromosome complement as well. *Dev Cell* 2010; **19**(3): 477-484.  
617
- 618 23. Yan J, Feng J, Schroer R, Li W, Skinner C, Schwartz CE *et al.* Analysis of the neuroligin 4Y gene in  
619 patients with autism. *Psychiatr Genet* 2008; **18**(4): 204-207.  
620
- 621 24. Margari L, Lamanna AL, Craig F, Simone M, Gentile M. Autism spectrum disorders in XYY  
622 syndrome: two new cases and systematic review of the literature. *Eur J Pediatr* 2014; **173**(3):  
623 277-283.  
624
- 625 25. Kopsida E, Stergiakouli E, Lynn PM, Wilkinson LS, Davies W. The Role of the Y Chromosome in  
626 Brain Function. *Open Neuroendocrinol J* 2009; **2**: 20-30.  
627
- 628 26. Kido T, Schubert S, Schmidtke J, Chris Lau YF. Expression of the human TSPY gene in the brains  
629 of transgenic mice suggests a potential role of this Y chromosome gene in neural functions. *J*  
630 *Genet Genomics* 2011; **38**(5): 181-191.  
631
- 632 27. Wu JB, Chen K, Li Y, Lau YF, Shih JC. Regulation of monoamine oxidase A by the SRY gene on  
633 the Y chromosome. *FASEB J* 2009; **23**(11): 4029-4038.  
634
- 635 28. Loke H, Harley V, Lee J. Biological factors underlying sex differences in neurological disorders.  
636 *Int J Biochem Cell Biol* 2015; **65**: 139-150.

- 637  
638 29. Czech DP, Lee J, Sim H, Parish CL, Vilain E, Harley VR. The human testis-determining factor SRY  
639 localizes in midbrain dopamine neurons and regulates multiple components of catecholamine  
640 synthesis and metabolism. *J Neurochem* 2012; **122**(2): 260-271.  
641
- 642 30. Knickmeyer R, Baron-Cohen S, Raggatt P, Taylor K, Hackett G. Fetal testosterone and empathy.  
643 *Horm Behav* 2006; **49**(3): 282-292.  
644
- 645 31. Chakrabarti B, Dudbridge F, Kent L, Wheelwright S, Hill-Cawthorne G, Allison C *et al.* Genes  
646 related to sex steroids, neural growth, and social-emotional behavior are associated with  
647 autistic traits, empathy, and Asperger syndrome. *Autism Res* 2009; **2**(3): 157-177.  
648
- 649 32. Schwarz E, Guest PC, Rahmoune H, Wang L, Levin Y, Ingudomnukul E *et al.* Sex-specific serum  
650 biomarker patterns in adults with Asperger's syndrome. *Mol Psychiatry* 2011; **16**(12): 1213-  
651 1220.  
652
- 653 33. Hu VW, Nguyen A, Kim KS, Steinberg ME, Sarachana T, Scully MA *et al.* Gene expression profiling  
654 of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal  
655 development and steroid biosynthesis. *PLoS One* 2009; **4**(6): e5775.  
656
- 657 34. Knickmeyer R, Baron-Cohen S, Fane BA, Wheelwright S, Mathews GA, Conway GS *et al.*  
658 Androgens and autistic traits: A study of individuals with congenital adrenal hyperplasia. *Horm*  
659 *Behav* 2006; **50**(1): 148-153.  
660
- 661 35. Wudy SA, Wachter UA, Homoki J, Teller WM. 17 alpha-hydroxyprogesterone, 4-  
662 androstenedione, and testosterone profiled by routine stable isotope dilution/gas  
663 chromatography-mass spectrometry in plasma of children. *Pediatr Res* 1995; **38**(1): 76-80.  
664
- 665 36. Bezdicikova M, Molikova R, Bebarova L, Kolar Z. Distribution of nuclear receptors for steroid  
666 hormones in the human brain: a preliminary study. *Biomed Pap Med Fac Univ Palacky Olomouc*  
667 *Czech Repub* 2007; **151**(1): 69-71.  
668
- 669 37. Simerly RB, Chang C, Muramatsu M, Swanson LW. Distribution of androgens and estrogen  
670 receptor mRNA-containing cells in the rat brain: an in situ hybridization study. *J Comp Neurol*  
671 1990; **294**(1): 76-95.  
672
- 673 38. Xu X, Coats JK, Yang CF, Wang A, Ahmed OM, Alvarado M *et al.* Modular genetic control of  
674 sexually dimorphic behaviors. *Cell* 2012; **148**(3): 596-607.  
675
- 676 39. Hammond J, Le Q, Goodyer C, Gelfand M, Trifiro M, LeBlanc A. Testosterone-mediated  
677 neuroprotection through the androgen receptor in human primary neurons. *J Neurochem*  
678 2001; **77**(5): 1319-1326.  
679
- 680 40. Pike CJ, Nguyen TV, Ramsden M, Yao M, Murphy MP, Rosario ER. Androgens cell signaling  
681 pathways involved in neuroprotective actions. *Horm Behav* 2008; **53**(5): 693-705.  
682

- 683 41. Hatanaka Y, Mukai H, Mitsunashi K, Hojo Y, Murakami G, Komatsuzaki Y *et al.* Androgens rapidly  
684 increases dendritic thorns of CA3 neurons in male rat hippocampus. *Biochem Biophys Res*  
685 *Commun* 2009; **381**(4): 728-732.  
686
- 687 42. Lustig RH. Sex hormone modulation of neural development in vitro. *Horm Behav* 1994; **28**(4):  
688 383-395.  
689
- 690 43. Massie CE, Adryan B, Barbosa-Morais NL, Lynch AG, Tran MG, Neal DE *et al.* New androgen  
691 receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Rep*  
692 2007; **8**(9): 871-878.  
693
- 694 44. Tan PY, Chang CW, Chng KR, Wansa KD, Sung WK, Cheung E. Integration of regulatory networks  
695 by NKX3-1 promotes androgens-dependent prostate cancer survival. *Mol Cell Biol* 2012; **32**(2):  
696 399-414.  
697
- 698 45. Bonora E, Lamb JA, Barnby G, Sykes N, Moberly T, Beyer KS *et al.* Mutation screening and  
699 association analysis of six candidate genes for autism on chromosome 7q. *Eur J Hum Genet*  
700 2005; **13**(2): 198-207.  
701
- 702 46. Marui T, Funatogawa I, Koishi S, Yamamoto K, Matsumoto H, Hashimoto O *et al.* Association of  
703 the neuronal cell adhesion molecule (NRCAM) gene variants with autism. *Int J*  
704 *Neuropsychopharmacol* 2009; **12**(1): 1-10.  
705
- 706 47. Boissart C, Poulet A, Georges P, Darville H, Julita E, Delorme R *et al.* Differentiation from human  
707 pluripotent stem cells of cortical neurons of the superficial layers amenable to psychiatric  
708 disease modeling and high-throughput drug screening. *Transl Psychiatry* 2013; **3**: e294.  
709
- 710 48. Adli M, Bernstein BE. Whole-genome chromatin profiling from limited numbers of cells using  
711 nano-ChIP-seq. *Nat Protoc* 2011; **6**(10): 1656-1668.  
712
- 713 49. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists  
714 using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**(1): 44-57.  
715
- 716 50. Gilissen C, Hahir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH *et al.* Genome  
717 sequencing identifies major causes of severe intellectual disability. *Nature* 2014; **511**(7509):  
718 344-347.  
719
- 720 51. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic de  
721 novo mutations in autism spectrum disorders. *Nature* 2012; **485**(7397): 242-245.  
722
- 723 52. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L *et al.* Convergence of genes and  
724 cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 2014; **94**(5):  
725 677-694.  
726
- 727 53. Garbett K, Ebert PJ, Mitchell A, Lintas C, Manzi B, Mirnics K *et al.* Immune transcriptome  
728 alterations in the temporal cortex of subjects with autism. *Neurobiol Dis* 2008; **30**(3): 303-311.  
729

- 730 54. Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA *et al.* Gene expression changes  
731 in children with autism. *Genomics* 2008; **91**(1): 22-29.  
732
- 733 55. Nishimura Y, Martin CL, Vazquez-Lopez A, Spence SJ, Alvarez-Retuerto AI, Sigman M *et al.*  
734 Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of  
735 autism and reveals shared pathways. *Hum Mol Genet* 2007; **16**(14): 1682-1698.  
736
- 737 56. Purcell AE, Jeon OH, Zimmerman AW, Blue ME, Pevsner J. Postmortem brain abnormalities of  
738 the glutamate neurotransmitter system in autism. *Neurology* 2001; **57**(9): 1618-1628.  
739
- 740 57. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S *et al.* Transcriptomic analysis of  
741 autistic brain reveals convergent molecular pathology. *Nature* 2011; **474**(7351): 380-384.  
742
- 743 58. Dietrich W, Gaba A, Zhegu Z, Bieglmayer C, Mairhofer M, Mikula M *et al.* Testosterone  
744 dependent androgen receptor stabilization and activation of cell proliferation in primary  
745 human myometrial microvascular endothelial cells. *Fertil Steril* 2011; **95**(4): 1247-1255 e1241-  
746 1242.  
747
- 748 59. Dodson RE, Gorski RA. Testosterone propionate administration prevents the loss of neurons  
749 within the central part of the medial preoptic nucleus. *J Neurobiol* 1993; **24**(1): 80-88.  
750
- 751 60. Forger NG. Control of cell number in the sexually dimorphic brain and spinal cord. *J*  
752 *Neuroendocrinol* 2009; **21**(4): 393-399.  
753
- 754 61. Badea T, Niculescu F, Soane L, Fosbrink M, Sorana H, Rus V *et al.* RGC-32 increases p34CDC2  
755 kinase activity and entry of aortic smooth muscle cells into S-phase. *J Biol Chem* 2002; **277**(1):  
756 502-508.  
757
- 758 62. Fargo KN, Galbiati M, Foeking EM, Poletti A, Jones KJ. Androgens regulation of axon growth  
759 and neurite extension in motoneurons. *Horm Behav* 2008; **53**(5): 716-728.  
760
- 761 63. MacLusky NJ, Hajszan T, Prange-Kiel J, Leranath C. Androgens modulation of hippocampal  
762 synaptic plasticity. *Neuroscience* 2006; **138**(3): 957-965.  
763
- 764 64. Marron TU, Guerini V, Rusmini P, Sau D, Brevini TA, Martini L *et al.* Androgens-induced neurite  
765 outgrowth is mediated by neuritin in motor neurones. *J Neurochem* 2005; **92**(1): 10-20.  
766
- 767 65. Reddy RC, Amodei R, Estill CT, Stormshak F, Meaker M, Roselli CE. Effect of Testosterone on  
768 Neuronal Morphology and Neuritic Growth of Fetal Lamb Hypothalamus-Preoptic Area and  
769 Cerebral Cortex in Primary Culture. *PLoS One* 2015; **10**(6): e0129521.  
770
- 771 66. Belgorosky A, Guercio G, Pepe C, Saraco N, Rivarola MA. Genetic and clinical spectrum of  
772 aromatase deficiency in infancy, childhood and adolescence. *Horm Res* 2009; **72**(6): 321-330.  
773
- 774 67. Wu MV, Manoli DS, Fraser EJ, Coats JK, Tollkuhn J, Honda S *et al.* Estrogen masculinizes neural  
775 pathways and sex-specific behaviors. *Cell* 2009; **139**(1): 61-72.  
776

- 777 68. Sarachana T, Xu M, Wu RC, Hu VW. Sex hormones in autism: androgens and estrogens  
778 differentially and reciprocally regulate RORA, a novel candidate gene for autism. *PLoS One*  
779 2011; **6**(2): e17116.  
780
- 781 69. Yerramilli-Rao P, Garofalo O, Whatley S, Leigh PN, Gallo JM. Androgens-controlled specific gene  
782 expression in neuroblastoma cells. *J Neurol Sci* 1995; **129 Suppl**: 131-135.  
783
- 784 70. Darville H, Poulet A, Rodet-Amsellem F, Chatrousse L, Pernelle J, Boissart C *et al.* Human  
785 Pluripotent Stem Cell-derived Cortical Neurons for High Throughput Medication Screening in  
786 Autism: A Proof of Concept Study in SHANK3 Haploinsufficiency Syndrome. *EBioMedicine* 2016;  
787 **9**: 293-305.  
788
- 789 71. Frijters R, Fleuren W, Toonen EJ, Tuckermann JP, Reichardt HM, van der Maaden H *et al.*  
790 Prednisolone-induced differential gene expression in mouse liver carrying wild type or a  
791 dimerization-defective glucocorticoid receptor. *BMC Genomics* 2010; **11**: 359.  
792
- 793 72. Asano Y, Kishida S, Mu P, Sakamoto K, Murohara T, Kadomatsu K. DRR1 is expressed in the  
794 developing nervous system and downregulated during neuroblastoma carcinogenesis.  
795 *Biochem Biophys Res Commun* 2010; **394**(3): 829-835.  
796
- 797 73. Schmidt MV, Schulke JP, Liebl C, Stiess M, Avrabos C, Bock J *et al.* Tumor suppressor down-  
798 regulated in renal cell carcinoma 1 (DRR1) is a stress-induced actin bundling factor that  
799 modulates synaptic efficacy and cognition. *Proc Natl Acad Sci U S A* 2011; **108**(41): 17213-  
800 17218.  
801
- 802 74. Masana M, Su YA, Liebl C, Wang XD, Jansen L, Westerholz S *et al.* The stress-inducible actin-  
803 interacting protein DRR1 shapes social behavior. *Psychoneuroendocrinology* 2014; **48**: 98-110.  
804
- 805 75. Shao L, Vawter MP. Shared gene expression alterations in schizophrenia and bipolar disorder.  
806 *Biol Psychiatry* 2008; **64**(2): 89-97.  
807
- 808 76. Hutcheson HB, Olson LM, Bradford Y, Folstein SE, Santangelo SL, Sutcliffe JS *et al.* Examination  
809 of NRCAM, LRRN3, KIAA0716, and LAMB1 as autism candidate genes. *BMC Med Genet* 2004; **5**:  
810 12.  
811
- 812 77. Sakurai T, Ramoz N, Reichert JG, Corwin TE, Kryzak L, Smith CJ *et al.* Association analysis of the  
813 NrCAM gene in autism and in subsets of families with severe obsessive-compulsive or self-  
814 stimulatory behaviors. *Psychiatr Genet* 2006; **16**(6): 251-257.  
815
- 816 78. Martin CL, Duvall JA, Ilkin Y, Simon JS, Arreaza MG, Wilkes K *et al.* Cytogenetic and molecular  
817 characterization of A2BP1/FOX1 as a candidate gene for autism. *Am J Med Genet B*  
818 *Neuropsychiatr Genet* 2007; **144B**(7): 869-876.  
819
- 820 79. Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M *et al.* A  
821 highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 2014;  
822 **159**(7): 1511-1523.  
823



- 824 80. Sakurai T. The role of NrCAM in neural development and disorders--beyond a simple glue in  
825 the brain. *Mol Cell Neurosci* 2012; **49**(3): 351-363.  
826
- 827 81. Moy SS, Nonneman RJ, Young NB, Demyanenko GP, Maness PF. Impaired sociability and  
828 cognitive function in Nrcam-null mice. *Behav Brain Res* 2009; **205**(1): 123-131.  
829
- 830 82. Coe BP, Girirajan S, Eichler EE. The genetic variability and commonality of neurodevelopmental  
831 disease. *Am J Med Genet C Semin Med Genet* 2012; **160C**(2): 118-129.  
832
- 833 83. Conacci-Sorrell ME, Ben-Yedidia T, Shtutman M, Feinstein E, Einat P, Ben-Ze'ev A. Nr-CAM is a  
834 target gene of the beta-catenin/LEF-1 pathway in melanoma and colon cancer and its  
835 expression enhances motility and confers tumorigenesis. *Genes Dev* 2002; **16**(16): 2058-2072.  
836
- 837 84. Caracci MO, Avila ME, De Ferrari GV. Synaptic Wnt/GSK3beta Signaling Hub in Autism. *Neural*  
838 *Plast* 2016; **2016**: 9603751.  
839
- 840 85. Yang F, Li X, Sharma M, Sasaki CY, Longo DL, Lim B *et al.* Linking beta-catenin to androgens-  
841 signaling pathway. *J Biol Chem* 2002; **277**(13): 11336-11344.  
842
- 843 86. Xu R, Shang C, Zhao J, Han Y, Liu J, Chen K *et al.* Knockdown of response gene to complement  
844 32 (RGC32) induces apoptosis and inhibits cell growth, migration, and invasion in human lung  
845 cancer cells. *Mol Cell Biochem* 2014; **394**(1-2): 109-118.  
846
- 847 87. Lu S, Jenster G, Epner DE. Androgens induction of cyclin-dependent kinase inhibitor p21 gene:  
848 role of androgen receptor and transcription factor Sp1 complex. *Mol Endocrinol* 2000; **14**(5):  
849 753-760.  
850
- 851 88. Yang SZ, Abdulkadir SA. Early growth response gene 1 modulates androgen receptor signaling  
852 in prostate carcinoma cells. *J Biol Chem* 2003; **278**(41): 39906-39911.  
853
- 854 89. Schultz MA, Hagan SS, Datta A, Zhang Y, Freeman ML, Sikka SC *et al.* Nrf1 and Nrf2 transcription  
855 factors regulate androgen receptor transactivation in prostate cancer cells. *PLoS One* 2014;  
856 **9**(1): e87204.  
857
- 858 90. Claessens F, Denayer S, Van Tilborgh N, Kerkhofs S, Helsen C, Haelens A. Diverse roles of  
859 androgen receptor (AR) domains in AR-mediated signaling. *Nucl Recept Signal* 2008; **6**: e008.  
860
- 861 91. Yu X, Gupta A, Wang Y, Suzuki K, Mirosevich J, Orgebin-Crist MC *et al.* Foxa1 and Foxa2 interact  
862 with the androgen receptor to regulate prostate and epididymal genes differentially. *Ann N Y*  
863 *Acad Sci* 2005; **1061**: 77-93.  
864
- 865 92. Augello MA, Hickey TE, Knudsen KE. FOXA1: master of steroid receptor function in cancer.  
866 *EMBO J* 2011; **30**(19): 3885-3894.  
867
- 868 93. Lee HJ, Hwang M, Chattopadhyay S, Choi HS, Lee K. Hepatocyte nuclear factor-3 alpha (HNF-  
869 3alpha) negatively regulates androgen receptor transactivation in prostate cancer cells.  
870 *Biochem Biophys Res Commun* 2008; **367**(2): 481-486.  
871
- 872 94. Coe BP, Girirajan S, Eichler EE. A genetic model for neurodevelopmental disease. *Curr Opin*  
873 *Neurobiol* 2012; **22**(5): 829-836.

874 **Figure legends**

875 **Figure 1. hNSCs expresses androgen receptor and responds to dihydrotestosterone**  
876 **treatment. A.** Androgen receptor (*AR*) mRNA expression in several cell lines assessed by  
877 RT-qPCR. **B.** Specific RT-PCR showing the two protein coding isoforms of *AR*  
878 (the large NM\_00004.3 and the short NM\_001011645.2). **C.** Western blots using specific  
879 antibodies raised against N-terminal (N-20) or C-terminal (C-19) domains of the *AR* (epitopes  
880 shown on the drawing). **D.** *AR* immunofluorescence staining of hNSCs treated 4h with the  
881 solvent DMSO or the dihydrotestosterone 100nM. Left panel: *AR* staining, green; right panel:  
882 merge with Hoechst DNA staining, blue. **E.** Western blot showing *AR* protein expression level  
883 after 24h of treatment with androgens. LNCaP: prostate cancer cell line, SH-SY5Y:  
884 neuroblastoma cell line, ND17: hNSCs differentiated 17 days into neurons, Neg: experiment  
885 negative control, NTD: N-terminal domain, DBD: DNA binding domain, LBD: ligand binding  
886 domain, DHT: dihydrotestosterone, Testo: testosterone.

887 **Figure 2. Top 5 DHT 100nM regulated genes: gene ontology and example of RGCC**  
888 **regulation validation. A.** Volcano plot showing the RNA-seq data (DHT 100nM treated hNSCs  
889 compared to control). Genes having a fold change above  $|1.2|$  and an adjusted p-value under  
890 0.05 are shown in red (see details of DHT 100nM regulated genes in **Supplementary**  
891 **Table S1.**) The large red dots indicate *top-five* genes-of-interest that display both high fold  
892 change as well as high statistical significance. **B.** Gene ontology (GO) term associated with the  
893 five selected genes (category Biological Process). **C.** RGCC androgens regulation at the mRNA  
894 level (left part) and protein level (right part) in hNSCs. Left part: RT-qPCR after 24h treatment  
895 with diverse androgens at several concentrations (dihydrotestosterone: DHT 10 and 100nM,  
896 testosterone: 10 and 100nM and R1881 1nM).  $7 < n < 12$ . Right part: western blot performed after

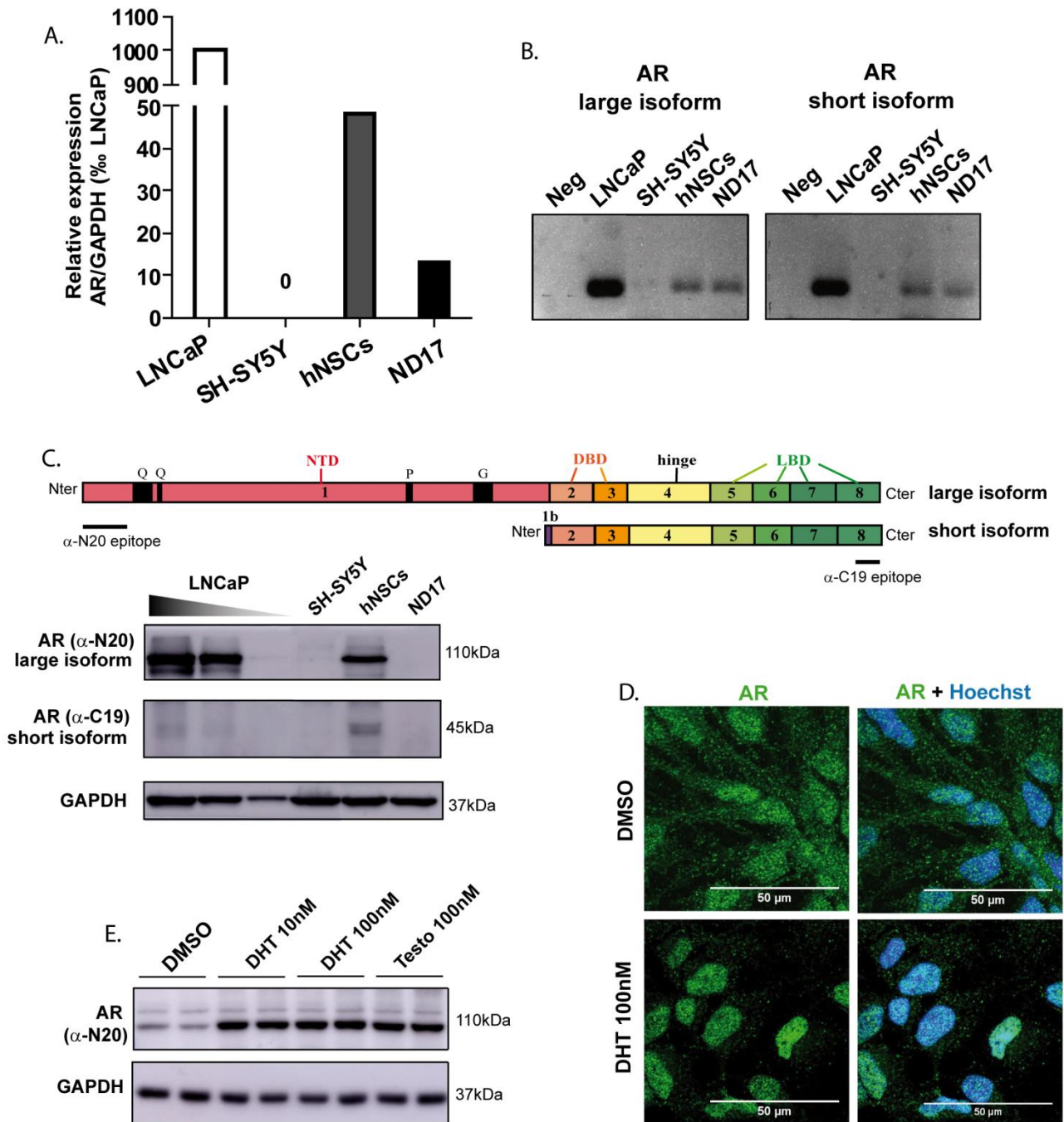
897 24h treatment with DMSO or DHT 100nM. **D.** *RGCC* mRNA expression in SH-SY5Y after  
898 DMSO or DHT 100nM treatment with or without AR overexpression. Lipo: lipofectamine,  
899 equivalent to the basal condition. n = 5-6. Results for other androgens-regulated genes (*RNF144B*,  
900 *NRCAM*, *TRIM22* and *FAM107A*) in **Supplementary Figure S4**. **E.** *RGCC* mRNA expression in  
901 hNSCs after DHT 100nM treatment without or with AR down-regulation by siRNA transfection.  
902 INF: INTERFERin equivalent to the basal condition. n = 3-4. Results for other androgens-regulated  
903 genes (*RNF144B*, *NRCAM*, *TRIM22* and *FAM107A*) in **Supplementary Figure S5**. Student's t test:  
904 ns: not significant; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001.

905 **Figure 3. DHT treatment increases cell proliferation, cell survival but has no effect on**  
906 **neurite outgrowth regulation.** **A.** Proliferation assay performed on hNSCs transfected with  
907 either scramble, *AR*, *RGCC* or *FAM107A* siRNA and then treated with DMSO or DHT 100nM.  
908 At each time point (days 1, 2, 3 and 4), cells were counting and data were normalized with the  
909 scramble siRNA transfection (treated with DMSO) as a control. n = 8-10; Student's t test  
910 (comparison with the scramble siRNA DMSO condition); \* p<0.05; \*\*\* p<0.001.  
911 **B.** Photographs of hNSCs differentiated into neurons during 17 days with either classical  
912 medium (upper part) or medium depleted in N2 and B27 (lower part). For each medium, cells  
913 were treated with DMSO (left part) or DHT 100nM (right part). The graph represents the  
914 quantification of percentage of wells with alive cells for each conditions. n = 30. Pictures at days  
915 3, 7, 10, 14 and 17 are presented in **Supplementary Figure S6**. **C.** Neurite outgrowth assays  
916 performed on neurons after 14 days of hNSC differentiation with DMSO (0.0031 to 0.4%),  
917 DHT (3.125 to 400nM), L-685458 (0.125 to 16µM) or retinoic acid (RA) treatments (4 to 500µM).  
918 Testosterone and R1881 treatment are shown in **Supplementary Figure S7**. Student's t test  
919 \* p<0.05; \*\* p<0.01; \*\*\* p<0.001.

920 **Figure 4. Focus on *NRCAM* promoter peaks.** **A.** Drawing of the *NRCAM* gene illustrating the  
921 location of CNVs within or encompassing *NRCAM* (from <sup>94</sup> and the Decipher database), SNPs  
922 significantly associated with autism (red stars <sup>45</sup>, purple stars <sup>46</sup>; details in **Supplementary**  
923 **Table S7**) and ChIP-seq peaks identified (with the transcription factors motifs and peaks  
924 identified by the ENCODE project). **B.** Screenshot of the AR peak identified in the *NRCAM* third  
925 intron with GRE/NR3C1 and FOXA1 motifs. Black stars show the significant peaks found in our  
926 ChIP-seq analysis. Repl.: Technical replicate. **C.** ChIP-qPCR validation of AR binding in the third  
927 intron of *NRCAM* gene. DNA enrichment is represented as percentage of input chromatin  
928 immunoprecipitated (IgG, immunoprecipitation negative control).

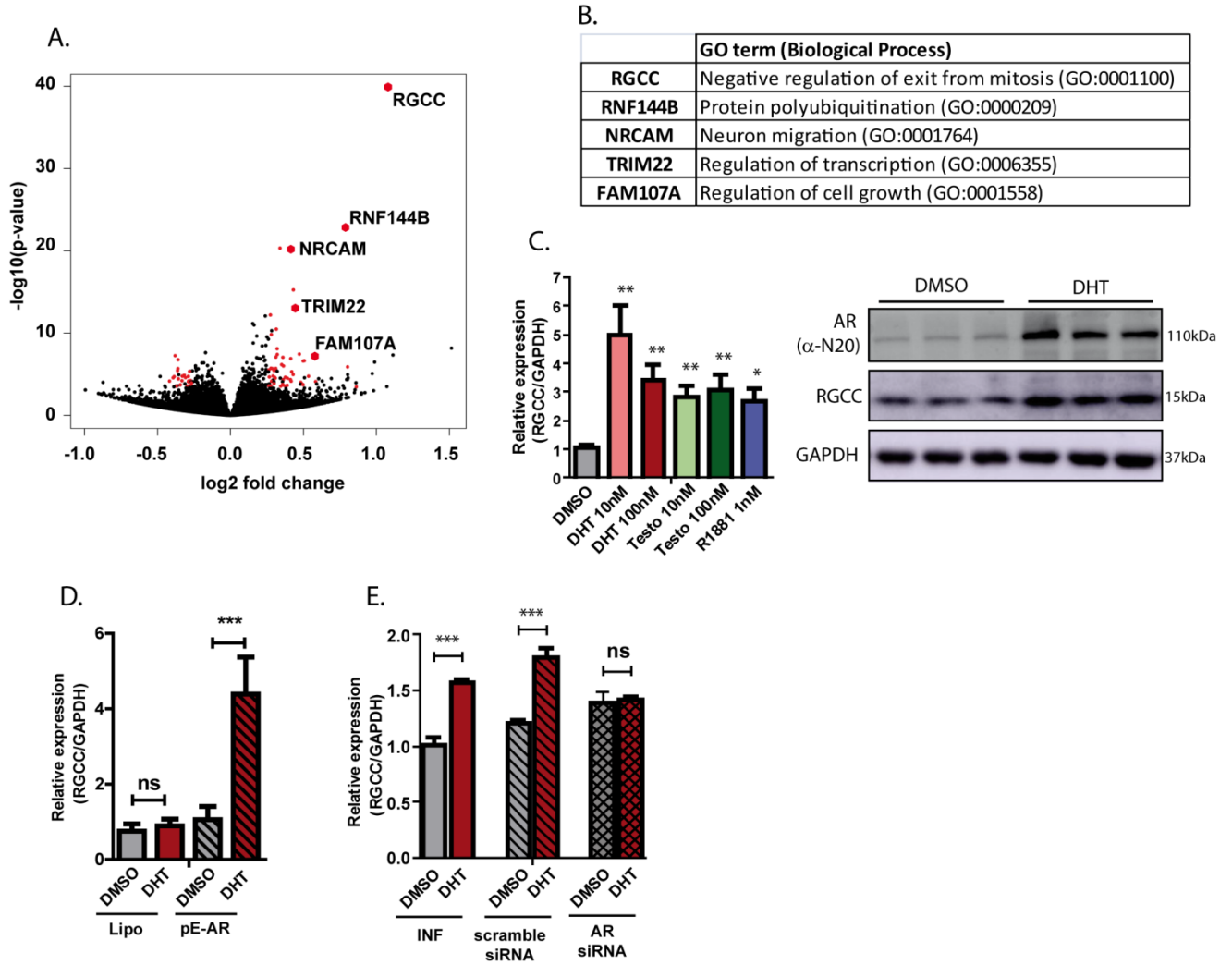
929

**Fig. 1 hNSCs express Androgen Receptor (AR) and responds to Dihydrotestosterone (DHT) treatment**



93

**Fig. 2 Top 5 DHT 100nM regulated genes: Gene Ontology and example of RGCC regulation validation.**



931

**Fig 3 DHT treatment increases cell proliferation, cell survival but has no effect on neurite outgrowth regulation**

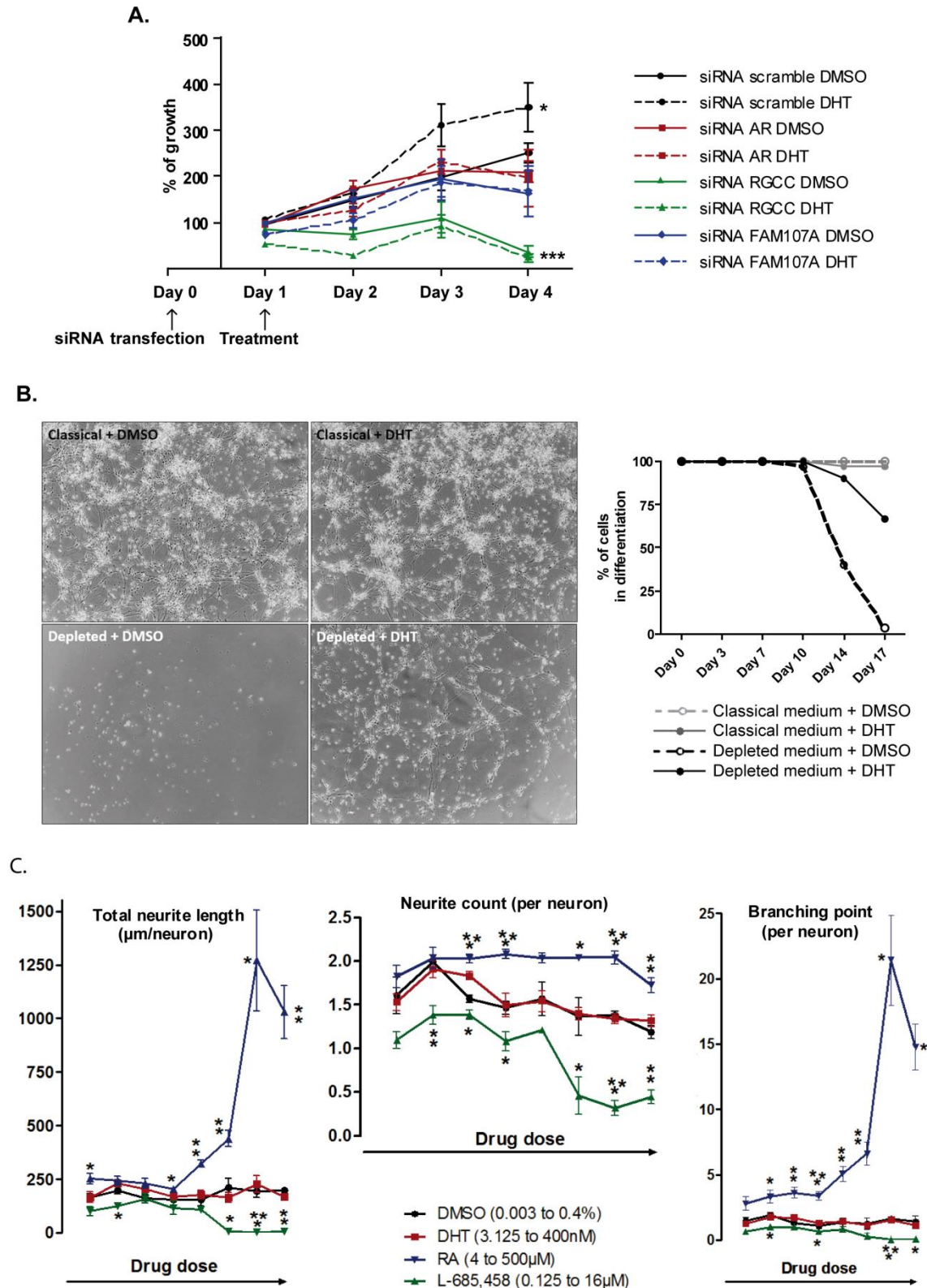
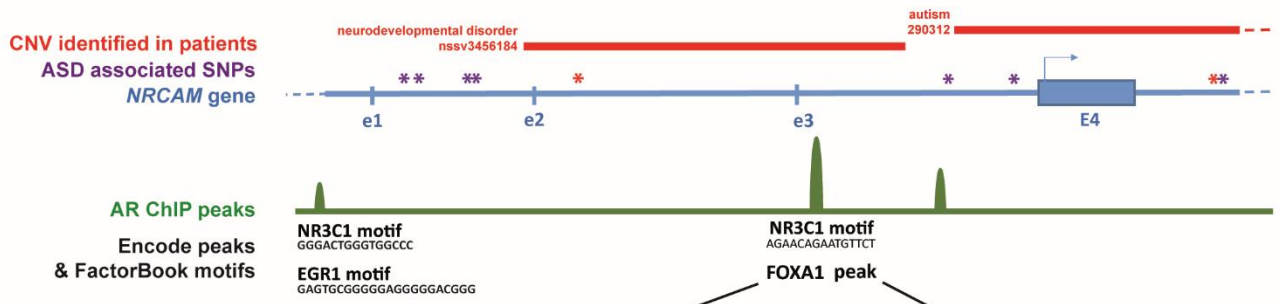
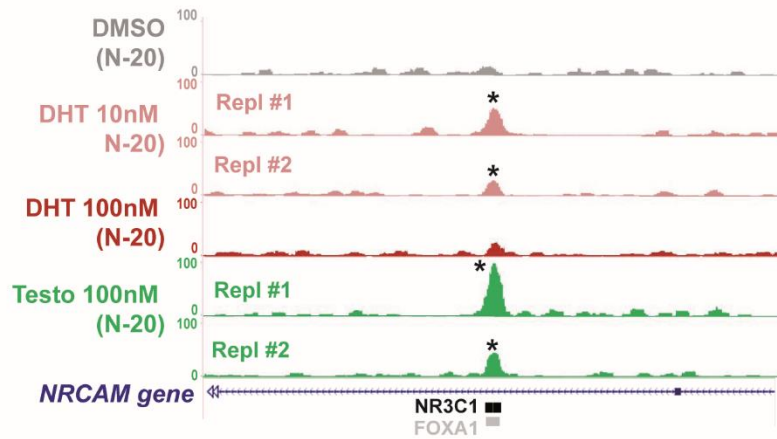


Figure 4.

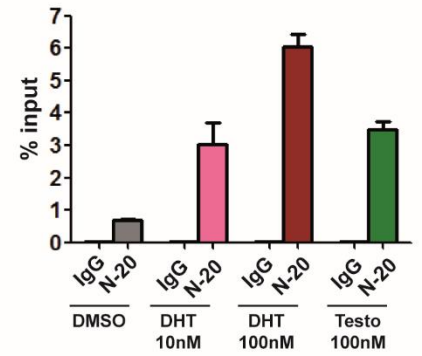
A.



B.



C.





1 **Tables**

**Table 1.**

**Comparison of DHT regulated genes (RNA-seq data, n = 77 genes)**

<b>Genes involved in ASD/ID:</b>	<b>Overlap</b>	<b>p-value</b>	<b>Genes</b>
- mutated in monogenic forms of ASD/ID n=651	6	2,87E-02	ATXN1, KCNQ3, LAMA2, LRAT, CUBN, PRODH,
- differentially expressed in ASD n=1660	15	1,17E-03	e.g.: FAMI07A, HES4, IGFBP5, IGFBPL1, NEFM, NRP1, RASSF4
- associated with ASD (GWAS) n=132	1	ns	TFPI
- associated with ASD (candidate gene) n=177	2	ns	NRCAM, TCN2

**Comparison of androgens bound genes (ChIP-seq data, n = 9 192 genes)**

<b>Genes involved in ASD/ID:</b>	<b>Overlap</b>	<b>p-value</b>	<b>Genes</b>
- mutated in monogenic forms of ASD/ID n=651	212	2,55E-03	e.g.: ARID1B, CUBN, DYRK1A, DOCK8, RAI1, STXBP1, SHANK2
- differentially expressed in ASD n=1660	490	2,23E-11	e.g.: FAMI07A, GADD45G, KIF2A, OXTR, SEMA5B, TCF4, VCAN
- associated with ASD (GWAS) n=132	47	ns	e.g.: CNTNAP2, CTNNA2, DCC, FOXP2, JARID2, KNCB1, RUNX1
- associated with ASD (candidate gene) n=177	55	1,82E-02	e.g.: CNTNAP2, BDNF, CREBBP, FOXP1, NRCAM, NRXN1, TSC2

2 **Table 1. Overlap between androgens-regulated genes and genes mutated or differentially expressed in ASD patients.**

3 For each comparison, the number of genes overlapping (column 'overlap') with examples (column 'genes') and the result of the

4 hypergeometric test (column 'p-value') are indicated. ns: not significant.

## 1 **Supplementary Information**

### 2 **Cells quality control**

3 To check the culture quality and prevent hNSCs derivation, quality control by immunostaining  
4 was performed every ten passages on hNSCs and neurons differentiated during 17 days (ND17)  
5 by assessing SOX2<sup>+</sup> (NSC marker) and HuC/D<sup>+</sup> (neuronal marker) cells percentages (see  
6 immunofluorescence methods for technical details).

### 7 **RNA isolation, reverse transcription-PCR and real-time quantitative PCR**

8 Total RNA were extracted from treated cells using TRI reagent (Molecular Research Center,  
9 Cincinnati, OH, USA), treated with DNase I recombinant (1U/μg of nucleic acids) (Roche, Bâle,  
10 Suisse) during 30 min at 37°C and then purify by using the RNeasy mini kit (Qiagen, Valencia,  
11 CA, USA). RNA levels and quality were quantified using a Nanodrop spectrophotometer and  
12 then with a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) for the RNA used for  
13 RNA-sequencing.

14 For reverse transcription-PCR (RT-PCR), 500ng to 1μg of total RNA was reverse transcribed  
15 into cDNA using random hexamers and SuperScript II reverse transcriptase according to  
16 manufacturer's recommendation. The cDNA were diluted ten times for the PCR reaction  
17 performed with specific primers to amplify either the large isoform NM\_00004.3 (AR\_RT\_F  
18 5'-ATCCCAGTCCCCTTGTGTC-3' and AR\_RT\_R 5'-GGTCTTCTGGGGTGGAAAGT-3')  
19 or the short isoform NM\_001011645.2 of AR (AR\_RT\_F2 5'-  
20 TACAGGGAACCAGGGAAACG-3' and AR\_RT\_R). Real-time PCR quantification (qPCR)  
21 was performed on LightCycler 480 II (Roche) using the QuantiTect SYBR Green PCR Master

22 Mix (Qiagen). All qPCR reactions were performed in triplicate. Primer sets are listed in  
23 **Supplementary Table S8**. Reaction specificity was controlled by post-amplification melting  
24 curve analysis. The relative expression of gene-of-interest vs *GAPDH* was calculated using the  
25  $2^{-(\Delta\Delta Ct)}$  method and a parametric Student's t-test was performed in order to compare  
26 androgens with DMSO treatment.

### 27 **ChIP-seq**

28 ChIP protocol was adapted from Massie and Mills *et al.* and Adli and Bernstein *et al.*<sup>1,2</sup>. hNSCs  
29 were cultured two days in depleted medium before adding DMSO or DHT 10nM, DHT 100nM  
30 or testosterone 100nM for 4h. 90 millions of cells were used for each treatment. DNA protein  
31 interactions were cross-linked using 0.4% formaldehyde for 10 min at room temperature, before  
32 quenching with a final concentration of 180mM glycine. Cells were harvested by scraping in  
33 PBS supplemented with 10% fetal calf serum and washed twice with PBS. Cellular lysates were  
34 isolated by incubating cells pellets (seven per condition) for 10 min at 4°C in 100µL SDS Lysis  
35 Buffer (50mM Tris-HCl pH8, 100nM NaCl, 5mM EDTA, 1% SDS) supplemented with protease  
36 inhibitor cocktail. Cellular lysates were diluted in 400µL of ChIP dilution buffer (17mM Tris-  
37 HCl pH8, 170mM NaCl, 1.2mM EDTA, 0.01% SDS, 1.1% Triton X-100) supplemented with  
38 protease inhibitor cocktail and sonicated for 20 times 1 min at 80% power in a Covaris E210  
39 AFA system (Covaris, MA, USA), combined (total volume 3.5mL) and insoluble debris were  
40 removed by centrifugation at 13 000 rpm for 5 min at 4°C. 1% of the supernatant was taken as  
41 total input control and the remainder was used for ChIP. For each ChIP reaction 20µl protein-A  
42 and 20µl protein-G sepharose beads (GE Healthcare, Little Chalfont, Royaume-Uni) were  
43 washed three times with 0.5% BSA in PBS, before blocking 3h at 4°C with 10mg/mL yeast  
44 tRNA (Invitrogen, Carlsbad, CA, USA). Beads were then centrifuged (13 000 rpm for 30s) and

45 used to pre-clear the chromatin with a 3-h-incubation at 4°C. Pre-cleared chromatin was  
46 incubated overnight with 5µg of specific rabbit polyclonal antibodies (rabbit IgG (Jackson  
47 immunoresearch, Baltimore, PA, USA), anti-AR N-20 or 3279.4) at 4°C with gentle agitation  
48 (110µg per immunoprecipitation). Chromatin-antibody complexes were combined with the  
49 blocked beads during 1h 30 min at 4°C with gentle agitation. The following day bead-antibody-  
50 protein-DNA complexes were washed once in low salt buffer (20mM Tris-HCl pH8, 150mM  
51 NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100), once in high salt buffer (20mM Tris-HCl  
52 pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100), twice in LiCl wash buffer  
53 (10mM Tris-HCl pH8, 0.25M LiCl, 1mM EDTA, 1% NP-40, 1% sodium deoxycholate), twice  
54 with Tris-EDTA at 4°C and eluted in twice with 100µl elution buffer (NaHCO<sub>3</sub> 0.1M, 1% SDS,  
55 5mM DTT) 15 min at 65°C with vortexing (600 rpm). Cross-links were reversed overnight at  
56 65°C by adding 10µL of NaCl 5M. RNA and proteins were degraded by adding 0.2mg/mL  
57 DNase-free RNase A, incubation for 30 min at 37°C, followed by addition of 1mg/mL  
58 Proteinase K (Sigma-Aldrich, St Louis, MO, USA) and incubation at 37°C for 2h. Genomic  
59 DNA was isolated using MinElute PCR Purification Kit (Qiagen) with two 20µL elutions.  
60 ChIP enrichment was tested by qPCR using 3µl of DNA and the remainder was used for library  
61 preparation.

## 62 **ChIP-qPCR**

63 We used quantitative PCR to confirm AR binding site in *NRCAM* third intron peak with specific  
64 primers (**Supplementary Table S8**) as described in the aforementioned 'real-time quantitative  
65 PCR'. For each condition (DMSO, DHT 10nM, DHT 100nM and testosterone 100nM), the input  
66 genomic DNA is used as a standard diluted 1:5, 1:25, 1:125 and 1:625 in triplicate.  
67 IgG (immunoprecipitation negative control) and N-20 ChIPed DNA were analyzed in triplicate.

68 Reaction specificity was controlled by post-amplification melting curve analysis. The *NRCAM*  
69 third intron occupancy results were calculated by normalizing ChIP value with the input  
70 ( $\Delta Ct = Ct[IP] - Ct[Input \times \text{dilution factor}]$ ) where the dilution factor consider the ratio between  
71 the volume used for the immunoprecipitation and the final folume after DNA purification. Then  
72 the % input ( $((2^{-\Delta Ct}) \times 100\%)$ ) were plotted for each immunoprecipitation and each condition to  
73 assess the relative enrichment in N-20 immunoprecipitations after androgens treatment.

#### 74 **RNA-seq and ChIP-seq librairies preparation**

75 For the RNA-seq, libraries of template molecules suitable for high throughput DNA sequencing  
76 were created using TruSeq™ RNA Sample Preparation v2 Kit. Briefly, mRNA were purified  
77 from 200ng of total RNA using poly-T oligo-attached magnetic beads and fragmented using  
78 divalent cations at 94°C for 8 min. The cleaved mRNA fragments were reverse transcribed to  
79 cDNA using random primers then the second strand of the cDNA was synthesized using DNA  
80 Polymerase I and RNase H. The double stranded cDNA fragments were blunted using T4 DNA  
81 polymerase, Klenow DNA polymerase and T4 PNK. A single 'A' nucleotide was added to the  
82 3' ends of the blunt DNA fragments using a Klenow fragment (3' to 5'exo minus) enzyme. The  
83 cDNA fragments were ligated to double stranded adapters using T4 DNA ligase. The ligated  
84 products were enriched by PCR amplification (30s at 98°C; [10s at 98°C, 30s at 60°C, 30s at  
85 72°C] x 12 cycles; 5 min at 72°C). Then surplus PCR primers were removed by purification  
86 using AMPure XP beads (Agencourt Biosciences, Fullerton, CA, USA).

87 Concerning the ChIP-seq librairies, they were prepared using NEXTflex ChIP-Seq Kit (#5143-  
88 02, Bioo Scientific, Austin, TX, USA) following the manufacturer's protocol (V12.10) with  
89 some modifications. Briefly, 4 to 10ng of ChIPed DNA or input DNA were end repaired using  
90 T4 DNA polymerase, Klenow DNA polymerase and T4 PNK, then size selected and cleaned-

91 up using Agencourt AMPure XP beads. A single 'A' nucleotide was added to the 3' ends of the  
92 blunt DNA fragments with a Klenow fragment (3' to 5'exo minus). The ends of the DNA  
93 fragments were ligated to double stranded barcoded DNA adapters (NEXTflex ChIP-Seq  
94 Barcodes - 6, #514120, Bioo Scientific) using T4 DNA ligase. The ligated products were  
95 enriched by PCR (2 min at 98°C; [30s at 98°C, 30s at 65°C, 60s at 72°C] x 14 cycles; 4 min at  
96 72°C) and cleaned-up using Agencourt AMPure XP beads.

### 97 **Bioinformatic analysis**

98 Reads were mapped onto the hg19 assembly of the human genome using Tophat v1.4.1<sup>3</sup> and  
99 the bowtie v0.12.7 aligner. Gene expression was quantified using HTSeq v0.5.3p5<sup>4</sup> and gene  
100 annotations from Ensembl release 72. Read counts were normalized across libraries with the  
101 method proposed by Anders and Huber<sup>5</sup>. Comparisons of interest were performed using the  
102 method proposed by Love et al.<sup>6</sup> implemented in the DESeq2 Bioconductor library (v1.0.19),  
103 adjusting for batch effect. Adjustment for multiple testing was performed with the Benjamini  
104 and Hochberg method. Genes weakly expressed (with less than 100 reads normalized and  
105 divided by gene length in kb for each condition being below 80th percentile) in one of the  
106 condition studied (DMSO or DHT 100nM 24h) were filtered out.

107 Concerning the ChIPseq analysis, reads were aligned to the human reference genome (hg19) using  
108 Bowtie<sup>7</sup> v0.12.8 using the following parameters '-m 1 --strata -best'. Peak calling was performed using  
109 MACS v1.4.2<sup>8</sup>. MACS was used using default parameters except for --nomodel along with the relevant  
110 shiftsize which was set as half the value of the fragment size estimated by Homer<sup>9</sup> v4.7.2  
111 makeTagDirectory. Peaks were annotated using Homer v4.7.2 with annotation extracted from Ensembl  
112 v75. Motifs searching of known motifs (Jaspar 2014 motif database) was made using FIMO<sup>10</sup> within  
113 peak sequences (+/-100nt around peak summits). FIMO results were then processed by a custom Perl

114 script that computed the number of occurrence of each motif. To assess the enrichment of motifs within  
115 the regions of interest, the same analysis was done  $n$  times ( $n = 100$ ) on randomly selected regions.  
116 Size distribution of the randomly selected regions was the same as for the peaks of interest.  
117 The significance of the motif occurrence was estimated through the computation of a Z-score. The Z-  
118 score was computed this way:  $z = (x-\mu)/\sigma$  where:  $x$  is the observed value (number of motif  
119 occurrence);  $\mu$  is the mean of the number of occurrences (computed on randomly selected data);  
120  $\sigma$  is the standard deviation of the number of occurrences of motifs (computed on randomly  
121 selected data). From the Z-scores, pvalues for each motif were computed. The P values were  
122 corrected for multi-testing with the method proposed by Benjamini et al. <sup>11</sup>. Statistical analysis  
123 was done with custom R scripts.

#### 124 **Western Blot**

125 Cells were lysed in RIPA buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.25% sodium  
126 deoxycholate, 1% NP-40) supplemented with protease inhibitor cocktail. 5 to 50 $\mu$ g of protein  
127 lysate were separated on 10% SDS-PAGE and transferred to polyvinylidene fluoride membrane.  
128 Membranes were blocked in 5% defatted milk diluted in tween 0.1% TBS followed by  
129 incubation with primary antibodies 2h at room temperature. Incubation with appropriate  
130 secondary HRP-labelled antibody (less than 1h) was followed by detection with Immobilon  
131 western chemiluminescent HRP substrate (Merck Millipore, Darmstadt, Germany).

#### 132 **Immunofluorescence**

133 Cells were fixed with 4% paraformaldehyde/PBS for 20 min, washed three times with PBS and  
134 incubated for 45min in blocking buffer (PBS, 10% fetal calf serum, 1% bovine serum albumin  
135 and 0.2% Triton X-100). Cells were then incubated with PerCP-Cy5.5-conjugated SOX2

136 monoclonal antibody in blocking buffer for 2h at room temperature or with polyclonal anti-AR  
137 N-20 or with monoclonal anti-HuC/D overnight at 4°C. After three washes in PBS, appropriate  
138 secondary antibodies conjugated to Alexa-488 were applied for 2h at room temperature. The  
139 cells were washed at least three times in PBS and visualized on an inverted confocal microscope  
140 (SP2UV, Leica, Wetzlar, Germany). For nuclear counter staining, cells were incubated in  
141 5µg/mL Hoechst 33342 for 10 min after immunostaining.

## 142 **Supplementary Figures Legends**

143 **Supplementary Figure S1. hNSCs quality control.** Immunostaining against SOX2 (NSC  
144 marker, in red, left panel) or HuC/D (neuronal marker, in green, right panel) performed on  
145 hNSCs and neurons differentiated during 17 days (N D17). The quantification shows a switch  
146 from SOX2<sup>+</sup> cells to HuC/D<sup>+</sup> cells during the differentiation, as expected by the protocol  
147 published by Boissart et al <sup>12</sup>. These data were confirmed by FACS (data not shown). n = 6

148 **Supplementary Figure S2. Steroid receptors, steroid enzymes expression and crosstalk**  
149 **with glucocorticoid receptor (GR) in hNSCs. A.** Steroid receptors and enzymes expression  
150 according to the number of DMSO normalized reads (mean of the ten samples analyzed by  
151 RNA-seq). AR: androgen receptor; GR: glucocorticoid receptor; ESR1: estrogen receptor alpha;  
152 ESR2: estrogen receptor beta; MR: mineralocorticoid receptor; PGR: progesterone receptor;  
153 SRD5A1: steroid-5-alpha-reductase; CYP19A1: estrogen synthetase; T: testosterone;  
154 E2: estradiol. We observed that *AR* and *GR* are the two major steroid receptors expressed in  
155 hNSCs (more than 300 normalized reads). **B.** Part of the steroid biosynthesis (downstream  
156 testosterone) showing the action of testosterone on two different pathways *via* androgen and/or  
157 estrogen receptors. Testosterone can be aromatized into estradiol by the estrogen synthetase,



158 also called aromatase or *CYP19A1*, and therefore be able to bind to estrogen receptors  
159 (*ESR1* and *ESR2*). We detected no expression of *CYP19A1* nor *ESR1* and *ESR2* in hNSCs,  
160 suggesting that no crosstalk of androgens with the estrogen pathway is possible in these cells.  
161 On the contrary, the 5-alpha-reductase (*SRD5A1*), the enzyme reducing testosterone into DHT,  
162 was well expressed in hNSCs. **C.** Western blot showing AR and GR protein expression level  
163 after 24h of treatment with androgens (DHT and Testo: testosterone) or glucocorticoid  
164 (Dex: dexamethasone). No significant change of GR level was observed after androgen or  
165 glucocorticoid treatments. In another hand, Dex treatment do not stabilize the AR protein since  
166 no difference was observed when compared with the DMSO condition.

167 **Supplementary Figure S3. Validation of *RNF144B*, *NRCAM*, *TRIM22* and *FAM107A***  
168 **regulation after androgens treatment in hNSCs. A.** Graphs showing mRNA level of  
169 *RNF144B*, *NRCAM*, *TRIM22*, *FAM107A* and *AR* after hNSCs 24h treatment with diverse  
170 androgens at several concentrations (dihydrotestosterone: DHT 10 and 100nM, testosterone:  
171 10 and 100nM and R1881 1nM). While androgens do not influence AR mRNA expression,  
172 they upregulate the four genes-of-interest.  $7 < n < 12$ . Student's t test: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ;  
173 \*\*\*  $p < 0.001$ . **B.** Western blots showing *RNF144B*, *NRCAM*, *TRIM22*, *FAM107A* and AR  
174 regulation at the protein level in hNSCs after 24h treatment with DMSO or DHT 100nM. DHT  
175 upregulates the four genes-of-interest at the protein level and stabilizes the AR protein. GAPDH  
176 is the reference protein and its level does not be influenced by DHT treatment.

177 **Supplementary Figure S4. DHT regulation *RNF144B*, *NRCAM*, *TRIM22* and *FAM107A***  
178 **in SH-5Y5Y overexpressing AR. A.** Androgen receptor (AR) expression in SH-SY5Y  
179 neuroblastoma cell lines before and after pE-AR transfection (left: graph showing AR mRNA  
180 expression assessed by RT-qPCR; right: western blot showing AR protein level).

181 pE-AR transfection leads to an efficient AR overexpression in SH-SY5Y, at the mRNA and  
182 protein level. **B.** Graphs showing mRNA level of *RNF144B*, *NRCAM*, *TRIM22* and *FAM107A*  
183 in SH-SY5Y after DMSO or DHT 100nM treatment with or without AR overexpression.  
184 Overexpression of AR in SH-SY5Y leads to a significant upregulation of the four genes-of-  
185 interest after DHT treatment. Lipo: lipofectamine (equivalent to the basal condition). n = 5-6.  
186 Student's t test: ns: not significant; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001.

187 **Supplementary Figure S5. Absence of DHT regulation of *RNF144B*, *NRCAM*, *TRIM22***  
188 **and *FAM107A* in hNSCs after AR knock-down. A.** Androgen receptor (AR) expression in  
189 hNSCs after transfection with scramble or AR siRNA (left: graph showing AR mRNA  
190 expression assessed by RT-qPCR; right: western blot showing AR protein level). AR siRNA  
191 transfection in hNSCs leads to a 80% decrease at the mRNA level and a 60% decrease at the  
192 protein level in basal condition while the use of a scramble siRNA did not affect AR expression.  
193 **B.** Graphs showing mRNA level of *RNF144B*, *NRCAM*, *TRIM22* and *FAM107A* in hNSCs after  
194 DMSO or DHT 100nM treatment with or without AR downregulation by siRNA transfection.  
195 Downregulation of AR in hNSCs abolishes the upregulation of the four genes-of-interest after  
196 DHT treatment. INF: INTERFERin (equivalent to the basal condition). n = 3-4. Student's t test:  
197 ns: not significant; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001.

198 **Supplementary Figure S6. DHT prevents neuron death during their differentiation under**  
199 **stress conditions.** Photographs of hNSCs during their differentiation into neurons (days 3, 7,  
200 10, 14 and 17) with classical (**A.**) or N2/B27 depleted medium (**B.**) supplemented with DMSO  
201 (upper part) or DHT 100nM (lower part).

202 **Supplementary Figure S7. Androgens (DHT, testosterone and R1881) have no effects on**  
203 **neurite outgrowth.** Neurite outgrowth assays performed on neurons after 14 days of hNSC  
204 differentiation with DMSO (0.0031 to 0.4%), DHT (3.125 to 400nM), testosterone (Testo; 3.125  
205 to 400nM), R1881 (3.125 to 400nM), L-685,458 (0.125 to 16 $\mu$ M) or retinoic acid (RA)  
206 treatments (4 to 500 $\mu$ M). The L-685,458 is a molecule known to inhibit neurite outgrowth  
207 (negative control) and the RA is known to enhance neurite outgrowth (positive control).  
208 Androgens treatment (with either DHT, testosterone or R1881) has no effect on neurite  
209 outgrowth in hNSCs. Student's t test: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

210 **Supplementary Figure S8. Validation of ChIP protocol and AR antibodies in LNCaP. A.**

211 Venn diagram illustrating the overlap between AR ChIP-seq peaks (Fold enrichment FE > 5)  
212 found in LNCaP (after N-20 immunoprecipitation or 3279.4 immunoprecipitation) and the AR  
213 ChIP-seq peaks from a previous published study<sup>13</sup>. Respectively 7 011 and 679 significantly  
214 enriched peaks were identified after filtering (FE > 5) with N-20 and homemade 3279.4  
215 antibody, with 410 peaks in common. Concerning the comparison with Tan dataset, we observed  
216 substantial overlaps (46.6% corresponding to 3 269 peaks for N-20, 68% corresponding to 461  
217 peaks for 3279.4 antibody). Top 30 peaks after N-20 and 3279.4 immunoprecipitation in LNCaP  
218 are described in **Supplementary Table S4. B.** Screenshots of AR ChIPseq peaks in *FKBP5*  
219 second intron from Tan et al.<sup>13</sup> (red profiles) and the present study (green profiles). Black stars  
220 show the significant peaks found. This peak is an example of common peak between our dataset  
221 and Tan dataset. **C.** Barplot showing the expected and the observed occurrences of the *top-six*  
222 significantly enriched motifs in DHT 100nM treated LNCaP after N-20 or 3279.4  
223 immunoprecipitation. Numbers on top of the barplots are motif enrichment p-values. This motif  
224 search validates our ChIP protocol and antibodies by revealing a clear enrichment of androgen

225 receptor response element (ARE)/ glucocorticoid receptor response element (GRE or NR3C1)  
226 sequences. Moreover, enrichments in binding motifs of several members of the FOX family,  
227 FOXA1, FOXA2, FOXP1 and FOXP2 were also observed in AR-bound sequences. If the two  
228 antibodies were therefore efficient to immunoprecipitate AR, N-20 antibody showed the best  
229 fold enrichment (N-20  $FE_{\max} = 82.18$  vs. 3279.4  $FE_{\max} = 37.37$ ) and was used for experiments  
230 on hNSCs.

231 **Supplementary Figure S9. Top 6 motifs found in AR immunoprecipitation genomic**  
232 **sequences in hNSCs treated with androgens.** Barplot showing the expected and the observed  
233 occurrences of the *top-six* significantly enriched motifs in androgens treated hNSCs after N-20  
234 immunoprecipitation. Numbers on top of the barplots are motif enrichment p-values. The AR  
235 and GRE/NR3C1 motifs, although not found in the “top-enriched” motif, were still slightly  
236 enriched (GRE/NR3C1: 86<sup>th</sup> motif, zscore 2.77 pvalue 0.006; AR 97<sup>th</sup> motif, zscore 1.95 pvalue  
237 0.05). Binding motifs for transcription factors of the FOX family were not significantly  
238 enriched.

## 239 **Supplementary Tables Legends**

240 **Supplementary Table S1. DHT significantly regulates 77 genes in hNSCs.** Analysis of DHT  
241 100nM 24 hours treated samples compared with ten control (DMSO treated) samples. For each  
242 genes is indicated the mean of reads (normalized and divided by gene length) obtained in the  
243 different RNA-seq. FC: fold change, comparison of mean reads obtained in DHT 100nM vs  
244 DMSO treatments. Adj p-value: p-value adjusted for multiple testing using Benjamini’s  
245 correction. The regulation with other androgen treatments DHT 10nM, testosterone 10 and

246 100nM and R1881 1nM (AR agonist) is also indicated (last column). The 29 grey highlighted  
247 genes were not considered because of their low read counts (< 100 reads normalized and divided  
248 by gene length for each condition). The red genes are the genes-of-interest that display both  
249 high FC as well as high statistical significance.

250 **Supplementary Table S2. Top 5 Diseases and Biological Functions from Ingenuity**  
251 **Pathways Analysis (IPA) with details of top 10 function annotations.** For each category and  
252 function annotation, number of genes with examples (columns 'number of genes' and 'genes')  
253 and results of Benjamini statistical test (column 'p-value') are indicated.

254 **Supplementary Table S3. Gene ontology (GO) analysis from Database for Annotation,**  
255 **Visualization and Integrated Discovery (DAVID).** For each GO term, fold enrichment,  
256 number of genes with examples (columns 'number of genes' and 'genes') and results of  
257 Benjamini statistical test (column 'pvalue') are indicated.

258 **Supplementary Table S4. Top 30 peaks after AR immunoprecipitations performed on**  
259 **DHT treated LNCaP.** The first worksheet contains N-20 antibody immunoprecipitation results  
260 and second worksheet the 3279.4 immunoprecipitation results. For each peak are indicated their  
261 coordinates (columns 'chr', 'start' and 'end'), results of Benjamini statistical test (column  
262 'pvalue'), their fold enrichment (compared to the input chromatin), annotation and the closest  
263 gene (columns 'ensembl gene id' and 'gene name'). Peaks were ranked according to their fold  
264 enrichment.

265 **Supplementary Table S5. AR N-20 immunoprecipitated peaks found in the 28 DHT**  
266 **regulated genes (from RNA-seq).** For each hNSCs N-20 immunoprecipitated peak are  
267 indicated their coordinates (columns 'chr', 'start' and 'end'), annotation and the closest gene

268 (columns 'ensembl gene id' and 'gene name'). Fold enrichment (FE) are also mentioned  
269 (for DHT 10nM, 100nM and testosterone 100nM treatments), DHT 10nM and testosterone  
270 100nM treatments having 2 technical replicates (Repl#1 and #2). Peaks were ranked according  
271 to the number of treatment in which they were found (column '#treatment'). The column  
272 'FactorBook transcription binding sites' indicates each enriched motif within each peak with  
273 peak signal value (out of 1 000) and motif score (corresponding to the score for the selected  
274 best-scoring motif sites among MEME-ChIP and FIMO tools transformed to  $-\log_{10}$ , nd: not  
275 determined) (from <http://www.factorbook.org/human/chipseq/tf/>).

276 **Supplementary Table S6. Retinoic acid (RA) significantly regulates 2 760 genes in hNSCs.**

277 Analysis of RA 1 $\mu$ M 24h treated hNSCs compared with control DMSO treated hNSCs (n = 3).  
278 For each gene is indicated their name and Ensembl id and the mean of reads (normalized and  
279 divided by gene length). FC: fold change, comparison of mean reads obtained in RA 1 $\mu$ M vs  
280 DMSO treatments. Adj p-value: p-value adjusted for multiple testing using Benjamini's  
281 correction. The 828 grey highlighted genes were not considered because of their low read counts  
282 (< 100 reads normalized and divided by gene length for each condition). Genes were ranked  
283 according to their statistical significance.

284 **Supplementary Table S7. SNPs in *NRCAM* significantly associated with ASD.** For each

285 SNP, their db id (column 'db SNP id'), coordinates, alleles and gene location are indicated. In  
286 the 'allele A/B' column, B represents the minor allele. These SNPs were analyzed in two  
287 association studies from Bonora *et al.*<sup>14</sup> and Marui *et al.*<sup>15</sup>.

288 **Supplementary Table S8. List and characteristics of primers used in PCR quantification.**

289 For each primer pair, the name of the forward (F) and reverse (R) primer with their sequences

290 and the amplicon size are indicated. Efficiency was calculated with the slope of a standard curve  
291 established with a serial dilution of a pool of samples. Primer were used when the efficiency  
292 was between 1.8 and 2.2 (equivalent to 80% - 120% PCR efficiency). nd: not determined

293 **Supplementary Table S9. Lists of genes used for the current study.** The first worksheet  
294 contains the list of 651 genes mutated in at least one ASD/ID patient combined from gene lists  
295 already published <sup>16-18</sup>. The second worksheet contains the list of 1 660 genes differentially  
296 expressed in autistic patients (data available from transcriptomic microarrays performed in  
297 blood, brain and lymphoblastoid cell lines <sup>19-24</sup>. The third and fourth worksheets contain genes  
298 collected by Autismkb from association studies from respectively candidate genes (177 genes)  
299 or genome-wide (132 genes) studies.

## 300 **Supplementary references**

- 301 1. Massie CE, Adryan B, Barbosa-Morais NL, Lynch AG, Tran MG, Neal DE *et al.* New androgen  
302 receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Rep*  
303 2007; **8**(9): 871-878.
- 304 2. Adli M, Bernstein BE. Whole-genome chromatin profiling from limited numbers of cells using  
305 nano-ChIP-seq. *Nat Protoc* 2011; **6**(10): 1656-1668.
- 306 3. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq.  
307 *Bioinformatics* 2009; **25**(9): 1105-1111.
- 308 4. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput  
309 sequencing data. *Bioinformatics* 2015; **31**(2): 166-169.
- 310 5. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*  
311 2010; **11**(10): R106.
- 312 6. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq  
313 data with DESeq2. *Genome Biol* 2014; **15**(12): 550.
- 314 7. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short  
315 DNA sequences to the human genome. *Genome Biol* 2009; **10**(3): R25.

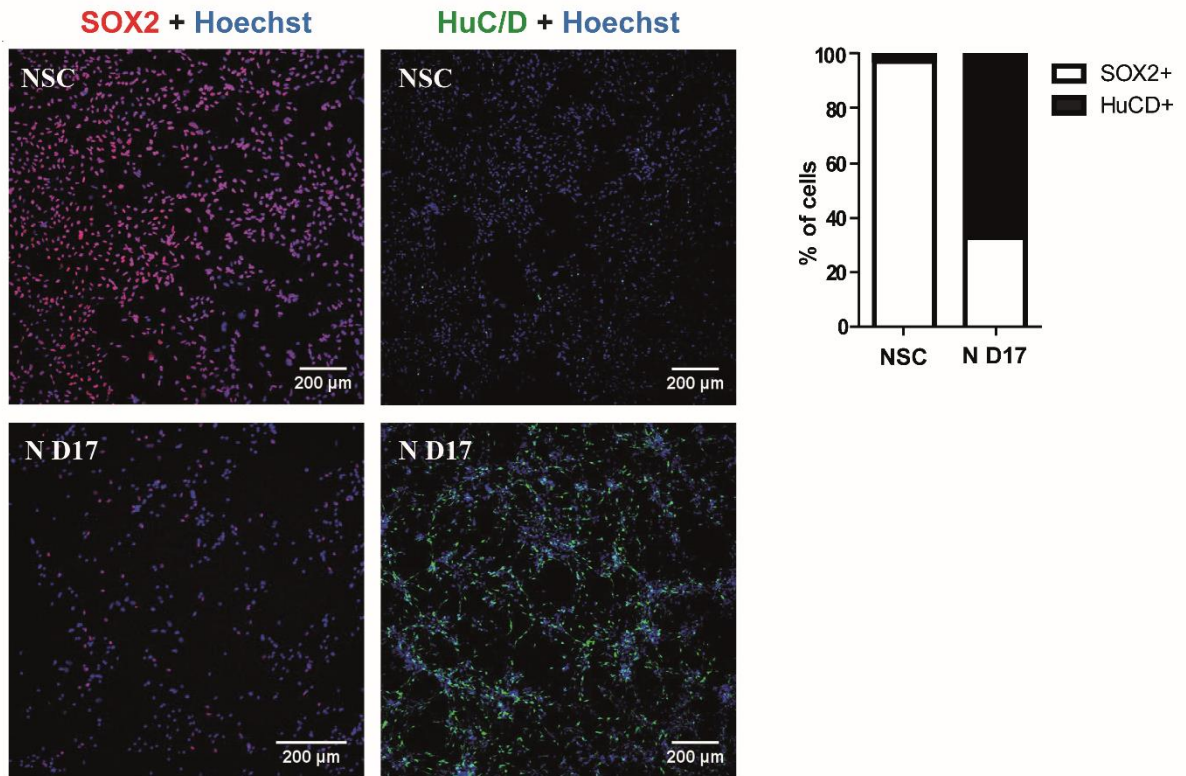
- 316 8. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE *et al.* Model-based analysis of  
317 ChIP-Seq (MACS). *Genome Biol* 2008; **9**(9): R137.
- 318 9. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P *et al.* Simple combinations of lineage-  
319 determining transcription factors prime cis-regulatory elements required for macrophage and  
320 B cell identities. *Mol Cell* 2010; **38**(4): 576-589.
- 321 10. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*  
322 2011; **27**(7): 1017-1018.
- 323 11. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior  
324 genetics research. *Behav Brain Res* 2001; **125**(1-2): 279-284.
- 325 12. Boissart C, Poulet A, Georges P, Darville H, Julita E, Delorme R *et al.* Differentiation from human  
326 pluripotent stem cells of cortical neurons of the superficial layers amenable to psychiatric  
327 disease modeling and high-throughput drug screening. *Transl Psychiatry* 2013; **3**: e294.
- 328 13. Tan PY, Chang CW, Chng KR, Wansa KD, Sung WK, Cheung E. Integration of regulatory networks  
329 by NKX3-1 promotes androgen-dependent prostate cancer survival. *Mol Cell Biol* 2012; **32**(2):  
330 399-414.
- 331 14. Bonora E, Lamb JA, Barnby G, Sykes N, Moberly T, Beyer KS *et al.* Mutation screening and  
332 association analysis of six candidate genes for autism on chromosome 7q. *Eur J Hum Genet*  
333 2005; **13**(2): 198-207.
- 334 15. Marui T, Funatogawa I, Koishi S, Yamamoto K, Matsumoto H, Hashimoto O *et al.* Association of  
335 the neuronal cell adhesion molecule (NRCAM) gene variants with autism. *Int J*  
336 *Neuropsychopharmacol* 2009; **12**(1): 1-10.
- 337 16. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH *et al.* Genome  
338 sequencing identifies major causes of severe intellectual disability. *Nature* 2014; **511**(7509):  
339 344-347.
- 340 17. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic de  
341 novo mutations in autism spectrum disorders. *Nature* 2012; **485**(7397): 242-245.
- 342 18. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L *et al.* Convergence of genes and  
343 cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 2014; **94**(5):  
344 677-694.
- 345 19. Garbett K, Ebert PJ, Mitchell A, Lintas C, Manzi B, Mirnics K *et al.* Immune transcriptome  
346 alterations in the temporal cortex of subjects with autism. *Neurobiol Dis* 2008; **30**(3): 303-311.
- 347 20. Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA *et al.* Gene expression changes  
348 in children with autism. *Genomics* 2008; **91**(1): 22-29.
- 349 21. Nishimura Y, Martin CL, Vazquez-Lopez A, Spence SJ, Alvarez-Retuerto AI, Sigman M *et al.*  
350 Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of  
351 autism and reveals shared pathways. *Hum Mol Genet* 2007; **16**(14): 1682-1698.



- 352 22. Purcell AE, Jeon OH, Zimmerman AW, Blue ME, Pevsner J. Postmortem brain abnormalities of  
353 the glutamate neurotransmitter system in autism. *Neurology* 2001; **57**(9): 1618-1628.
- 354 23. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S *et al.* Transcriptomic analysis of  
355 autistic brain reveals convergent molecular pathology. *Nature* 2011; **474**(7351): 380-384.
- 356 24. Hu VW, Nguyen A, Kim KS, Steinberg ME, Sarachana T, Scully MA *et al.* Gene expression profiling  
357 of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal  
358 development and steroid biosynthesis. *PLoS One* 2009; **4**(6): e5775.
- 359

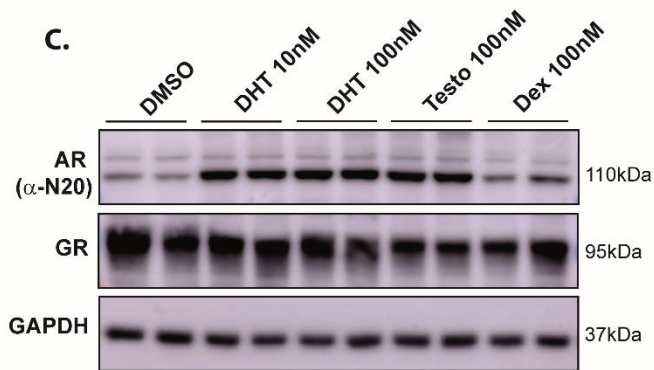
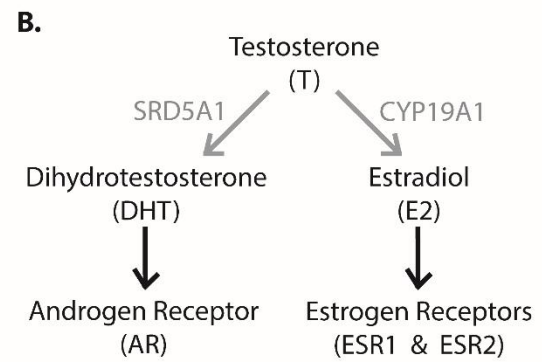
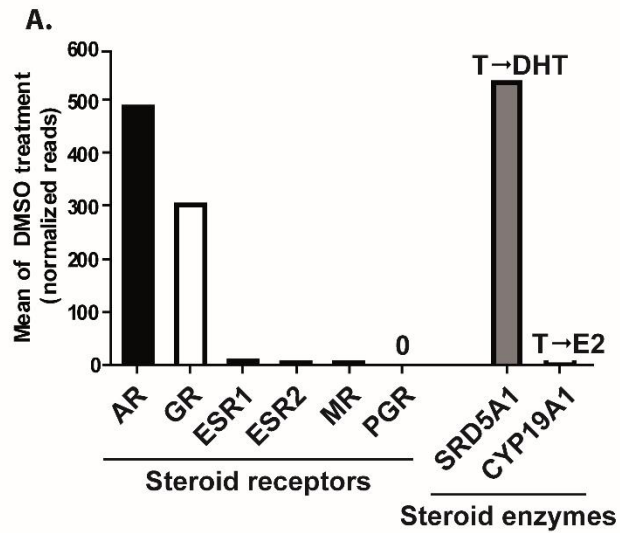
1 **Supplementary Figures**

**Supplementary Figure S1**



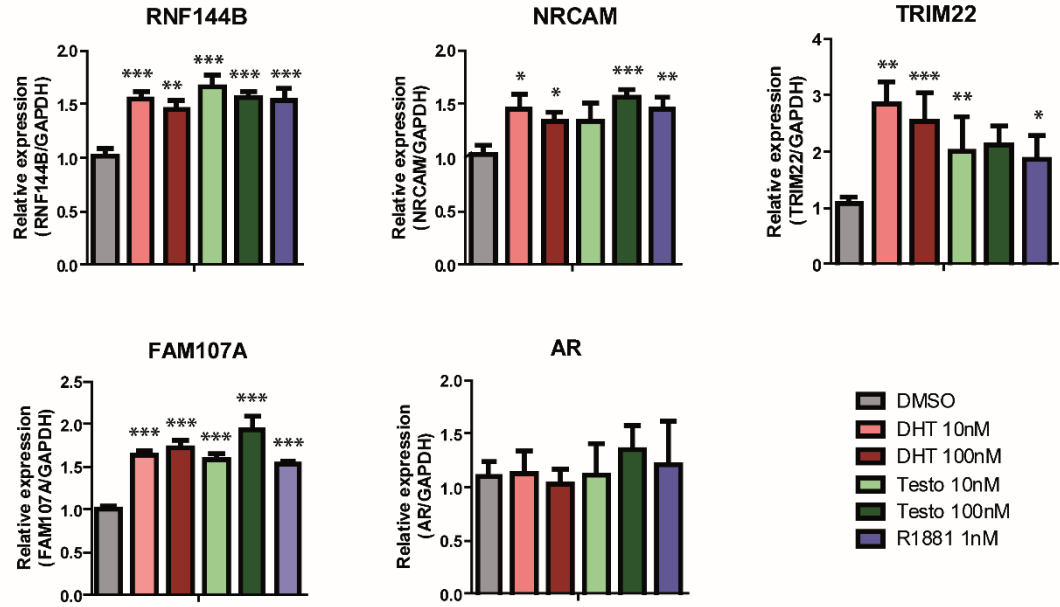
2

Supplementary Figure S2

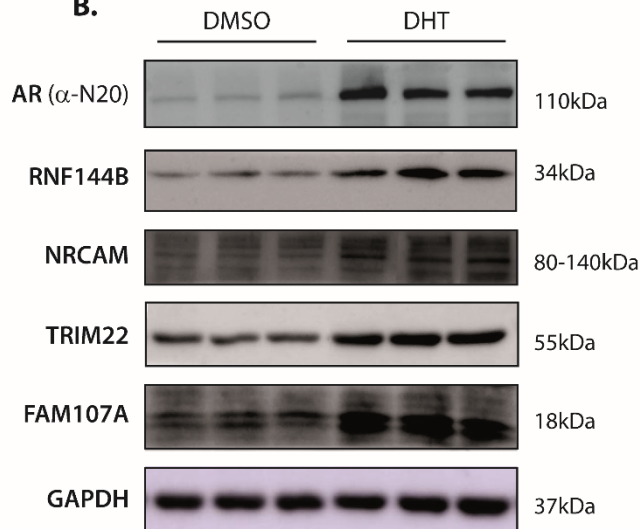


Supplementary Figure S3

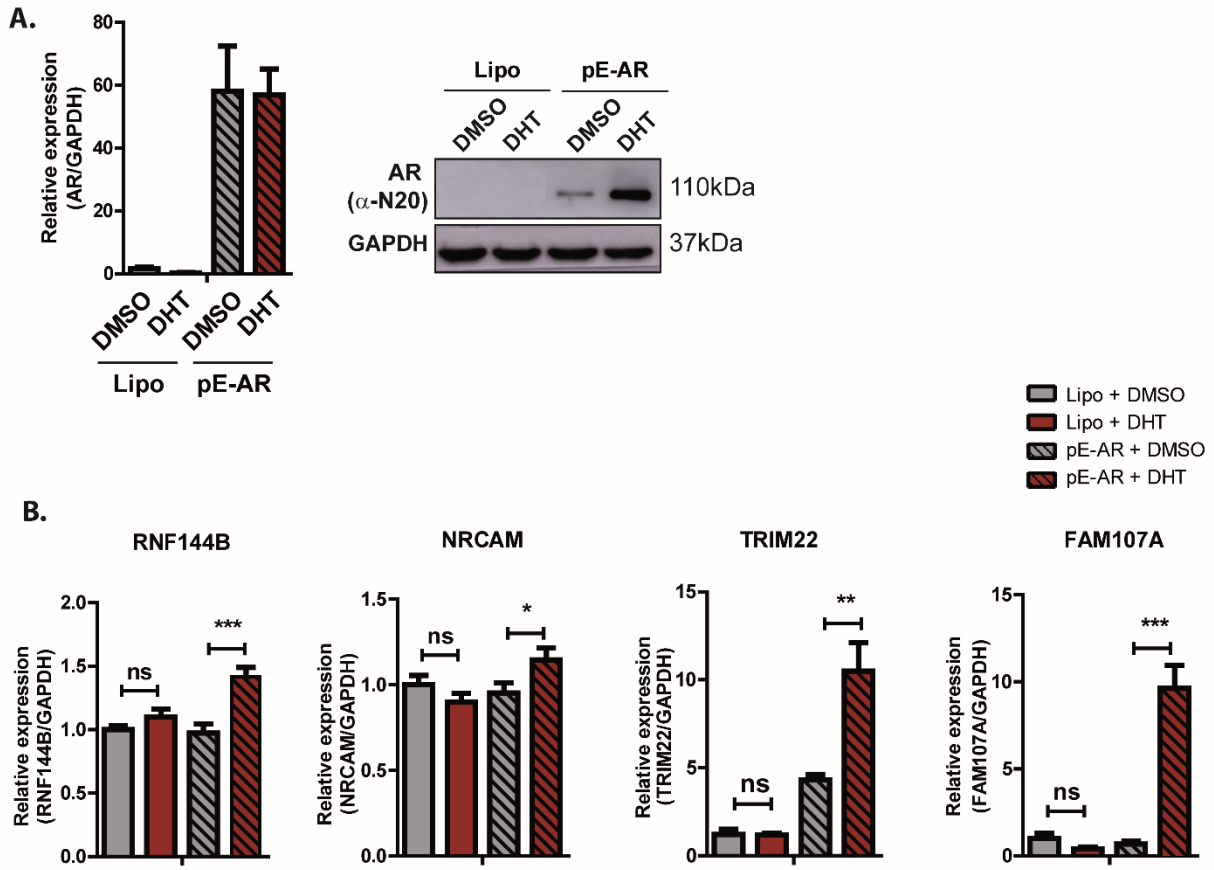
A.



B.

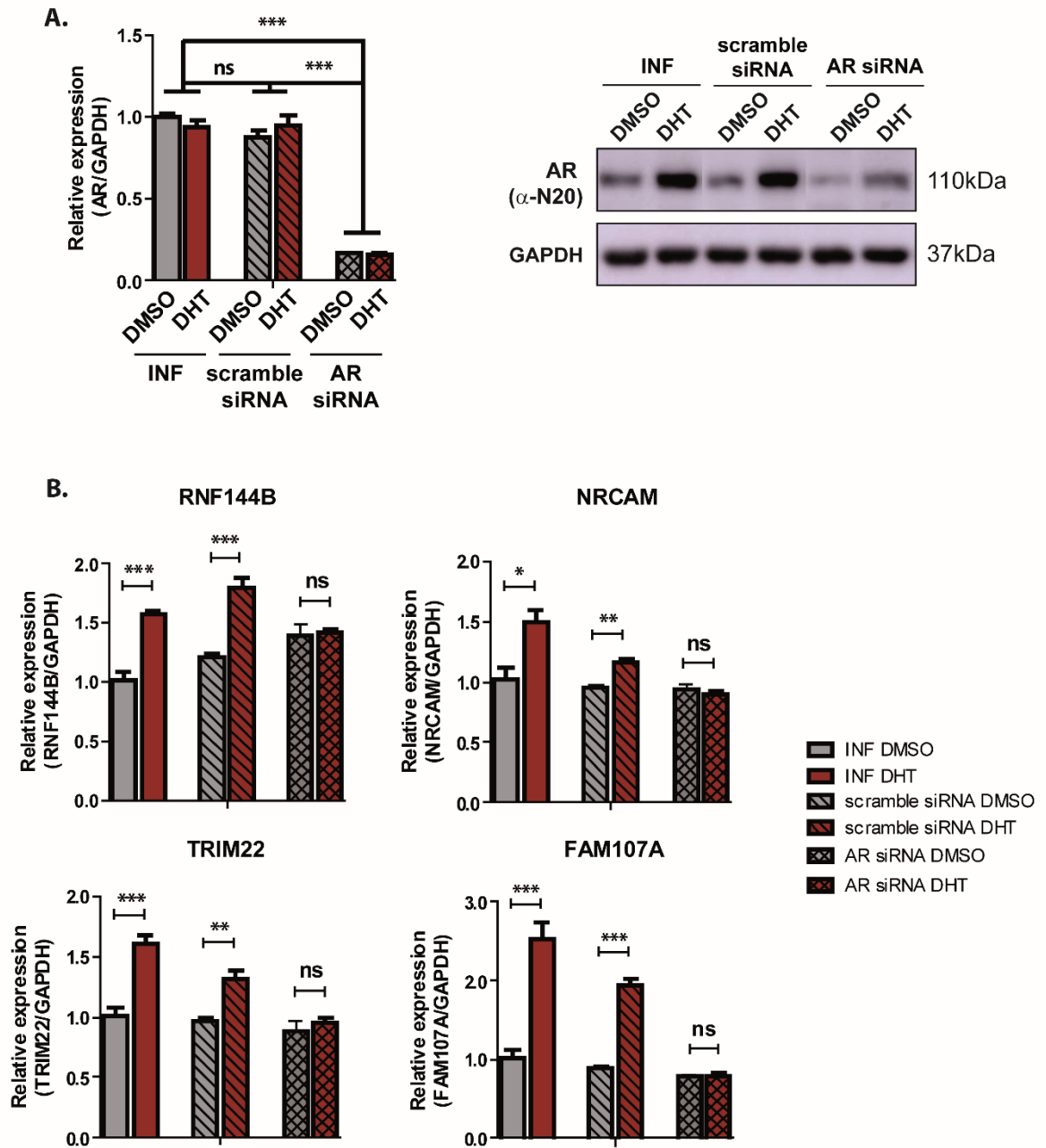


Supplementary Figure S4



5

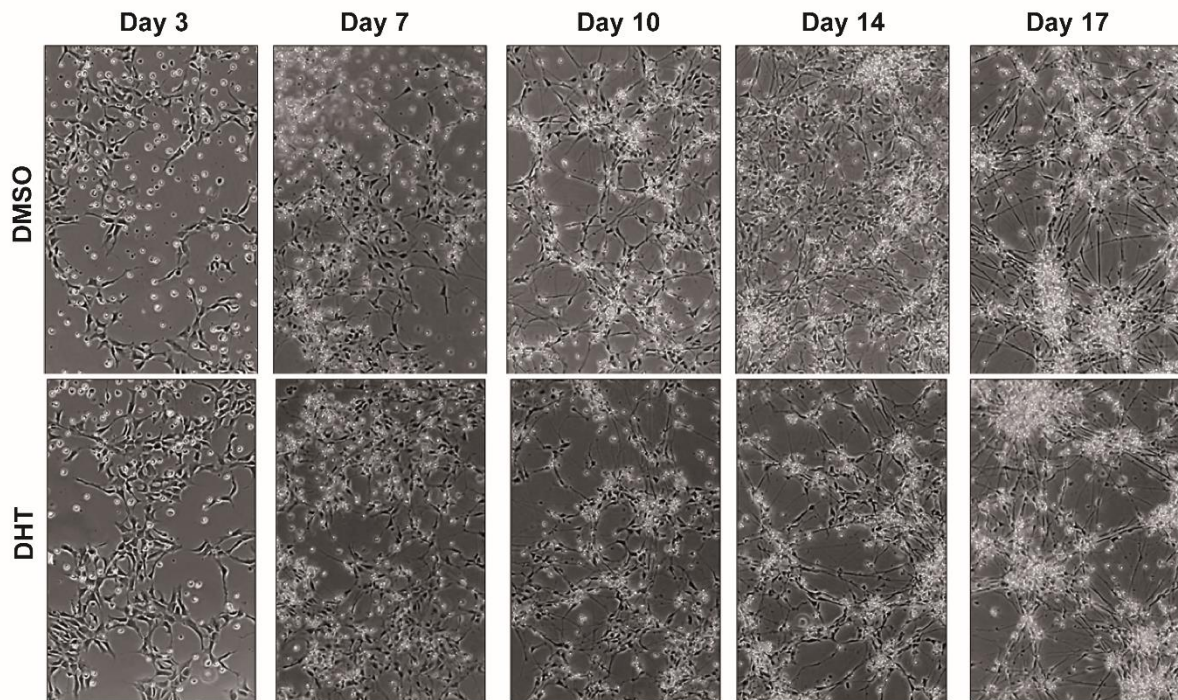
Supplementary Figure S5



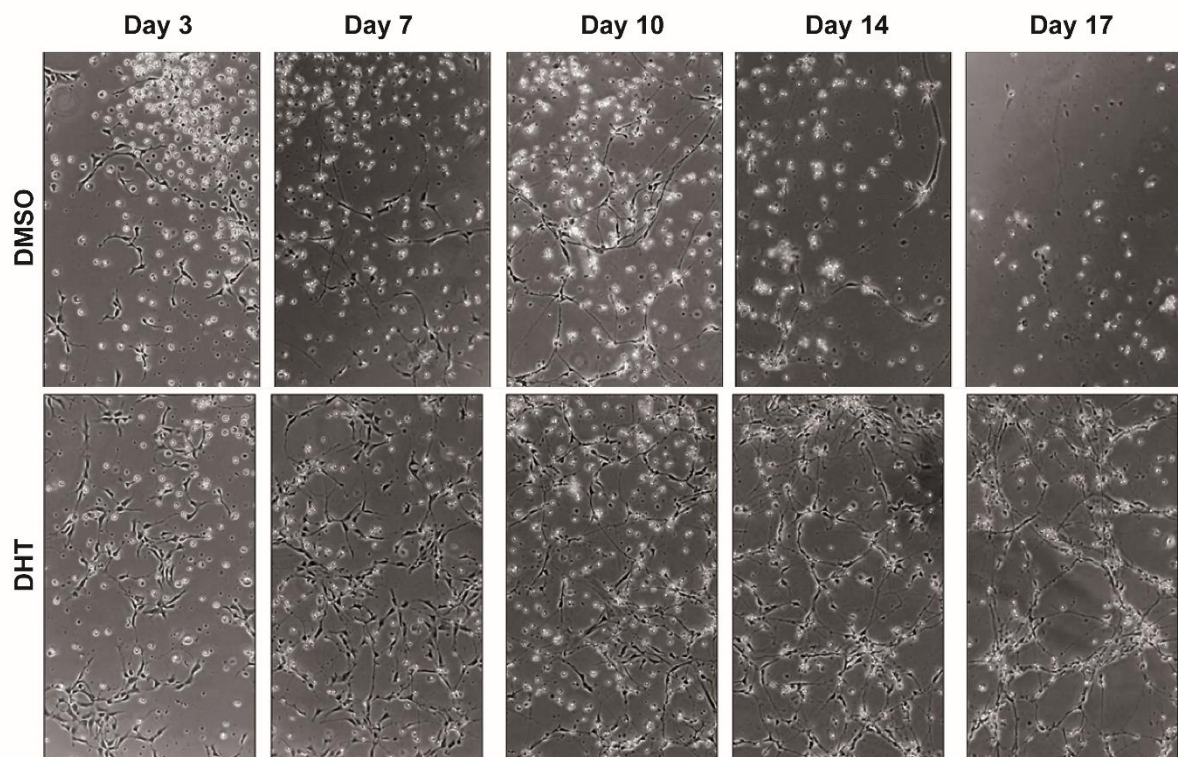


Supplementary Figure S6

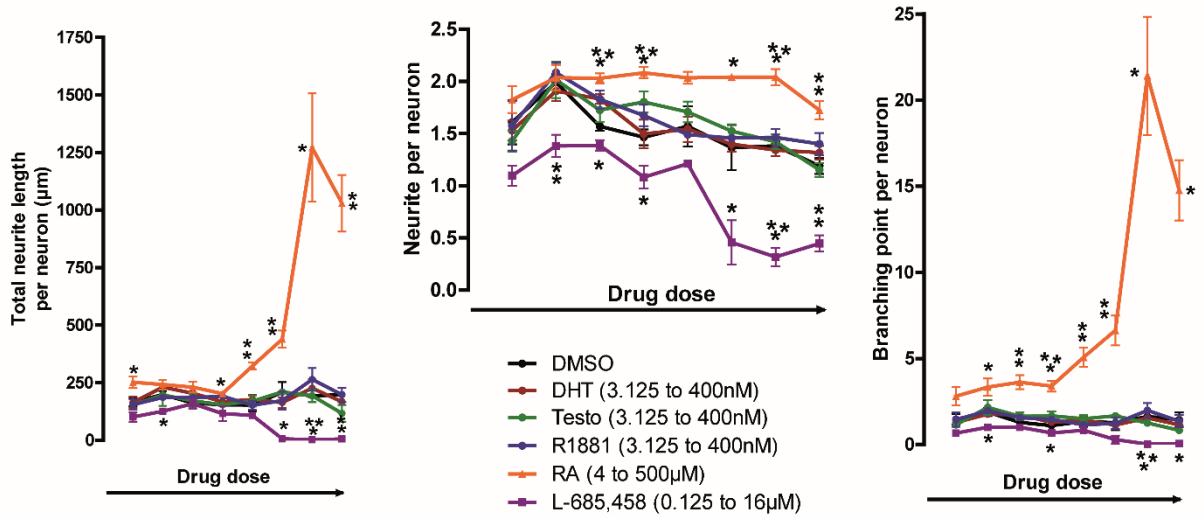
A. Classical medium



B. Depleted medium



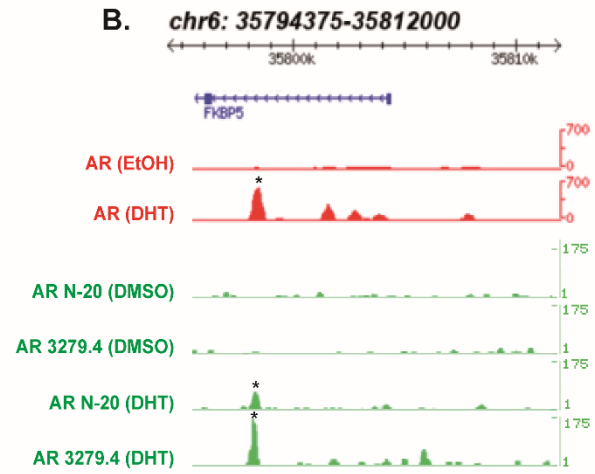
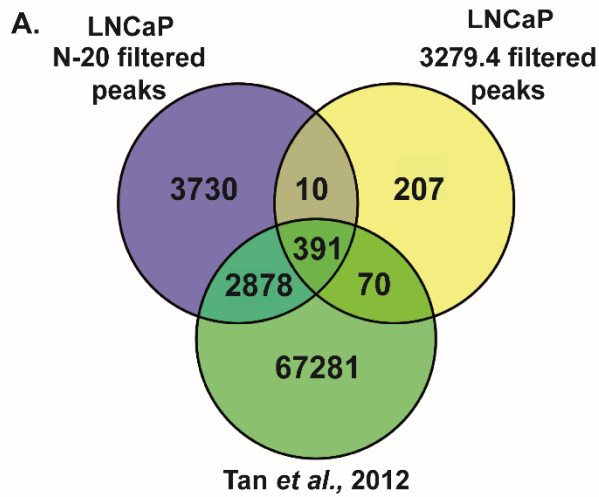
Supplementary Figure S7



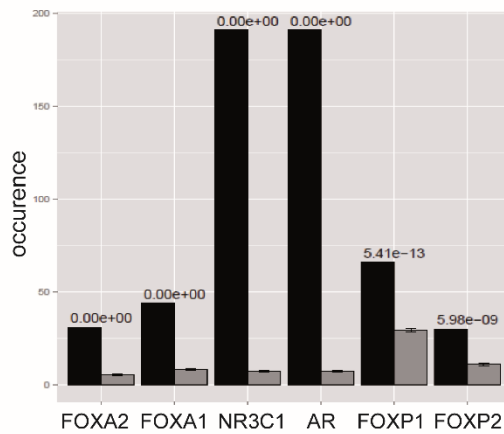
8



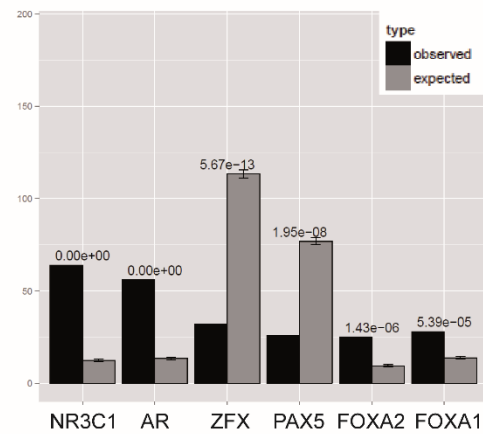
Supplementary Figure S8



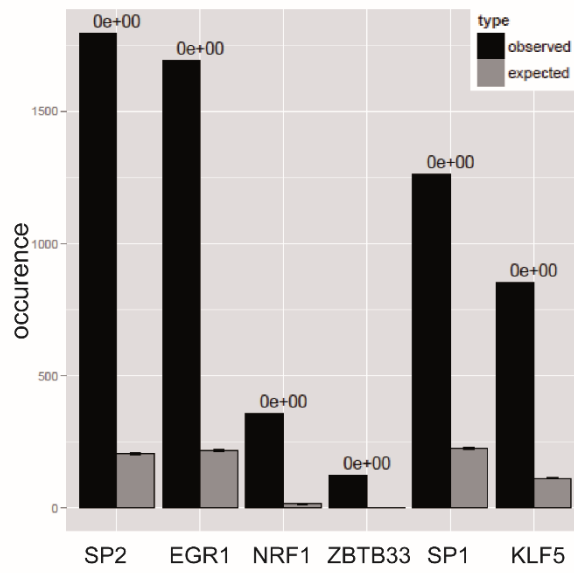
**C.** LNCaP DHT N-20 Top 6 motifs



LNCaP DHT 3279.4 Top 6 motifs



Supplementary Figure S9



1

## 2 Supplementary Tables

Supplementary Table S1. DHT significantly regulates 77 genes in hNSCs

Ensembl gene ID	Gene name	Mean reads DMSO	Mean reads DHT 100nM	FC	Adj p-value	Significant difference in other androgens treatments
ENSG00000102760	<b>RGCC</b>	127	287	<b>2,11</b>	<b>5,22E-36</b>	4/4
ENSG00000137393	<b>RNF144B</b>	125	216	<b>1,74</b>	<b>3,94E-19</b>	4/4
ENSG00000091129	<b>NRCAM</b>	5373	7169	<b>1,33</b>	<b>6,73E-17</b>	4/4
ENSG00000132274	<b>TRIM22</b>	495	721	<b>1,35</b>	<b>8,08E-10</b>	4/4
ENSG00000168309	<b>FAM107A</b>	222	370	<b>1,49</b>	<b>1,15E-04</b>	4/4
ENSG00000196569	<b>LAMA2</b>	120	186	<b>1,41</b>	<b>2,46E-04</b>	4/4
ENSG00000115461	<b>IGFBP5</b>	6557	17752	<b>1,74</b>	<b>9,74E-04</b>	4/4
ENSG00000121207	<b>LRAT</b>	618	807	<b>1,21</b>	<b>3,46E-09</b>	2/4
ENSG00000174306	<b>ZHX3</b>	1326	1667	<b>1,24</b>	<b>1,17E-07</b>	2/4
ENSG00000133401	<b>PDZD2</b>	261	404	<b>1,39</b>	<b>6,05E-05</b>	2/4
ENSG00000149591	<b>TAGLN</b>	2405	2198	<b>-1,28</b>	<b>1,90E-04</b>	2/4
ENSG00000124440	<b>HIF3A</b>	165	196	<b>1,20</b>	<b>7,45E-03</b>	2/4
ENSG00000104722	<b>NEFM</b>	3587	4777	<b>1,24</b>	<b>9,11E-03</b>	2/4
ENSG00000112232	<b>KHDRBS2</b>	245	204	<b>-1,20</b>	<b>2,39E-02</b>	2/4
ENSG00000182022	<b>CHST15</b>	132	182	<b>1,35</b>	<b>4,31E-12</b>	1/4
ENSG00000137142	<b>IGFBPL1</b>	1853	2457	<b>1,28</b>	<b>1,15E-04</b>	1/4
ENSG00000077943	<b>ITGA8</b>	956	1436	<b>1,45</b>	<b>6,54E-03</b>	1/4
ENSG00000148677	<b>ANKRD1</b>	429	389	<b>-1,29</b>	<b>4,22E-02</b>	1/4
ENSG00000107551	<b>RASSF4</b>	862	1073	<b>1,27</b>	<b>6,14E-17</b>	0/4
ENSG00000166483	<b>WEE1</b>	5196	6414	<b>1,21</b>	<b>4,48E-07</b>	0/4
ENSG00000150551	<b>LYPD1</b>	3276	3605	<b>1,25</b>	<b>1,46E-05</b>	0/4
ENSG00000141744	<b>PNMT</b>	271	318	<b>1,25</b>	<b>6,46E-05</b>	0/4
ENSG00000188290	<b>HES4</b>	3046	2383	<b>-1,30</b>	<b>8,26E-05</b>	0/4
ENSG00000185432	<b>METTL7A</b>	166	192	<b>1,24</b>	<b>1,15E-04</b>	0/4
ENSG00000164106	<b>SCRG1</b>	95	118	<b>1,33</b>	<b>1,48E-04</b>	0/4
ENSG00000144810	<b>COL8A1</b>	271	433	<b>1,42</b>	<b>2,36E-04</b>	0/4
ENSG00000164929	<b>BAALC</b>	3492	4402	<b>1,28</b>	<b>5,71E-04</b>	0/4
ENSG00000003436	<b>TFPI</b>	160	224	<b>1,29</b>	<b>7,69E-04</b>	0/4
ENSG00000006468	<b>ETV1</b>	5037	5591	<b>1,21</b>	<b>1,34E-03</b>	0/4
ENSG00000002587	<b>HS3ST1</b>	312	394	<b>1,24</b>	<b>1,38E-03</b>	0/4
ENSG00000145246	<b>ATP10D</b>	456	560	<b>1,22</b>	<b>1,38E-03</b>	0/4

**Supplementary Table S2.**

**Top 5 Diseases and Biological Functions from Ingenuity Pathways Analysis (IPA) with details of top 10 function annotations. In bold are indicated FAM107A and NRCAM and their relative functions**

Function Annotation	p-value	Number of genes	Genes
<b>Category: Nervous System Development and Function</b>	<b>5,52E-08</b>	<b>23</b>	
Abnormal morphology of spinal nerve	5,52E-08	5	ALPL, ERBB3, ETV1, NRARP, NRP1
Coordination	2,32E-06	8	ALPL, ATXN1, CALB1, CNTN6, ETV1, GRIA2, <b>NRCAM</b> , PLCB4
<b>Abnormal morphology of nervous system</b>	5,81E-06	14	ALPL, CALB1, CDH11, CNTN6, ERBB3, ETV1, GRIA2, ID1, KHDRBS2, LAMA2, NRARP, <b>NRCAM</b> , NRP1, PLCB4
<b>Morphogenesis of neurites</b>	5,99E-05	9	ATXN1, CYR61, <b>FAM107A</b> , ID1, MYO5B, NEFM, <b>NRCAM</b> , NRP1, WEE1
<b>Branching of neurites</b>	8,23E-05	7	ATXN1, CYR61, <b>FAM107A</b> , MYO5B, NEFM, <b>NRCAM</b> , NRP1
<b>Dendritic growth and branching</b>	1,54E-04	6	ATXN1, CYR61, <b>FAM107A</b> , MYO5B, NEFM, <b>NRCAM</b>
<b>Abnormal morphology of cerebellum</b>	1,74E-04	5	CALB1, ERBB3, KHDRBS2, <b>NRCAM</b> , PLCB4
<b>Foliation of cerebellum</b>	5,16E-04	3	KHDRBS2, <b>NRCAM</b> , PLCB4
Abnormal morphology of ventral root	6,73E-04	2	ERBB3, ETV1
<b>Abnormal morphology of brain</b>	8,50E-04	8	CALB1, CNTN6, ERBB3, GRIA2, ID1, KHDRBS2, <b>NRCAM</b> , PLCB4
<b>Category: Organismal Development</b>	<b>6,94E-06</b>	<b>33</b>	
Abnormal morphology of head	6,94E-06	15	ALPL, CALB1, CNTN6, COL8A1, CYP1B1, ERBB3, GRIA2, ID1, ITGA8, KHDRBS2, LRAT, <b>NRCAM</b> , PKDCC, PLCB4, TFPI
Formation of lung	3,68E-05	8	ABCA1, ATXN1, CDH11, ELK3, HIF3A, ID1, ITGA8, PKDCC
Abnormal morphology of cerebellum	1,74E-04	5	CALB1, ERBB3, KHDRBS2, <b>NRCAM</b> , PLCB4
Proliferation of mammary cells	3,89E-04	3	ERBB3, ID1, IGFBP5
Foliation of cerebellum	5,16E-04	3	KHDRBS2, <b>NRCAM</b> , PLCB4
Abnormal morphology of anterior segment of the eye	6,20E-04	4	COL8A1, CYP1B1, ID1, <b>NRCAM</b>
Size of body	6,39E-04	12	ABCA1, ALPL, ERBB3, ETV1, GRIA2, ID1, IGFBP5, LAMA2, <b>NRCAM</b> , PKDCC, PLCB4, SSFA2

**Supplementary Table S3. Gene ontology (GO) analysis from Database for Annotation, Visualization And Integrated Discovery (DAVID)**

Category	Term	Fold Enrichment	PValue	Number of genes	Genes
GOTERM_BP_FAT	GO:0007155~ Cell adhesion	3,67	6,11E-04	11	AJUBA, CDH11, COL4A6, COL8A1, CNTN6, CYR61, ITGA8, LAMA2, <b>NRCAM</b> , NRP1, PDZD2,
GOTERM_BP_FAT	GO:0022610~ Biological adhesion	3,66	6,18E-04	11	AJUBA, CDH11, COL4A6, COL8A1, CNTN6, CYR61, ITGA8, LAMA2, <b>NRCAM</b> , NRP1, PDZD2,
GOTERM_BP_FAT	GO:0001558~ Regulation of cell growth	7,21	1,32E-03	6	CYR61, <b>FAM107A</b> , IGFBP5, IGFBPL1, <b>NRCAM</b> , NRP1,
GOTERM_MF_FAT	GO:0019842~ Vitamin binding	8,76	2,36E-03	5	CALB1, CRABP2, CUBN, LRAT, TCN2,
GOTERM_MF_FAT	GO:0005520~ insulin-like growth factor binding	27,33	5,14E-03	3	
GOTERM_CC_FAT	GO:0005576~ extracellular region	1,97	7,23E-03	17	
GOTERM_MF_FAT	GO:0019838~ growth factor binding	8,68	1,04E-02	4	
GOTERM_MF_FAT	GO:0016564~ transcription repressor activity	4,32	1,15E-02	6	
GOTERM_BP_FAT	GO:0007517~ muscle organ development	5,53	1,19E-02	5	
GOTERM_BP_FAT	GO:0040008~ regulation of growth	4,10	1,42E-02	6	
GOTERM_BP_FAT	GO:0043569~ negative regulation of insulin-like growth factor receptor signaling pathway	116,62	1,67E-02	2	
GOTERM_BP_FAT	GO:0015889~ cobalamin transport	116,62	1,67E-02	2	
GOTERM_CC_FAT	GO:0044421~ extracellular region part	2,42	1,82E-02	10	
GOTERM_BP_FAT	GO:0030155~ regulation of cell adhesion	6,81	2,00E-02	4	

**Supplementary Table S4.**

**Top 30 peaks after AR N-20 and 3279.4 immunoprecipitation performed on DHT treated LNCaP**

<b>N-20 antibody immunoprecipitation</b>							
<b>Chr</b>	<b>Start</b>	<b>End</b>	<b>-10*LOG10 (pvalue)</b>	<b>Fold Enrichment</b>	<b>Annotation</b>	<b>Ensembl Gene ID</b>	<b>Gene Name</b>
chr10	123879995	123881069	2942,73	82,18	Distant-promoter	ENSG00000138162	TACC2
chr1	9495950	9496912	1121,41	41,65	Intergenic	ENSG00000252956	RNA5SP40
chr11	129894846	129895663	1067,94	39,42	Intergenic	ENSG00000255262	TCEB2P2
chr11	114049601	114051514	1907,17	38,34	promoter-TSS	ENSG00000109906	<b>ZBTB16</b>
chr2	102511960	102513064	977,66	38,34	Intergenic	ENSG00000233404	FLJ20373
chr12	64983602	64984821	846,88	36,07	Intergenic	ENSG00000223294	SNORD83
chr12	111505008	111506107	869,94	33,86	Distant-promoter	ENSG00000221259	<b>CUX2</b>
chr16	78156520	78157737	750,44	31,79	Distant-promoter	ENSG00000186153	WVOX
chr8	124417384	124418278	1130,94	31,12	intron	ENSG00000254294	IMPDH1P6
chr12	52519829	52521223	1242,85	30,95	Intergenic	ENSG00000265804	AC078864.2
chr7	107452024	107453079	866,79	30,43	Intergenic	ENSG00000229626	PIGCP2
chr1	32967312	32968105	855,83	29,75	intron	ENSG00000273274	ZBTB8B
chr6	49632127	49633342	566,14	29	Intergenic	ENSG00000112077	RHAG
chr11	15689773	15690750	637,74	28,63	intron	ENSG00000254695	RP11-396020.1
chr14	36397978	36399138	767,33	28,57	intron	NA	NA
chr12	73447519	73448501	869,57	28,26	Intergenic	ENSG00000258294	RP11-314D7.1
chr11	70585594	70586723	671,3	27,89	intron	ENSG00000207196	Y_RNA
chr4	106831783	106832957	678,55	27,76	promoter-TSS	ENSG00000168743	NPNT
chr11	61462762	61463366	658,49	27,14	intron	ENSG00000134780	DAGLA
chr11	26931722	26932824	542,56	25,66	Intergenic	ENSG00000176971	FIBIN
chr15	89672512	89673781	596,07	25,32	intron	ENSG00000259699	HMGB1P8
chr5	172438131	172439223	565,45	25,3	Distant-promoter	ENSG00000212402	SNORA74B
chr11	70823841	70824910	584,65	24,93	Distant-promoter	ENSG00000162105	SHANK2
chr13	63295633	63296739	559,23	24,54	intron	ENSG00000228669	LINC00448
chr1	9613387	9614469	466,19	24,41	intron	NA	NA
chr2	234861674	234862822	539,85	24,07	promoter-TSS	ENSG00000144481	TRPM8
chr15	91585008	91585724	587,67	23,91	Distant-promoter	NA	NA
chr7	157665782	157666561	627,81	23,8	intron	NA	NA
chr6	110125442	110126667	431,64	23,74	intron	ENSG00000112367	FIG4
chr2	100615148	100616127	492,66	23,29	intron	ENSG00000144218	AFF3
<b>3279.4 antibody immunoprecipitation</b>							
chr10	123880031	123881378	3100	37,37	Distant-promoter	ENSG00000138162	TACC2
chr11	129895176	129895716	3100	28,75	Intergenic	ENSG00000255262	TCEB2P2
chr8	124417432	124418171	3100	21,67	intron	ENSG00000254294	IMPDH1P6
chr12	52520181	52521195	3100	20,76	Intergenic	ENSG00000265804	AC078864.2

Supplementary Table S5. AR-N-20immunoprecipitated peaks found in the 28 DHT regulated genes (from RNA-seq) part 1/2

Chr	Start	End	Annotation	Ensembl Gene ID	Gene Name	FE DHT 10nM N20 Repl#1	FE DHT 10nM N20 Repl#2	FE DHT 100nM N20	FE Testo 100nM N20 Repl#1	FE Testo 100nM N20 Repl#2	#Treatment
chr7	107949253	107950043	intron	ENSG00000091129	<b>NRCAM</b>	10,79	6,36		13,27	10,21	4
chr20	39878252	39879122	intron	ENSG00000174306	ZHX3	7,31	4,21		6,19	6,11	3
chr11	9590302	9591203	Distant-promoter	ENSG00000166483	WEE1	4,59	2,87		10,56	5,7	2
chr1	86048640	86049848	3'	ENSG00000142871	CYR61		4,04		5,23	5,3	2
chr20	9047990	9049205	promoter-TSS	ENSG00000101333	PLCB4				5,62		1
chr14	23451204	23452333	promoter-TSS	ENSG00000129474	AJUBA	2,92	3,83		5,1		1
chr12	56475553	56476371	Distant-promoter	ENSG00000065361	ERBB3	5,21					1
chr9	140213371	140214019	Distant-promoter	ENSG00000198435	NRARP				5,62		1
chr10	135143054	135143947	Distant-promoter	ENSG00000130643	CALY		5,66				1
chr10	15590018	15590980	exon	ENSG00000077943	ITGA8			5,34			1
chr8	133425540	133426486	intron	ENSG00000184156	KCNQ3				5,83		1
chr9	140197032	140198313	promoter-TSS	ENSG00000198435	NRARP		2,59		5,29		1
chr17	37824518	37825353	intron	ENSG00000141744	PNMT		6,74		4,96		1
chr17	45055695	45057250	exon	ENSG00000179673	RPRML	3,29			5,55		1
chr3	58567104	58568003	Distant-promoter	ENSG00000168309	<b>FAM107A</b>					5,59	1
chr7	107902091	107903039	Distant-promoter	ENSG00000091129	<b>NRCAM</b>					5,26	1
chr7	108095988	108096672	intron	ENSG00000091129	<b>NRCAM</b>	3,68			5,57		1
chr1	21892298	21893107	Distant-promoter	ENSG00000162551	ALPL					5,29	1
chr11	47621339	47622072	Distant-promoter	ENSG00000172247	C1QTNF4					7,61	1
chr16	65080225	65081197	intron	ENSG00000140937	CDH11					5,59	1

Supplementary Table S5. AR N-20immunoprecipitated peaks found in the 28 DHT regulated genes (from RNA-seq) part 2/2

Gene Name	FactorBook transcription binding sites (signal/motif score)									
	NR3C1	FOXA1	FOXA2	FOXP2	SP1	EGR1	NRF1			
<b>NRCAM</b>	1000/2,89	173/nd								
ZHX3										
WEE1	1000/2,49	1000/nd + 560/nd	1000/nd	339/nd + 123/nd	1000/nd + 187/nd					
CYR61										
PLCB4	200/nd + 176/nd	202/nd			670/2,23	269/nd	242/3,31			
AJUBA				164/nd	248/2,84	390/2,77				
ERBB3		821/1,68	586/1,68		1000/nd					
NRARP				209/nd		168/nd				
CALY						152				
ITGA8		187/1,04								
KCNQ3				188/nd						
NRARP						438/2,87 + 184/nd				
PNMT						329/3,49				
RPRML						941/5,21				
<b>FAM107A</b>										
<b>NRCAM</b>										
<b>NRCAM</b>	543/1.43									
ALPL										
C1QTNF4										
CDH11										



**Supplementary Table S6. Retinoic acid (RA) significantly regulates 2 760 genes in hNSCs**

Ensembl gene id	Gene name	Mean reads DMSO	Mean reads RA 1 $\mu$ M	FC	Adj. p-value
ENSG00000130600	<b>H19</b>	26	1253	<b>43,31</b>	<b>0</b>
ENSG00000162496	<b>DHRS3</b>	192	6833	<b>34,04</b>	<b>0</b>
ENSG00000005102	<b>MEOX1</b>	31	842	<b>24,83</b>	<b>0</b>
ENSG00000085563	<b>ABCB1</b>	340	8907	<b>24,68</b>	<b>0</b>
ENSG00000118432	<b>CNR1</b>	357	7967	<b>21,15</b>	<b>0</b>
ENSG00000151892	<b>GFRA1</b>	512	7794	<b>15,06</b>	<b>0</b>
ENSG00000077092	<b>RARB</b>	226	2460	<b>10,66</b>	<b>0</b>
ENSG00000179981	<b>TSHZ1</b>	141	1171	<b>8,22</b>	<b>0</b>
ENSG00000064651	<b>SLC12A2</b>	866	3279	<b>3,77</b>	<b>0</b>
ENSG00000164929	<b>BAALC</b>	3604	12274	<b>3,40</b>	<b>0</b>
ENSG00000108852	<b>MPP2</b>	2717	7725	<b>2,84</b>	<b>0</b>
ENSG00000145423	<b>SFRP2</b>	4578	12761	<b>2,79</b>	<b>0</b>
ENSG00000175093	<b>SPSB4</b>	3763	8810	<b>2,34</b>	<b>0</b>
ENSG00000123983	<b>ACSL3</b>	9030	20044	<b>2,22</b>	<b>0</b>
ENSG00000179314	<b>WSCD1</b>	5245	2649	<b>-1,98</b>	<b>0</b>
ENSG00000129757	<b>CDKN1C</b>	4031	975	<b>-4,11</b>	<b>0</b>
ENSG00000120075	<b>HOXB5</b>	110	817	<b>7,33</b>	<b>5,27E-292</b>
ENSG00000182742	<b>HOXB4</b>	297	1606	<b>5,34</b>	<b>5,97E-290</b>
ENSG00000108511	<b>HOXB6</b>	659	2621	<b>3,96</b>	<b>4,82E-287</b>
ENSG00000145911	<b>N4BP3</b>	112	480	<b>4,25</b>	<b>6,84E-287</b>
ENSG00000188042	<b>ARL4C</b>	1191	2214	<b>1,86</b>	<b>2,48E-284</b>
ENSG00000117394	<b>SLC2A1</b>	44608	64527	<b>1,45</b>	<b>4,82E-275</b>
ENSG00000167244	<b>IGF2</b>	11	1058	<b>71,22</b>	<b>1,44E-273</b>
ENSG00000114115	<b>RBP1</b>	2043	21116	<b>9,97</b>	<b>4,86E-266</b>
ENSG00000112773	<b>FAM46A</b>	36	492	<b>12,91</b>	<b>2,90E-260</b>
ENSG00000171004	<b>HS6ST2</b>	1269	3438	<b>2,70</b>	<b>1,71E-258</b>
ENSG00000113361	<b>CDH6</b>	291	3095	<b>10,25</b>	<b>1,76E-253</b>
ENSG00000114982	<b>KANSL3</b>	2978	6209	<b>2,08</b>	<b>7,31E-249</b>
ENSG00000174721	<b>FGFBP3</b>	12568	4674	<b>-2,68</b>	<b>4,87E-247</b>
ENSG00000092969	<b>TGFB2</b>	561	1603	<b>2,85</b>	<b>3,74E-242</b>
ENSG00000129474	<b>AJUBA</b>	14972	23333	<b>1,56</b>	<b>3,72E-237</b>
ENSG00000148180	<b>GSN</b>	831	4339	<b>5,15</b>	<b>8,42E-232</b>
ENSG00000103260	<b>METRN</b>	2953	11131	<b>3,74</b>	<b>1,11E-231</b>
ENSG00000168309	<b>FAM107A</b>	507	69	<b>-7,13</b>	<b>1,80E-231</b>
ENSG00000170542	<b>SERPINB9</b>	27	274	<b>9,84</b>	<b>5,19E-230</b>
ENSG00000120068	<b>HOXB8</b>	1712	6634	<b>3,85</b>	<b>3,36E-225</b>

**Supplementary Table S7. SNPs in *NRCAM* significantly associated with ASD**

db SNP id	Coordinates	Allele A/B	Gene location	Ref.
rs1859767	chr7:108448129	A/T	Intron 1	Marui et al., 2009
rs1034825	chr7:108443311	T/C	Intron 1	Marui et al., 2009
rs2300045	chr7:108423044	T/C	Intron 1	Marui et al., 2009
rs2300043	chr7:108422901	T/C	Intron 1	Marui et al., 2009
rs722519	chr7:108386349	A/G	Intron 2	Bonora et al., 2005
rs1269655	chr7:108262189	G/C	Intron 3	Marui et al., 2009
rs1269642	chr7:108247285	T/G	Intron 3	Marui et al., 2009
rs1269622	chr7:108213143	A/G	Intron 4	Bonora et al., 2005
rs1269621	chr7:108209714	A/G	Intron 4	Marui et al., 2009

10

**Supplementary Table S8. List and characteristics of primers used in PCR quantification**

Primer name	Primer sequence (5' to 3')	Amplicon size (bp)	Efficiency
AR_RT_F AR_RT_R	CCTGGCTTCCGCAACTTACAC GGACTTGTGCATGCGGTACTC	168	1.87
FAM107A_RT_F FAM107A_RT_R	CCTCTCGGAGTCACCAGGA GCTTCCAGCTCCTCCTTCTT	148	2.05
GAPDH_RT_F GAPDH_RT_R	GAAGGTGAAGGTCGGAGTC TTGATGGCAACAATATCCACTT	96	1.96
NRCAM_RT_F NRCAM_RT_R	CTCAGCTTTCCCTTCGCTCATGAT CCAAGCTTTTCCTGGACCCGTAAT	120	1.91
RGC32_RT_F RGC32_RT_R	TCTCTCTGCCACTGTCACTC CCCTCTGGCAGCAGATTCTA	195	1.93
RNF144B_RT_F RNF144B_RT_R	GACAGTCAGCCTATTGTCCTGC TCATTGCGTTCGATATAAACCCG	110	1.96
TRIM22_RT_F TRIM22_RT_R	AGGCAGTGCCACTTCGAATGTT ATTGGCAGCCGAAGACACCAAA	125	2
NRCAM_ChIP_F NRCAM_ChIP_R	GAGAGCACTGCTCAGACTTCATGT CAGCATTTCCTGGAAGCTGAATGAC	127	nd

11

**Supplementary Table S9a.**

**List of genes mutated in at least one ASD/ID patient used for the current study.**

This dataset contains **651** genes. We compiled a combination of gene lists already published<sup>16-18</sup>.

12

Gene name
ABCD1
ACSL4
ADSL
AFF2
AGA
AGTR2
AHI1
AIPL1
ALDH5A1
ALDH7A1
ALG12
ALG3
ALG6
ALG8
ALG9
ANKH
AP1S2
AP4E1
AP4M1
ARFGEF2
ARG1
ARHGEF6
ARHGEF9
ARL13B
ARL6
ARX
ASPM
ATP6AP2
ATP6V0A2
ATP7A
ATR
ATRX
AVPR2
BBS1
BBS10

13

**Supplementary Table S9b.**

**List of genes differentially expressed in autistic patients data available from microarray performed in blood, brain and lymphoblastoid cell lines.**

This dataset contains 1 660 genes collected from microarray studies performed in blood, brain and lymphoblastoid cell lines <sup>19-24</sup>.

14

Gene name
AAA1
AASDH
ABCA1
ABCB9
ABCC5
ABCC8
ABCG4
ABHD3
ABLIM2
ABTB1
ACADSB
ACHE
ACIN1
ACOT7
ACSBG1
ACSL4
ACSL5
ACTA1
ACTL6B
ACTN1
ACTR1B
ACY3
ADAMTS9
ADARB1
ADCYAP1
ADI1
ADM
ADORA2B
ADPRHL1
AEBP1
AEN
AES
AGA
AGPAT9

15

**Supplementary Table S9c.**

**List of genes associated with ASD (candidate genes association study from AutismKB) used for the current study.**

This dataset contains 177 genes. Autismkb collected all the candidate genes association studies in the dataset, including both family-based association studies and case-control based association studies.

16

<b>Gene name</b>
AANAT
ABI2
ACCN1
ADA
ADORA2A
AHI1
APC
APOE
AR
ARNT2
ASMT
ATP10A
ATP2B2
AVPR1A
BAIAP2
BDNF
BLMH
C4B
CACNA1G
CACNB1
CCL20
CCL25
CDH6
CNIH3
CNTFR
CNTNAP2
COMT
COPG2
CREBBP
CRHR1
CYP11B1
DAPPER1
DDC

17

**Supplementary Table S9d. List of genes associated with ASD (GWAS from AutismKB) used for the current study.**

This dataset contains **132** genes. Autismkb collected all the genome-wide association studies in the dataset, including both family-based studies and case-control based studies.

18

<b>Gene name</b>
ABCC1
ADRA1B
AHCYL2
APCDD1
ARHGEF26
C10orf11
C16orf68
CACNA1I
CCDC148
CCDC64
CDH10
CDH17
CDH22
CDH8
CDH9
CDK14
CENPC1
CHRM5
CNGA3
CNTNAP2
CPA1
CPA2
CTNNA2
CTNNA3
CUX1
DAB1
DCC
DHRS3
DMD
DNAJC10
DNER
DOCK4
DRD3
EDEM1

19

20



### **III. Discussion and perspectives**

A discussion has already been done in the above-cited paper but some points were not approached due to a lack of space. A more extensive discussion concerning this paper results will be done here, without detailing again the one already addressed.

#### **1. The case of testosterone-only regulated genes**

For this study, we focused on DHT because it is the most powerful natural ligand for the AR. We performed several replicates but we realized that the number of genes regulated by DHT never achieved the one we obtained with the unique testosterone treatment we did. These testosterone-regulated genes deserve our attention. According to our transcriptomic results, 152 genes are regulated after a testosterone 10nM treatment and 228 after a testosterone 100nM treatment (fold change > |1.2|, adjusted p-value < 0.05). About 46% of these testosterone-regulated genes are also regulated by DHT. The other genes appear regulated only with testosterone revealing a possible ligand-specific gene expression regulation by the AR in hNSCs. The ChIP-seq data are consistent with these observations since more genomic sequences are bound by AR after testosterone treatment (with 100nM, 10,276 sequences corresponding to 8,442 genes) than after DHT treatment (with 10 and 100nM, 1,513 sequences corresponding to 1,475 genes). Moreover, an *in vitro* study showed that DHT and testosterone could trigger differential gene regulation by allowing AR to bind different selective ARE (Hsiao *et al.* 2000). At the organ level, it is also known that DHT and testosterone could lead to different responses even if they bind and activate the same receptor: testosterone is the most important androgen in muscle (where there is no 5- $\alpha$ -reductase to convert it into DHT) and DHT is the main actor in genitals during development. In our cellular model, it is clear that testosterone and DHT regulate overlapping but different sets of genes. Interestingly, many testosterone-regulated genes are implicated in brain functioning, some being involved in neuron-glia communication within the cerebellum (*e.g DNER*), playing a role in neurite outgrowth (*e.g MAP1B, RELN, DCX, DCC*) or in neuronal migration (*e.g DNER, RELN, DCX*).



Moreover, several testosterone-regulated genes we identified are mutated in brain disorders. The first example is the *RELN* gene encoding a glycoprotein which determine the final location of migrating neurons in the developing brain and is also essential for synaptic physiology in adulthood (Beffert *et al.* 2005; Weeber *et al.* 2002). This gene is already known to be regulated by testosterone. Recent experimental data suggests that testosterone is able to promote *Reln* demethylation leading to its possible upregulation, although no direct evidence of hormonal influence on *Reln* promoter methylation was obtained (da Silva *et al.* 2015). Moreover, it has been reported that adult male heterozygous *reeler* mice (rl/+), haploinsufficient for reelin, have reduced numbers of Purkinje cells, in comparison to female rl/+ mice and wild-type mice of either sex. In this model, it seems that this male-selective Purkinje cells loss could be the result of an interaction between reelin deficiency and steroids during cerebellar development (Biamonte *et al.* 2009; Hadj-Sahraoui *et al.* 1996). Therefore, these mice represent an interesting experimental model to study interactions between genes and sexual hormones during brain development. In addition, *RELN* has been found altered in patients with different forms of NDD (Folsom & Fatemi 2013). *RELN* maps to the *AUTS1* locus on 7q22, and at this time over 40 distinct rare variants that would alter the protein sequence have been identified, four of which occurred *de novo* (Lammert & Howell, 2016). Rare, homozygous, null mutations in *RELN* lead to lissencephaly with cerebellar hypoplasia accompanied by developmental delay and epilepsy (Hong *et al.* 2000).

A second example is the *DCX* gene coding for a protein involved in microtubule stabilization (Shmueli *et al.* 2001). There are several *DCX*-related disorders resulting from aberrant migration of neurons during development of the cortex that have been reported. They include lissencephaly (or “smooth brain” lacking of brain gyri and sulci) and subcortical band heterotopia (also known as “double cortex syndrome” because of the appearance of an extra layer of neurons under the normal gray matter of brain cortex). Lissencephaly occurs usually in males while subcortical band heterotopia is primarily found in females. This difference of phenotype between sexes could be due to the fact that: 1) the *DCX* gene is located on the X-chromosome and 2) *DCX* is regulated by testosterone.

The last example is the *DCC* gene encoding a functional receptor for netrin. *DCC* is known as an axon guidance receptor when stimulated by netrin-1 (Li *et al.* 2004). Heterozygous loss of function mutations in *DCC* result in congenital mirror movements (Srour *et al.* 2010). Marsh *et al.* described *DCC* mutations in four multigenerational families

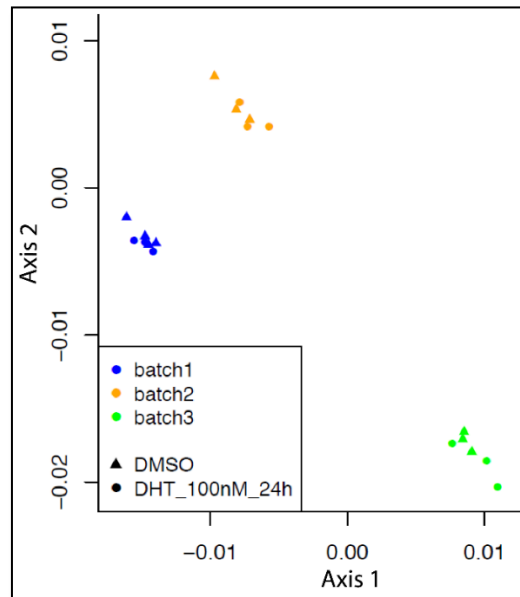
with individuals presenting with mirror movements and/or isolated agenesis of the corpus callosum (ACC). In parallel, a review of the phenotype of individuals carrying *DCC* mutations showed that, in individuals with truncating variants, males exhibit mirror movements more frequently than females while ACC is more often present in females. My results showed an upregulation of *DCC* by testosterone in hNSCs. This suggests that this upregulation during development in males could, in this specific case, protect them against ACC that may occur when the amount of *DCC* mRNA/protein falls below a threshold, which would occur more frequently in females. In this way, the phenotypic variability could be determined by genetic, epigenetic and/or environmental factors among which hormonal context (**Annexe 2**) (Marsh *et al.*, in revision in Nat. Genet.).

In these three examples, the modulation of the expression of these genes by the testosterone is linked to the variable expressivity of the genetic mutations since the phenotype is different between males and females carrying the same genetic anomaly. Indeed, it is clear that testosterone-regulated genes could be of interest to explain some sex-specific phenotype variabilities occurring in several NDD and/or brain malformations. Saying that, it is now clear that testosterone treatment replicates in hNSCs should have been done in order to go further in the characterization of these testosterone-only regulated genes. Moreover, it would also be interesting to understand why they are not regulated by DHT and what could be the consequences during brain development. To summarize, we used multiple androgens (testosterone, DHT and R1881) at several doses to obtain a global view of androgens regulation in hNSCs and simplify our gene selection. However, this strategy gave us much more complex results than we were expected.

## **2. Cellular model variability**

The cellular model (hNSCs) we used is particularly relevant for this study because: 1) they model fetal brain; 2) they express AR at the mRNA and protein level at higher level than differentiated neurons, which could indicate that androgens action is more important in these precursor cells than in mature neurons and 3) one can more readily obtain the cell number necessary for RNA-seq and ChIP-seq experiments. Nevertheless, their main drawback is their variability. Even though they are defined as a “stable population of self-renewing cells” (which can be checked by immunofluorescence quality controls), they tend to be in an unstable state and fluctuate until they receive a clear differentiation signal.

Since no other XY hNSCs were available to perform biological replicates with various individuals, we decided to do treatment replicates with the unique cell line at our disposal. Finally, three batches of hNSCs were used for DHT 100nM 24 hours treatment and the RNA-seq results clearly reveal the unstable state we expected (**Figure 24**). However, a specific bioinformatics analysis taking into account this batch variation was carried out and the list of genes obtained was, by this way, very robust.



**Figure 24. Factorial correspondence analysis.** The correspondence analysis was calculated on variance stabilized data calculated by the method proposed by (Anders & Huber, 2010). The first factorial plan is represented on this figure.

### 3. DHT/androgens-regulated genes and their expression within the brain

Concerning the choice of the androgen to use, we preferred DHT because this natural metabolite of testosterone is known to activate in a more potent manner the AR (higher affinity and lower dissociation with the AR, higher stability of the DHT-AR-DNA complex). Moreover, the enzyme converting testosterone to DHT (*SRD5A1*) is well expressed in hNSCs suggesting that this metabolite could be the predominant androgen in this model of developing brain. Finally, our microarray results performed on differentiated neurons confirmed that DHT regulates more genes than testosterone (**Table 6**, page 180).

The DHT physiological concentration is 10nM and we used a 10-fold higher concentration to potentiate the gene expression regulation. Surprisingly, only low changes in few genes were identified in hNSCs after DHT 100nM 24 hours treatment. We could now think to use different DHT doses but also several time points to find the optimal conditions in order to obtain an exhaustive list of DHT-regulated genes in hNSCs. Preliminary results revealed a higher number of genes regulated with DHT physiological concentration and shorter time point but these results need to be replicated and studied deeper in the future.

Another important point is that DHT appears to regulate different genes in hNSCs and in differentiated neurons since there were only four genes found regulated in the hNSCs RNA-seq and in the neurons microarray (*CYP11B1*, *ELK3*, *DTX3L* and *OLFML2B*). We did not study deeply these genes because the microarray experiment was only performed once. In addition, the regulation of these genes display only a borderline significance (adjusted p-value between 0.012 and 0.046) revealing that they could be false positive.

Among the 77 DHT-regulated genes in hNSCs, we validated the top seven genes (*RGCC*, *RNF144B*, *NRCAM*, *TRIM22*, *FAM107A*, *LAMA2* and *IGFBP5*). These genes were selected because they were regulated by the different androgens (DHT, testosterone and R1881) with the most significant adjusted p-value. In order to go further, we would like to assess the potential implication of these seven best genes during human brain development. I looked at the Atlas of the Developing Human Brain (Brainspan accessible at [www.brainspan.org](http://www.brainspan.org)) giving RNA-seq and exon microarray data profiling in up to sixteen cortical and subcortical structures of developing human brain (at 15 to 21 post-conceptual weeks). While *TRIM22*, *RNF144B* and *LAMA2* appear to be lowly expressed ( $-\log_2$  RPKM  $< 4$ ), *NRCAM*, *IGFBP5*, *RGCC* and *FAM107A* seem highly expressed ( $-\log_2$  RPKM  $> 6.8$ ), as it has also been shown in mouse brain. To complete these data, we are currently checking their expression and co-expression with AR (at the mRNA and protein level respectively by *in situ* hybridization or ISH and immunohistochemistry) on human fetal brain slides (in collaboration with Dr F. Guimiot at Hôpital Robert Debré, Paris). In a first trial, I chose to look at the period of the first testosterone surge occurring during *in utero* development in male fetuses (the beginning of the second trimester of pregnancy). Male fetuses (n = 3, between 17 and 20 weeks of amenorrhea) and one female fetus (16 weeks of amenorrhea) collected after late spontaneous abortions were used to optimize ISH protocol on brainstem, cerebellum, frontal cortex and hippocampus slides. The preliminary results in ISH showed

low signals with *AR*, *FAM107A* and *RNF144B* mRNA sense probes. In parallel, the first results concerning AR immunohistochemistry showed that AR protein is present in choroid plexus, tanycytes from ventricle and layers I, II and II of the cerebral cortex. These first results are encouraging and need to be replicated and refined after protocols optimization.

## 4. Pathway analysis

### i. Cell proliferation and survival

Due to the low number of genes, the pathway analysis of the androgens-regulated genes was unfruitful. However, the bibliographic study of the top seven genes revealed that they are involved in interesting pathways playing a role in ID and/or ASD physiopathology. As an example, several genes have been involved in cell proliferation and survival (regulation of cell growth, mitosis or apoptosis) such as *RGCC*, *RNF144B*, *TRIM22*, *FAM107A* and *IGFBP5*. *RGCC*, also known as *RGC32*, plays a role in cell cycle activation since it is a cell cycle regulatory factor that mediates cell proliferation (Badea *et al.* 2002). *FAM107A*, also known as *TU3A* or *DRR1*, is a stress-responding protein involved in cellular survival (even if contradictory effects are documented) (Asano *et al.* 2010; Liu *et al.* 2009). *RNF144B* (or *IBRDC2*) and *TRIM22* (or *STATF50*) are both E3 ubiquitin ligases (Duan *et al.* 2008; Ng *et al.* 2003). While *RNF144B* regulates the level of BAX (a pro-apoptotic protein) and protects cells from unprompted BAX activation (Benard *et al.* 2010), *TRIM22* is a p53 target gene that plays an important role in cell proliferation (Obad *et al.* 2004). Finally *IGFBP5*, an IGF activity regulator protein, triggers cell cycle arrest and apoptosis (with increased expression of BAX and decreased expression of the anti-apoptotic BCL) (Butt *et al.* 2003). Androgens seem therefore to be implicated in the regulation of cell number by both increasing cell proliferation and protecting against cell death. Interestingly, previous observations showed that regulation of cell death by androgens might play a role in establishing sexual dimorphism of the brain in mice, some regions having more cells in one sex than in the other (Forger 2009). Here, we may have identified genes regulated by androgens implicated in this phenomenon: *RGCC* and *FAM107A* (confirming by cellular proliferation experiment) but also *RNF144B*, *TRIM22* and *IGFBP5*. In parallel, it is known that autistic patients present an abnormal kinetics of brain growth with macrocephaly during childhood. This early brain overgrowth seems to be sex-dependent since it has been observed more in boys (especially with developmental regression)

than in girls (Nordahl *et al.* 2011). Some exceptions do exist since certain ID and autistic patients present with microcephaly. In any case, autism and more generally NDD, are very often associated with abnormal brain growth (either micro or macrocephaly), suggesting a misregulation of neuronal precursors proliferation during brain development.

## ii. Other pathways related to brain development and ID/ASD

In another hand, some of the genes regulated by androgens present more obvious implications in brain development by regulating diverse crucial cellular processes. *NRCAM* is the best example since this neural cell adhesion molecule is involved in neuronal migration, axonal guidance, neurite outgrowth, dendritic spine remodeling, myelination or node of Ranvier formation among others (Custer *et al.* 2003; Demyanenko *et al.* 2011; Treubert & Brümmendorf, 1998). Another example is *FAM107A*, encoding a protein located at the synapse (mainly presynaptic synapses). This protein is also an actin binding protein promoting bundling and stabilization of actin filaments which impacts actin-dependent neurite outgrowth (Schmidt *et al.* 2011). In the animal model, there are evidences showing that alteration of these proteins level mediated functions relevant to symptoms observed in autism: *Nrcam*-null mice show impaired sociability, learning and memory (Moy *et al.* 2009) and virus-induced hippocampal *Fam107a* upregulation in mice improves spatial memory (Schmidt *et al.* 2011) and, in lateral septum, increases sociability (Masana *et al.* 2014).

Interestingly, *NRCAM* is known to be regulated by the Wnt/ $\beta$ -catenin pathway, an essential signaling for normal embryonic development of the central nervous system. The  $\beta$ -catenin is a protein involved in both cell-cell adhesion and transcription. The membrane form is an adhesion molecule acting as an adaptor protein which anchors the actin cytoskeleton to coordinate migration, adhesion and differentiation of adjacent cells. The cytosolic  $\beta$ -catenin is a transcription factor activated by Wnt which stimulates Wnt target genes expression (Williams & Casanova, 2011). In parallel, several Wnt/ $\beta$ -catenin target genes have been described to be autism susceptibility factors. Polymorphisms in *WNT2* and *EN2* genes for instance have been associated with autism and *de novo* pathogenic variations were also identified in *WNT2* and *CTNNB1* genes (coding for the  $\beta$ -catenin) in patients with autism (De Ferrari & Moon, 2006; O'Roak *et al.* 2012). Moreover, *in utero* exposure to teratogens, such as valproic acid, causes higher incidence of autism in the offspring and valproic acid has been shown to activate Wnt-dependent gene expression.

Several studies demonstrate that AR and  $\beta$ -catenin interact together in the presence of DHT. Liganded AR shuttles  $\beta$ -catenin to the nucleus and this nuclear interaction may modulate their transcriptional activity (Pawlowski *et al.* 2002; Song *et al.* 2003; Yang *et al.* 2002). Nevertheless, we did not observe any translocation of  $\beta$ -catenin into the nucleus after DHT treatment in hNSCs, suggesting that, in this cell type, liganded AR does not shuttle  $\beta$ -catenin to the nucleus. However, the interaction between AR and  $\beta$ -catenin has to be tested in hNSCs. In addition, there is only a small overlap between the genes regulated by androgens in hNSCs and genes known to be  $\beta$ -catenin targets (Nusse, 2009): only two genes (*NRCAM* and *CYR61*, a gene encoding a secreted protein also involved in cellular adhesion) were found to be common.

### **5. Link between genes regulated by androgens and genes involved in ID and/or ASD**

The 77 androgens-regulated genes found by RNA-seq appear slightly enriched in genes mutated in monogenic form of ID and/or ASD (hypergeometric law  $p = 0.0287$ ). One of these genes is *LAMA2* in which mutations cause congenital muscular dystrophy associated in less than 10% of the cases with ID (Xiong *et al.* 2015). Moreover, recurrent *de novo* events have been found in *LAMA2* in schizophrenic patients (Xu *et al.* 2012). No pathogenic variants in the six others top-regulated genes have been described in monogenic forms of ID and/or ASD until now (neither in PubMed nor DECIPHER databases). However, numerous deletions encompassing *NRCAM* are reported in individuals with autism, ID or global developmental delay in the DECIPHER database (database of genomic variations and phenotypes in humans using ensembl resources, reporting CNV and SNV identified in more than 18,000 children affected by a potential genetic condition including 14,000 children with severe undiagnosed developmental disorders from the Deciphering Developmental Disorders project (DDD)) and by the team of S. Jacquemont (Jacquemont *et al.* 2014). To be as updated as possible, I submitted the seven top-regulated genes on GeneMatcher which is "a freely accessible website designed to enable connections between clinicians and researchers from around the world who share an interest in the same gene" (Sobreira *et al.* 2015) (accessible at [www.genematcher.org](http://www.genematcher.org)). Up to now, I got an answer from several researchers concerning variations in *NRCAM* (at least two rare variants in patients with neurological condition) but none of them were functionally validated and

their pathogenicity is still uncertain. Two other genes were found in GeneMatcher (*IGFBP5*, one hit and *LAMA2*, three hits), but I did not receive any details concerning the type of the anomaly and the patient phenotype. Finally, *RGCC*, *RNF144B*, *TRIM22* and *FAM107A* did not get any hit.

As the highly penetrant monogenic forms of ID and/or ASD do not present any male excess, we do not expect to find pathogenic variants in these androgens-regulated genes. However, in the context of the oligogenic/multifactorial model of NDD, some risk factors, less penetrant (with phenotype variability between sexes and high sex bias), might be expected instead. In this model, some additional factors (genetic or environmental) would be required to develop the disease (such as fetal androgens). Up to date, none of the androgens-regulated genes have been identified as risk factor for autism. Nevertheless, androgens could regulate genes differentially expressed in autistic patients (because of other genetic and/or environmental factors) and this male-specific androgen regulation could exacerbate this abnormal gene expression and leads to the increased susceptibility of ASD in males.

## **6. Non-genomic actions of androgens in hNSCs**

In our study, we looked at the best-known androgens' effects namely the transcription regulation by the AR and its co-regulators binding within the promoter or enhancers of target genes. This corresponds to the direct genomic effect of AR. However, there are other androgens' effects involving a non-genomic pathway independently of the nuclear AR. Indeed, after androgens' binding to the cytosolic AR, some signaling pathways can already be modulated (such as MAPK or PI3K pathways). Moreover, androgens might bind other membrane receptors (including the AR itself) and recruit the intracellular calcium or activate the PKA and PKC. To be as exhaustive as possible, it will be important to characterize this non-genomic regulation of androgens in hNSCs. Because this kind of regulation could be faster and take place within the few seconds after androgens exposition, we could first have a look to shorter treatment timing (around 30 minutes) and do protein expression profiling in complement to transcriptomic analysis. In parallel, we could also analyze the AR co-regulators in order to have a global view of AR roles in hNSCs and maybe find more connections with genes/proteins and pathways involved in brain development and/or NDD.





# General discussion

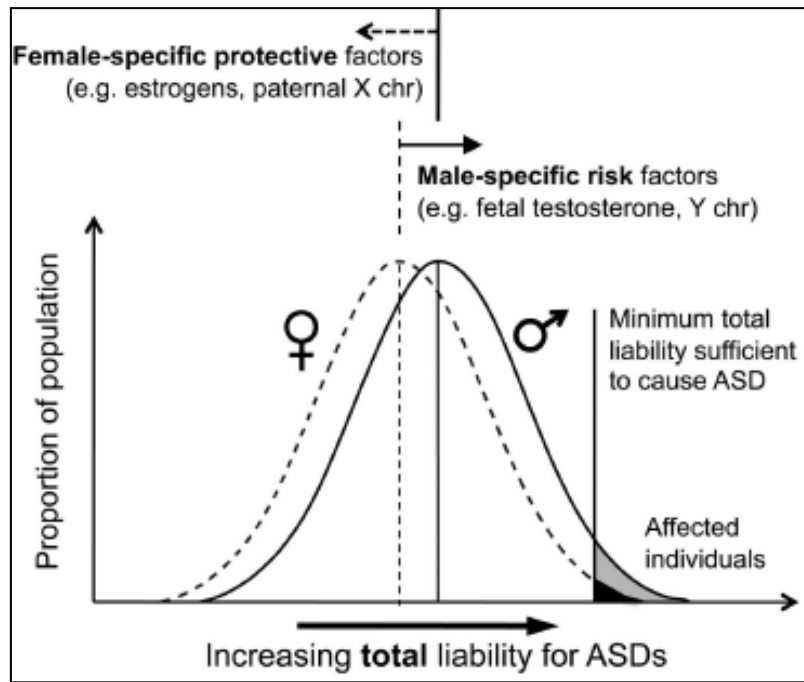
---



## **I. Monogenic versus multifactorial**

It is now well accepted that most of the NDD present a male sex bias. Concerning ID (whose predominant etiology model appears monogenic but with a significant proportion that may be oligogenic or polygenic), a part of the explanation comes from the causative mutations found in genes on the X chromosome. However, as we saw in the chapter 1 of this work and in accordance with the literature, this could account for only one third of the sex bias observed in ID (Mandel & Chelly, 2004). It is even less for ASD that is known to be, in most of the cases, a multifactorial condition with genetic and environmental factors involved. Consequently, the relationship between this male excess and ASD (but also, to a lesser extent, ID) has attracted a variety of research ranging from clinical and etiological to neurobiological studies.

Although genetic processes are involved, they do not provide an explanation for the sex difference in NDD, especially in autism. In fact, while the first twin studies suggested only minimal non-shared environmental contribution to ASD (90% heritability), a recent study estimates that genetic factors explain only half of the liability for autism (50% heritability). In other words, genetic and non-genetic influences appear similarly important for ASD, suggesting that interactions between multiple genes, epigenetic factors and exposure to environmental modifiers may contribute to variable expression of autism-related traits (Sandin *et al.* 2014). Indeed, ASD are complex conditions regrouping monogenic, oligo/polygenic and multifactorial forms. The multifactorial model was the original model proposed to explain ASD etiology. Since that time, studies on sex bias in ASD add some complexity to the original multifactorial model with a sex-differential liability. Evidence suggests that the liability distribution for females could be shifted away from the threshold required for reaching affected status by female-specific protective factors. On the contrary, the liability distribution for males could be shifted closer to the same threshold by male-specific risk factors. These factors can be either genetic (*e.g.* X chromosome gene protective effects) (Skuse, 2000) or environmental (*e.g.* prenatal hormones) (Auyeung *et al.* 2013; Baron-Cohen *et al.* 2015) (**Figure 25**).



**Figure 25. Multifactorial sex-differential liability model.**

Female-specific factors shift females' total liability distribution (including genetic, environmental, and other factors) away from and male-specific factors shift males' distribution toward a single threshold (Werling & Geschwind, 2013).

## **II. Link between genes and environment**

The multifactorial model suggests that genetic and environmental factors are closely entwined: increasing evidence demonstrates a complex interaction between sex hormones (mainly prenatal androgens) and genetic and epigenetic factors in the establishment of sex differences in brain and behavior. Sex hormones exert organizational (epigenetic) effects which, coupled with differential neural expression of genes on X and Y chromosomes, lead to sexually dimorphic brain development. Early prenatal development is a critical period during which these genetic and epigenetic mechanisms relevant to autism take place (Baron-Cohen *et al.* 2015; Parikshak *et al.* 2013). Therefore, an abnormal interplay between genes and sex hormones during brain development (including epigenetic) could explain the widespread differential susceptibility to pathology of males and females. Much work is needed to definitively identify the most critical players, but studies need to go beyond comparing patients at genetic or environmental levels alone.

My work on hNSCs, a cellular model of brain development, showed that androgens can regulate a small set of genes involved in crucial pathways for normal brain development (such as proliferation and cell survival that are essential for an optimal brain growth and organization). These androgens-regulated genes could be part of the above-mentioned critical players. By replicating this study in hNSCs collecting from many individuals, we could see if the genetic background influences androgens' response. These findings might help to draw a link between genetic and environmental factors during normal male brain development, which could then be useful to understand the role of each of these factors in ASD etiology. Moreover, we need to go further to elucidate the precise mechanisms by which slight changes in the expression of androgens-regulated genes, identified during my project, can modulate the ASD phenotype. For that, it would be informative to further consider the moderating effect of androgens in a model of autism with non-penetrant genetic alterations (such as patients' cells). One can think to the patients carrying *SHANK1* deletion or point mutation since the phenotype is different between males (autism) and females (anxiety and shyness). Another example presenting the same kind of phenotypic variability is the 16p11.2 CNV: individuals with the same 16p11.2 genetic risk factor can be profoundly autistic to essentially unaffected.

### **III. Clarifying recent findings and guiding future research directions**

Findings on sex bias in ID and ASD are complex and do not always relate to each other in a straightforward manner. Lai *et al.* summarized these findings in a four-level conceptual framework to simplify future research directions with four distinct but deeply interlinked questions (Lai, Lombardo, *et al.* 2015):

- “How are ID and ASD defined and diagnosed in males and females?”
- “What are the similarities and differences between males and females with ASD?”
- “How is the risk to develop ID and ASD linked to sex?”
- “What developmental mechanisms of ID and ASD are implicated by sexual differentiation?”

## 1. ID and ASD definition and diagnosis

“How are ID and ASD defined and diagnosed in males and females?” is of major importance since how these conditions are defined and identified substantially affects all aspects of our understanding. This implicitly asks the question “What are the similarities and differences between males and females with ASD?”. For autistic patients, females tend to be more severely affected, and to have cognitive abnormalities (such as ID) in a higher proportion than males and with more harmful genetic factors involved. Nevertheless, this increased severity in female should not be generalized. Indeed, recent investigations suggest that females diagnosed with autism tend to have different biological characteristics (such as brain abnormalities: a greater enlargement of the amygdala is observed among affected females than in affected males, hippocampus and cerebellar enlargement noting among affected males are not observed among the female, etc.) rather than be “more severe” (Kirkovski *et al.* 2013; Lai, Baron-Cohen, *et al.* 2015). Moreover, participant sampling is very often biased towards males which decreases the data available for females. Therefore, the diagnostic tools developed until now could be inappropriate for females which could explain their underdiagnosis. In the future, it is necessary to use comparable male/female group-size study designs and to develop new ways to objectively qualify and quantify ASD presentations (by developing diagnostic criteria sensitive to females) in order to 1) assist identification in females and 2) define diagnostic features that are dependent or independent of sex.

## 2. “How is the risk to develop ID and ASD linked to sex?”

In other words, “Is the genetic and environmental factors contribution different between sexes?”. It seems that most of the factors (genetic and environmental) identified until now for ASD have no sex-specific effects as they could cause ASD in both sexes. However, in some cases, the phenotype severity could be different between sexes for a given factor such as *SHANK1* anomalies which lead to autism in boys but result in anxiety and shyness in girls. Nevertheless, there are evidences of quantitative sex differences. It has been shown that females affected by ASD present an excess of deleterious autosomal CNV and private deleterious SNV compared to affected males. Although ASD females showed higher mutational burden and lower cognition, the excess mutational burden remained,

even after adjustment for those cognitive differences. These results strongly suggest that females have an increased etiological burden (Jacquemont *et al.* 2014). One can therefore postulate that there are monogenic forms of ASD with both males and females affected (and that do not present a male excess) and more complex forms in which there are more males affected than females. In a near future, it should be interesting to use data from multiplex families to compare male-only, sex-mixed, and female-only families in order to evaluate the implication of each factor type (genetic *versus* environmental).

### **3. A sex-specific developmental trajectory**

“What developmental mechanisms of ID and ASD are implicated by sexual differentiation?”. This last question concerns the specific etiological and developmental mechanisms that are different between males and females. More importantly, it implicitly asks whether the underlying mechanisms could influence the susceptibility to develop NDD. Some longitudinal epidemiological studies are needed to evaluate if the developmental trajectories differ between males and females (be it physical, neuroanatomical, cognitive, or behavioral) and test whether there are crosstalks between factors associated with normative sexual differentiation (genetic, epigenetic, and prenatal environmental factors) and molecular/cellular mechanisms related to ASD. My work analyzed the gene regulation by a male-specific environmental factor (the androgens’ exposition) in a human cellular model (hNSCs) and revealed pathways and cellular processes in which these androgens-regulated genes are involved. Even if further work is needed, this study focusing on androgens’ roles in normal male brain development partially answers the question.

## **IV. Potential implications in therapeutics and opening to other diseases**

A better comprehension of the mechanisms underlying risk and resilience to ASD between the sexes may be of fundamental importance to identify innovative therapeutic approaches for these conditions. Consequently, a better understanding of the interplay between genes and sex hormones in ASD is needed and relevant signaling pathways and critical windows of susceptibility have to be identified. Androgens-regulated genes during



human brain development may be some of these innovative therapeutic targets/pathways (for example, the androgens-regulated genes we identified in hNSCs). However, animal models are essential to confirm these regulations and to address the underlying mechanisms *in vivo*. Unfortunately, there are still few studies addressing the central and peripheral hormonal profile and the status of brain sex steroid-related pathways in rodent genetic model for ASD. Moreover, in rodents, most of the behavior studies have been performed on males only and the comparison male-female has never been performed for these monogenic forms of ASD. This kind of comparison would be even more informative in animal model carrying genetic predisposing factors for ASD such as *SHANK1* anomalies or 16p11.2 CNV already evoked earlier. In general, literature strongly argues for the need of further studies aimed at the evaluation of the advantages and disadvantages conferred by biological sex under pathological conditions of neurodevelopmental origin (for example, in animal models available for ID and ASD) (de la Torre-Ubieta *et al.* 2016; Romano *et al.* 2016). Much effort will be expected to unravel the potential therapeutic efficacy of drugs targeting sex hormones-related signaling pathways in ASD of well-established genetic origin.

The research challenges are considerable but the potential rewards are greater because they will shed light on some of the key causal processes involved in the etiology of ID and ASD. Moreover, some of the underlying hormonal mechanisms that has been (and will be) identified in this project are relevant to a broader category of disorders presenting with sex bias (and for which links with sex steroids have been documented, **Table 7**), whether it is neurodevelopmental (*i.e.* ADHD, Tourette syndrome, Hirschsprung disease...) or neurodegenerative (*i.e.* Parkinson's disease, amyotrophic lateral sclerosis...).

	Condition	Sex ratio (male:female)	Link with sex steroids
<b>Neurodevelopmental</b>	<b>ADHD</b>	2:1 to 10:1	<p>In human:</p> <ul style="list-style-type: none"> <li>- Association between the CAG and GGC polymorphisms of the AR and ADHD (Comings <i>et al.</i> 1999)</li> <li>- Fetal testosterone possibly implicated based the hypermasculinization of ADHD patients (McFadden <i>et al.</i> 2005)</li> </ul> <p>In animal model:</p> <ul style="list-style-type: none"> <li>- Early androgens exposure enhances expression of ADHD-like phenotypes (King <i>et al.</i> 2000)</li> </ul>
	<b>Tourette Syndrome</b>	4:1 to 6:1	<p>In human:</p> <ul style="list-style-type: none"> <li>- Tics increase in intensity during puberty</li> <li>- Fetal testosterone possibly implicated based on study of play preferences and spatial skills (Alexander &amp; Peterson, 2004)</li> <li>- Anti-androgen treatment modestly improves symptoms (Peterson <i>et al.</i> 1998)</li> </ul>
	<b>Hirschsprung's disease</b>	8:1	In human, no deleterious effects on testosterone on enteric nervous system development (Vohra <i>et al.</i> 2007) but no more studies up to now.
<b>Neurodegenerative</b>	<b>Parkinson's disease</b>	1.5:1	<p>In human:</p> <ul style="list-style-type: none"> <li>- Low testosterone blood level results in an increased susceptibility to the disease in men (Nitkowska <i>et al.</i> 2015)</li> <li>- Supplemental testosterone failed to show significant improvement in motor or non-motor outcome (Smith &amp; Dahodwala, 2014)</li> </ul>
	<b>Amyotrophic lateral sclerosis</b>	1.3:1 to 2:1	<p>In human:</p> <ul style="list-style-type: none"> <li>- May be due to a loss of AR (Weiner, 1980)</li> <li>- AR mRNA reduced in affected motoneurons (Nakamura <i>et al.</i> 1997)</li> <li>- Serum level of free testosterone reduced in patients (Militello <i>et al.</i> 2002)</li> <li>- Association between CAG and GGC polymorphisms of the AR and the disease (Garofalo <i>et al.</i> 1993)</li> </ul>

**Table 7.** Evidence for the effect of sex steroids in male biased neuropsychiatric conditions.



# Annexes

---



## **Annexe 1.**

**Paper:** Efficient strategy for the molecular diagnosis of intellectual disability using targeted high throughput sequencing

Redin C, Gérard B, Lauer J, Herenger Y, Muller J, **Quartier A**, *et al.* J Med Genet. 2014.

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

Contribution: Minor. Performed functional validations of potential causative mutations found in *FMRI* (two families) and *PQBPI* (one family).

*NB:* According to their size, the supplementary figures and tables are not enclosed.





OPEN ACCESS

ORIGINAL ARTICLE

# Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing

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

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2014-102554>).

For numbered affiliations see end of article.

## Correspondence to

Amélie Piton, Translational Medicine & Neurogenetic, 67 400 Illkirch, France; [amelie.piton@igbmc.fr](mailto:amelie.piton@igbmc.fr)  
Jean-Louis Mandel, Translational Medicine & Neurogenetic, 67 400 Illkirch, France; [jlmandel@igbmc.fr](mailto:jlmandel@igbmc.fr)

Received 1 June 2014  
Revised 31 July 2014  
Accepted 3 August 2014

## ABSTRACT

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

**Methods** We report the alternative strategy of targeted high-throughput sequencing of 217 genes in which mutations had been reported in patients with ID or autism as the major clinical concern. We analysed 106 patients with ID of unknown aetiology following array-CGH analysis and other genetic investigations. Ninety per cent of these patients were males, and 75% sporadic cases.

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

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

## INTRODUCTION

Intellectual disability (ID) is a common neurodevelopmental disorder reported in 1.5–2% of children and adolescents.<sup>1,2</sup> ID is defined by significant limitations in both intellectual functioning and adaptive behaviour with onset before the age of 18. Different classes of ID are conventionally defined upon IQ values (severe or profound, <35; moderate, 35–49 and mild, 50–70). However, in routine genetic practice, clinical assessment mainly based on records of developmental history, speech acquisition and patients' autonomy is used for classification in such subcategories.

Causes of ID can be environmental, genetic or multifactorial. Single genetic events are thought to account for a majority of cases, varying from large chromosomal anomalies or copy number variants (CNVs) affecting several genes to point mutations in single genes. These latter monogenic forms are characterised by an extreme genetic heterogeneity, with a hundred genes described as implicated in X-linked ID (XLID), and more associated to autosomal-recessive or autosomal-dominant forms. Altogether there are more than 500 genes proposed to cause ID with high penetrance when mutated<sup>3–9</sup> underlying a phenotypic heterogeneity of the same extent in both severity and associated symptoms. This genetic heterogeneity has long limited the diagnostic offer for patients and families, which

**To cite:** Redin C, Gérard B, Lauer J, et al. *J Med Genet* Published Online First: [please include Day Month Year] doi:10.1136/jmedgenet-2014-102554



was often restricted to fragile-X (MIM 300624) testing, array-CGH (comparative genomic hybridization) analysis and generic metabolic tests (see online supplementary figure S1). It may be complemented by sequencing a few genes associated to a specific syndrome evoked by patients' phenotype, yet the diagnostic yield remains low (1–2% for the recurrent fragile-X mutation; 10–15% for array-CGH and chromosomal analyses, higher in highly syndromic patients).<sup>10–12</sup> A majority of patients remain therefore without molecular diagnosis, while it is of crucial importance for establishing recurrence risks and providing genetic counselling in the family. Moreover, such diagnosis often has direct consequences for the medical prognosis of patients or their optimised healthcare, and even (yet in still a minority of cases) can indicate specific therapeutic options.

To obviate this low diagnostic yield, we developed the simultaneous targeted sequencing of protein-coding exons of 217 genes associated with ID or autism spectrum disorders (ASDs) as primary clinically significant feature: 99 located on the X-chromosome, 118 on the autosomes. We report here the results of such strategy on a cohort of 106 ID patients with or without associated autistic-like features, negative for array-CGH, fragile-X and other specific genetic analyses. A causal mutation was detected in 25% of these patients, regardless the severity of their cognitive impairment. We illustrate cases in which the molecular diagnosis was immediately established, as well as other more complex situations. This highlights the challenge of interpreting variants generated by NGS technologies, already from targeted approaches restricted to a few hundred genes. This work demonstrates that a targeted sequencing approach is highly efficient for the diagnosis of ID, but also allows refining the clinical spectrum associated with mutations in certain genes, and confirming or questioning the involvement of other genes in cognitive disorders.

## METHODS

### Cohort of patients

DNA samples from 106 patients were addressed for testing through clinical geneticists from 16 public hospitals in France. Inclusion criteria for the patients were to be negative for the recurrent fragile-X mutation and for pathogenic CNVs via array-CGH testing, and availability of DNA samples from family members in order to conclude upon molecular findings. Additional specific genetic investigations had been performed on a majority of patients (on average: two genes tested per patient, [table 1](#)). Presence of multiple major congenital anomalies or suspected mitochondrial/peroxisomal disorders was an exclusion criterion. Clinical data were recorded following a standardised clinical record highlighting prenatal history, developmental history, neurological and behavioural disorders. ID severity was assessed by medical geneticists upon clinical evaluation and was not a discriminating inclusion criterion, although we encouraged inclusion of probands with moderate or severe ID. This study was approved by the local Ethics Committee of the Strasbourg University Hospital (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale (CCPRB)). For all patients, a written informed consent for genetic testing was obtained from their legal representative.

### Targeted genes and capture design

The 217 selected genes include 99 genes associated to XLID and 118 genes located on autosomes, implicated in dominant (45), recessive (66) or complex (7) forms of ID (see online supplementary table S8, further justification on gene selection is given in online supplementary methods). We targeted all

protein-coding exons of these genes, including 20 bp of intronic flanking sequences. The overall size of targeted regions is 1.034 Mbp. Corresponding 120 bp RNA baits were designed using SureDesign (<https://earray.chem.agilent.com/suredesign/>).

### Library preparation and sequencing

DNA samples were extracted from peripheral blood or saliva. Sequencing libraries were prepared as described previously,<sup>13</sup> performing individual in-solution SureSelect capture reaction for each DNA sample (Agilent, Santa Clara, California, USA). Paired-end sequencing (2×101-bp) was performed on an Illumina HiSeq 2000/2500, multiplexing up to 16 samples per sequencing lane.

### Bioinformatic pipeline and variant ranking

Read mapping, variant calling and annotation was performed as described previously.<sup>13</sup> Detected variants (short indels and single nucleotide variants (SNVs)) were ranked by VaRank (an in-house developed script), which incorporates the annotations retrieved by alamut-HT (putative effect on the protein, conservation scores, splice site predictions, allelic frequency in the 106 patients and in control cohorts such as Exome Variant Server (EVS) or 1000 genomes). Candidate variants were selected when harbouring a frequency compatible with the incidence of the disease: expectedly accounting for less than 0.1% of all ID cases, aka resulting in a disease frequency <0.002%.<sup>14</sup> Using the EVS population as a subset of the general population, candidate variants were thus retrieved when reported in EVS: with a minor allele frequency lower than 0.45% for variants in autosomal-recessive genes (ie, with a frequency of homozygotes <0.002%), in no more than one carrier for variants in autosomal-dominant genes or in no more than one male for variants in X-linked genes (as we cannot exclude, although it is unlikely, that a particular carrier from the general population might have mild ID). We concomitantly excluded variants present more than twice in the cohort of 106 patients. Remaining variants predicted as potentially pathogenic and fitting with the mode of inheritance associated to the affected gene were further tested for validation (see online supplementary table S9).

### CNVs detection pipeline

Putative heterozygous/homozygous/hemizygous structural variants or CNVs were highlighted using the previously described method based on a depth-of-coverage comparison between the index sample and eight other random samples from the same sequencing lane.<sup>13</sup> For the X-chromosome, coverage was normalised according to the number of X-chromosomes of the patient.

### Mutation validation

All candidate mutations were validated by Sanger sequencing and co-segregation analyses were performed as extensively as possible. Putative splicing mutations were confirmed either using a minigene *in vitro* assay with the SPL3B plasmid as described previously<sup>15</sup> or using patients' fibroblasts or blood RNA when available. For apparent *de novo* variants, pedigree concordance was checked using polymorphic microsatellite markers (PowerPlex 16HS System, Promega). Mutations were considered as certainly causative when no doubt remained regarding their pathogenicity and an unambiguous diagnosis could thus be established. Such mutations co-segregated with the disease status in the family and were either truncating mutations or missense mutations that had been previously

**Table 1** Description of the cohort of 106 patients with intellectual disability (ID) and global diagnostic results

	Cohort (n=106)	With conclusive genetic diagnosis (n=26)			Yield (per category)
		XLID	ADID	Total	
<b>Gender</b>					
Male	96 (91%)	14	8	22	22/96 (23%)
Female	10 (9%)	2	2	4	4/10 (40%)
<i>Total</i>	<i>106</i>	<i>16</i>	<i>10</i>	<i>26</i>	<i>26/106 (25%)</i>
<b>Age</b>					
[0–10]	57 (54%)	9	5	14	14/57 (25%)
[10–20]	31 (29%)	5	4	9	9/31 (29%)
>20	18 (17%)	2	1	3	3/18 (17%)
<b>Sporadic cases *</b>					
Female	8 (7%)	2	2	4	4/8 (50%)
Male	72 (68%)	7	7	14	14/72 (19%)
<i>Total</i>	<i>80 (75%)</i>	<i>9</i>	<i>9</i>	<i>18</i>	<i>18/80 (23%)</i>
<b>Familial history</b>					
Male sib-pairs	12	5	0	5	5/12 (42%)
Possible XLID†	8	2	1	3	3/8 (38%)
Other (non-X-linked)	6	0	0	0	0/6 (0%)
<i>Total</i>	<i>26 (25%)</i>	<i>7</i>	<i>1</i>	<i>8</i>	<i>8/26 (31%)</i>
Consanguinity	3 (3%)	0	0	0	0/3 (0%)
<b>ID severity</b>					
Mild/borderline	12 (11%)	3	0	3	3/12 (25%)
Moderate	49 (46%)	4	6	10	10/49 (20%)
Severe	45 (42%)	9	4	13	13/45 (29%)
<b>Comorbidity</b>					
Microcephaly (<–2 SD)	14 (13%)	3	2	5	5/14 (36%)
Epilepsy	28 (26%)	5	2	6	6/28 (21%)
Autistic traits	34 (32%)	7	2	9	9/34 (26%)
Hypotonia	36 (34%)	4	4	8	8/36 (22%)
<b>Previous exploration</b>					
CGH	106 (100%)				
Fragile-X	105 (99%)				
Karyotype	99 (96%)				
# Other genetic tests (mean per patient)	2				
MRI	53 (50%)				
Metabolism‡	82 (77%)				

\*No familial first degree ID.

†Affected male relatives.

‡At least one biochemical test performed.

XLID, X-linked ID; in *italics*: total number per category.

convincingly published or that we confirmed with functional analyses. Mutations were considered as potentially causative when they appeared to co-segregate with ID in available members of the family and were predicted to be damaging, but further functional studies are needed to prove unambiguously their pathogenicity.

## RESULTS

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

Our strategy allowed generating a high-quality sequencing dataset, with a mean depth of coverage of 350× and an average per patient of 97.7% of targeted regions being well covered (>40×; see online supplementary table S1). Such coverage ensures a sensitivity of 99–99.9% of detecting SNVs and indels at any allelic state (Illumina Technical Note). We further assessed the sensitivity of SNV detection by comparing allelic states of SNPs detected by SNP-array (Affymetrix SNP Array 6.0) with the corresponding sites located in the targeted sequencing data of two patients and found no false-negatives after Sanger

sequencing validation of six SNPs showing discrepant allelic states between both methods (452 SNPs analysed). Interestingly, for these six SNPs, Sanger sequencing results were always in favour of targeted sequencing data suggesting a much higher accuracy (data not shown). No false positive was detected out of the 80 candidate variants located in well-covered regions that were tested for confirmation by Sanger sequencing.

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

Very few regions (a total of 3.9 kb, only 1.8 kb being protein coding) appear consistently poorly covered (coverage <40× in >90% of the samples; see online supplementary table S3). Those are mainly first exons or highly GC-rich regions that are a well-known burden in such capture strategy.

## Cognitive and behavioural genetics

## Cohort description and diagnostic yield

Patients harboured various degrees of cognitive impairment, although with a higher proportion of moderate or severe forms (46% and 42% respectively; [table 1](#)). The cohort was highly enriched in males. Among male probands, 68% were sporadic cases, the remaining had familial history of cognitive impairment mainly evocative of an X-linked mode of

transmission ([table 1](#)). We detected certainly causative mutations in 26/106 patients ([table 2](#); see online supplementary figures S2–S19), leading to an overall diagnostic yield of 25% for the entire cohort (from 23% for sporadic cases to 31% for familial cases). Unexpectedly, the diagnostic yield appears unrelated to the severity of ID in patients ([table 1](#)).

**Table 2** List of all causative/possibly causative mutations identified in our cohort

Patient ID	Sex	Gene	Mutation	Inheritance	Mode of inheritance	Degree of ID	Consistency with classic phenotype	See online supplementary figure #
Certainly-causative mutations								
APN-58	M	<i>DYRK1A</i>	chr21:g.38858865C>T; c.613C>T; p.Arg205*; htz	De novo	AD	++	Yes	S2
APN-87	M	<i>DYRKA1</i>	chr21:g.38858873del; c.621_624delinsGAA; p.Glu208Asnfs*3; htz	De novo	AD	++	Yes	S2
APN-63	M	<i>GRIN1</i>	chr9:g.140056661C>G; c.1733C>G; p.Pro578Arg; htz	De novo	AD	+++	Yes	S3
APN-14	M	<i>MED13L</i>	chr12:g.116406845_116406852del; c.6118_6125del; p.Gly2040Asnfs*32; htz	De novo	AD	++	Partially	S4
APN-46	M	<i>RAI1</i>	chr17:g.17698594_17698598del; c.2332_2336del; p.Gly778Glnfs*7; htz	De novo	AD	++	Partially	S5
APN-122	F	<i>SHANK3</i>	chr:g.51159168_51159183dup; c.2955_2970dup; p.Pro992Argfs*325; htz	De novo	AD	+++	Yes	S6
APN-38	M	<i>SLC2A1</i>	chr1:g.43395407G>A; c.724C>T; p.Gln242*; htz	De novo	AD	+++	Yes	S7
APN-139	M	<i>SYNGAP1</i>	chr6:g.33414346G>A; c.3583-6G>A; p.Val1195Alafs*27; htz; splice	De novo	AD	++	Yes	S8
APN-41	M	<i>TCF4</i>	chr18:g.53017622_53017625del; c.514_517del; p.Lys172Phefs*61; htz	De novo	AD	+++	Yes	S9
APN-117	F	<i>TCF4</i>	chr18:g.53017619G>A; c.520C>T; p.Arg174*; htz	De novo	AD	++	No	S9
APN-138	M	<i>ATRX</i>	<b>chrX: g.76972632G&gt;A; c.109C&gt;T; p.Arg37* (rs122445108); hemz</b>	Inherited (Ma)	XL	+++	Yes	S10
APN-137	M	<i>CUL4B</i>	chrX: g.119681009_119681010del; c.811_812del; p.Gln271Aspfs*11; hemz	Inherited (Ma)	XL	+++	Partially	S11
APN-42	M	<i>DMD</i>	chrX:g.31164440del; c.10889del; p.Arg3630Glnfs*27; hemz	Inherited (Ma)	XL	++	No	S12
APN-26	M	<i>FMR1</i>	Last exon deletion; hemz	Inherited (Ma)*	XL	+++	Partially	S13
APN-113	M	<i>HCFC1; (ATRX)</i>	<b>chrX:g.153230153G&gt;A; c.218C&gt;T; p.Ala73Val; hemz;</b> (chrX:g.76939735G>C, c.1013C>G, p.Ser338Cys; hemz)	Inherited (Ma) Inherited (Ma)	XL; (XL)	+++	Yes (partially)	3
APN-82	M	<i>IL1RAPL1</i>	chrX:g.29935696_29935705del; c.894_903del; p.Trp299Thrfs*18; hemz	Inherited (Ma)	XL	++	Yes	S14
APN-68	M	<i>IQSEC2</i>	chrX:g.53268395G>A; c.3097C>T; p.Gln1033*; hemz	De novo	XL	+++	Yes	S15
APN-34	M	<i>KDM5C</i>	chrX:g.53228250C>G; c.2152G>C; p.Ala718Pro; hemz	De novo	XL	++	Partially	S16
APN-135	M	<i>KDM5C</i>	chrX:g.53240784dup; c.1296dup; p.Glu433*; hemz	Inherited (Ma)	XL	++	Partially	S16
APN-16	M	<i>MAOA</i>	chrX:g.43590942_43590943delinsTT; c.797_798delinsTT; p.Cys266Phe; hemz	Inherited (Ma)	XL	+/-	Yes	42
APN-130	F	<i>MECP2</i>	<b>chrX: g.153296363G&gt;A; c.916C&gt;T; p.Arg306Cys (rs28935468); htz</b>	De novo	XL	+++	Partially	S17
APN-142	F	<i>MECP2</i>	<b>chrX:g.153296777G&gt;A; c.502C&gt;T; p.Arg168* (rs61748421); htz</b>	De novo	XL	+++	Partially	S17
APN-3	M	<i>MECP2</i>	Complex rearrangement of exon 4; hemz	Inherited (Ma)	XL	+++	Yes	S17
APN-105	M	<i>PHF8; (DOCK8)</i>	chrX:g.54028583C>G; c.1249+5G>C; p.Tyr406Phefs*24; hemz; (chr9:g.407035G>T; c.3496G>T; p.Glu1166*; htz)	Inherited (Ma); (de novo)	XL; (AD)	+	Partially (No)	S18
APN-43	M	<i>SLC9A6</i>	chrX:g.135080258_135080262del; c.526-9_526-5del; p.?; splice disrupted; hemz	Inherited (Ma)	XL	+	Yes	Masurel-Paulet <i>et al</i> , in preparation
APN-110	M	<i>SLC16A2</i>	<b>chrX:g.73749067T&gt;C; c.1412T&gt;C ; p.Leu471Pro (rs122455132); hemz</b>	Inherited (Ma)	XL	+++	Yes	S19

Continued

Table 2 Continued

Patient ID	Sex	Gene	Mutation	Inheritance	Mode of inheritance	Degree of ID	Consistency with classic phenotype	See online supplementary figure #
Possibly causative mutations								
APN-131	M	<i>SLC2A1</i> ; ( <i>ANKRD11</i> )	chr1:g.43392779del; c.1412delG; p.Gly471Glufs*37; htz; (chr16:g.89348867G>T; c.4083C>A; p.His1361Gln; htz)	Inherited (Pa); Inherited (Ma)	AD; AD	+++	Partially	S7
APN-101	M	<i>TCF4</i>	chr18:g.52899907C>T; c.1487-5G>A; p.Arg495_Gly496insAla?.; htz	De novo ? †	AD	++	No	S9
APN-99	M	<i>NLGN3</i>	chrX:g.70389249C>T; c.1849C>T; p.Arg617Trp; hemz	Inherited (Ma)	XL	+++	Yes	S20
APN-70	M	<i>PQBP1</i>	chrX:g.48760294C>T; c.731C>T; p.Pro244Leu; hemz	Inherited (Ma)	XL	++	No	S21

\*Present in the three brothers. Mother untested, but most probably maternally inherited.

†Absent from the mother, deceased father (untested).

In bold: mutations previously reported in other patients.

–: no ID, +: mild ID, ++: moderate ID, +++: severe ID.

AD, autosomal dominant; F, female; hemz, hemizygous; htz, heterozygous; ID, intellectual disability; M, male; Ma, maternally-inherited; MAOA, Monoamine Oxidase A enzyme; Pa, paternally inherited; XL, X-linked.

Sixteen mutations are located in genes of the X-chromosome: 14 point mutations or small indels (in *ATRX*, *CUL4B*, *DMD*, *HCFC1*, *IL1RAPL1*, *IQSEC2*, *KDM5Cx2*, *MECP2x2*, *MAOA*, *PHF8*, *SLC9A6*, *SLC16A2*), as well as two larger pathogenic events (one hemizygous complex rearrangement in *MECP2*, one hemizygous exon deletion in *FMR1*; see online supplementary figures S17 and S13). We identified 10 de novo point mutations or small indels in genes involved in autosomal-dominant/haploinsufficient forms of ID (in *DYRK1Ax2*, *GRIN1*, *MED13L*, *RAI1*, *SHANK3*, *SYNGAP1*, *SLC2A1*, *TCF4x2*). In four other patients, we identified potentially causative mutations (in *NLGN3*, *PQBP1*, *SLC2A1* and *TCF4*; see online supplementary figures S20, S21, S7 and S9, respectively), whose implication in cognitive impairment has to be further confirmed. Finally, missense variants that appeared at first likely to be pathogenic, notably based on the very high evolutionary conservation of the affected residues or on previous publications, appeared excluded as causal after further segregation analysis (in *FLNA*, *FMR1*, *HUWE1* or *MECP2*, see online supplementary figure S22).

### Atypical type of mutations

Among the 26 certainly causative mutations identified, some were surprising by the nature of the mutation itself. In a boy with severe encephalopathy, epilepsy, hypotonia and microcephaly, we detected a highly complex rearrangement in exon #4 of *MECP2*, involving a 139 bp deletion flanked by the insertion of two sequences in inverted orientation derived from intron #2 still keeping the reading frame downstream of the rearrangement. Such event is inherited from the proband's mother who presents with speech delay and dyslexia (see online supplementary figure S17). This exon is known to be the one accumulating most mutations in patients, especially the 3' half, which is a recombination hotspot and the target of several deletions/duplications and less frequently inversions.<sup>16–18</sup>

We report here for the first time an intragenic deletion affecting *FMR1* outside of the promoter/exon #1 region that is the target of the fragile-X syndrome CGG-expansion. Very few point mutations have been reported in coding regions. We identified a complete deletion of the last exon of *FMR1* in one patient and his two affected brothers unevenly presenting with

clinical features of fragile-X syndrome (see online supplementary figure S13).

We also identified a patient carrying a maternally inherited 10 bp deletion causing a frameshift in exon #7 of *IL1RAPL1*, while unexpectedly his affected brother bears a de novo deletion of the full exon highlighting that affected relatives may carry distinct mutations. We propose that small 10 bp deletion may have created a sequence conformation favouring further instability and leading to the larger deletion observed in the second brother (see online supplementary figure S14). Indeed, a large proportion of *IL1RAPL1* causative mutations are intragenic exon deletions or pericentric inversions,<sup>19</sup> supporting that *IL1RAPL1* region is highly susceptible to recombination events.

Lastly, we identified a de novo Pro578Arg missense mutation in *GRIN1* in a male proband with severe ID, hypotonia, feeding disorders and very poor speech but no epilepsy, a phenotype similar to that associated to *GRIN1* missense mutations in some patients.<sup>20–21</sup> A maternal uncle presented with similar features except for the poor speech and hypotonia (see online supplementary figure S3), initially suggesting an associated X-linked mode of inheritance. The finding of a de novo deleterious missense mutation excludes this latter hypothesis, further highlighting the prevalence of phenocopies in cognitive disorders.

### Genotype–phenotype correlations: from expected to unexpected

The majority of certainly causative mutations were identified in patients whose clinical phenotype was retrospectively consistent with previous reports (table 2). For instance, the proband carrying a truncating mutation in *IQSEC2* presents with severe ID, no speech, motor developmental delay, severe epilepsy, strabismus and autistic features (see online supplementary figure S15), which matches the recently proposed clinical spectrum associated to mutations in this gene.<sup>22–23</sup> Likewise, *DYRK1A* was originally found disrupted by translocations or deleted in several patients with ID and microcephaly.<sup>24–26</sup> More recently, truncating mutations in this gene were shown to cause ID associated with primary microcephaly (sometimes borderline at –2 SD), growth retardation, developmental delay, facial dysmorphic traits, seizures and major feeding difficulties, with or without



associated autism.<sup>27–31</sup> We report here two novel de novo truncating mutations in patients with similar clinical features but no epilepsy (see online supplementary figure S2). Nonetheless, in a few other cases (eg, mutations in *RAI1* or *MECP2*), the probands lacked some clinical features, thus the corresponding diagnosis of Smith-Magenis (MIM 182290) or Rett (MIM 312750) syndrome was not evoked by experienced clinical geneticists (see online supplementary figures S5 and S17). For instance, among the three patients with *MECP2* mutations, one female had the classic Rett phenotype, the other female proband presented with non-classic Rett phenotype (no regression episode, no hand-flapping), while the rarity of reports of *MECP2* encephalopathies in adolescent males precluded suspicion of the involvement of that gene in a patient (APN-3) who had undergone rather extensive prior genetic testing

A few detected mutations were unexpected as they were detected in patients whose phenotype did not match previous descriptions. For instance, mutations in *TCF4* mainly cause Pitt-Hopkins syndrome (PHS, MIM 610954) characterised by severe motor retardation, absence of speech, characteristic dysmorphic traits, autistic features, intestinal problems and hyperventilation.<sup>32</sup> We here describe truncating *TCF4* mutations in two patients, one with clinical manifestations highly suggestive of PHS, the other with less syndromic manifestations and no dysmorphic traits (see online supplementary figure S9). *TCF4* mutations were already reported in patients with non-syndromic ID, suggesting that such mutations were likely to be underdiagnosed.<sup>33</sup>

Another patient and his affected brother both carry a distal frameshift mutation in *DMD* affecting the major muscle transcript encoding the dystrophin protein associated to Duchenne or Becker muscular dystrophy (DMD, MIM 310200; BMD, MIM 300376), and the brain-specific isoform Dp71. The index case presents with moderate ID, psychomotor retardation, no speech, behavioural disorders, dysmorphic traits but strikingly no muscular phenotype (see online supplementary figure S12). His brother harbours a milder phenotype with additional hypotonia and cerebellar dysplasia. Both harbour borderline-high CPK levels. The association of cognitive impairment with DMD/BMD has been extensively reported and correlated to truncating mutations affecting Dp71, yet never in the absence of a muscular phenotype.<sup>34–39</sup> Our findings extend the recent report of a large family with affected males carrying an in-frame single amino acid deletion associated to mild ID and no muscular phenotype.<sup>40</sup>

### Confirmation of candidate genes for cognitive disorders

Some selected genes were only candidate ID or ASD genes at the time of the design, with single pieces of evidence in the literature. The identification of additional mutations in patients with similar phenotype definitively confirms their implication in cognitive disorders.

We reported a damaging missense affecting the function of the Monoamine Oxidase A enzyme (MAOA),<sup>41–42</sup> which replicated for the first time in 20 years the implication of MAOA in autism/ID associated to significant behavioural disorders.<sup>41–42</sup> We also identified a probably pathogenic missense variant in *NLGN3* in a male and his cousin, both presenting with ID and autism (see online supplementary figure S20). In silico predictions, high conservation of the mutated residue across all neuro-ligin paralogs, and familial analysis are altogether in favour of a pathogenicity of this missense change. A definitive functional effect of this missense still needs to be demonstrated to clearly establish the diagnosis. The implication of this gene was never

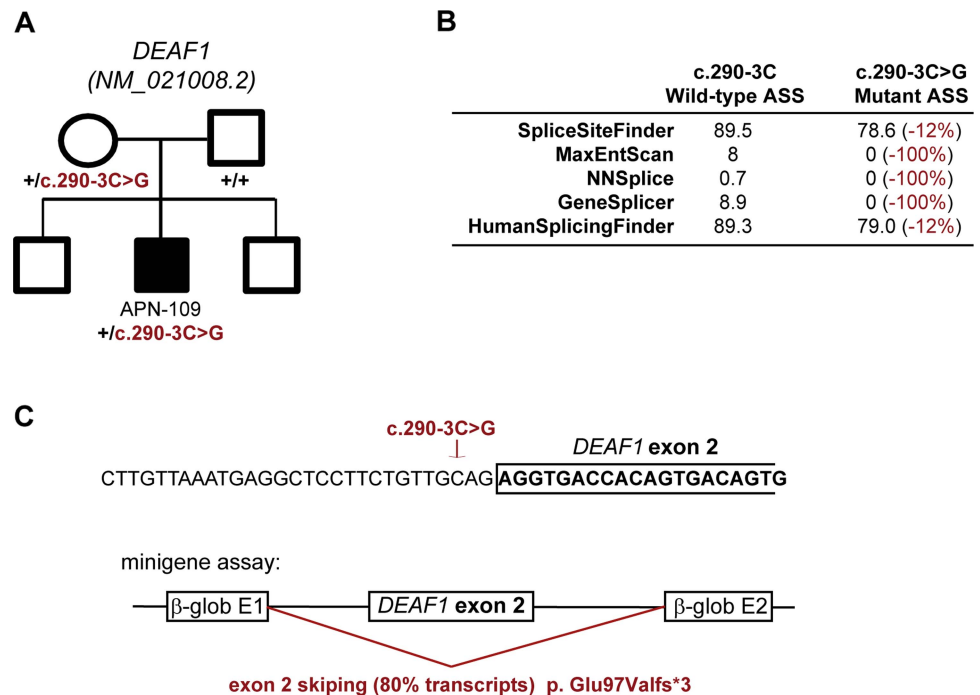
replicated since the initial publication,<sup>43</sup> although screened in several cohorts with comparable phenotypes.<sup>44–50</sup>

Similarly, we identified a novel truncating point mutation in *MED13L* confirming the implication of this gene in ID (see online supplementary figure S4). Disruption of *MED13L* was initially associated with transposition of the great arteries (TGA), associated to ID in a single case with a chromosomal translocation.<sup>51</sup> A homozygous missense mutation was then identified in two siblings from a consanguineous family presenting with non-syndromic ID, suggesting an implication of the gene in autosomal-recessive forms of ID justifying its selection in our panel.<sup>5</sup> A total of five patients were more recently described with de novo intragenic CNVs or point mutations affecting *MED13L*, delineating a recognisable *MED13L*-haploinsufficiency syndrome characterised by hypotonia, moderate ID, variable cardiac defects, facial hypotonia and dysmorphic traits.<sup>52–53</sup> Our patient with the *MED13L* mutation presents with concordant phenotype such as an open mouth appearance and muscular hypotonia but no cardiac defects. Due to the initially proposed autosomal-recessive mode of inheritance associated to ID, at first we did not consider this heterozygous truncating mutation as causative, as it may also have been the case for a heterozygous splicing mutation identified earlier in a large-scale exome sequencing study in one male with ASD.<sup>28</sup> Altogether, these findings suggest that ID associated to *MED13L*-haploinsufficiency syndrome is a relatively frequent condition.<sup>5–53</sup>

### Ambiguous mode of inheritance in ID-associated genes: the example of *DEAF1*

As for *MED13L*, the mode of inheritance associated to some genes is ambiguously described in literature. We identified an heterozygous variant affecting splicing in *DEAF1* inherited from the asymptomatic mother in a patient presenting with severe ID, developmental delay, poor speech, pain resistance, dysmorphic features and aggressive behaviour (figure 1), while the gene had been proposed as associated to autosomal-dominant forms of ID.<sup>7–8</sup> The recent report of two additional individuals carrying de novo missense mutations narrowed the associated phenotype to moderate/severe ID, speech impairment, behavioural problems, high pain threshold, dysmorphic features and abnormal walking pattern, hence highly similar to the one of our proband.<sup>54</sup> Although in vitro validation studies suggest that the reported missense variants lead to an impaired function of *DEAF1*, the authors concluded that they presumably act as dominant-negatives incapacitating both normal and mutant proteins since truncating variants had been observed in asymptomatic individuals.<sup>54</sup> In parallel, a homozygous missense mutation clustering in the same SAND-domain with all three de novo missense mutations was reported in members of a consanguineous family presenting with ID, microcephaly and white matter abnormalities, therefore suggesting a possible autosomal-recessive mode of inheritance.<sup>55</sup> The pathogenic mechanism associated to *DEAF1* mutations is therefore unclear. Due to the highly similar clinical features of the herein reported proband and of probands carrying de novo missense mutations, the splice variant detected here may contribute to the phenotype of our patient, possibly through a recessive mode of inheritance (ie, acting in *trans* with another heterozygous variant) since haploinsufficiency appears tolerated in healthy individuals. Altogether those findings either suppose a similar phenotype for autosomal-dominant and autosomal-recessive mode of inheritance associated to *DEAF1* mutations, or a universal autosomal-recessive mode of inheritance with a second variant that has not

**Figure 1** Inherited disrupting splice variant in *DEAF1*: what contribution to the phenotype? (A) Pedigree showing the maternally inherited splice variant in *DEAF1* (c.290-3C>G); (B) prediction scores for the effect of the herein described variant on splicing (prediction scores for acceptor splice sites (ASS) as computed by SpliceSite Finder, MaxEntScan, NNSplice, GeneSplicer and Human Splicing Finder for the consensus ASS with either the wild-type or the mutated allele); (C) localisation of the variant, and its resulting effect on splicing in vitro (minigene construct): 90% of abnormal transcripts: 80% with entire exon #2 skipped, and 10% using the alternative ASS c.290-16 both leading to a frameshift and a premature stop codon.



been yet identified, alike what was finally proven for Thrombocytopenia-absent radius (TAR) syndrome for instance.<sup>56</sup>

### Unsupportive evidences for proposed ID-associated mutations and genes

The identification in patients of a previously reported ID mutation should be considered with caution: indeed, as already discussed in previous publications, some 'false-positive' mutations are present among variants annotated as 'pathogenic' in dbSNP, OMIM or in published reports.<sup>13 14</sup> We here question the causative effect of a few previously proposed mutations (see online supplementary table S4). For instance, one missense variant identified in *FLNA* (c.3872C>T, p.Pro1291Leu) was previously reported in a patient with FG syndrome (MIM 300321).<sup>57</sup> This variant was identified here in a patient with a different phenotype (see online supplementary figure S22) and is also reported in one male in EVS (presumably not presenting with cognitive disorders), raising doubt about its pathogenicity.

The identification of non-segregating truncating variants (ie, detected both in patients and healthy relatives) can also challenge the implication of genes in X-linked and autosomal-dominant forms of ID. In one family, we identified a frameshift variant in *SRPX2* in a male proband. It is most likely inherited from the deceased asymptomatic maternal grandfather since it was also detected in the mother and in three maternal aunts yet absent from the maternal grandmother. Despite recent functional evidences regarding the role of *SRPX2* in brain development,<sup>58 59</sup> its definitive implication in cognitive disorders has already been questioned following the presence of the initially proposed mutations in EVS<sup>14</sup> and in control individuals,<sup>60</sup> and subsequently to the identification of a missense mutation in *GRIN2A* co-segregating with the epileptic status in the initial *SRPX2* family.<sup>61</sup> In another family, we identified a nonsense variant in *SHROOM4* in a male proband yet also in his unaffected brothers, a finding that further challenges the

implication of *SHROOM4* in X-linked cognitive disorders (see online supplementary table S5, figure 2).<sup>14</sup>

### Patients carrying probably pathogenic variants in two ID-associated genes

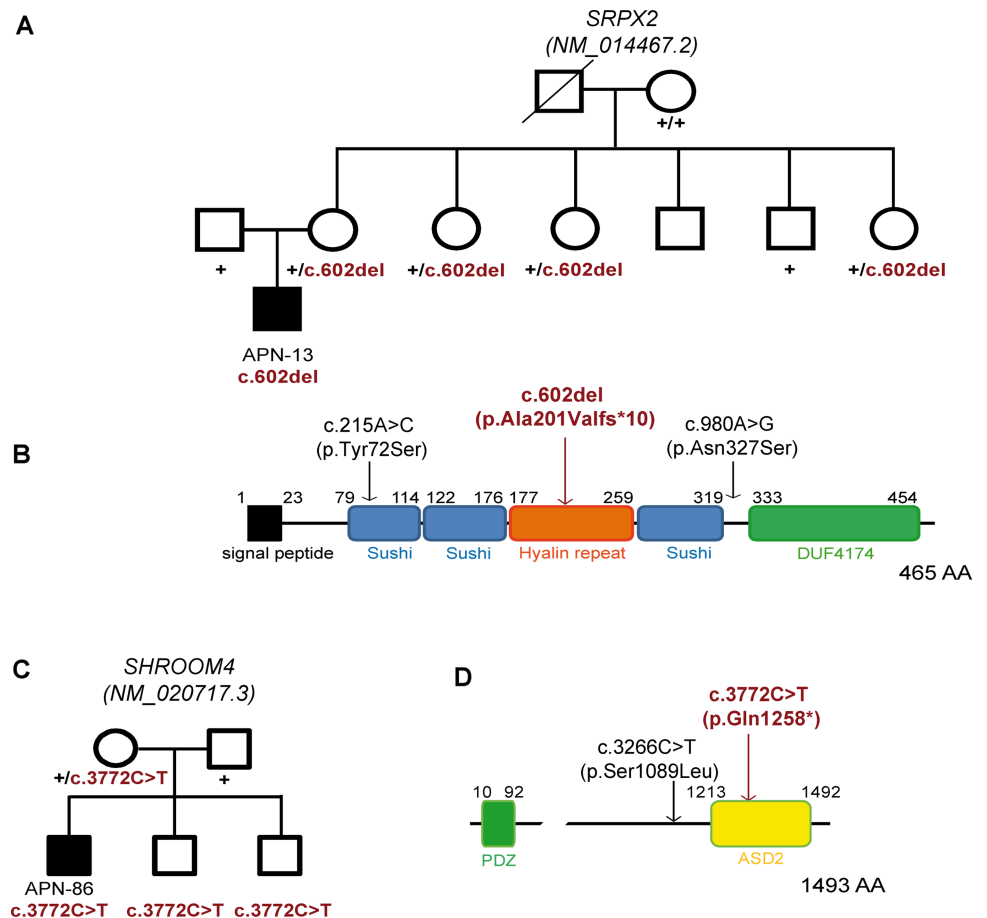
In three unrelated patients, we identified candidate variants in two separate genes, requiring the evaluation of different scenarios: (a) one single contributor while the second variant is innocuous, (b) one major contributor while the second variant acts as a modifier, and (c) both variants are implicated in the phenotype and have a synergistic effect.

In one male proband, we identified a maternally inherited splice variant leading to a frameshift in the X-linked gene *PHF8*, together with a de novo truncating variant in *DOCK8* (see online supplementary figure S18). Considering both the weak evidences implicating *DOCK8* in autosomal-dominant ID (two probands reported with translocations disrupting the gene), the clinical features consistent with a *PHF8* mutation (even in the absence of cleft lip/palate) and the X-inactivation bias identified in the mother, the *PHF8* variant alone is most likely to be responsible for the phenotype leaving the *DOCK8* variant as probably innocuous.<sup>62 63</sup>

In another male proband, two possibly causative missense variants were identified in the XLID genes *ATRX* and *HCFC1* (figure 3). Patients' phenotype was not evocative of an *ATRX* mutation (no dysmorphic traits, no urogenital abnormalities, absence of Heinz bodies), but perfectly matched the recent description of cobalamin-X metabolic disorder (*cbIX*, MIM 309541<sup>64</sup>). We concluded to a causative effect of the *HCFC1* mutation (being one of the *cbIX* recurrent mutations), but cannot exclude a contributory effect of the *ATRX* variant based on current co-segregation data.

In the third family, the proband carries one distal truncating variant in *SLC2A1* and one missense variant in *ANKRD11*, each inherited from one asymptomatic parent. Some mutations in *SLC2A1* have already been associated to GLUT-1 deficiency (MIM 606777/612126) with incomplete penetrance.<sup>65</sup> The

**Figure 2** Truncating variants not or ambiguously co-segregating with ID. (A) Pedigree of patient APN-13 carrying a frameshift variant in *SRPX2* (c.602del, p.Ala201Valfs\*10) demonstrating the likely inheritance from the asymptomatic (deceased) maternal grandfather, yet a putative germinal mosaicism of a de novo mutation cannot be excluded; (B) predicted functional domains of *SRPX2* from Pfam indicating locations of the herein identified mutation (in red) and those previously described. DUF4174: domain of unknown function; (C) pedigree showing the non-segregating nonsense variant in *SHROOM4* (c.3772C>T; p.Gln1258\*) in the family of patient APN-86; (D) location of the premature stop codon, which would disrupt the ASD2 domain of the protein.



proband presents with evocative symptoms of both *SLC2A1* and *ANKRD11* mutations (major hypotonia along with skeletal abnormalities) that might suggest a synergistic contribution of both variants to the phenotype (see online supplementary figure S7). No conclusion could be unambiguously drawn in this case, but such a di- or oligogenic mode of inheritance has already been proposed in neurodevelopmental disorders.<sup>66–68</sup>

## DISCUSSION

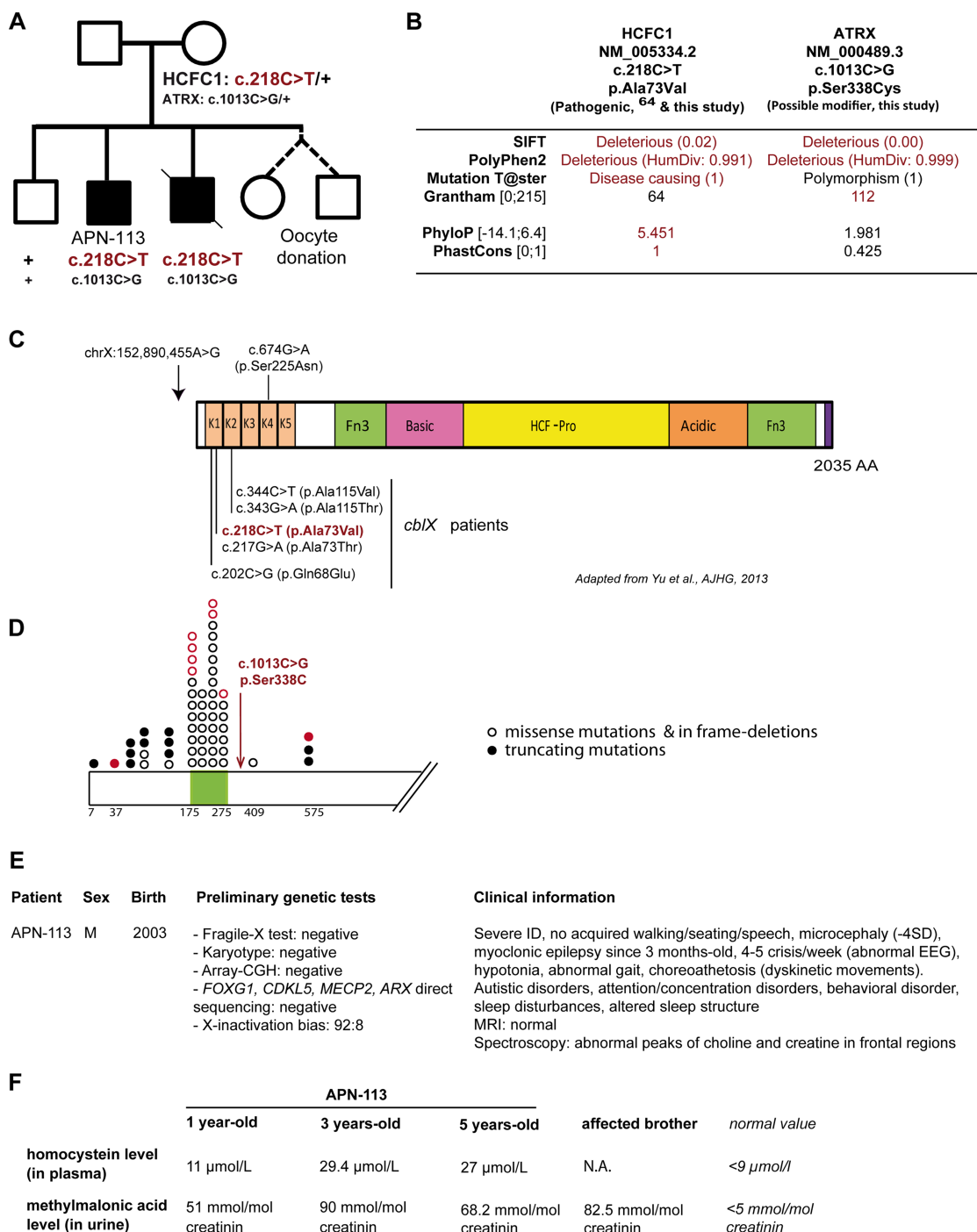
Targeted sequencing of 217 genes in a cohort of 106 patients with unknown genetic aetiology of ID led to a conclusive diagnostic yield of 25%. A majority of causative mutations identified are located in XLID genes, which can be partially explained by our male-enriched cohort. When excluding familial cases, X-linked mutations are found in 7/72 sporadic male cases, matching the proposed figure that 10% of males with sporadic ID carry mutations in X-linked genes.<sup>69</sup> Combining the results of recent trio-exome analyses leads to similar proportions (X-linked causative mutations in 7/66 males with sporadic ID<sup>3 8</sup>). The detection rate of X-linked mutations in our cohort is—as expected—higher in males with familial history of ID, but is also significant in females (2/10, both with *MECP2* mutations). Also unsurprisingly, mutations in autosomal-dominant/haploinsufficient ID genes are mostly found within sporadic cases.

Although we did not expect patients to harbour mutations in the same genes because of the small size of the cohort and the extensive genetic heterogeneity of ID, we had a few genes hit more than once: *MECP2* (despite being screened in 11% of patients prior to inclusion), *KDM5C*, *DYRK1A* and *TCF4*. Our results, intersected with the ones from other studies, highlight mutations within common genes (*MECP2*, *CUL4B*, *IL1RAPL1*,

*IQSEC2*, *KDM5C*, *SLC9A6*, *SLC16A2*, *DYRK1A*, *SLC2A1*, *SYNGAP1* or *TCF4*), suggesting that they are more frequently mutated in ID patients.<sup>3 7–9 20 28–30 70 71</sup> If mutations in such ID genes are confirmed to account for a substantial number of patients, introducing as a diagnostic step massive multiplexed resequencing of such genes in large cohorts may be considered.<sup>29</sup>

The extensive genetic and phenotypic heterogeneity of ID is the major hindrance for obtaining a precise molecular diagnosis. Direct sequencing strategies consisting in sequentially screening candidate genes are being replaced in diagnostic laboratories by more high-throughput NGS-based strategies: multiplex targeted sequencing of a few genes in large cohorts, selective targeted sequencing of up to several hundred genes, exome sequencing, while whole-genome sequencing (WGS), although powerful, notably for genome rearrangements and for CG-rich exons, remains in the research domain. Exome or full-genome strategies are more exhaustive alternatives and very attractive issues in the field of genetic diagnosis. Their universal approaches, whatever the clinical features, enhance the technical management of the workflow. However, the actual coverage proposed with the exome or full-genome strategies is still frequently insufficient (see online supplementary table S6), which may result in missing mutations.<sup>9 29</sup> Also, the finding of a putative mutation in a gene never associated before to a given pathology marks the start of a research endeavour to validate the finding but is not per se a diagnostic result.

Targeted sequencing appears more appealing for well-defined pathologies in which most implicated genes have been uncovered or in clinically homogeneous entities (ie, Bardet-Biedl and related ciliopathies,<sup>13</sup> retinal dystrophies,<sup>72</sup> hearing loss,<sup>73</sup> etc).



**Figure 3** Patient carrying two probably damaging missense variants in *HCFC1* and *ATRX*: one causative mutation and one modifier variant? (A) Family tree of patient APN-113: proband carries a missense mutation in *HCFC1* (c.218C>T; p.Ala73Val) already reported in two patients with *cbfX* as well as another missense variant in *ATRX* (c.1013C>G, p.Ser338Cys). Both variants are maternally inherited, absent from the unaffected brother, but carried by the younger brother who died of sudden death at 2 months old; (B) associated predictions for the recurrent pathogenic missense mutation in *HCFC1*<sup>64</sup> and the possible modifier in *ATRX*, showing putative pathogenicity and moderate nucleotide conservation; (C) representation of *HCFC1* and its domains: kelch domains (K1–K5), Fn3 (fibronectin type 3), basic domain, HCF-proteolysis repeats (HCF-pro), acidic domain and nuclear localisation signal (NLS) domain. The initial mutations involved in milder non-syndromic ID are indicated above: a regulatory variant in the 5' UTR of *HCFC1* was identified by targeted massive parallel resequencing in a family with probable X-linked ID (XLID) (MRX3), which had for long remained unsolved. This variant was disrupting the functional binding site of the transcription factor YY1 within the *HCFC1* promoter region, leading to an upregulation of its expression in lymphoblastoid cells.<sup>74</sup> Subsequent screening of additional unsolved families identified one single co-segregating missense variant (c.674G>A; p.Ser225Asn) in *HCFC1*. The phenotype of both patients was rather mild: non-syndromic mild to moderate ID. In the bottom are indicated the mutations recently described in *cbfX* patients<sup>64</sup>; (D) the *ATRX* missense variant is located close but outside of the hotspot for disease-causing missense mutations (in the zinc-finger binding domain, in green) reported in patients with *ATRX* mutations, in red: mutations reported independently in at least two patients. Mutations are indicated when affecting residues 1–700 (reported in OMIM, ClinVar or in<sup>75</sup>), the rest of the protein is not represented; (E) Clinical information regarding APN-113 and previous genetic explorations, including X-inactivation status in the mother. *cbfX* patients present with the same very severe phenotype: severe ID, early infantile epilepsy, choreoathetosis, microcephaly and more variable muscular hypotonia. Three of them are reported with early death in infancy<sup>64</sup>; (F) biochemical abnormalities observed in the two affected brothers are similar to those previously observed.<sup>64</sup>



We show here that it is also a powerful alternative for diagnostic purposes in ID (see online supplementary table S6). The positive diagnostic yield in our cohort of 106 patients is of 25% overall and 21% for sporadic cases, which is similar to the initial reports with the trio-exome strategy (32 highly likely causative mutations in a total of 151 patients, a 21% yield; see online supplementary table S7<sup>3 8</sup>). Indeed, most of the mutations identified by exome sequencing affect genes included in our panel and would thus have been detected with our strategy. Our diagnostic yield is slightly lower than the 30% yield (on 211 patients) we now calculated from pooling the three published exome studies (including also<sup>29</sup> and using the updated data for<sup>2</sup> very recently reported in Gilissen *et al*<sup>9</sup>), a difference that can be accounted for by the recently recognised ID genes missing in our panel (see online supplementary table S7). Our relatively high diagnostic yield is not due to an overestimation regarding the pathogenicity of the identified variants as we were highly stringent regarding the classification of certainly causative mutations (eg, we did not include all probably damaging missense variants located in XLID genes that were detected in males). The high depth of coverage and the smaller portion of poorly covered regions achieved with our strategy ensure a high sensitivity and specificity of detecting pathogenic events in the regions of interest (SNVs, indels, but also CNVs, a topic that had not been addressed in previous exome studies). Also, the relative ease of variant analysis and smaller number of follow-up studies for candidate variants may contribute to this significant yield. As fewer candidate variants are identified per patient with this approach, they can also be analysed more thoroughly (putative effect on splicing for variants not affecting canonical splice sites, predictive impact on the protein through structural modelling, etc). With an even broader panel of genes (including, for instance, genes involved in ID associated to multiple congenital anomalies but found mutated in patients with a milder presentation), one can expect that the resulting diagnostic yield will be higher.

The limited number of sequenced genes with the targeted approach—restricted to those involved in cognitive disorders—also avoids the controversial issues raised by incidental findings. Nonetheless, the targeted sequencing approach will miss newly identified genes and theoretically prevents data reanalysis (ie, incorporating novel findings regarding genetics of ID). The major value of targeted sequencing for ID is that it should allow the application of such test to a much higher proportion of patients awaiting molecular diagnosis given the significantly lower cost of sequencing, but especially of data analysis, storage and interpretation. It should thus generate much more promptly large amounts of data regarding the spectrum of mutations and phenotypes associated to the many genes implicated in ID, and thus considerably increase our knowledge on the specific condition associated with each gene.

The substantial proportion of patients that remain without molecular diagnosis with either strategy raises several issues: whether a large fraction of ID-associated genes remain to be discovered, whether many mutations are missed because they are located in non-coding regions or lastly whether more complex genetic scenarios are implicated such as variants with reduced penetrance, oligogenic and/or multifactorial modes of inheritance.<sup>66–68</sup> In particular, many disorders associated with autosomal-dominant inheritance are associated with incomplete penetrance and high intrafamilial phenotypic variability, although such scenario is mostly excluded from trio-studies that focus on de novo mutations. The recent results from Gilissen *et al* provide initial response on this issue: a substantial

proportion of such previously negative patients seem to carry pathogenic CNVs that could not be detected by array-CGH studies (and were not tested in the exome studies), other patients carry mutations located in newly identified genes or, mutations in known genes yet that had been missed by previous bioinformatic pipelines.<sup>9</sup> With an estimated cumulative diagnostic yield of 62% (although derived from a small cohort of 50 patients, for WGS analysis), whole-genome analyses of larger cohorts of patients may shed some more light regarding the contribution of the abovementioned hypotheses to explain the remaining proportion of patients without molecular diagnosis.

#### Author affiliations

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

**Acknowledgements** This study was supported by grants and fellowships from Fondation pour la Recherche Médicale, Agence de Biomedecine and Fondation Jerome Lejeune, APLM and CREGEMES. We thank the students who were involved in this project: Sébastien Kirsch, Audrey Creppy, Inès Bekkour and Grace Gan. We thank Nadège Calmels, Valérie Biancalana, Elsa Nourisson and all other members of the Genetic Diagnostic Laboratory of the Nouvel Hôpital Civil (Strasbourg) for their help with patients' DNA sample selection and preparation. We thank Cecile Pizot for the development of VaRank. We thank Damien Sanlaville, Christine Coubes, Delphine Héron, Sophie Naudion, James Lespinasse and Marie-Line Bichon for their contribution to the recruitment of patients, and all medical interns or genetic counsellors who participated in this project. Finally, we warmly thank all patients and their families for their implication in this study.

**Contributors** Study concept and design: CR, J-LM, AP. Clinical genetics investigations: AM-P, MW, GL, SE-C, YA, MB, DB, EC, HD, BD, M-AD, VD-G, EF, MF, CF, AG, SL, MM-D, DM-C, DL, GM, AP, SS, CT-R, JT, MD-F, DG, PS, PE, BI and LO-F. Acquisition, analysis and interpretation of data: CR, BG, JL, YH, JM, AQ, BJ, J-LM and AP. Drafting of the manuscript: CR, J-LM and AP. Critical revision of the manuscript for important intellectual content: BG and JM. Obtained funding: J-LM and AP. Administrative, technical or material support: NH, MD, CF, VG, SLG, MP and SV. Study supervision: BG, J-LM, AP.

**Competing interests** None.

**Patient consent** Obtained.

**Ethics approval** Comité de Protection des Personnes, France.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** All the variants identified during the course of this study will be submitted to the variant database ClinVar: <http://www.ncbi.nlm.nih.gov/clinvar/>.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev* 2002;8:117–34.
- Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. *Res Dev Disabil* 2011;32:419–36.
- de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, de Vries P, Gilissen C, del Rosario M, Hoischen A, Scheffer H, de Vries BB, Brunner HG, Veltman JA, Vissers LE. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 2012;367:1921–9.
- Lubs HA, Stevenson RE, Schwartz CE. Fragile X and X-linked intellectual disability: four decades of discovery. *Am J Hum Genet* 2012;90:579–90.
- Najmabadi H, Hu H, Garshashi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamal P, Zecha A, Mohseni M, Puttmann L, Vahid LN, Jensen C, Moheb LA, Bienek M, Larti F, Mueller I, Weissmann R, Darvish H, Wrogemann K, Hadavi V, Lipkowitz B, Esmaeili-Nieh S, Wiczorek D, Kariminejad R, Firouzabadi SG, Cohen M, Fattahi Z, Rost I, Mojahedi F, Hertzberg C, Dehghan A, Rajab A, Banavandi MJ, Hoffer J, Falah M, Musante L, Kalscheuer V, Ullmann R, Kuss AW, Tzschach A, Kahrizi K, Ropers HH. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 2011;478:57–63.
- Ropers HH. Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet* 2010;11:161–87.
- Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA. A de novo paradigm for mental retardation. *Nat Genet* 2010;42:1109–12.
- Rauch A, Wiczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T, Albrecht B, Bartholdi D, Beygo J, Di Donato N, Dufke A, Cremer K, Hempel M, Horn D, Hoyer J, Joset P, Popke A, Moog U, Riess A, Thiel CT, Tzschach A, Wiesener A, Wohlheber E, Zweier C, Ekici AB, Zink AM, Rump A, Meisinger C, Grallert H, Sticht H, Schenck A, Engels H, Rappold G, Schroock E, Wieacker P, Riess O, Meitinger T, Reis A, Strom TM. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 2012;380:1674–82.
- Gilissen CH-KJ, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A, Leach R, Klein R, Tearle R, Bo T, Pfundt R, Yntema HG, de Vries BB, Kleefstra T, Brunner HG, Vissers LE, Veltman JA. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014;511:344–7.
- Biancalana V, Beldjord C, Taillandier A, Szpiro-Tapia S, Cusin V, Gerson F, Philippe C, Mandel JL. Five years of molecular diagnosis of Fragile X syndrome (1997–2001): a collaborative study reporting 95% of the activity in France. *Am J Med Genet A* 2004;129A:218–24.
- Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, Zenker M, Hüffmeier U, Thiel C, Rüschemdorf F, Nürnberg P, Reis A, Trautmann U. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A* 2006;140:2063–74.
- van Karnebeek CD, Jansweijer MC, Leenders AG, Offringa M, Hennekam RC. Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness. *Eur J Hum Genet* 2005;13:6–25.
- Redin C, Le Gras S, Mhamdi O, Geoffroy V, Stoetzel C, Vincent MC, Chiurazzi P, Lacombe D, Ouertani I, Petit F, Till M, Verloes A, Jost B, Chaabouni HB, Dollfus H, Mandel JL, Muller J. Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alstrom syndromes. *J Med Genet* 2012;49:502–12.
- Piton A, Redin C, Mandel JL. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am J Hum Genet* 2013;93:368–83.
- Piton A, Jouan L, Rochefort D, Dobrzaniecka S, Lachapelle K, Dion PA, Gauthier J, Rouleau GA. Analysis of the effects of rare variants on splicing identifies alterations in GABAA receptor genes in autism spectrum disorder individuals. *Eur J Hum Genet* 2013;21:749–56.
- Archer HL, Whately SD, Evans JC, Ravine D, Huppke P, Kerr A, Bunyan D, Kerr B, Sweeney E, Davies SJ, Reardon W, Horn J, MacDermot KD, Smith RA, Magee A, Donaldson A, Crow Y, Hermon G, Miedzybrodzka Z, Cooper DN, Lazarou L, Butler R, Sampson J, Pilz DT, Laccione F, Clarke AJ. Gross rearrangements of the MECP2 gene are found in both classical and atypical Rett syndrome patients. *J Med Genet* 2006;43:451–6.
- Lebo RV, Ikuta T, Milunsky JM, Milunsky A. Rett syndrome from quintuple and triple deletions within the MECP2 deletion hotspot region. *Clin Genet* 2001;59:406–17.
- Ravn K, Nielsen JB, Skjeldal OH, Kerr A, Hulten M, Schwartz M. Large genomic rearrangements in MECP2. *Hum Mutat* 2005;25:324.
- Behnecke A, Hinderhofer K, Bartsch O, Numann A, Ipach ML, Damatova N, Haaf T, Dufke A, Riess O, Moog U. Intragenic deletions of IL1RAPL1: report of two cases and review of the literature. *Am J Med Genet A* 2011;155A:372–9.
- Epi KC, Epilepsy Phenome/Genome P, Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, Epstein MP, Glauser T, Goldstein DB, Han Y, Heinzen EL, Hitomi Y, Howell KB, Johnson MR, Kuzniecky R, Lowenstein DH, Lu YF, Madou MR, Marson AG, Mefford HC, Esmaeili Nieh S, O'Brien TJ, Ottman R, Petrovski S, Poduri A, Ruzzo EK, Scheffer IE, Sherr EH, Yuskaitis CJ, Abou-Khalil B, Alldredge BK, Bautista JF, Berkovic SF, Boro A, Cascino GD, Consalvo D, Crumrine P, Devinsky O, Dlugos D, Epstein MP, Fiol M, Fountain NB, French J, Friedman D, Geller EB, Glauser T, Glynn S, Haut SR, Hayward J, Helmers SL, Joshi S, Kanner A, Kirsch HE, Knowlton RC, Kossoff EH, Kuperman R, Kuzniecky R, Lowenstein DH, McGuire SM, Motika PV, Novotny EJ, Ottman R, Paolicchi JM, Parent JM, Park K, Poduri A, Scheffer IE, Shellhaas RA, Sherr EH, Shih JJ, Singh R, Sivén J, Smith MC, Sullivan J, Lin Thio L, Venkat A, Vining EP, Von Allmen GK, Weisenberg JL, Widdess-Walsh P, Winawer MR. De novo mutations in epileptic encephalopathies. *Nature* 2013;501:217–21.
- Hamdan FF, Gauthier J, Araki Y, Lin DT, Yoshizawa Y, Higashi K, Park AR, Spiegelman D, Dobrzaniecka S, Piton A, Tomitori H, Daoud H, Massicotte C, Henrien E, Diallo O, Group SD, Shekarabi M, Marineau C, Shevell M, Maranda B, Mitchell G, Nadeau A, D'Anjou G, Vanasse G, Vanase M, Sroum M, Lafreniere RG, Drapeau P, Lacaille JC, Kim E, Lee JR, Igarashi K, Hugarin RL, Rouleau GA, Michaud JL. Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am J Hum Genet* 2011;88:306–16.
- Shoubridge C, Tarpey PS, Abidi F, Ramsden SL, Rujirabanjerd S, Murphy JA, Boyle J, Shaw M, Gardner A, Proos A, Puusepp H, Raymond FL, Schwartz CE, Stevenson RE, Turner G, Field M, Walkonis RS, Harvey RJ, Hackett A, Futreal PA, Stratton MR, Geck J. Mutations in the guanine nucleotide exchange factor IQSEC2 cause nonsyndromic intellectual disability. *Nat Genet* 2010;42:486–8.
- Tran Mau-Them F, Willems M, Albrecht B, Sanchez E, Puechberty J, Ende S, Schneider A, Ruiz Pallares N, Missirian C, Rivier F, Girard M, Holder M, Manouvrier S, Touitou I, Lefort G, Sarda P, Moncla A, Drunat S, Wiczorek D, Genevieve D. Expanding the phenotype of IQSEC2 mutations: truncating mutations in severe intellectual disability. *Eur J Hum Genet* 2014;22:289–92.
- Moller RS, Kubart S, Hoeltzenbein M, Heye B, Vogel I, Hansen CP, Menzel C, Ullmann R, Tommerup N, Ropers HH, Tumer Z, Kalscheuer VM. Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. *Am J Hum Genet* 2008;82:1165–70.
- van Bon BW, Hoischen A, Hehir-Kwa J, de Brouwer AP, Ruivenkamp C, Gijbbers AC, Marcellis CL, de Leeuw N, Veltman JA, Brunner HG, de Vries BB. Intragenic deletion in DYRK1A leads to mental retardation and primary microcephaly. *Clin Genet* 2011;79:296–9.
- 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:113–19.
- Courcet JB, Faivre L, Malzac P, Masurel-Paulet A, Lopez E, Callier P, Lambert L, Lemesle M, Thevenon J, Gigot N, Duplomb L, Ragon C, Marle N, Mosca-Boidron AL, Huet F, Philippe C, Moncla A, Thauvin-Robinet C. The DYRK1A gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. *J Med Genet* 2012;49:731–6.
- Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekritsky M, Pradhan K, Ghiban E, Kramer M, Parla J, Demeter R, Fulton LL, Fulton RS, Magrini VJ, Ye K, Darnell JC, Darnell RB, Mardis ER, Wilson RK, Schatz MC, McCombie WR, Wigler M. De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012;74:285–99.
- O'Roak BJ, Vives L, Fu W, Egerton 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. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012;338:1619–22.
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012;485:246–50.
- Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA,

## Cognitive and behavioural genetics

- Eng CM. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* 2013;369:1502–11.
- 32 Amiel J, Rio M, de Pontual L, Redon R, Malan V, Boddaert N, Plouin P, Carter NP, Lyonnet S, Munnich A, Colleaux L. Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am J Hum Genet* 2007;80:988–93.
- 33 Hamdan FF, Daoud H, Patry L, Dionne-Laporte A, Spiegelman D, Dobrzyńska S, Rouleau GA, Michaud JL. Parent-child exome sequencing identifies a de novo truncating mutation in TCF4 in non-syndromic intellectual disability. *Clin Genet* 2013;83:198–200.
- 34 Bushby KM, Appleton R, Anderson LV, Welch JL, Kelly P, Gardner-Medwin D. Deletion status and intellectual impairment in Duchenne muscular dystrophy. *Dev Med Child Neurol* 1995;37:260–9.
- 35 Daoud F, Angeard N, Demerre B, Martie I, Benyaou R, Leturcq F, Cossee M, Deburgrave N, Saillour Y, Tuffery S, Urtizberea A, Toutain A, Echenne B, Frischman M, Mayer M, Desguerre I, Estournet B, Reveillere C, Penisson B, Cuisset JM, Kaplan JC, Heron D, Rivier F, Chelly J. Analysis of Dp71 contribution in the severity of mental retardation through comparison of Duchenne and Becker patients differing by mutation consequences on Dp71 expression. *Hum Mol Genet* 2009;18:3779–94.
- 36 Emery AE, Skinner R, Holloway S. A study of possible heterogeneity in Duchenne muscular dystrophy. *Clin Genet* 1979;15:444–9.
- 37 Kerr TP, Sewry CA, Robb SA, Roberts RG. Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay? *Hum Genet* 2001;109:402–7.
- 38 Taylor PJ, Betts GA, Maroulis S, Gillissen C, Pedersen RL, Mowat DR, Johnston HM, Buckley MF. Dystrophin gene mutation location and the risk of cognitive impairment in Duchenne muscular dystrophy. *PLoS ONE*;5:e8803.
- 39 Zellweger H, Niedermeyer E. Central nervous system manifestations in childhood muscular dystrophy (CMD). I. Psychometric and electroencephalographic findings. *Ann Paediatr* 1965;205:25–42.
- 40 de Brouwer AP, Nabuurs SB, Verhaar IE, Oudakker AR, Hordijk R, Yntema HG, Hordijk-Hos JM, Voeselek K, de Vries BB, van Essen T, Chen W, Hu H, Chelly J, den Dunnen JT, Kalscheuer VM, Aartsma-Rus AM, Hamel BC, van Bokhoven H, Kleefstra T. A 3-base pair deletion, c.9711\_9713del, in DMD results in intellectual disability without muscular dystrophy. *Eur J Hum Genet* 2014;22:480–5.
- 41 Brunner HG, Nelen M, Breakefield XO, Ropers HH, van Oost BA. Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science* 1993;262:578–80.
- 42 Piton A, Poquet H, Redin C, Masurel A, Lauer J, Muller J, Thevenon J, Herenger Y, Chancotte S, Bonnet M, Poinot JM, Huet F, Thauvin-Robinet C, Jaeger AS, Le Gras S, Jost B, Gerard B, Peoc'h K, Launay JM, Faivre L, Mandel JL. 20 ans apres: a second mutation in MAOA identified by targeted high-throughput sequencing in a family with altered behavior and cognition. *Eur J Hum Genet* 2014;22:776–83.
- 43 Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, Bourgeron T. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 2003;34:27–9.
- 44 Blasi F, Bacchelli E, Pesaresi G, Carone S, Bailey AJ, Maestrini E. Absence of coding mutations in the X-linked genes neuroligin 3 and neuroligin 4 in individuals with autism from the IMGSAC collection. *Am J Med Genet B Neuropsychiatr Genet* 2006;141B:220–1.
- 45 Gauthier J, Bonnel A, St-Onge J, Karemera L, Laurent S, Mottron L, Fombonne E, Joobor R, Rouleau GA. NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population. *Am J Med Genet B Neuropsychiatr Genet* 2005;132B:74–5.
- 46 Liu Y, Du Y, Liu W, Yang C, Wang H, Gong X. Lack of association between NLGN3, NLGN4, SHANK2 and SHANK3 gene variants and autism spectrum disorder in a Chinese population. *PLoS One* 2013;8:e56639.
- 47 Vincent JB, Kolozsvari D, Roberts WS, Bolton PF, Gurling HM, Scherer SW. Mutation screening of X-chromosomal neuroligin genes: no mutations in 196 autism probands. *Am J Med Genet B Neuropsychiatr Genet* 2004;129B:82–4.
- 48 Wermter AK, Kamp-Becker I, Strauch K, Schulte-Korne G, Remschmidt H. 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. *Am J Med Genet B Neuropsychiatr Genet* 2008;147B:535–7.
- 49 Yan J, Oliveira G, Coutinho A, Yang C, Feng J, Katz C, Sram J, Bockholt A, Jones IR, Craddock N, Cook EH Jr, Vicente A, Sommer SS. Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. *Mol Psychiatry* 2005;10:329–32.
- 50 Ylisaukko-oja T, Rehnstrom K, Auranen M, Vanhala R, Alen R, Kempas E, Ellonen P, Turunen JA, Makkonen I, Riikonen R, Nieminen-von Wendt T, von Wendt L, Peltonen L, Jarvela I. Analysis of four neuroligin genes as candidates for autism. *Eur J Hum Genet* 2005;13:1285–92.
- 51 Muncie N, Jung C, Rudiger H, Ulmer H, Roeth R, Hubert A, Goldmuntz E, Driscoll D, Goodship J, Schon K, Rappold G. Missense mutations and gene interruption in PROSIT240, a novel TRAP240-like gene, in patients with congenital heart defect (transposition of the great arteries). *Circulation* 2003;108:2843–50.
- 52 Asadollahi R, Oneda B, Sheth F, Azzarello-Burri S, Baldinger R, Joset P, Latal B, Knirsch W, Desai S, Baumer A, Houge G, Andrieux J, Rauch A. Dosage changes of MED13L further delineate its role in congenital heart defects and intellectual disability. *Eur J Hum Genet* 2013;21:1100–4.
- 53 van Haelst MM, Monroe GR, Duran K, van Binsbergen E, Breur JM, Giltay JC, van Haften G. Further confirmation of the MED13L haploinsufficiency syndrome. *Eur J Hum Genet* 2014 Apr 30. Published Online First. doi: 10.1038/ejhg.2014.69
- 54 Vulto-van Silfhout AT, Rajamanickam S, Jensik PJ, Vergult S, de Rocker N, Newhall KJ, Raghavan R, Reardon SN, Jarrett K, McIntyre T, Bulinski J, Ownby SL, Huggenvik JI, McKnight GS, Rose GM, Cai X, Willaert A, Zweier C, Ende S, de Ligt J, van Bon BW, Lugtenberg D, de Vries PF, Veltman JA, van Bokhoven H, Brunner HG, Rauch A, de Brouwer AP, Carvill GL, Hoischen A, Mefford HC, Eichler EE, Vissers LE, Menten B, Collard MW, de Vries BB. Mutations Affecting the SAND Domain of DEAF1 Cause Intellectual Disability with Severe Speech Impairment and Behavioral Problems. *Am J Hum Genet* 2014;94:649–61.
- 55 Faqeih EA, Al-Owain M, Colak D, Kenana R, Al-Yafee Y, Al-Dosary M, Al-Saman A, Albalawi F, Al-Sarar D, Domiaty D, Daghestani M, Kaya N. Novel homozygous DEAF1 variant suspected in causing white matter disease, intellectual disability, and microcephaly. *Am J Med Genet A* 2014;164A:1565–70.
- 56 Albers CA, Paul DS, Schulze H, Freson K, Stephens JC, Smethurst PA, Jolley JD, Cvejic A, Kostadima M, Bertone P, Breuning MH, Deblin N, Deloukas P, Favier R, Fiedler J, Hobbs CM, Huang N, Hurler ME, Kiddle G, Krapels I, Nurden P, Ruivenkamp CA, Sambrook JG, Smith K, Stemple DL, Strauss G, Thys C, van Geet C, Newbury-Ecob R, Ouwelhand WH, Ghevaert C. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet* 2012;44:435–9. S1–2.
- 57 Unger S, Mainberger A, Spitz C, Bahr A, Zeschinig C, Zabel B, Superti-Furga A, Morris-Rosendahl DJ. Filamin A mutation is one cause of FG syndrome. *Am J Med Genet A* 2007;143A:1876–9.
- 58 Salmi M, Bruneau N, Cillario J, Lozovaya N, Massacrier A, Buhler E, Cloarec R, Tsintsadze T, Watrin F, Tsintsadze V, Zimmer C, Villard C, Lafitte D, Cardoso C, Bao L, Lesca G, Rudolf G, Muscatelli F, Pauly V, Khalilov I, Durbec P, Ben-Ari Y, Burnashev N, Represa A, Szepietowski P. Tubacin prevents neuronal migration defects and epileptic activity caused by rat Srxp2 silencing in utero. *Brain* 2013;136(Pt 8):2457–73.
- 59 Sia GM, Clem RL, Huganir RL. The human language-associated gene SRPX2 regulates synapse formation and vocalization in mice. *Science* 2013;342:987–91.
- 60 Reinthaler EM, Lal D, Jurkowski W, Feucht M, Steinbock H, Gruber-Sedlmayr U, Ronen GM, Geldner J, Haberlandt E, Neophytou B, Hahn A, Altmuller J, Thiele H, Toliat MR, Lerche H, Nurnberg P, Sander T, Neubauber BA, Zimprich F. Analysis of ELP4, SRPX2, and interacting genes in typical and atypical rolandic epilepsies. *Epilepsia* 2014 Aug. Published Online First. doi: 10.1111/epi.12712.
- 61 Lesca G, Rudolf G, Bruneau N, Lozovaya N, Labalme A, Boutry-Kryza N, Salmi M, Tsintsadze T, Addis L, Motte J, Wright S, Tsintsadze V, Michel A, Doummar D, Lascelles K, Strug L, Waters P, de Bellesize J, Vrielynck P, de Saint Martin A, Ville D, Ryvlin P, Arzimanoglou A, Hirsch E, Vincent A, Pal D, Burnashev N, Sanlaville D, Szepietowski P. GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat Genet* 2013;45:1061–6.
- 62 Griggs BL, Ladd S, Saul RA, DuPont BR, Srivastava AK. Deducator of cytokinesis 8 is disrupted in two patients with mental retardation and developmental disabilities. *Genomics* 2008;91:195–202.
- 63 Laumonier F, Holbert S, Ronce N, Faravelli F, Lenzner S, Schwartz CE, Lespinasse J, Van Esch H, Lacombe D, Goizet C, Phan-Dinh Tuy F, van Bokhoven H, Frys JP, Chelly J, Ropers HH, Moraine C, Hamel BC, Briault S. Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J Med Genet* 2005;42:780–6.
- 64 Yu HC, Sloan JL, Scharer G, Brebner A, Quintana AM, Achilly NP, Manoli I, Coughlin CR II, Geiger EA, Schneck U, Watkins D, Suomalainen T, Van Hove JL, Fowler B, Baumgartner MR, Rosenblatt DS, Venditti CP, Shaikh TH. An X-linked cobalamin disorder caused by mutations in transcriptional coregulator HCFC1. *Am J Hum Genet* 2013;93:506–14.
- 65 Leen WG, Klepper J, Verbeek MM, Leferink M, Hofste T, van Engelen BG, Wevers RA, Arthur T, Bahi-Buisson N, Ballhausen D, Bekhof J, van Bogaert P, Carrilho I, Chabrol B, Champion MP, Coldwell J, Clayton P, Donner E, Evangelidou A, Ebinger F, Farrell K, Forsyth RJ, de Goede CG, Gross S, Grunewald S, Holthausen H, Jayawant S, Lachlan K, Laugel V, Leppig K, Lim MJ, Mancini G, Marina AD, Martorell L, McMenamin J, Meuwissen ME, Mundy H, Nilsson NO, Panzer A, Poll-The BT, Rauscher C, Rouselle CM, Sandvig I, Scheffner T, Sheridan E, Simpson N, Sykora P, Tomlinson R, Trounce J, Webb D, Weschke B, Scheffer H, Willemsen MA. Glucose transporter-1 deficiency syndrome: the expanding clinical and genetic spectrum of a treatable disorder. *Brain* 2010;133(Pt 3):655–70.
- 66 Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, Vives L, Walsh T, McCarthy SE, Baker C, Mefford HC, Kidd JM, Browning SR, Browning BL, Dickel DE, Levy DL, Ballif BC, Platky K, Farber DM, Gowans GC, Wetherbee JJ, Asamoah A, Weaver DD, Mark PR, Dickerson J, Garg BP, Ellingwood SA, Smith R, Banks VC, Smith W, McDonald MT, Hoo JJ, French BN, Hudson C, Johnson JP, Ozmores JR, Moeschler JB, Surti U, Escobar LF, El-Khechen D, Gorski JL, Kussmann J,

- Salbert B, Lacassie Y, Biser A, McDonald-McGinn DM, Zackai EH, Deardorff MA, Shaikh TH, Haan E, Friend KL, Fichera M, Romano C, Gecz J, DeLisi LE, Sebat J, King MC, Shaffer LG, Eichler EE. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 2010;42:203–9.
- 67 Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G, Konyukh M, Chaste P, Ey E, Rastam M, Anckarsater H, Nygren G, Gillberg IC, Melke J, Toro R, Regnault B, Fauchereau F, Mercati O, Lemièrre N, Skuse D, Poot M, Holt R, Monaco AP, Jarvela I, Kantojärvi K, Vanhala R, Curran S, Collier DA, Bolton P, Chiochetti A, Klauck SM, Poustka F, Freitag CM, Waltes R, Kopp M, Duketis E, Bacchelli E, Minopoli F, Ruta L, Battaglia A, Mazzone L, Maestrini E, Sequeira AF, Oliveira B, Vicente A, Oliveira G, Pinto D, Scherer SW, Zelenika D, Delepine M, Lathrop M, Bonneau D, Guinchat V, Devillard F, Assouline B, Mouren MC, Leboyer M, Gillberg C, Boeckers TM, Bourgeron T. Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet* 2012;8:e1002521.
- 68 Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, Hawes A, Lewis L, Akbar H, Varghese R, Boerwinkle E, Gibbs RA, Zoghbi HY. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum Mol Genet* 2011;20:3366–75.
- 69 Mandel JL, Chelly J. Monogenic X-linked mental retardation: is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations. *Eur J Hum Genet* 2004;12:689–93.
- 70 Nava C, Lamari F, Heron D, Mignot C, Rastetter A, Keren B, Cohen D, Faudet A, Bouteiller D, Gilleron M, Jacqueline A, Whalen S, Afenjar A, Perisse D, Laurent C, Dupuits C, Gautier C, Gerard M, Huguet G, Cailliet S, Leheup B, Leboyer M, Gillberg C, Delorme R, Bourgeron T, Brice A, Depienne C. Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. *Transl Psychiatry* 2012;2:e179.
- 71 Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C, Dicks E, Menzies A, Stephens P, Blow M, Greenman C, Xue Y, Tyler-Smith C, Thompson D, Gray K, Andrews J, Barthorpe S, Buck G, Cole J, Dunmore R, Jones D, Maddison M, Mironenko T, Turner R, Turrell K, Varian J, West S, Widaa S, Wray P, Teague J, Butler A, Jenkinson A, Jia M, Richardson D, Shepherd R, Wooster R, Tejada MI, Martinez F, Carvill G, Goliath R, de Brouwer AP, van Bokhoven H, Van Esch H, Chelly J, Raynaud M, Ropers HH, Abidi FE, Srivastava AK, Cox J, Luo Y, Mallya U, Moon J, Parnau J, Mohammed S, Tolmie JL, Shoubridge C, Corbett M, Gardner A, Haan E, Rujirabanjerd S, Shaw M, Vandeleur L, Fullston T, Easton DF, Boyle J, Partington M, Hackett A, Field M, Skinner C, Stevenson RE, Bobrow N, Turner G, Schwartz CE, Gecz J, Raymond FL, Futreal PA, Stratton MR. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet* 2009;41:535–43.
- 72 Simpson DA, Clark GR, Alexander S, Silvestri G, Willoughby CE. Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa. *J Med Genet* 2011;48:145–51.
- 73 Brownstein Z, Friedman LM, Shahin H, Oron-Karni V, Kol N, Abu Rayyan A, Parzefall T, Lev D, Shalev S, Frydman M, Davidov B, Shohat M, Rahile M, Lieberman S, Levy-Lahad E, Lee MK, Shomron N, King MC, Walsh T, Kanaan M, Avraham KB. Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in Middle Eastern families. *Genome Biol* 2011;12:R89.
- 74 Huang L, Jolly LA, Willis-Owen S, Gardner A, Kumar R, Douglas E, Shoubridge C, Wiczorek D, Tzschach A, Cohen M, Hackett A, Field M, Froyen G, Hu H, Haas SA, Ropers HH, Kalscheuer VM, Corbett MA, Gecz J. A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. *Am J Hum Genet* 2012;91:694–702.
- 75 Gibbons RJ, Wada T, Fisher CA, Malik N, Mitson MJ, Steensma DP, Fryer A, Goudie DR, Krantz ID, Traeger-Synodinos J. Mutations in the chromatin-associated protein ATRX. *Hum Mutat* 2008;29:796–802.





## Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing

Claire Redin, Bénédicte Gérard, Julia Lauer, et al.

*J Med Genet* published online August 28, 2014

doi: 10.1136/jmedgenet-2014-102554

---

Updated information and services can be found at:

<http://jmg.bmj.com/content/early/2014/08/28/jmedgenet-2014-102554.full.html>

---

*These include:*

**Data Supplement**

"Supplementary Data"

<http://jmg.bmj.com/content/suppl/2014/08/29/jmedgenet-2014-102554.DC1.html>

**References**

This article cites 72 articles, 13 of which can be accessed free at:

<http://jmg.bmj.com/content/early/2014/08/28/jmedgenet-2014-102554.full.html#ref-list-1>

**Open Access**

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

**P<P**

Published online August 28, 2014 in advance of the print journal.

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

Advance online articles have been peer reviewed, accepted for publication, edited and typeset, but have not yet appeared in the paper journal. Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

---

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>

## **Annexe 2.**

### **Paper:** Mutations in *DCC* Cause Agenesis of the Corpus Callosum

Marsh APL, Heron D, Edwards T, **Quartier A**, Galea C, Nava C, *et al.*, in revision in Nat. Genet.

Aim of the study: Report of *DCC* mutations in four multigenerational families with individuals presenting with isolated agenesis of the corpus callosum (ACC) and/or mirror movements. Moreover, *DCC* sequencing was performed in 70 unrelated individuals with ACC including 46 with normal cognitive development. In parallel, analyzing of the phenotype of individuals with *DCC* mutations reported in the literature was carried out to assess the penetrance of mirror movements and ACC. This showed that, in individuals with truncating variants, males exhibit mirror movements more frequently than females while ACC is more often present in females. *DCC* is upregulated by testosterone in hNSCs, a cellular model of brain development. In this way, the phenotypic variability could be determined by genetic, epigenetic and/or environmental factors among which hormonal context.

Contribution: Minor. Performed testosterone treatment and *DCC* mRNA expression analysis on human neural stem cells

*NB:* According to their size, the supplementary figures and tables are not enclosed.



1 **Mutations in *DCC* Cause Agenesis of the Corpus Callosum**

2

3 Ashley PL Marsh<sup>1,2</sup> †, Delphine Heron<sup>3,4,5</sup> †, Timothy J Edwards<sup>6,7</sup> †, **Angélique Quartier**<sup>8</sup>,  
 4 Charles Galea<sup>9</sup>, Caroline Nava<sup>3,10</sup>, Agnès Rastetter<sup>10</sup>, Marie-Laure Moutard<sup>11,12,13</sup>, Vicki  
 5 Anderson<sup>14</sup>, Pierre Bitoun<sup>15</sup>, Jens Bunt<sup>6</sup>, Anne Faudet<sup>3</sup>, Catherine Garel<sup>16</sup>, Greta Gillies<sup>1</sup>, Ilan  
 6 Gobius<sup>6</sup>, Justine Guegan<sup>17</sup>, Solveig Heide<sup>3,4</sup>, Boris Keren<sup>3,10</sup>, Fabien Lesne<sup>3</sup>, Vesna Lukic<sup>18</sup>,  
 7 Simone A Mandelstam<sup>2,19,20</sup>, George Mcgillivray<sup>21</sup>, Alissandra McIlroy<sup>14</sup>, Aurélie Méneret<sup>10,22</sup>  
 8 , Cyril Mignot<sup>3,4,5</sup>, Laura R Morcom<sup>6</sup>, Sylvie Odent<sup>23,24</sup>, Annalisa Paolino<sup>6</sup>, Kate Pope<sup>1</sup>, Florence  
 9 Riant<sup>25</sup>, Gail Robinson<sup>26</sup>, Megan Spencer-Smith<sup>14,27</sup>, Myriam Srouf<sup>28,29</sup>, Sarah E Stephenson<sup>1,2</sup>,  
 10 Rick Tankard<sup>30,31</sup>, Oriane Trouillard<sup>10</sup>, Quentin Welniarz<sup>10,32</sup>, Amanda Wood<sup>33</sup>, Alexis  
 11 Brice,<sup>3,10</sup> Guy Rouleau<sup>29,34</sup>, Tania Attié-Bitach<sup>35</sup>, Martin B Delatycki<sup>1,2,36</sup>, Jean-Louis Mandel<sup>8,37</sup>,  
 12 David J Amor<sup>1,2,17</sup>, Emmanuel Roze<sup>10,22</sup>, Amélie Piton<sup>8,37</sup>, Melanie Bahlo<sup>30,31</sup>, Thierry Billette  
 13 de Villemeur<sup>5,11,12,38</sup>, Elliott H Sherr<sup>39</sup>, Richard J Leventer<sup>2,40,41</sup>, Linda J Richards<sup>6,42</sup> §\*, Paul J  
 14 Lockhart<sup>1,2</sup> §\*, Christel Depienne<sup>3,8,10,37</sup> §\*

15

16 <sup>1</sup> Bruce Lefroy Centre for Genetic Health Research, Murdoch Childrens Research Institute,  
 17 Royal Children's Hospital, Parkville, Victoria 3052, Australia.

18 <sup>2</sup> Department of Paediatrics, University of Melbourne, Parkville, Victoria 3052, Australia.

19 <sup>3</sup> AP-HP, Hôpital de la Pitié-Salpêtrière, Département de Génétique, F-75013, Paris, France.

20 <sup>4</sup> Groupe de Recherche Clinique (GRC) "déficience intellectuelle et autisme" UPMC, 75013  
 21 Paris, France.

22 <sup>5</sup> Centre de Référence "déficiences intellectuelles de causes rares", 75013 Paris, France.

23 <sup>6</sup> The University of Queensland, Queensland Brain Institute, St Lucia Campus, Brisbane, 4067,  
 24 Australia.



25 <sup>7</sup> The University of Queensland, School of Medicine, St Lucia Campus, Brisbane, 4067,  
26 Australia.

27 <sup>8</sup> Département de Médecine translationnelle et Neurogénétique, IGBMC, CNRS UMR  
28 7104/INSERM U964/Université de Strasbourg, 67400 Illkirch, France.

29 <sup>9</sup> Medicinal Chemistry and Drug Delivery, Disposition and Dynamics (D4), Monash Institute of  
30 Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia.

31 <sup>10</sup> INSERM, U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S  
32 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France

33 <sup>11</sup> AP-HP, Hôpital Trousseau, service de neuropédiatrie, 75012 Paris, France.

34 <sup>12</sup> UPMC, GRC ConCer-LD, Sorbonne université, Paris France.

35 <sup>13</sup> Centre de référence "neurogénétique", Paris France.

36 <sup>14</sup> Murdoch Childrens Research Institute, Parkville, Victoria 3052, Australia.

37 <sup>15</sup> Génétique Médicale, CHU Paris Nord, Hôpital Jean Verdier, 93140 Bondy, France.

38 <sup>16</sup> AP-HP, GHUEP, Hôpital Armand-Trousseau, Service de Radiologie, 75012 Paris, France.

39 <sup>17</sup> iCONICS facility, ICM, 75013, Paris, France.

40 <sup>18</sup> Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal  
41 Parade, Parkville, Victoria 3052, Australia.

42 <sup>19</sup> The Florey Institute of Neuroscience and Mental Health, Melbourne, Victoria, Australia.

43 <sup>20</sup> Department of Radiology, University of Melbourne, Royal Children's Hospital, Parkville,  
44 Victoria 3052, Australia.

45 <sup>21</sup> Victorian Clinical Genetics Services, Murdoch Childrens Research Institute, Parkville,  
46 Victoria 3052, Australia.

47 <sup>22</sup> AP-HP, Hôpital de la Pitié-Salpêtrière, Département de Neurologie, F-75013, Paris, France.

- 48 <sup>23</sup> Service de Génétique Clinique, Centre de référence CLAD-Ouest, CHU Rennes, 35000  
49 Rennes, France.
- 50 <sup>24</sup> UMR 6290 CNRS, IGDR Institut de Génétique et développement de Rennes, Université de  
51 Rennes1, Rennes France.
- 52 <sup>25</sup> AP-HP, Groupe Hospitalier Saint Louis, Lariboisière, Fernand Widal, Laboratoire de  
53 Génétique, 75010 Paris, France.
- 54 <sup>26</sup> Neuropsychology Research Unit, School of Psychology, The University of Queensland,  
55 Brisbane QLD 4072, Australia.
- 56 <sup>27</sup> School of Psychological Sciences and Monash Institute of Cognitive and Clinical  
57 Neurosciences, Monash University, Clayton Campus, Clayton Victoria 3800, Australia.
- 58 <sup>28</sup> Department of Pediatrics, Montreal Children's Hospital, McGill University, Montréal,  
59 Quebec, H4A 3J1, Canada.
- 60 <sup>29</sup> Department of Neurology and Neurosurgery, McGill University Health Center, Montreal,  
61 Quebec, H3A 2B4 Canada.
- 62 <sup>30</sup> Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical  
63 Research, Parkville 3052 VIC, Australia
- 64 <sup>31</sup> Department of Medical Biology, The University of Melbourne, Parkville, 3052, Victoria,  
65 Australia.
- 66 <sup>32</sup> Neuroscience Paris Seine, CNRS UMR8246, Inserm U1130, Sorbonne Universités, UPMC  
67 UM119 75013 Paris, France
- 68 <sup>33</sup> School of Psychology, University of Birmingham, West Midlands B15 2TT, United-  
69 Kingdom.
- 70 <sup>34</sup> Montreal Neurological Institute and Hospital, McGill University, Montréal, Quebec H3A  
71 2B4, Canada.

72 <sup>35</sup> INSERM U1163, Laboratory of Embryology and Genetics of Congenital Malformations,  
73 Paris Descartes University, Sorbonne Paris Cité and Imagine Institute, 75015 Paris, France;  
74 Département de Génétique, Hôpital Necker - Enfants Malades, Assistance Publique - Hôpitaux  
75 de Paris, 75015 Paris, France.

76 <sup>36</sup> Department of Clinical Genetics, Austin Health, Heidelberg, Victoria, 3084, Australia.

77 <sup>37</sup> Laboratoires de génétique, Institut de génétique médicale d'Alsace, Hôpitaux  
78 Universitaires de Strasbourg, 67 000 Strasbourg, France.

79 <sup>38</sup> INSERM U1141, 75019 Paris, France.

80 <sup>39</sup> Department of Neurology, UCSF Benioff Children's Hospital, San Francisco, California,  
81 USA.

82 <sup>40</sup> Neuroscience Research Group, Murdoch Childrens Research Institute, Parkville, Victoria  
83 3052, Australia.

84 <sup>41</sup> Department of Neurology, University of Melbourne, Royal Children's Hospital, Parkville,  
85 Victoria 3052, Australia.

86 <sup>42</sup> The University of Queensland, School of Biomedical Sciences, St Lucia Campus, Brisbane,  
87 4067, Australia.

88

89 † These authors contributed equally to this work.

90 § Equal last and senior authors

91 \* Corresponding authors. E-mail: depiennc@igbmc.fr (CD), paul.lockhart@mcri.edu.au (PJL),  
92 richards@uq.edu.au (LJR)

93

94

95

96 **ABSTRACT**

97 Brain malformations involving the corpus callosum are common in children with developmental  
98 disabilities. We report that *DCC* mutations cause isolated agenesis of the corpus callosum  
99 (ACC), without intellectual disability, in four families and five sporadic individuals. *DCC*  
100 mutations result in variable dominant phenotypes with reduced penetrance including mirror  
101 movements and ACC associated with a favorable developmental prognosis. Phenotype  
102 modifiers include the type and location of mutation and sex of the individual.

103

104 **MAIN TEXT**

105 The corpus callosum (CC) is the main cerebral commissure in placental mammals with a key  
106 role in communication between the brain hemispheres<sup>1</sup>. Formation of the CC is a complex  
107 process involving ligands such as those in the Netrin, Ephrin, Semaphorin and Slit families and  
108 their receptors<sup>2</sup>. Agenesis of the corpus callosum (ACC) is the complete or partial absence of  
109 the CC. This frequent brain malformation affects ~1/4,000 newborns and 3-5% of children with  
110 intellectual disability (ID)<sup>3,4</sup> and is a common cause of late pregnancy termination<sup>5</sup>. Mutations  
111 in many genes cause syndromes with ID and ACC, whereas the genetics of isolated ACC remain  
112 poorly understood<sup>3,6,7</sup>. The netrin receptor *Dcc* plays a critical role in CC development in mice  
113 by guiding callosal axons at the midline<sup>8</sup>. While mutations in *DCC* have been associated with  
114 congenital mirror movements (MM) in humans,<sup>9</sup> they have not been described in individuals  
115 with ACC.

116

117 We investigated four multigenerational families with individuals presenting with ACC, MM or  
118 both phenotypes segregating as autosomal dominant traits (Fig. 1a). Neuroimaging and clinical  
119 studies confirmed that complete or partial ACC was isolated in most cases (Fig. 1b, Fig. S4) and

120 associated with a range of intellectual abilities (normal-borderline impaired); additionally,  
121 specific cognitive impairments, including language delay or visuospatial deficits, were  
122 documented (Table S1). MRI tractography using probabilistic constrained spherical  
123 deconvolution identified reduced crossing of descending corticospinal tract projections at the  
124 pyramidal decussation in all affected individuals in families 2 and 4 with either ACC and MM  
125 (n=5) or MM only (n=2) (Fig. 1c, Fig. S1). The tractography results for other commissural  
126 fibers, including the decussation of the superior cerebellar peduncles, anterior commissure,  
127 posterior commissure and optic chiasm, were comparable between the affected individuals and  
128 controls.

129  
130 Linkage analysis and exome sequencing of three affected individuals in family 1 identified two  
131 shared candidate variants in the 16 linkage regions (Fig. S2, Table S2) including a truncating  
132 mutation in *DCC* (NM\_005215.3:c.925delA, p.Thr309Profs\*26). For family 2, 48 candidate  
133 variants were identified in 28 linkage regions (Fig. S3, Table S3), including a missense variant  
134 (c.2378T>G, p.Val793Gly) in *DCC*. The previously-reported nonsense mutation (c.823C>T,  
135 p.Arg275\*) in *DCC* segregated with MM in five individuals of family 3<sup>10</sup>. Further investigation  
136 revealed two additional female mutation carriers with ACC and a male carrier with MM who  
137 had a thin rostrum. Direct screening in family 4 identified a heterozygous *DCC* missense variant  
138 (c.2414G>A, p.Gly805Glu). All four *DCC* mutations were absent from public databases,  
139 including 1000 Genomes and ExAC, and segregated with ACC and/or MM in all available  
140 individuals tested (Fig. 1a). In addition, we sequenced *DCC* in 70 unrelated individuals with  
141 ACC including 46 with normal cognitive development. Five individuals, all with isolated  
142 complete ACC, had at least one heterozygous missense variant altering a conserved amino acid  
143 of *DCC* (Fig. 1d, Fig. S4 and S5, Table S4). Analysis of all available imaging in mutation-

144 positive individuals with complete ACC also showed absence of the hippocampal commissure  
145 and cingulate gyri, and dysmorphic lateral ventricles (usually colpocephaly) as would be  
146 expected (detailed in the supplementary materials and methods). Apart from this, no consistent  
147 additional brain malformations were seen.

148

149 We analyzed the phenotype of individuals with *DCC* mutations reported in the literature and in  
150 this study (Table S5) to assess the penetrance of MM and ACC. Of the 88 individuals with *DCC*  
151 mutations identified to date, 50 had MM; among the 39 who had brain imaging, 19 exhibited  
152 ACC. Excluding the index individuals from the analysis, the penetrance of MM was estimated  
153 to be 42% and the penetrance of ACC to be 26% (Table S6). Overall, males (n=31) exhibited  
154 MM more frequently than females (n=19, male:female ratio=1.8, p=0.0027, Fisher's Exact test;  
155 Table S6) while, in individuals with truncating *DCC* variants, ACC was more often present in  
156 females (n=7) than males (n=1, male:female ratio=0.2). Sex differences in CC anatomy have  
157 been associated with testosterone levels during prenatal brain development<sup>11-13</sup>; therefore, we  
158 tested the effect of androgens on *DCC* expression. Independent analysis by RNAseq and RT-  
159 qPCR demonstrated a significant dose-dependent increase in *DCC* expression in human neural  
160 stem cells treated with 10 nM or 100 nM testosterone (Fig. 1e, Fig. S6). Since variants  
161 introducing a premature stop codon generally result in haploinsufficiency due to nonsense  
162 mediated decay of the mutant mRNA, it is possible that ACC may occur when the amount of  
163 *DCC* mRNA/protein falls below a threshold level during CC development, which would occur  
164 more frequently in females. However, given the incomplete penetrance observed in both sexes,  
165 the phenotypic outcome must also be influenced by additional genetic, epigenetic and/or  
166 environmental factors. Interestingly, families 1 and 3, in which a majority of females display  
167 ACC, are both of North African background, supporting the hypothesis of genetic modifiers.

168 Contrary to truncating variants, missense mutant proteins are usually present in the cell and can  
169 interfere with the function of the wildtype protein, potentially resulting in differing phenotypes  
170 compared to haploinsufficiency for the same protein. Binding of Netrin-1 to DCC results in  
171 intracellular homodimerization or heterodimerization with UNC5, another axon guidance  
172 receptor, and is critical for both the chemoattractive and chemorepulsive properties of the  
173 signaling complexes<sup>15,16</sup>. The Netrin-1 binding region involves the 4th, 5th and 6th fibronectin  
174 type III-like domains of DCC<sup>14,15</sup>, therefore amino acid substitutions in this binding region may  
175 compromise DCC function. Five of the eight *DCC* missense variants identified in individuals  
176 with ACC are located in the Netrin-1 binding region (Fig. 1d), which represents a considerable  
177 enrichment compared to missense variants located in this domain in ExAC (5/74, 6.7% *versus*  
178 519/~60000, 0.86%;  $p=5 \times 10^{-4}$  (all rare variants) or 284/~60000, 0.47%;  $p=3 \times 10^{-5}$  (rare variants  
179 predicted to be damaging by SIFT), Fisher's exact test (Table S7)). Given the reduced  
180 penetrance and mild phenotype of *DCC*-related ACC, it is possible that some individuals  
181 described in ExAC have pathogenic *DCC* mutations and undiagnosed ACC.

182  
183 Modelling of *DCC* missense variants revealed that the amino acid substitutions in families 2  
184 and 4, both located within the *DCC*/netrin-1 binding interface, are predicted to be most  
185 disruptive. The p.Val793Gly substitution abolishes a hydrophobic interaction with Thr147 of  
186 Netrin-1 while p.Gly805Glu introduces a highly unfavorable charged moiety within a  
187 hydrophobic pocket, disrupting interaction with Leu113 of Netrin-1 (Fig. 1f). The predicted  
188 effects of the three substitutions within the Netrin-1 binding region but outside the binding  
189 interface (Fig. S7-11) are consistent with *in vitro* studies demonstrating that even conservative  
190 mutations to residues in this binding region can disrupt *DCC* dimerization, Netrin-1 binding and  
191 axon guidance<sup>14</sup>.

192

193 In addition to the effect of sex hormones and the type and location of *DCC* mutations,  
194 developmental differences between the CC and corticospinal tract may also contribute to the  
195 variable ACC/MM phenotypes. Callosal and sub-cerebrally projecting pyramidal neurons of the  
196 cortex are specified at early stages of development and the molecular identity of each population  
197 directly affects its axonal connectivity<sup>16</sup>. While corticospinal axons utilise *DCC*/netrin-1  
198 signalling to reach the midline, callosal axons use *DCC*/netrin-1 chemoattraction to attenuate  
199 *ROBO1*/*SLIT-2*-mediated chemorepulsion to approach and cross the midline<sup>17</sup>. Therefore, a  
200 *DCC* mutation may differentially affect commissural versus subcerebral axon trajectories,  
201 leading to ACC, MM or both. MM were consistently associated with reduced crossing of  
202 descending corticospinal tract projections at the pyramidal decussation in this study as well as  
203 in individuals with *RAD51*-related MM<sup>18</sup>, suggesting that *DCC*-mediated MM are primarily the  
204 result of corticospinal tract decussation abnormalities.

205

206 In conclusion, our results provide compelling evidence that *DCC* mutations cause isolated ACC  
207 in humans, in addition to the previously-reported MM phenotype. The factors determining the  
208 phenotypic variability are complex and likely include the hormonal context during development,  
209 the type and location of *DCC* mutation, and the genetic background of the individual. Although  
210 the full spectrum of phenotypes associated with *DCC* mutations remains to be fully  
211 characterized, individuals described in this study have an intellectual quotient within the  
212 normal/borderline range. Heterozygous mutations in *DCC* therefore appear to result in isolated  
213 ACC with a mild phenotype and favorable cognitive outcomes, contrasting with the pejorative  
214 developmental outcomes associated with syndromic ACC. Given the high frequency of *DCC*  
215 mutations detected in our cohorts, this observation has prenatal diagnostic and parental



216 counselling implications for fetuses with ACC as the condition currently has unclear  
217 prognostication. Our data suggests that the prenatal detection of a pathogenic *DCC* mutation  
218 indicates a lower risk of an abnormal neurodevelopmental outcome.

219  
220 **URLs.** 1000 Genomes Project, <http://www.1000genomes.org/>; Exome Variant Server,  
221 <http://exac.broadinstitute.org/>; SIFT, <http://sift.jcvi.org/>; PolyPhen-2,  
222 <http://genetics.bwh.harvard.edu/pph2/>.

223

## 224 **ACKNOWLEDGMENTS AND FUNDING**

225 We thank the families and the Lefroy family for their participation in and support of this study.

226 We thank the DNA and cell bank of the ICM (Paris, France) for DNA extraction, Sinead Eyre

227 (QBI) for study co-ordination and Mike Kean (RCH) and Marc Seal (MCRI) for assistance with

228 MRI protocols and scanning. This work was funded in part by National Health and Medical

229 Research Council (NHMRC) Australia Project Grants (GNT1059666, GNT631466,

230 GNT1064174, GNT1048849, GNT1104455, GNT1064174), the Agence Nationale de la

231 Recherche (ANR Blanc CILAXCAL, ANR Blanc HARTaGeNe), Assistance Publique des

232 Hôpitaux de Paris (APHP), the “programme hospitalier de recherche clinique” (PHRC)

233 ACCREM, and the “Investissements d’Avenir” programme ANR-10-IAIHU-06 (IHU-A-ICM).

234 APLM, TJE and LM are supported by an Australian Postgraduate Award and AP is supported

235 by a QBI PhD scholarship. SH and AQ are respectively supported by a master and a doctoral

236 grant from the Fondation pour la recherche médicale (FRM). MB is supported by an NHMRC

237 Senior Research Fellowship and an NHMRC Program Grant (APP1054618). EHS is supported

238 by a grant from the NIH, 2R01NS058721 and RJL is supported by a Melbourne Childrens

239 Clinician Scientist Fellowship. LJR is supported by an NMHRC Principal Research Fellowship,

240 and P.J.L. is supported by a NHMRC Career Development Fellowship (GNT1032364). CD and  
241 CN are members of the Bio-Psy Labex. This work has been supported by the Victorian  
242 Government's Operational Infrastructure Support Program and Australian Government  
243 NHMRC IRIISS.

244

#### 245 **AUTHOR CONTRIBUTIONS**

246 Ashley Marsh - formulation of theory and prediction, contributions to experimental conception  
247 and design, acquisition, analysis and/or interpretation of data and drafting the article and  
248 revising it critically for important intellectual content

249 Delphine Héron - contributions to experimental conception and design, interpretation of data  
250 and revising the article critically for important intellectual content

251 Timothy Edwards - formulation of theory and prediction, contributions to experimental  
252 conception and design, acquisition, analysis and/or interpretation of data and drafting the article  
253 and revising it critically for important intellectual content

254 Charles Galea - formulation of theory and prediction and acquisition, analysis and/or  
255 interpretation of data

256 Angélique Quartier- acquisition, analysis and interpretation of data

257 Caroline Nava - contributions to experimental conception and design, acquisition, analysis and  
258 interpretation of data, and revising the article critically for important intellectual content

259 Agnès Rastetter - acquisition, analysis and interpretation of data

260 Marie-Laure Moutard - acquisition, analysis and interpretation of data and revising the article  
261 critically for important intellectual content

262 Vicki Anderson - acquisition, analysis and/or interpretation of data and drafting the article or  
263 revising it critically for important intellectual content

- 264 Pierre Bitoun - acquisition, analysis and interpretation of data
- 265 Jens Bunt - acquisition, analysis and/or interpretation of data
- 266 Anne Faudet - acquisition, analysis and/or interpretation of data
- 267 Catherine Garel - acquisition, analysis and/or interpretation of data
- 268 Greta Gillies - acquisition, analysis and interpretation of data
- 269 Ilan Gobius - acquisition, analysis and interpretation of data
- 270 Justine Guegan - analysis of data
- 271 Solveig Heide - acquisition, analysis and/or interpretation of data
- 272 Boris Keren - acquisition, analysis and/or interpretation of data
- 273 Fabien Lesne - acquisition, analysis and/or interpretation of data
- 274 Vesna Lukic - acquisition, analysis and/or interpretation of data
- 275 Simone Mandelstam- acquisition, analysis and/or interpretation of data and drafting the article
- 276 or revising it critically for important intellectual content
- 277 George McGillivray - acquisition, analysis and/or interpretation of data and drafting the article
- 278 or revising it critically for important intellectual content
- 279 Alissandra McIlroy - acquisition, analysis and/or interpretation of data
- 280 Aurélie Meneret - acquisition, analysis and/or interpretation of data
- 281 Cyril Mignot - acquisition, analysis and/or interpretation of data, and revising the article
- 282 critically for important intellectual content
- 283 Laura Morcom - acquisition, analysis and/or interpretation of data
- 284 Sylvie Odent - acquisition, analysis and/or interpretation of data
- 285 Annalisa Paolino - acquisition, analysis and/or interpretation of data
- 286 Kate Pope - acquisition, analysis and/or interpretation of data
- 287 Florence Riant - acquisition, analysis and/or interpretation of data

- 288 Gail Robinson - acquisition, analysis and/or interpretation of data
- 289 Megan Spencer-Smith - acquisition, analysis and/or interpretation of data
- 290 Myriam Srour - acquisition, analysis and/or interpretation of data
- 291 Sarah Stephenson - contributions to experimental conception and design
- 292 Rick Tankard - acquisition, analysis and/or interpretation of data
- 293 Oriane Trouillard - acquisition, analysis and/or interpretation of data
- 294 Quentin Welniarz - acquisition, analysis and/or interpretation of data
- 295 Amanda Wood - acquisition, analysis and/or interpretation of data
- 296 Alexis Brice - acquisition, analysis and/or interpretation of data and revising the article critically  
297 for important intellectual content
- 298 Guy Rouleau - acquisition, analysis and/or interpretation of data and revising the article  
299 critically for important intellectual content
- 300 Tania Attié-Bitach- contributions to experimental design, and revising the article critically for  
301 important intellectual content
- 302 Martin Delatycki - drafting the article or revising it critically for important intellectual content
- 303 Jean Louis Mandel - contributions to experimental conception and design, interpretation of data  
304 and revising the article critically for important intellectual content
- 305 David Amor - drafting the article or revising it critically for important intellectual content
- 306 Emmanuel Roze - acquisition, analysis and/or interpretation of data and revising the article  
307 critically for important intellectual content
- 308 Amélie Piton - contributions to experimental conception and design, interpretation of data and  
309 revising the article critically for important intellectual content
- 310 Melanie Bahlo - acquisition, analysis and/or interpretation of data and drafting the article or  
311 revising it critically for important intellectual content

312 Thierry Billette de Villemeur - acquisition, analysis and/or interpretation of data and revising  
313 the article critically for important intellectual content

314 Elliott Sherr - formulation of theory and prediction, acquisition, analysis and/or interpretation  
315 of data and drafting the article and revising it critically for important intellectual content

316 Richard Leventer - formulation of theory and prediction, acquisition, analysis and/or  
317 interpretation of data and drafting the article and revising it critically for important intellectual  
318 content

319 Linda Richards - formulation of theory and prediction, contributions to experimental conception  
320 and design, acquisition, analysis and/or interpretation of data and drafting the article and  
321 revising it critically for important intellectual content

322 Paul Lockhart - formulation of theory and prediction, contributions to experimental conception  
323 and design, acquisition, analysis and/or interpretation of data and drafting the article and  
324 revising it critically for important intellectual content

325 Christel Depienne - formulation of theory and prediction, contributions to experimental  
326 conception and design, acquisition, analysis and/or interpretation of data and drafting the article  
327 and revising it critically for important intellectual content

328

### 329 **COMPETING FINANCIAL INTERESTS**

330 The authors declare no competing financial interests.

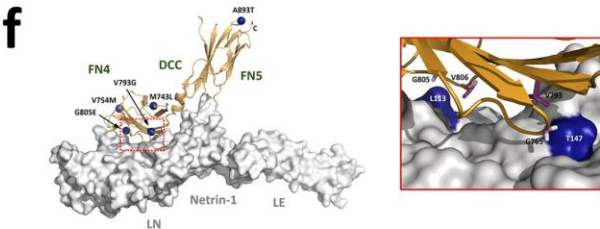
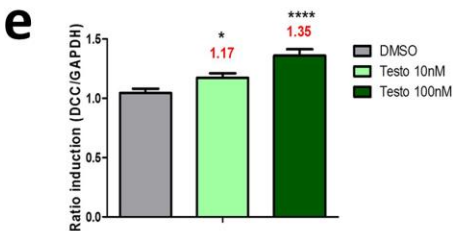
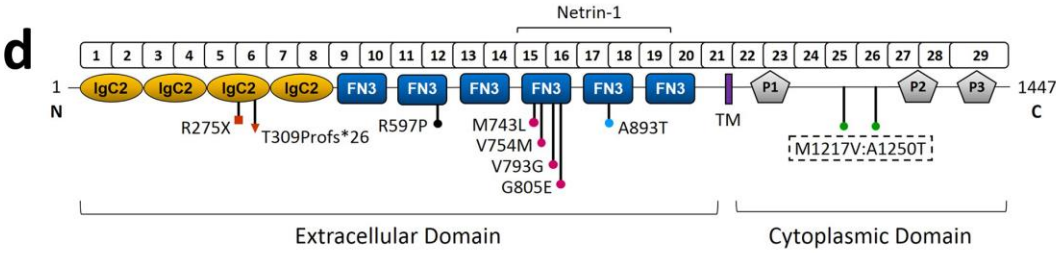
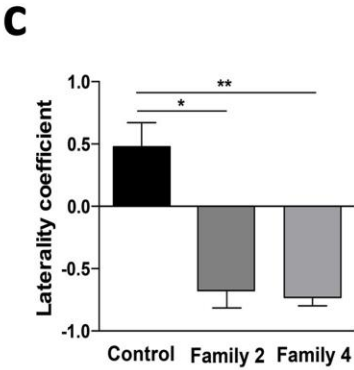
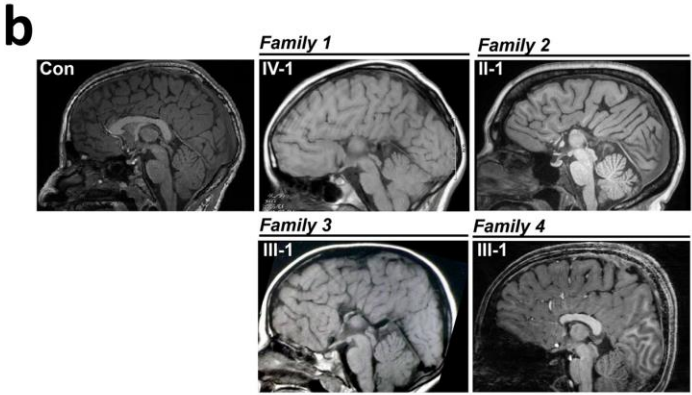
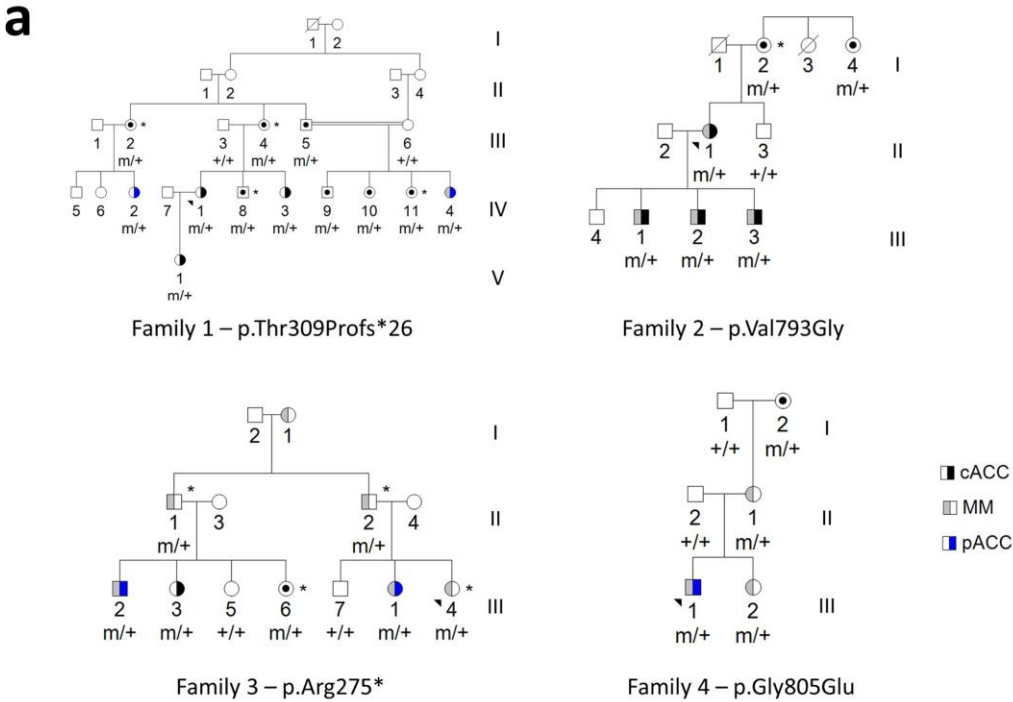
331

332 **REFERENCES**

- 333 1. Lindwall, C., Fothergill, T. & Richards, L.J. *Curr Opin Neurobiol* **17**, 3-14 (2007).
- 334 2. Chedotal, A. *Curr Opin Neurobiol* **21**, 68-75 (2011).
- 335 3. Paul, L.K. *et al. Nat Rev Neurosci* **8**, 287-99 (2007).
- 336 4. Glass, H.C., Shaw, G.M., Ma, C. & Sherr, E.H. *Am J Med Genet A* **146a**, 2495-500 (2008).
- 337 5. Rouleau, C. *et al. Arch Dis Child Fetal Neonatal Ed* **96**, F360-4 (2011).
- 338 6. Edwards, T.J., Sherr, E.H., Barkovich, A.J. & Richards, L.J. *Brain* **137**, 1579-613 (2014).
- 339 7. Sotiriadis, A. & Makrydimas, G. *Am J Obstet Gynecol* **206**, 337.e1-5 (2012).
- 340 8. Fazeli, A. *et al. Nature* **386**, 796-804 (1997).
- 341 9. Srour, M. *et al. Science* **328**, 592 (2010).
- 342 10. Meneret, A. *et al. Neurology* **82**, 1999-2002 (2014).
- 343 11. Ardekani, B.A., Figarsky, K. & Sidtis, J.J. *Cereb Cortex* **23**, 2514-20 (2013).
- 344 12. Moffat, S.D., Hampson, E., Wickett, J.C., Vernon, P.A. & Lee, D.H. *Brain Res* **767**, 297-
- 345 304 (1997).
- 346 13. Chura, L.R. *et al. Psychoneuroendocrinology* **35**, 122-32 (2010).
- 347 14. Finci, L.I. *et al. Neuron* **83**, 839-49 (2014).
- 348 15. Xu, K. *et al. Science* **344**, 1275-9 (2014).
- 349 16. Greig, L.C., Woodworth, M.B., Galazo, M.J., Padmanabhan, H. & Macklis, J.D. *Nat Rev*
- 350 *Neurosci* **14**, 755-69 (2013).
- 351 17. Fothergill, T. *et al. Cereb Cortex* **24**, 1138-51 (2014).
- 352 18. Gallea, C. *et al. Brain* **136**, 3333-46 (2013).
- 353

354 **Figure legend**355 **Figure 1** Characterization of *DCC* mutations.

356 **a.** All available family samples were analysed; m=mutation; black dot=mutation carrier;  
357 blue=partial ACC; black=complete ACC and grey=MM. \* indicates neuroimaging data for  
358 mutation carrier or individual with MM. **b.** Midsagittal MRI of control and family  
359 proband/representative individual (1-2=complete ACC; 3=near complete ACC with thin  
360 rostrum and genu remaining and 4=partial ACC with absence of the rostrum and genu). **c.**  
361 Group-wise comparison of laterality coefficient in both families (family 2, n=4; family 4, n=3)  
362 were compared to controls (n=6). For each individual, a laterality coefficient for the  
363 corticospinal tract was calculated as the ratio of the difference between the numbers of crossed  
364 and uncrossed streamlines to the total number of streamlines. Right and left coefficients were  
365 averaged to find the laterality coefficient of each subject. Greater positive values indicate more  
366 crossed and negative values more uncrossed streamlines (mean +/- S.D, \*=p<0.05; \*\*=p<0.01;  
367 two-tailed Mann-Whitney U-test). **d.** Protein domain structure depicting the location of the *DCC*  
368 truncation (red square and triangle) and missense mutations (colored dots). The Netrin-1 binding  
369 region is indicated, IgC2, immunoglobulin-like type C2 domain; FN, fibronectin type III-like  
370 domain; TM, transmembrane domain; P1-3, conserved motifs. **e.** RT-qPCR demonstrating  
371 upregulation of *DCC* expression in human NSCs treated with 10 nM (n=46) or 100 nM (n=48)  
372 compared with DMSO-treated cells (n=52). (mean +/- S.D, \*=p<0.05; \*\*\*\*=p<0.0001; two-  
373 sided parametric Student's t-test). **f.** Structural model with *DCC* the orange ribbon and netrin-1  
374 as a white solvent accessible surface, mutations are represented as blue spheres. Expansion of  
375 the binding interface (red dotted box) with Netrin-1 residues colored blue and critical *DCC*  
376 residues represented as purple sticks. LN = laminin domain, LE = laminin-type EGF-like  
377 domain, FN = fibronectin type III-like domain. (Protein Data Bank ID: 4PLO).







# References

---



- Aarnisalo, Palvimo & Jänne, 1998. CREB-binding protein in androgen receptor-mediated signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 95(5), pp.2122–2127.
- Abrahams *et al.*, 2013. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Molecular Autism*, 4(1), p.36.
- Ahlbom, Prins & Ceccatelli, 2001. Testosterone protects cerebellar granule cells from oxidative stress-induced cell death through a receptor mediated mechanism. *Brain Research*, 892(2), pp.255–262.
- Ahrens-Fath *et al.*, 2005. Androgen receptor function is modulated by the tissue-specific AR45 variant. *The FEBS journal*, 272(1), pp.74–84.
- Alexander & Peterson, 2004. Testing the prenatal hormone hypothesis of tic-related disorders: gender identity and gender role behavior. *Development and Psychopathology*, 16(2), pp.407–420.
- Amiel *et al.*, 2007. Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *American Journal of Human Genetics*, 80(5), pp.988–993.
- Anders & Huber, 2010. Differential expression analysis for sequence count data. *Genome Biology*, 11(10), p.R106.
- Anderson *et al.*, 2008. Examination of association to autism of common genetic variation in genes related to dopamine. *Autism Research: Official Journal of the International Society for Autism Research*, 1(6), pp.364–369.
- Anney *et al.*, 2010. A genome-wide scan for common alleles affecting risk for autism. *Human Molecular Genetics*, 19(20), pp.4072–4082.
- Anon, Le principe de la différenciation sexuelle - Base documentaire en SVT - Ressources pour les enseignants - Ressources lycée. *Assistance scolaire personnalisée et gratuite - ASP*. Available at: [http://www.assistancescolaire.com/enseignant/lycee/ressources/base-documentaire-ensvt/le-principe-de-la-differenciation-sexuelle-t\\_t203i02](http://www.assistancescolaire.com/enseignant/lycee/ressources/base-documentaire-ensvt/le-principe-de-la-differenciation-sexuelle-t_t203i02)
- Ariani *et al.*, 2008. FOXP1 is responsible for the congenital variant of Rett syndrome. *American Journal of Human Genetics*, 83(1), pp.89–93.
- Aronson, Hagberg & Gillberg, 1997. Attention deficits and autistic spectrum problems in children exposed to alcohol during gestation: a follow-up study. *Developmental Medicine and Child Neurology*, 39(9), pp.583–587.
- Asano *et al.*, 2010. DRR1 is expressed in the developing nervous system and downregulated during neuroblastoma carcinogenesis. *Biochemical and Biophysical Research Communications*, 394(3), pp.829–835.
- Auyeung *et al.*, 2006. Foetal testosterone and the child systemizing quotient. *European Journal of Endocrinology*, 155(suppl\_1), pp.S123–S130.
- Auyeung *et al.*, 2010. Foetal testosterone and autistic traits in 18 to 24-month-old children. *Molecular Autism*, 1, p.11.
- Auyeung, Lombardo & Baron-Cohen, 2013. Prenatal and postnatal hormone effects on the human brain and cognition. *Pflugers Archiv: European Journal of Physiology*, 465(5), pp.557–571.
- Avdjieva-Tzavella *et al.*, 2012. Analysis of the genes encoding neuroligins NLGN3 and NLGN4 in Bulgarian patients with autism. *Genetic Counseling (Geneva, Switzerland)*, 23(4), pp.505–511.
- Azad *et al.*, 2003. Testosterone treatment enhances regional brain perfusion in hypogonadal men. *The Journal of Clinical Endocrinology and Metabolism*, 88(7), pp.3064–3068.

- Badea *et al.*, 2002. RGC-32 increases p34CDC2 kinase activity and entry of aortic smooth muscle cells into S-phase. *The Journal of Biological Chemistry*, 277(1), pp.502–508.
- Bagasra *et al.*, 2013. Role of perfumes in pathogenesis of autism. *Medical Hypotheses*, 80(6), pp.795–803.
- Bagatell & Bremner, 2003. *Androgens in Health and Disease*, Springer Science & Business Media.
- Bagheri-Fam *et al.*, 2011. Defective survival of proliferating Sertoli cells and androgen receptor function in a mouse model of the ATR-X syndrome. *Human Molecular Genetics*, 20(11), pp.2213–2224.
- Bailey *et al.*, 1995. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological Medicine*, 25(1), pp.63–77.
- Bailey *et al.*, 1998. Autism: the phenotype in relatives. *Journal of Autism and Developmental Disorders*, 28(5), pp.369–392.
- Banerjee-Basu & Packer, 2010. SFARI Gene: an evolving database for the autism research community. *Disease Models & Mechanisms*, 3(3–4), pp.133–135.
- Bao *et al.*, 2006. A direct androgenic involvement in the expression of human corticotropin-releasing hormone. *Molecular Psychiatry*, 11(6), pp.567–576.
- Bao & Swaab, 2010. Sex differences in the brain, behavior, and neuropsychiatric disorders. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 16(5), pp.550–565.
- Bardsley *et al.*, 2013. 47,XXY syndrome: clinical phenotype and timing of ascertainment. *The Journal of Pediatrics*, 163(4), pp.1085–1094.
- Baron-Cohen, 2002. The extreme male brain theory of autism. *Trends in Cognitive Sciences*, 6(6), pp.248–254.
- Baron-Cohen & Wheelwright, 2004. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2), pp.163–175.
- Baron-Cohen, Lutchmaya & Knickmeyer, 2006. *Prenatal Testosterone in Mind*. MIT Press. Available at: <https://mitpress.mit.edu/books/prenatal-testosterone-mind>
- Baron-Cohen *et al.*, 2001. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), pp.5–17.
- Baron-Cohen *et al.*, 2003. The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1430), pp.361–374.
- Baron-Cohen *et al.*, 2011. Why are autism spectrum conditions more prevalent in males? *PLoS biology*, 9(6), p.e1001081.
- Baron-Cohen *et al.*, 2013. Do girls with anorexia nervosa have elevated autistic traits? *Molecular Autism*, 4(1), p.24.
- Baron-Cohen *et al.*, 2015. Elevated fetal steroidogenic activity in autism. *Molecular Psychiatry*, 20(3), pp.369–376.
- Basel-Vanagaite *et al.*, 2006. The CC2D1A, a member of a new gene family with C2 domains, is involved in autosomal recessive non-syndromic mental retardation. *Journal of Medical Genetics*, 43(3), pp.203–210.
- Bassell & Warren, 2008. Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron*, 60(2), pp.201–214.

- Basu, Kollu & Banerjee-Basu, 2009. AutDB: a gene reference resource for autism research. *Nucleic Acids Research*, 37(Database issue), pp.D832-836.
- Baumgartner *et al.*, 1999. Predictors of skeletal muscle mass in elderly men and women. *Mechanisms of Ageing and Development*, 107(2), pp.123–136.
- Beattie *et al.*, 1975. Role of chronic low-level lead exposure in the aetiology of mental retardation. *Lancet (London, England)*, 1(7907), pp.589–592.
- Beffert *et al.*, 2005. Modulation of synaptic plasticity and memory by Reelin involves differential splicing of the lipoprotein receptor Apoer2. *Neuron*, 47(4), pp.567–579.
- Begeer *et al.*, 2013. Sex differences in the timing of identification among children and adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 43(5), pp.1151–1156.
- Beilin *et al.*, 2000. Effect of the androgen receptor CAG repeat polymorphism on transcriptional activity: specificity in prostate and non-prostate cell lines. *Journal of Molecular Endocrinology*, 25(1), pp.85–96.
- Belgorosky *et al.*, 2009. Genetic and clinical spectrum of aromatase deficiency in infancy, childhood and adolescence. *Hormone Research*, 72(6), pp.321–330.
- Bellinger, Stiles, & Needleman, 1992. Low-level lead exposure, intelligence and academic achievement: a long-term follow-up study. *Pediatrics*, 90(6), pp.855–861.
- Benard *et al.*, 2010. IBRD2, an IBR-type E3 ubiquitin ligase, is a regulatory factor for Bax and apoptosis activation. *The EMBO journal*, 29(8), pp.1458–1471.
- Berenbaum & Resnick, 1997. Early androgen effects on aggression in children and adults with congenital adrenal hyperplasia. *Psychoneuroendocrinology*, 22(7), pp.505–515.
- Berenbaum & Snyder, 1995. Early hormonal influences on childhood sex-typed activity and playmate preferences: Implications for the development of sexual orientation. *Developmental Psychology*, 31(1), pp.31–42.
- Berenbaum, Duck & Bryk, 2000. Behavioral effects of prenatal versus postnatal androgen excess in children with 21-hydroxylase-deficient congenital adrenal hyperplasia. *The Journal of Clinical Endocrinology and Metabolism*, 85(2), pp.727–733.
- Berg, 1989. DNA binding specificity of steroid receptors. *Cell*, 57(7), pp.1065–1068.
- Berletch *et al.*, 2011. Genes that escape from X inactivation. *Human Genetics*, 130(2), pp.237–245.
- Betancur, 2011. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Research*, 1380, pp.42–77.
- Bezdicikova *et al.*, 2007. Distribution of nuclear receptors for steroid hormones in the human brain: a preliminary study. *Biomedical Papers of the Medical Faculty of the University Palacký, Olomouc, Czechoslovakia*, 151(1), pp.69–71.
- Biamonte *et al.*, 2009. Interactions between neuroactive steroids and reelin haploinsufficiency in Purkinje cell survival. *Neurobiology of Disease*, 36(1), pp.103–115.
- Biancalana *et al.*, 1992. Moderate instability of the trinucleotide repeat in spino bulbar muscular atrophy. *Human Molecular Genetics*, 1(4), pp.255–258.
- Biancalana *et al.*, 2005. FMR1 premutations associated with fragile X-associated tremor/ataxia syndrome in multiple system atrophy. *Archives of Neurology*, 62(6), pp.962–966.
- Billuart *et al.*, 1998. Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation. *Nature*, 392, n° 6679, pp. 923-926.
- Bing *et al.*, 2014. DHT Inhibits the A $\beta$ 25–35-Induced Apoptosis by Regulation of Seladin-1, Survivin, XIAP, bax, and bcl-xl Expression Through a Rapid PI3-K/Akt Signaling in C6 Glial Cell Lines. *Neurochemical Research*, 40(1), pp.41–48.

- Bishop *et al.*, 2011. Autism, language and communication in children with sex chromosome trisomies. *Archives of Disease in Childhood*, 96(10), pp.954–959.
- Black & Grant 2014. *DSM-5 guidebook : the essential companion to the Diagnostic and statistical manual of mental disorders, fifth edition*, Washington, DC : American Psychiatric Association.
- Blasi *et al.*, 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: The Official Publication of the International Society of Psychiatric Genetics*, 141B(3), pp.220–221.
- Bodega *et al.*, 2006. Influence of intermediate and uninterrupted FMR1 CGG expansions in premature ovarian failure manifestation. *Human Reproduction (Oxford, England)*, 21(4), pp.952–957.
- Boissart *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, p.e294.
- Bolton *et al.*, 2007. Cell- and gene-specific regulation of primary target genes by the androgen receptor. *Genes & Development*, 21(16), pp.2005–2017.
- Boorjian *et al.*, 2009. Expression and significance of androgen receptor coactivators in urothelial carcinoma of the bladder. *Endocrine-Related Cancer*, 16(1), pp.123–137.
- Boyle & Kaufmann, 2010. The behavioral phenotype of FMR1 mutations. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, 154C(4), pp.469–476.
- Bremer *et al.*, 2011. Copy number variation characteristics in subpopulations of patients with autism spectrum disorders. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 156(2), pp.115–124.
- Brinkmann *et al.*, 1989. The human androgen receptor: domain structure, genomic organization and regulation of expression. *Journal of Steroid Biochemistry*, 34(1–6), pp.307–310.
- Brooks *et al.*, 1998. A cell culture model for androgen effects in motor neurons. *Journal of Neurochemistry*, 70(3), pp.1054–1060.
- Brown *et al.*, 1989. Androgen receptor locus on the human X chromosome: regional localization to Xq11-12 and description of a DNA polymorphism. *American Journal of Human Genetics*, 44(2), pp.264–269.
- Bruining *et al.*, 2009. Psychiatric characteristics in a self-selected sample of boys with Klinefelter syndrome. *Pediatrics*, 123(5), pp.e865-870.
- Bruining *et al.*, 2010. The parent-of-origin of the extra X chromosome may differentially affect psychopathology in Klinefelter syndrome. *Biological Psychiatry*, 68(12), pp.1156–1162.
- Brunner *et al.*, 1993. X-linked borderline mental retardation with prominent behavioral disturbance: phenotype, genetic localization, and evidence for disturbed monoamine metabolism. *American Journal of Human Genetics*, 52(6), pp.1032–1039.
- Butler *et al.*, 2005. Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *Journal of Medical Genetics*, 42(4), pp.318–321.
- Butler, Leigh & Gallo, 2001. Androgen-induced up-regulation of tubulin isoforms in neuroblastoma cells. *Journal of Neurochemistry*, 78(4), pp.854–861.
- Butt *et al.*, 2003. Insulin-like growth factor-binding protein-5 inhibits the growth of human breast cancer cells in vitro and in vivo. *The Journal of Biological Chemistry*, 278(32), pp.29676–29685.
- Cannell, 2008. Autism and vitamin D. *Medical Hypotheses*, 70(4), pp.750–759.

- Carrel & Willard, 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, 434(7031), pp.400–404.
- Casavant, 2006. Antibiotic not linked to autism. *Medical Hypotheses*, 66(3), p.678.
- Casey *et al.*, 2012. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Human Genetics*, 131(4), pp.565–579.
- Cassidy *et al.*, 1983. Family studies in tuberous sclerosis. Evaluation of apparently unaffected parents. *JAMA*, 249(10), pp.1302–1304.
- Castellani *et al.*, 2009. Autism and immunity: revisited study. *International Journal of Immunopathology and Pharmacology*, 22(1), pp.15–19.
- Castoria *et al.*, 2003. Androgen-stimulated DNA synthesis and cytoskeletal changes in fibroblasts by a nontranscriptional receptor action. *The Journal of Cell Biology*, 161(3), pp.547–556.
- Chahrour *et al.*, 2012. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS genetics*, 8(4), p.e1002635.
- Chakrabarti & Fombonne, 2001. Pervasive developmental disorders in preschool children. *JAMA*, 285(24), pp.3093–3099.
- Chakrabarti *et al.*, 2009. Genes related to sex steroids, neural growth, and social-emotional behavior are associated with autistic traits, empathy, and Asperger syndrome. *Autism Research: Official Journal of the International Society for Autism Research*, 2(3), pp.157–177.
- Chapman *et al.*, 2006. Fetal testosterone and empathy: evidence from the empathy quotient (EQ) and the “reading the mind in the eyes” test. *Social Neuroscience*, 1(2), pp.135–148.
- Chaste *et al.*, 2007. Mutation screening of the ARX gene in patients with autism. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 144B(2), pp.228–230.
- Chaste *et al.*, 2015. A genome-wide association study of autism using the Simons Simplex Collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? *Biological Psychiatry*, 77(9), pp.775–784.
- Chaudhry *et al.*, 2015. Phenotypic spectrum associated with PTCHD1 deletions and truncating mutations includes intellectual disability and autism spectrum disorder. *Clinical Genetics*, 88(3), pp.224–233.
- Chen *et al.*, 2012. Androgen pathway stimulates microRNA-216a transcription to suppress the tumor suppressor in lung cancer-1 gene in early hepatocarcinogenesis. *Hepatology (Baltimore, Md.)*, 56(2), pp.632–643.
- Chen *et al.*, 2016. Maternal autoimmune diseases and the risk of autism spectrum disorders in offspring: A systematic review and meta-analysis. *Behavioural Brain Research*, 296, pp.61–69.
- Chih *et al.*, 2004. Disorder-associated mutations lead to functional inactivation of neuroligins. *Human Molecular Genetics*, 13(14), pp.1471–1477.
- Chisu *et al.*, 2006. Effects of testosterone on differentiation and oxidative stress resistance in C1300 neuroblastoma cells. *Neuro Endocrinology Letters*, 27(6), pp.807–812.
- Choate, Slayden & Resko, 1998. Immunocytochemical localization of androgen receptors in brains of developing and adult male rhesus monkeys. *Endocrine*, 8(1), pp.51–60.
- Chodak *et al.*, 1992. Nuclear localization of androgen receptor in heterogeneous samples of normal, hyperplastic and neoplastic human prostate. *The Journal of Urology*, 147(3 Pt 2), pp.798–803.
- Choong & Wilson, 1998. Trinucleotide repeats in the human androgen receptor: a molecular basis for disease. *Journal of Molecular Endocrinology*, 21(3), pp.235–257.



- Christensen *et al.*, 2013. Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA*, 309(16), pp.1696–1703.
- Chubykin *et al.*, 2007. Activity-dependent validation of excitatory versus inhibitory synapses by neuroligin-1 versus neuroligin-2. *Neuron*, 54(6), pp.919–931.
- Claessens *et al.*, 1996. The androgen-specific probasin response element 2 interacts differentially with androgen and glucocorticoid receptors. *The Journal of Biological Chemistry*, 271(32), pp.19013–19016.
- Claessens *et al.*, 2001. Selective DNA binding by the androgen receptor as a mechanism for hormone-specific gene regulation. *The Journal of Steroid Biochemistry and Molecular Biology*, 76(1–5), pp.23–30.
- Claessens *et al.*, 2008. Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nuclear Receptor Signaling*, 6, p.e008.
- Clépet *et al.*, 1993. The human SRY transcript. *Human Molecular Genetics*, 2(12), pp.2007–2012.
- Cleutjens *et al.*, 1996. Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specific antigen promoter. *The Journal of Biological Chemistry*, 271(11), pp.6379–6388.
- Cleutjens *et al.*, 1997. An androgen response element in a far upstream enhancer region is essential for high, androgen-regulated activity of the prostate-specific antigen promoter. *Molecular Endocrinology (Baltimore, Md.)*, 11(2), pp.148–161.
- Coffee *et al.*, 2009. Incidence of fragile X syndrome by newborn screening for methylated FMR1 DNA. *American Journal of Human Genetics*, 85(4), pp.503–514.
- Coffey & Robson, 2012. Regulation of the androgen receptor by post-translational modifications. *The Journal of Endocrinology*, 215(2), pp.221–237.
- Cohen *et al.*, 2003. Association of autism severity with a monoamine oxidase A functional polymorphism. *Clinical Genetics*, 64(3), pp.190–197.
- Cohen *et al.*, 2011. Autism severity is associated with child and maternal MAOA genotypes. *Clinical Genetics*, 79(4), pp.355–362.
- Collaer & Hines, 1995. Human behavioral sex differences: a role for gonadal hormones during early development? *Psychological Bulletin*, 118(1), pp.55–107.
- Collins *et al.*, 2010. Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. *American Journal of Medical Genetics. Part A*, 152A(10), pp.2512–2520.
- Comings *et al.*, 1999. Association of the androgen receptor gene (AR) with ADHD and conduct disorder. *Neuroreport*, 10(7), pp.1589–1592.
- Comoletti *et al.*, 2004. The Arg451Cys-neuroligin-3 mutation associated with autism reveals a defect in protein processing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 24(20), pp.4889–4893.
- Compagnone & Mellon, 2000. Neurosteroids: biosynthesis and function of these novel neuromodulators. *Frontiers in Neuroendocrinology*, 21(1), pp.1–56.
- Cooper *et al.*, 2011. A copy number variation morbidity map of developmental delay. *Nature Genetics*, 43(9), pp.838–846.
- Cornish, Turk & Hagerman, 2008. The fragile X continuum: new advances and perspectives. *Journal of intellectual disability research: JIDR*, 52(Pt 6), pp.469–482.
- Cossée *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: EJHG*, 14(4), pp.418–425.

- Cowan, Diamond & Welch, 2003. Polyglutamine protein aggregation and toxicity are linked to the cellular stress response. *Human Molecular Genetics*, 12(12), pp.1377–1391.
- Crepel *et al.*, 2010. DISC1 duplication in two brothers with autism and mild mental retardation. *Clinical Genetics*, 77(4), pp.389–394.
- Croen *et al.*, 2005. Maternal autoimmune diseases, asthma and allergies, and childhood autism spectrum disorders: a case-control study. *Archives of Pediatrics & Adolescent Medicine*, 159(2), pp.151–157.
- Croft, Larsen & Banerjee-Basu, 2011. Genetic Heterogeneity of Autism Spectrum Disorders. In Deutsch, ed. *Autism Spectrum Disorders: The Role of Genetics in Diagnosis and Treatment*. InTech. Available at: <http://www.intechopen.com/books/autism-spectrum-disorders-the-role-of-genetics-in-diagnosis-and-treatment/genetic-heterogeneity-of-autism-spectrum-disorders>
- Curran *et al.*, 2011. No association between a common single nucleotide polymorphism, rs4141463, in the MACROD2 gene and autism spectrum disorder. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 156B(6), pp.633–639.
- Cuscó *et al.*, 2009. Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. *Human Molecular Genetics*, 18(10), pp.1795–1804.
- Custer *et al.*, 2003. The role of the ankyrin-binding protein NrCAM in node of Ranvier formation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(31), pp.10032–10039.
- Czech *et al.*, 2012. The human testis-determining factor SRY localizes in midbrain dopamine neurons and regulates multiple components of catecholamine synthesis and metabolism. *Journal of Neurochemistry*, 122(2), pp.260–271.
- da Silva *et al.*, 2015. Testosterone Depletion Induces Demethylation of Murine Reelin Promoter CpG Dinucleotides: A Preliminary Study. *BioMed Research International*, 2015, p.286369.
- da Silva Dal Pizzol, Knop, & Mengue, 2006. Prenatal exposure to misoprostol and congenital anomalies: systematic review and meta-analysis. *Reproductive Toxicology (Elmsford, N.Y.)*, 22(4), pp.666–671.
- Dar *et al.*, 2014. N-terminal domain of the androgen receptor contains a region that can promote cytoplasmic localization. *The Journal of Steroid Biochemistry and Molecular Biology*, 139, pp.16–24.
- Davies *et al.*, 1997. Androgen insensitivity with mental retardation: a contiguous gene syndrome? *Journal of Medical Genetics*, 34(2), pp.158–160.
- Davis *et al.*, 1992. Autism and developmental abnormalities in children with perinatal cocaine exposure. *Journal of the National Medical Association*, 84(4), pp.315–319.
- Davis *et al.*, 2008. Pax6 3' deletion results in aniridia, autism and mental retardation. *Human Genetics*, 123(4), pp.371–378.
- De Boulle *et al.*, 1993. A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nature Genetics*, 3(1), pp.31–35.
- De Cock, Maas & van de Bor, 2012. Does perinatal exposure to endocrine disruptors induce autism spectrum and attention deficit hyperactivity disorders? Review. *Acta Paediatrica (Oslo, Norway: 1992)*, 101(8), pp.811–818.
- De Ferrari & Moon, 2006. The ups and downs of Wnt signaling in prevalent neurological disorders. *Oncogene*, 25(57), pp.7545–7553.

- De Jaco *et al.*, 2006. A mutation linked with autism reveals a common mechanism of endoplasmic reticulum retention for the alpha,beta-hydrolase fold protein family. *The Journal of Biological Chemistry*, 281(14), pp.9667–9676.
- De Kretser *et al.*, 2004. The role of activin, follistatin and inhibin in testicular physiology. *Molecular and Cellular Endocrinology*, 225(1–2), pp.57–64.
- De la Torre-Ubieta *et al.*, 2016. Advancing the understanding of autism disease mechanisms through genetics. *Nature Medicine*, 22(4), pp.345–361.
- De Ligt *et al.*, 2012. Diagnostic exome sequencing in persons with severe intellectual disability. *The New England Journal of Medicine*, 367(20), pp.1921–1929.
- Demyanenko *et al.*, 2011. NrCAM deletion causes topographic mistargeting of thalamocortical axons to the visual cortex and disrupts visual acuity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(4), pp.1545–1558.
- Denayer *et al.*, 2010. The rules of DNA recognition by the androgen receptor. *Molecular Endocrinology (Baltimore, Md.)*, 24(5), pp.898–913.
- DeStefano 2007. Vaccines and autism: evidence does not support a causal association. *Clinical Pharmacology and Therapeutics*, 82(6), pp.756–759.
- Devriendt *et al.*, 1997. Skewed X-chromosome inactivation in female carriers of dyskeratosis congenita. *American Journal of Human Genetics*, 60(3), pp.581–587.
- Dewing *et al.*, 2006. Direct regulation of adult brain function by the male-specific factor SRY. *Current biology: CB*, 16(4), pp.415–420.
- Dodson & Gorski, 1993. Testosterone propionate administration prevents the loss of neurons within the central part of the medial preoptic nucleus. *Journal of Neurobiology*, 24(1), pp.80–88.
- Doesburg *et al.*, 1997. Functional in vivo interaction between the amino-terminal, transactivation domain and the ligand binding domain of the androgen receptor. *Biochemistry*, 36(5), pp.1052–1064.
- DonCarlos *et al.*, 2003. Androgen receptor immunoreactivity in forebrain axons and dendrites in the rat. *Endocrinology*, 144(8), pp.3632–3638.
- DonCarlos *et al.*, 2006. Novel cellular phenotypes and subcellular sites for androgen action in the forebrain. *Neuroscience*, 138(3), pp.801–807.
- Donnelly *et al.*, 2000. Female with autistic disorder and monosomy X (Turner syndrome): parent-of-origin effect of the X chromosome. *American Journal of Medical Genetics*, 96(3), pp.312–316.
- Dorrington & Armstrong, 1975. Follicle-stimulating hormone stimulates estradiol-17beta synthesis in cultured Sertoli cells. *Proceedings of the National Academy of Sciences of the United States of America*, 72(7), pp.2677–2681.
- Duan *et al.*, 2008. Identification of TRIM22 as a RING finger E3 ubiquitin ligase. *Biochemical and Biophysical Research Communications*, 374(3), pp.502–506.
- Dunn, Nisula & Rodbard, 1981. Transport of steroid hormones: binding of 21 endogenous steroids to both testosterone-binding globulin and corticosteroid-binding globulin in human plasma. *The Journal of Clinical Endocrinology and Metabolism*, 53(1), pp.58–68.
- Durand *et al.*, 2006. Expression and genetic variability of PCDH11Y, a gene specific to Homo sapiens and candidate for susceptibility to psychiatric disorders. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 141B(1), pp.67–70.
- Dworzynski *et al.*, 2012. How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? *Journal of the American Academy of Child and Adolescent Psychiatry*, 51(8), pp.788–797.

- Edwards *et al.*, 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, 12(2), pp.241–253.
- Enstrom, Van de Water, & Ashwood, 2009. Autoimmunity in autism. *Current Opinion in Investigational Drugs (London, England: 2000)*, 10(5), pp.463–473.
- Falkenstein & Wehling, 2000. Nongenomically initiated steroid actions. *European Journal of Clinical Investigation*, 30 Suppl 3, pp.51–54.
- Fallon, 2005. Could one of the most widely prescribed antibiotics amoxicillin/clavulanate “augmentin” be a risk factor for autism? *Medical Hypotheses*, 64(2), pp.312–315.
- Fang *et al.*, 1996. Hsp90 regulates androgen receptor hormone binding affinity in vivo. *The Journal of Biological Chemistry*, 271(45), pp.28697–28702.
- Fargo *et al.*, 2008. Androgen regulation of axon growth and neurite extension in motoneurons. *Hormones and Behavior*, 53(5), pp.716–728.
- Fassio *et al.*, 2011. SYN1 loss-of-function mutations in autism and partial epilepsy cause impaired synaptic function. *Human Molecular Genetics*, 20(12), pp.2297–2307.
- Fichera *et al.*, 2002. A new MRXS locus maps to the X chromosome pericentromeric region: a new syndrome or narrow definition of Sutherland-Haan genetic locus? *Journal of Medical Genetics*, 39(4), pp.276–280.
- Fichera *et al.*, 2005. Skewed X-inactivation in a family with mental retardation and PQBP1 gene mutation. *Clinical Genetics*, 67(5), pp.446–447.
- Finegan, Bartleman & Wong, 1989. A window for the study of prenatal sex hormone influences on postnatal development. *The Journal of Genetic Psychology*, 150(1), pp.101–112.
- Folsom & Fatemi, 2013. The involvement of Reelin in neurodevelopmental disorders. *Neuropharmacology*, 68, pp.122–135.
- Folstein & Rutter, 1977. Genetic influences and infantile autism. *Nature*, 265(5596), pp.726–728.
- Fombonne *et al.*, 2006. Pervasive developmental disorders in Montreal, Quebec, Canada: prevalence and links with immunizations. *Pediatrics*, 118(1), pp.e139-150.
- Fombonne, Elsabbagh & Clarke, 2012. *Epidemiology of autism*, Centre of Excellence for Early Childhood Development and Strategic Knowledge Cluster on Early Child Development.
- Forger, 2009. Control of cell number in the sexually dimorphic brain and spinal cord. *Journal of Neuroendocrinology*, 21(4), pp.393–399.
- Freedman *et al.*, 1988. The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature*, 334(6182), pp.543–546.
- Gandin *et al.*, 2015. Excess of runs of homozygosity is associated with severe cognitive impairment in intellectual disability. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 17(5), pp.396–399.
- Ganong 2005. *Physiologie médicale*, De Boeck Supérieur.
- Gao & McPhaul, 1998. Functional Activities of the A and B Forms of the Human Androgen Receptor in Response to Androgen Receptor Agonists and Antagonists. *Molecular Endocrinology*, 12(5), pp.654–663.
- Garofalo *et al.*, 1993. Androgen receptor gene polymorphisms in amyotrophic lateral sclerosis. *Neuromuscular disorders: NMD*, 3(3), pp.195–199.
- Garolla *et al.*, 2005. Molecular analysis of the androgen receptor gene in testicular cancer. *Endocrine-Related Cancer*, 12(3), pp.645–655.
- Gaugler *et al.*, 2014. Most genetic risk for autism resides with common variation. *Nature Genetics*, 46(8), pp.881–885.

- Gauthier *et al.*, 2003. Mutation screening of FOXP2 in individuals diagnosed with autistic disorder. *American Journal of Medical Genetics. Part A*, 118A(2), pp.172–175.
- Gauthier *et al.*, 2005. NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 132B(1), pp.74–75.
- Germanaud *et al.*, 2011. The Renpenning syndrome spectrum: new clinical insights supported by 13 new PQBP1-mutated males. *Clinical Genetics*, 79(3), pp.225–235.
- Giarelli *et al.*, 2010. Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and Health Journal*, 3(2), pp.107–116.
- Gilman *et al.*, 2011. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron*, 70(5), pp.898–907.
- Gioeli & Paschal, B.M., 2012. Post-translational modification of the androgen receptor. *Molecular and Cellular Endocrinology*, 352(1–2), pp.70–78.
- Giovannucci *et al.*, 1997. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 94(7), pp.3320–3323.
- Girirajan *et al.*, 2011. Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLoS genetics*, 7(11), p.e1002334.
- Glasson *et al.*, 2004. Perinatal factors and the development of autism: a population study. *Archives of General Psychiatry*, 61(6), pp.618–627.
- Glessner *et al.*, 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, 459(7246), pp.569–573.
- Goines & Van de Water, 2010. The immune system's role in the biology of autism. *Current Opinion in Neurology*, 23(2), pp.111–117.
- Goldstein & Sengelaub, 1992. Timing and duration of dihydrotestosterone treatment affect the development of motoneuron number and morphology in a sexually dimorphic rat spinal nucleus. *The Journal of Comparative Neurology*, 326(1), pp.147–157.
- Golzio *et al.*, 2012. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature*, 485(7398), pp.363–367.
- Gong *et al.*, 2004. Association between the FOXP2 gene and autistic disorder in Chinese population. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 127B(1), pp.113–116.
- Gong *et al.*, 2008. Analysis of X chromosome inactivation in autism spectrum disorders. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 147B(6), pp.830–835.
- Gonzalez-Mantilla *et al.*, 2016. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *JAMA psychiatry*, 73(3), pp.275–283.
- Gorski, 1991. Father-to-daughter transmission of focal dermal hypoplasia associated with nonrandom X-inactivation: support for X-linked inheritance and paternal X chromosome mosaicism. *American Journal of Medical Genetics*, 40(3), pp.332–337.
- Gottlieb *et al.*, 2012. The androgen receptor gene mutations database: 2012 update. *Human Mutation*, 33(5), pp.887–894.
- Goulioumis *et al.*, 2011. Androgen receptor in laryngeal carcinoma: could there be an androgen-refractory tumor? *ISRN oncology*, 2011, p.180518.

- Goy & Deputte, 1996. The effects of diethylstilbestrol (DES) before birth on the development of masculine behavior in juvenile female rhesus monkeys. *Hormones and Behavior*, 30(4), pp.379–386.
- Goy, Bercovitch & McBair, 1988. Behavioral masculinization is independent of genital masculinization in prenatally androgenized female rhesus macaques. *Hormones and Behavior*, 22(4), pp.552–571.
- Graf *et al.*, 2004. Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell*, 119(7), pp.1013–1026.
- Grant, 2003. Synapse signalling complexes and networks: machines underlying cognition. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 25(12), pp.1229–1235.
- Grasso *et al.*, 2012. The mutational landscape of lethal castration-resistant prostate cancer. *Nature*, 487(7406), pp.239–243.
- Gregory, He & Wilson, 2001. The putative androgen receptor-A form results from in vitro proteolysis. *Journal of Molecular Endocrinology*, 27(3), pp.309–319.
- Gregory *et al.*, 2009. Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC medicine*, 7, p.62.
- Grino, Griffin & Wilson, 1990. Testosterone at high concentrations interacts with the human androgen receptor similarly to dihydrotestosterone. *Endocrinology*, 126(2), pp.1165–1172.
- Grønsvov *et al.*, 2011. A nonsense mutation in FMR1 causing fragile X syndrome. *European journal of human genetics: EJHG*, 19(4), pp.489–491.
- Grozeva *et al.*, 2015. Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Human Mutation*, 36(12), pp.1197–1204.
- Guilmatre *et al.*, 2009. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Archives of General Psychiatry*, 66(9), pp.947–956.
- Guo *et al.*, 2002. Nongenomic testosterone calcium signaling. Genotropic actions in androgen receptor-free macrophages. *The Journal of Biological Chemistry*, 277(33), pp.29600–29607.
- Hadj-Sahraoui *et al.*, 1996. Gender effect on Purkinje cell loss in the cerebellum of the heterozygous reeler mouse. *Journal of Neurogenetics*, 11(1–2), pp.45–58.
- Hagerman, Amiri & Cronister, 1991. Fragile X checklist. *American Journal of Medical Genetics*, 38(2–3), pp.283–287.
- Hallmayer *et al.*, 2011. Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of General Psychiatry*, 68(11), pp.1095–1102.
- Ham *et al.*, 1988. Characterization of response elements for androgens, glucocorticoids and progestins in mouse mammary tumour virus. *Nucleic Acids Research*, 16(12), pp.5263–5276.
- Hamann *et al.*, 2014. Brain responses to sexual images in 46,XY women with complete androgen insensitivity syndrome are female-typical. *Hormones and Behavior*, 66(5), pp.724–730.
- Hamdan, Gauthier, *et al.*, 2009. Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. *The New England Journal of Medicine*, 360(6), pp.599–605.
- Hamdan, Piton, *et al.*, 2009. De novo STXBP1 mutations in mental retardation and nonsyndromic epilepsy. *Annals of Neurology*, 65(6), pp.748–753.
- Hamdan *et al.*, 2010. De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *American Journal of Human Genetics*, 87(5), pp.671–678.

- Hamel *et al.*, 1994. Mental retardation, congenital heart defect, cleft palate, short stature, and facial anomalies: a new X-linked multiple congenital anomalies/mental retardation syndrome: clinical description and molecular studies. *American Journal of Medical Genetics*, 51(4), pp.591–597.
- Hammond *et al.*, 1997. Fragile X syndrome and deletions in FMR1: new case and review of the literature. *American Journal of Medical Genetics*, 72(4), pp.430–434.
- Hammond *et al.*, 2001. Testosterone-mediated neuroprotection through the androgen receptor in human primary neurons. *Journal of Neurochemistry*, 77(5), pp.1319–1326.
- Handt *et al.*, 2014. Point mutation frequency in the FMR1 gene as revealed by fragile X syndrome screening. *Molecular and Cellular Probes*, 28(5–6), pp.279–283.
- Hanssen & Fryns, J.P., 1995. Cowden syndrome. *Journal of Medical Genetics*, 32(2), pp.117–119.
- Hatanaka *et al.*, 2009. Androgen rapidly increases dendritic thorns of CA3 neurons in male rat hippocampus. *Biochemical and Biophysical Research Communications*, 381(4), pp.728–732.
- He *et al.*, 1999. Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH(2)-terminal domain. *The Journal of Biological Chemistry*, 274(52), pp.37219–37225.
- He *et al.*, 2013. Krüppel-like factor 8 is a novel androgen receptor co-activator in human prostate cancer. *Acta Pharmacologica Sinica*, 34(2), pp.282–288.
- Heard & Disteché, 2006. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes & Development*, 20(14), pp.1848–1867.
- Heemers & Tindall, 2007. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocrine Reviews*, 28(7), pp.778–808.
- Helleday *et al.*, 1993. Personality characteristics and platelet MAO activity in women with congenital adrenal hyperplasia (CAH). *Psychoneuroendocrinology*, 18(5–6), pp.343–354.
- Henningsson *et al.*, 2009. Possible association between the androgen receptor gene and autism spectrum disorder. *Psychoneuroendocrinology*, 34(5), pp.752–761.
- Henshall *et al.*, 2001. Altered expression of androgen receptor in the malignant epithelium and adjacent stroma is associated with early relapse in prostate cancer. *Cancer Research*, 61(2), pp.423–427.
- Higgins *et al.*, 2004. A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation. *Neurology*, 63(10), pp.1927–1931.
- Hines, Ahmed & Hughes, 2003. Psychological outcomes and gender-related development in complete androgen insensitivity syndrome. *Archives of Sexual Behavior*, 32(2), pp.93–101.
- Hong *et al.*, 2000. Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human RELN mutations. *Nature Genetics*, 26(1), pp.93–96.
- Housley *et al.*, 1990. Evidence that the conserved region in the steroid binding domain of the glucocorticoid receptor is required for both optimal binding of hsp90 and protection from proteolytic cleavage. A two-site model for hsp90 binding to the steroid binding domain. *The Journal of Biological Chemistry*, 265(22), pp.12778–12781.
- Hsiao *et al.*, 1999. The linkage of Kennedy's neuron disease to ARA24, the first identified androgen receptor polyglutamine region-associated coactivator. *The Journal of Biological Chemistry*, 274(29), pp.20229–20234.
- Hsiao *et al.*, 2000. Differential regulation of testosterone vs. 5 $\alpha$ -dihydrotestosterone by selective androgen response elements. *Molecular and Cellular Biochemistry*, 206(1–2), pp.169–175.

- Hsing *et al.*, 2000. Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China. *Cancer Research*, 60(18), pp.5111–5116.
- Hu *et al.*, 2009. Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal development and steroid biosynthesis. *PloS One*, 4(6), p.e5775.
- Hu *et al.*, 2016. X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Molecular Psychiatry*, 21(1), pp.133–148.
- Hultman, Sparén & Cnattingius, 2002. Perinatal risk factors for infantile autism. *Epidemiology (Cambridge, Mass.)*, 13(4), pp.417–423.
- Hutton, 2016. Does Rubella Cause Autism: A 2015 Reappraisal? *Frontiers in Human Neuroscience*, 10, p.25.
- Igarashi *et al.*, 1992. Strong correlation between the number of CAG repeats in androgen receptor genes and the clinical onset of features of spinal and bulbar muscular atrophy. *Neurology*, 42(12), pp.2300–2302.
- Imperato-McGinley & Zhu, 2002. Androgens and male physiology the syndrome of 5alpha-reductase-2 deficiency. *Molecular and Cellular Endocrinology*, 198(1–2), pp.51–59.
- Inozemtseva, Matute & Juárez, 2008. Learning disabilities spectrum and sexual dimorphic abilities in girls with congenital adrenal hyperplasia. *Journal of Child Neurology*, 23(8), pp.862–869.
- Iossifov *et al.*, 2012. De novo gene disruptions in children on the autistic spectrum. *Neuron*, 74(2), pp.285–299.
- Iossifov *et al.*, 2014. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515(7526), pp.216–221.
- Irvine *et al.*, 1995. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. - PubMed - NCBI. Available at: [zotero://attachment/667/](https://pubmed.ncbi.nlm.nih.gov/7666667/)
- Isaacs, 1984. Antagonistic effect of androgen on prostatic cell death. *The Prostate*, 5(5), pp.545–557.
- Iwasaki & Thomsen, 2014. The splicing factor PQBP1 regulates mesodermal and neural development through FGF signaling. *Development (Cambridge, England)*, 141(19), pp.3740–3751.
- Jacobs *et al.*, 1997. Turner syndrome: a cytogenetic and molecular study. *Annals of Human Genetics*, 61(Pt 6), pp.471–483.
- Jacquemont *et al.*, 2011. Clinical utility gene card for: fragile X mental retardation syndrome, fragile X-associated tremor/ataxia syndrome and fragile X-associated primary ovarian insufficiency. *European journal of human genetics: EJHG*, 19(9).
- Jacquemont *et al.*, 2014. A higher mutational burden in females supports a “female protective model” in neurodevelopmental disorders. *American Journal of Human Genetics*, 94(3), pp.415–425.
- Jamain *et al.*, 2002. Y chromosome haplogroups in autistic subjects. *Molecular Psychiatry*, 7(2), pp.217–219.
- Jamain *et al.*, 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genetics*, 34(1), pp.27–29.
- Jenster *et al.*, 1991. Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. *Molecular Endocrinology (Baltimore, Md.)*, 5(10), pp.1396–1404.
- Jiménez-Barrón *et al.*, 2015. Genome-wide variant analysis of simplex autism families with an integrative clinical-bioinformatics pipeline. *Cold Spring Harbor Molecular Case Studies*, 1(1), p.a000422.



- Johnson *et al.*, 2011. *Biologie - Version luxe*, De Boeck Supérieur.
- Kalscheuer *et al.*, 2003. Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. *Nature Genetics*, 35(4), pp.313–315.
- Kawata *et al.*, 2008. Steroid receptor signalling in the brain--lessons learned from molecular imaging. *Journal of Neuroendocrinology*, 20(6), pp.673–676.
- Kaya *et al.*, 2012. A novel X-linked disorder with developmental delay and autistic features. *Annals of Neurology*, 71(4), pp.498–508.
- Kelleher & Bear, 2008. The autistic neuron: troubled translation? *Cell*, 135(3), pp.401–406.
- Kennedy, Alter & Sung, 1968. Progressive proximal spinal and bulbar muscular atrophy of late onset. A sex-linked recessive trait. *Neurology*, 18(7), pp.671–680.
- Khan *et al.*, 2006. Multifunction steroid receptor coactivator, E6-associated protein, is involved in development of the prostate gland. *Molecular Endocrinology (Baltimore, Md.)*, 20(3), pp.544–559.
- Kido *et al.*, 2011. Expression of the human TSPY gene in the brains of transgenic mice suggests a potential role of this Y chromosome gene in neural functions. *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, 38(5), pp.181–191.
- Kimura *et al.*, 1993. Immunocytochemical localization of androgen receptor with polyclonal antibody in paraffin-embedded human tissues. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society*, 41(5), pp.671–678.
- King *et al.*, 2000. Early androgen treatment decreases cognitive function and catecholamine innervation in an animal model of ADHD. *Behavioural Brain Research*, 107(1–2), pp.35–43.
- Kirkovski, Enticott & Fitzgerald, 2013. A review of the role of female gender in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 43(11), pp.2584–2603.
- Kishino, Lalande & Wagstaff, 1997. UBE3A/E6-AP mutations cause Angelman syndrome. *Nature Genetics*, 15(1), pp.70–73.
- Kleefstra *et al.*, 2004. Genotype-phenotype studies in three families with mutations in the polyglutamine-binding protein 1 gene (PQBP1). *Clinical Genetics*, 66(4), pp.318–326.
- Knickmeyer & Baron-Cohen, 2006. Fetal testosterone and sex differences in typical social development and in autism. *Journal of Child Neurology*, 21(10), pp.825–845.
- Knickmeyer, Baron-Cohen, Fane, *et al.*, 2006. Androgens and autistic traits: A study of individuals with congenital adrenal hyperplasia. *Hormones and Behavior*, 50(1), pp.148–153.
- Knickmeyer, Baron-Cohen, Raggatt, *et al.*, 2006. Fetal testosterone and empathy. *Hormones and Behavior*, 49(3), pp.282–292.
- Knickmeyer *et al.*, 2005. Foetal testosterone, social relationships, and restricted interests in children. *Journal of Child Psychology and Psychiatry*, 46(2), pp.198–210.
- Koller *et al.*, 2004. Recent developments in low-level lead exposure and intellectual impairment in children. *Environmental Health Perspectives*, 112(9), pp.987–994.
- Komuro, Saeki & Kato, 1999a. Association of two nuclear proteins, Npw38 and NpwBP, via the interaction between the WW domain and a novel proline-rich motif containing glycine and arginine. *The Journal of Biological Chemistry*, 274(51), pp.36513–36519.
- Komuro, Saeki & Kato, 1999b. Npw38, a novel nuclear protein possessing a WW domain capable of activating basal transcription. *Nucleic Acids Research*, 27(9), pp.1957–1965.
- Kooy *et al.*, 1999. CAG repeat contraction in the androgen receptor gene in three brothers with mental retardation. *American Journal of Medical Genetics*, 85(3), pp.209–213.

- Kousteni *et al.*, 2003. Kinase-mediated regulation of common transcription factors accounts for the bone-protective effects of sex steroids. *The Journal of Clinical Investigation*, 111(11), pp.1651–1664.
- Krumm *et al.*, 2014. A de novo convergence of autism genetics and molecular neuroscience. *Trends in Neurosciences*, 37(2), pp.95–105.
- Krumm *et al.*, 2015. Excess of rare, inherited truncating mutations in autism. *Nature Genetics*, 47(6), pp.582–588.
- Kuil & Mulder, 1994. Mechanism of antiandrogen action: conformational changes of the receptor. *Molecular and Cellular Endocrinology*, 102(1–2), pp.R1-5.
- Kuiper *et al.*, 1993. Localization and hormonal stimulation of phosphorylation sites in the LNCaP-cell androgen receptor. *The Biochemical Journal*, 291 ( Pt 1), pp.95–101.
- Kujawa, Emeric & Jones, 1991. Testosterone differentially regulates the regenerative properties of injured hamster facial motoneurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 11(12), pp.3898–3906.
- Kujawa, Jacob & Jones, 1993. Testosterone regulation of the regenerative properties of injured rat sciatic motor neurons. *Journal of Neuroscience Research*, 35(3), pp.268–273.
- Kunde *et al.*, 2011. The X-chromosome-linked intellectual disability protein PQBP1 is a component of neuronal RNA granules and regulates the appearance of stress granules. *Human Molecular Genetics*, 20(24), pp.4916–4931.
- Kuo *et al.*, 2015. Genome-Wide Association Study for Autism Spectrum Disorder in Taiwanese Han Population. *PloS One*, 10(9), p.e0138695.
- Kurotaki *et al.*, 2002. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nature Genetics*, 30(4), pp.365–366.
- La Spada *et al.*, 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, 352(6330), pp.77–79.
- Lai, Baron-Cohen & Buxbaum, 2015. Understanding autism in the light of sex/gender. *Molecular Autism*, 6, p.24.
- Lai, Lombardo, *et al.*, 2015. Sex/gender differences and autism: setting the scene for future research. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54(1), pp.11–24.
- Lammert & Howell, 2016. RELN Mutations in Autism Spectrum Disorder. *Frontiers in Cellular Neuroscience*, 10, p.84.
- Landrigan, 2010. What causes autism? Exploring the environmental contribution. *Current Opinion in Pediatrics*, 22(2), pp.219–225.
- Landrigan *et al.*, 2004. Children's health and the environment: public health issues and challenges for risk assessment. *Environmental Health Perspectives*, 112(2), pp.257–265.
- Langley, Zhou, & Wilson, 1995. Evidence for an anti-parallel orientation of the ligand-activated human androgen receptor dimer. *The Journal of Biological Chemistry*, 270(50), pp.29983–29990.
- Larsson *et al.*, 2005. Risk factors for autism: perinatal factors, parental psychiatric history, and socioeconomic status. *American Journal of Epidemiology*, 161(10), pp.916-925-928.
- Laumonnier *et al.*, 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), pp.552–557.
- Lauritsen *et al.*, 2002. Investigation of two variants in the DOPA decarboxylase gene in patients with autism. *American Journal of Medical Genetics*, 114(4), pp.466–470.

- Le Meur *et al.*, 2010. MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *Journal of Medical Genetics*, 47(1), pp.22–29.
- Leblond *et al.*, 2012. Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. *PLoS genetics*, 8(2), p.e1002521.
- Leen *et al.*, 2010. Glucose transporter-1 deficiency syndrome: the expanding clinical and genetic spectrum of a treatable disorder. *Brain: A Journal of Neurology*, 133(Pt 3), pp.655–670.
- Lejeune, Turpin & Gautier, 1959. [Chromosomal diagnosis of mongolism]. *Archives Françaises De Pédiatrie*, 16, pp.962–963.
- Lemay *et al.*, 2003. A rational approach to the child with mental retardation for the paediatrician. *Paediatrics & Child Health*, 8(6), pp.345–356.
- Lenski *et al.*, 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. *American Journal of Human Genetics*, 74(4), pp.777–780.
- Leonard & Wen 2002. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Mental Retardation and Developmental Disabilities Research Reviews*, 8(3), pp.117–134.
- Leranth, Petnehazy & MacLusky, 2003. Gonadal hormones affect spine synaptic density in the CA1 hippocampal subfield of male rats. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(5), pp.1588–1592.
- Levin *et al.*, 2001. Persistent behavioral consequences of neonatal chlorpyrifos exposure in rats. *Brain Research. Developmental Brain Research*, 130(1), pp.83–89.
- Levy *et al.*, 2011. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*, 70(5), pp.886–897.
- Li *et al.*, 2004. Activation of FAK and Src are receptor-proximal events required for netrin signaling. *Nature neuroscience*, 7(11), pp.1213–1221.
- Li *et al.*, 2005. Absence of causative mutations and presence of autism-related allele in FOXP2 in Japanese autistic patients. *Brain & Development*, 27(3), pp.207–210.
- Libbey *et al.*, 2005. Autistic disorder and viral infections. *Journal of Neurovirology*, 11(1), pp.1–10.
- Lim *et al.*, 2013. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron*, 77(2), pp.235–242.
- Lin *et al.*, 2004. Regulation of androgen receptor signaling by PTEN (phosphatase and tensin homolog deleted on chromosome 10) tumor suppressor through distinct mechanisms in prostate cancer cells. *Molecular Endocrinology (Baltimore, Md.)*, 18(10), pp.2409–2423.
- Lin *et al.*, 2009. Integrated expression profiling and CHIP-seq analyses of the growth inhibition response program of the androgen receptor. *PLoS One*, 4(8), p.e6589.
- Liu *et al.*, 2009. Induction of tumor inhibition and apoptosis by a candidate tumor suppressor gene DRR1 on 3p21.1. *Oncology Reports*, 22(5), pp.1069–1075.
- Liu *et al.*, 2013. Lack of association between NLGN3, NLGN4, SHANK2 and SHANK3 gene variants and autism spectrum disorder in a Chinese population. *PLoS One*, 8(2), p.e56639.
- Lobaccaro *et al.*, 1993. Androgen receptor gene mutation in male breast cancer. *Human Molecular Genetics*, 2(11), pp.1799–1802.
- Lord, Rutter & Le Couteur, 1994. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), pp.659–685.

- Lord *et al.*, 2000. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), pp.205–223.
- Lorenz, Garcia-Segura & DonCarlos, 2005. Cellular phenotype of androgen receptor-immunoreactive nuclei in the developing and adult rat brain. *The Journal of Comparative Neurology*, 492(4), pp.456–468.
- Losel *et al.*, 2003. Nongenomic steroid action: controversies, questions, and answers. *Physiological Reviews*, 83(3), pp.965–1016.
- Lossi *et al.*, 2002. Abnormal expression of the KLF8 (ZNF741) gene in a female patient with an X;autosome translocation t(X;21)(p11.2;q22.3) and non-syndromic mental retardation. *Journal of Medical Genetics*, 39(2), pp.113–117.
- Loy, Sim & Yong, 2003. Filamin-A fragment localizes to the nucleus to regulate androgen receptor and coactivator functions. *Proceedings of the National Academy of Sciences of the United States of America*, 100(8), pp.4562–4567.
- Lu & Cantor, 2012. Allowing for sex differences increases power in a GWAS of multiplex Autism families. *Molecular Psychiatry*, 17(2), pp.215–222.
- Lubahn, Joseph, Sar *et al.*, 1988. The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate. *Molecular Endocrinology (Baltimore, Md.)*, 2(12), pp.1265–1275.
- Lubahn, Joseph, Sullivan *et al.*, 1988. Cloning of human androgen receptor complementary DNA and localization to the X chromosome. *Science (New York, N.Y.)*, 240(4850), pp.327–330.
- Lubs, Stevenson & Schwartz, 2012. Fragile X and X-linked intellectual disability: four decades of discovery. *American Journal of Human Genetics*, 90(4), pp.579–590.
- Lubs *et al.*, 2006. Golabi-Ito-Hall syndrome results from a missense mutation in the WW domain of the PQBP1 gene. *Journal of Medical Genetics*, 43(6), p.e30.
- Lugenbeel *et al.*, 1995. Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. *Nature Genetics*, 10(4), pp.483–485.
- Luo *et al.*, 2014. Cryptic FMR1 mosaic deletion in a phenotypically normal mother of a boy with fragile X syndrome: case report. *BMC medical genetics*, 15, p.125.
- Luo *et al.*, 2015. Mutational analyses of the FMR1 gene in Chinese pediatric population of fragile x suspects: low tolerance for point mutation. *Journal of Child Neurology*, 30(6), pp.803–806.
- Lutchmaya, Baron-Cohen & Raggatt, 2001. Foetal testosterone and vocabulary size in 18- and 24-month-old infants. *Infant Behavior and Development*, 24(4), pp.418–424.
- Lutchmaya, Baron-Cohen & Raggatt, 2002. Foetal testosterone and eye contact in 12-month-old human infants. Available at: <http://www.sciencedirect.com.gate2.inist.fr/science/article/pii/S0163638302000942>
- Ma *et al.*, 2009. A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Annals of Human Genetics*, 73(Pt 3), pp.263–273.
- MacLean *et al.*, 2008. Impaired skeletal muscle development and function in male, but not female, genomic androgen receptor knockout mice. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 22(8), pp.2676–2689.
- MacLusky *et al.*, 2006. Androgen modulation of hippocampal synaptic plasticity. *Neuroscience*, 138(3), pp.957–965.
- Makarova *et al.*, 2004. A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *The EMBO Journal*, 23(12), pp.2381–2391.

- Malandrini *et al.*, 2001. PAX6 mutation in a family with aniridia, congenital ptosis, and mental retardation. *Clinical Genetics*, 60(2), pp.151–154.
- Mandel & Chelly, 2004. Monogenic X-linked mental retardation: is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations. *European journal of human genetics: EJHG*, 12(9), pp.689–693.
- Mandy *et al.*, 2012. Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. *Journal of Autism and Developmental Disorders*, 42(7), pp.1304–1313.
- Manolagas, Kousteni & Jilka, 2002. Sex steroids and bone. *Recent Progress in Hormone Research*, 57, pp.385–409.
- Manson, 2008. Prenatal exposure to sex steroid hormones and behavioral/cognitive outcomes. *Metabolism: Clinical and Experimental*, 57 Suppl 2, pp.S16-21.
- Margari *et al.*, 2014. Autism spectrum disorders in XYY syndrome: two new cases and systematic review of the literature. *European Journal of Pediatrics*, 173(3), pp.277–283.
- Marron *et al.*, 2005. Androgen-induced neurite outgrowth is mediated by neuritin in motor neurones. *Journal of Neurochemistry*, 92(1), pp.10–20.
- Marshall *et al.*, 2008. Structural variation of chromosomes in autism spectrum disorder. *American Journal of Human Genetics*, 82(2), pp.477–488.
- Martínez-Garay *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: EJHG*, 15(1), pp.29–34.
- Marui *et al.*, 2005. No association of FOXP2 and PTPRZ1 on 7q31 with autism from the Japanese population. *Neuroscience Research*, 53(1), pp.91–94.
- Masana *et al.*, 2014. The stress-inducible actin-interacting protein DRR1 shapes social behavior. *Psychoneuroendocrinology*, 48, pp.98–110.
- Matsumoto *et al.*, 2008. Androgen receptor functions in male and female physiology. *The Journal of Steroid Biochemistry and Molecular Biology*, 109(3–5), pp.236–241.
- Mayer *et al.*, 1998. The Y-chromosomal genes SRY and ZFY are transcribed in adult human brain. *Neurogenetics*, 1(4), pp.281–288.
- McAbee & DonCarlos, 1998. Ontogeny of region-specific sex differences in androgen receptor messenger ribonucleic acid expression in the rat forebrain. *Endocrinology*, 139(4), pp.1738–1745.
- McAbee & DonCarlos, 1999. Regulation of androgen receptor messenger ribonucleic acid expression in the developing rat forebrain. *Endocrinology*, 140(4), pp.1807–1814.
- McFadden *et al.*, 2005. Physiological evidence of hypermasculinization in boys with the inattentive type of attention-deficit/hyperactivity disorder (ADHD). *Clinical Neuroscience Research*, 5(5–6), pp.233–245.
- McPhaul, 2002. Androgen receptor mutations and androgen insensitivity. *Molecular and Cellular Endocrinology*, 198(1–2), pp.61–67.
- McRae *et al.*, 2016. Prevalence, phenotype and architecture of developmental disorders caused by de novo mutation. *bioRxiv*, p.49056.
- Mendola *et al.*, 2002. Environmental factors associated with a spectrum of neurodevelopmental deficits. *Mental Retardation and Developmental Disabilities Research Reviews*, 8(3), pp.188–197.
- Menon, Yates & Bochar, 2010. Regulation of androgen-responsive transcription by the chromatin remodeling factor CHD8. *Molecular Endocrinology (Baltimore, Md.)*, 24(6), pp.1165–1174.

- Mhatre *et al.*, 1993. Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nature Genetics*, 5(2), pp.184–188.
- Michaelson *et al.*, 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell*, 151(7), pp.1431–1442.
- Mikhailov *et al.*, 2014. Screening of NLGN3 and NLGN4X genes in Thai children with autism spectrum disorder. *Psychiatric Genetics*, 24(1), pp.42–43.
- Militello *et al.*, 2002. The serum level of free testosterone is reduced in amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, 195(1), pp.67–70.
- Miller *et al.*, 2005. Autism associated with conditions characterized by developmental errors in early embryogenesis: a mini review. *International Journal of Developmental Neuroscience: The Official Journal of the International Society for Developmental Neuroscience*, 23(2–3), pp.201–219.
- Milsted *et al.*, 2004. Regulation of tyrosine hydroxylase gene transcription by Sry. *Neuroscience Letters*, 369(3), pp.203–207.
- Mizuguchi *et al.*, 2014. Mutations in the PQBP1 gene prevent its interaction with the spliceosomal protein U5-15 kD. *Nature Communications*, 5, p.3822.
- Moffat *et al.*, 2002. Longitudinal assessment of serum free testosterone concentration predicts memory performance and cognitive status in elderly men. *The Journal of Clinical Endocrinology and Metabolism*, 87(11), pp.5001–5007.
- Mohamed *et al.*, 2015. Assessment of Hair Aluminum, Lead, and Mercury in a Sample of Autistic Egyptian Children: Environmental Risk Factors of Heavy Metals in Autism. *Behavioural Neurology*, 2015, p.545674.
- Moilanen *et al.*, 1997. The presence of a transcription activation function in the hormone-binding domain of androgen receptor is revealed by studies in yeast cells. *FEBS letters*, 412(2), pp.355–358.
- Molinari *et al.*, 2002. Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation. *Science (New York, N.Y.)*, 298(5599), pp.1779–1781.
- Mong, Glaser & McCarthy, 1999. Gonadal steroids promote glial differentiation and alter neuronal morphology in the developing hypothalamus in a regionally specific manner. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 19(4), pp.1464–1472.
- Moore *et al.*, 2000. A clinical study of 57 children with fetal anticonvulsant syndromes. *Journal of Medical Genetics*, 37(7), pp.489–497.
- Moore *et al.*, 2010. An alternative splicing network links cell-cycle control to apoptosis. *Cell*, 142(4), pp.625–636.
- Morrow *et al.*, 2008. Identifying autism loci and genes by tracing recent shared ancestry. *Science (New York, N.Y.)*, 321(5886), pp.218–223.
- Moss & Howlin, 2009. Autism spectrum disorders in genetic syndromes: implications for diagnosis, intervention and understanding the wider autism spectrum disorder population. *Journal of intellectual disability research: JIDR*, 53(10), pp.852–873.
- Moy *et al.*, 2009. Impaired sociability and cognitive function in Nrcam-null mice. *Behavioural Brain Research*, 205(1), pp.123–131.
- Muhle, Trentacoste & Rapin, 2004. The genetics of autism. *Pediatrics*, 113(5), pp.e472–486.
- Müller & Skakkebaek, 1992. The prenatal and postnatal development of the testis. *Baillière's Clinical Endocrinology and Metabolism*, 6(2), pp.251–271.

- Munkley *et al.*, 2015. The androgen receptor controls expression of the cancer-associated sTn antigen and cell adhesion through induction of ST6GalNAc1 in prostate cancer. *Oncotarget*, 6(33), pp.34358–34374.
- Myrick *et al.*, 2014. Fragile X syndrome due to a missense mutation. *European journal of human genetics: EJHG*, 22(10), pp.1185–1189.
- Naftolin *et al.*, 1975. The formation of estrogens by central neuroendocrine tissues. *Recent Progress in Hormone Research*, 31, pp.295–319.
- Nagamani *et al.*, 1979. Maternal and amniotic fluid steroids throughout human pregnancy. *American Journal of Obstetrics and Gynecology*, 134(6), pp.674–680.
- Najmabadi *et al.*, 2011. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*, 478(7367), pp.57–63.
- Nakamura *et al.*, 1997. The reduction of androgen receptor mRNA in motoneurons of X-linked spinal and bulbar muscular atrophy. *Journal of the Neurological Sciences*, 150(2), pp.161–165.
- Nanson, 1992. Autism in fetal alcohol syndrome: a report of six cases. *Alcoholism, Clinical and Experimental Research*, 16(3), pp.558–565.
- Nass & Baker, 1991. Learning disabilities in children with congenital adrenal hyperplasia. *Journal of Child Neurology*, 6(4), pp.306–312.
- Naumova *et al.*, 1998. Parental origin-dependent, male offspring-specific transmission-ratio distortion at loci on the human X chromosome. *American Journal of Human Genetics*, 62(6), pp.1493–1499.
- Neale *et al.*, 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, 485(7397), pp.242–245.
- Needham *et al.*, 2000. Differential interaction of steroid hormone receptors with LXXLL motifs in SRC-1a depends on residues flanking the motif. *The Journal of Steroid Biochemistry and Molecular Biology*, 72(1–2), pp.35–46.
- Neggers, 2014. Increasing prevalence, changes in diagnostic criteria, and nutritional risk factors for autism spectrum disorders. *ISRN nutrition*, 2014, p.514026.
- Nemoto *et al.*, 1994. Dimerization characteristics of the DNA- and steroid-binding domains of the androgen receptor. *The Journal of Steroid Biochemistry and Molecular Biology*, 50(5–6), pp.225–233.
- Newbury *et al.*, 2002. FOXP2 is not a major susceptibility gene for autism or specific language impairment. *American Journal of Human Genetics*, 70(5), pp.1318–1327.
- Newschaffer *et al.*, 2007. The epidemiology of autism spectrum disorders. *Annual Review of Public Health*, 28, pp.235–258.
- Ng *et al.*, 2003. p53RFP, a p53-inducible RING-finger protein, regulates the stability of p21WAF1. *Oncogene*, 22(28), pp.4449–4458.
- Ng *et al.*, 2010. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature Genetics*, 42(9), pp.790–793.
- Ngan *et al.*, 2009. Microarray coupled to quantitative RT-PCR analysis of androgen-regulated genes in human LNCaP prostate cancer cells. *Oncogene*, 28(19), pp.2051–2063.
- Nguyen & Distèche, 2006. High expression of the mammalian X chromosome in brain. *Brain Research*, 1126(1), pp.46–49.
- Nguyen, Yao & Pike, 2005. Androgens activate mitogen-activated protein kinase signaling: role in neuroprotection. *Journal of Neurochemistry*, 94(6), pp.1639–1651.
- Nguyen, Yao & Pike, 2007. Flutamide and cyproterone acetate exert agonist effects: induction of androgen receptor-dependent neuroprotection. *Endocrinology*, 148(6), pp.2936–2943.

- Nguyen *et al.*, 2010. Androgens Selectively Protect Against Apoptosis in Hippocampal Neurons. *Journal of Neuroendocrinology*, 22(9), pp.1013–1022.
- Ni *et al.*, 2013. Androgen induces a switch from cytoplasmic retention to nuclear import of the androgen receptor. *Molecular and Cellular Biology*, 33(24), pp.4766–4778.
- Nieschlag, Behre & Nieschlag, 2012. *Testosterone: Action, Deficiency, Substitution*, Cambridge University Press.
- Nitkowska *et al.*, 2015. Prolactin and sex hormones levels in males with Parkinson's disease. *Acta Neurologica Scandinavica*, 131(6), pp.411–416.
- Noor *et al.*, 2010. Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. *Science Translational Medicine*, 2(49), p.49ra68.
- Nordahl *et al.*, 2011. Brain enlargement is associated with regression in preschool-age boys with autism spectrum disorders. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), pp.20195–20200.
- Nordeen *et al.*, 1985. Androgens prevent normally occurring cell death in a sexually dimorphic spinal nucleus. *Science (New York, N.Y.)*, 229(4714), pp.671–673.
- Nusse, 2009. A list of target genes of Wnt/beta-catenin signaling. Available at: [https://web.stanford.edu/group/nusselab/cgi-bin/wnt/target\\_genes](https://web.stanford.edu/group/nusselab/cgi-bin/wnt/target_genes).
- Obad *et al.*, 2004. Staf50 is a novel p53 target gene conferring reduced clonogenic growth of leukemic U-937 cells. *Oncogene*, 23(23), pp.4050–4059.
- Obendorf *et al.*, 2007. FoxG1, a member of the forkhead family, is a corepressor of the androgen receptor. *The Journal of Steroid Biochemistry and Molecular Biology*, 104(3–5), pp.195–207.
- Oberlé *et al.*, 1991. New polymorphism and a new chromosome breakpoint establish the physical and genetic mapping of DXS369 in the DXS98-FRAXA interval. *American Journal of Medical Genetics*, 38(2–3), pp.336–342.
- Ornoy, Weinstein-Fudim & Ergaz, 2015. Prenatal factors associated with autism spectrum disorder (ASD). *Reproductive Toxicology (Elmsford, N.Y.)*, 56, pp.155–169.
- O'Roak *et al.*, 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics*, 43(6), pp.585–589.
- O'Roak *et al.*, 2012. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, 485(7397), pp.246–250.
- Orwoll & Klein, 1995. Osteoporosis in men. *Endocrine Reviews*, 16(1), pp.87–116.
- Ostatníková *et al.*, 2002. Salivary testosterone levels in preadolescent children. *BMC pediatrics*, 2, p.5.
- Ozonoff *et al.*, 2011. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics*, 128(3), pp.e488–495.
- Paniagua *et al.*, 1991. Ultrastructure of the aging human testis. *Journal of Electron Microscopy Technique*, 19(2), pp.241–260.
- Pankhurst & McLennan, 2012. Inhibin B and anti-Müllerian hormone/Müllerian-inhibiting substance may contribute to the male bias in autism. *Translational Psychiatry*, 2, p.e148.
- Parikshak *et al.*, 2013. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*, 155(5), pp.1008–1021.
- Pawlowski *et al.*, 2002. Liganded androgen receptor interaction with beta-catenin: nuclear co-localization and modulation of transcriptional activity in neuronal cells. *The Journal of Biological Chemistry*, 277(23), pp.20702–20710.
- Peixoto *et al.*, 2012. Transsynaptic signaling by activity-dependent cleavage of neuroligin-1. *Neuron*, 76(2), pp.396–409.



- Pelekanou *et al.*, 2013. Androgen receptors in early and castration resistant prostate cancer: friend or foe? *Hormones (Athens, Greece)*, 12(2), pp.224–235.
- Pereira de Jésus-Tran *et al.*, 2006. Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. *Protein Science: A Publication of the Protein Society*, 15(5), pp.987–999.
- Peterson *et al.*, 1998. A double-blind, placebo-controlled, crossover trial of an antiandrogen in the treatment of Tourette's syndrome. *Journal of Clinical Psychopharmacology*, 18(4), pp.324–331.
- Petrij *et al.*, 1995. Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature*, 376(6538), pp.348–351.
- Pietras & Szego, 1977. Specific binding sites for oestrogen at the outer surfaces of isolated endometrial cells. *Nature*, 265(5589), pp.69–72.
- Pinto *et al.*, 2010. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, 466(7304), pp.368–372.
- Pinto *et al.*, 2014. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics*, 94(5), pp.677–694.
- Piton, Redin & Mandel, 2013. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *American Journal of Human Genetics*, 93(2), pp.368–383.
- Piton *et al.*, 2008. Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Human Molecular Genetics*, 17(24), pp.3965–3974.
- Piton *et al.*, 2011. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Molecular Psychiatry*, 16(8), pp.867–880.
- Piton *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: EJHG*, 22(6), pp.776–783.
- Plante *et al.*, 1996. Elevated androgen, brain development and language/learning disabilities in children with congenital adrenal hyperplasia. *Developmental Medicine and Child Neurology*, 38(5), pp.423–437.
- Pomerantz & Sholl, 1987. Analysis of sex and regional differences in androgen receptors in fetal rhesus monkey brain. *Brain Research*, 433(1), pp.151–154.
- Poukka *et al.*, 1999. Ubc9 interacts with the androgen receptor and activates receptor-dependent transcription. *The Journal of Biological Chemistry*, 274(27), pp.19441–19446.
- Puy *et al.*, 1995. Immunocytochemical detection of androgen receptor in human temporal cortex characterization and application of polyclonal androgen receptor antibodies in frozen and paraffin-embedded tissues. *The Journal of Steroid Biochemistry and Molecular Biology*, 55(2), pp.197–209.
- Ramsden, Shin & Pike, 2003. Androgens modulate neuronal vulnerability to kainate lesion. *Neuroscience*, 122(3), pp.573–578.
- Rasalam *et al.*, 2005. Characteristics of fetal anticonvulsant syndrome associated autistic disorder. *Developmental Medicine and Child Neurology*, 47(8), pp.551–555.
- Raskin *et al.*, 2009. Conditional inactivation of androgen receptor gene in the nervous system: effects on male behavioral and neuroendocrine responses. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(14), pp.4461–4470.

- Rauch *et al.*, 2012. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet (London, England)*, 380(9854), pp.1674–1682.
- Reddy, 2005. Cytogenetic abnormalities and fragile-X syndrome in Autism Spectrum Disorder. *BMC medical genetics*, 6, p.3.
- Reddy *et al.*, 2015. Effect of Testosterone on Neuronal Morphology and Neuritic Growth of Fetal Lamb Hypothalamus-Preoptic Area and Cerebral Cortex in Primary Culture. *PloS One*, 10(6), p.e0129521.
- Redin *et al.*, 2014. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *Journal of Medical Genetics*, 51(11), pp.724–736.
- Reichelt, Rodgers & Clapcote, 2012. The role of neurexins in schizophrenia and autistic spectrum disorder. *Neuropharmacology*, 62(3), pp.1519–1526.
- Rejeb *et al.*, 2011. A novel frame shift mutation in the PQBP1 gene identified in a Tunisian family with X-linked mental retardation. *European Journal of Medical Genetics*, 54(3), pp.241–246.
- Renpenning *et al.*, 1962. Familial sex-linked mental retardation. *Canadian Medical Association Journal*, 87, pp.954–956.
- Rhind *et al.*, 2014. An examination of autism spectrum traits in adolescents with anorexia nervosa and their parents. *Molecular Autism*, 5(1), p.56.
- Riazuddin *et al.*, 2016. Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. *Molecular Psychiatry*.
- Ricciardelli *et al.*, 2005. Androgen receptor levels in prostate cancer epithelial and peritumoral stromal cells identify non-organ confined disease. *The Prostate*, 63(1), pp.19–28.
- Richards *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: Official Journal of the American College of Medical Genetics*, 17(5), pp.405–424.
- Riegman *et al.*, 1991. The promoter of the prostate-specific antigen gene contains a functional androgen responsive element. *Molecular Endocrinology (Baltimore, Md.)*, 5(12), pp.1921–1930.
- Ritvo *et al.*, 1989. The UCLA-University of Utah epidemiologic survey of autism: recurrence risk estimates and genetic counseling. *The American Journal of Psychiatry*, 146(8), pp.1032–1036.
- Roberts *et al.*, 2013. Perinatal air pollutant exposures and autism spectrum disorder in the children of Nurses' Health Study II participants. *Environmental Health Perspectives*, 121(8), pp.978–984.
- Robinson *et al.*, 1977. Amniotic fluid androgens and estrogens in midgestation. *The Journal of Clinical Endocrinology and Metabolism*, 45(4), pp.755–761.
- Roche, Hoare & Parker, 1992. A consensus DNA-binding site for the androgen receptor. *Molecular Endocrinology (Baltimore, Md.)*, 6(12), pp.2229–2235.
- Roelfsema *et al.*, 2005. Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. *American Journal of Human Genetics*, 76(4), pp.572–580.
- Rogers *et al.*, 2008. Finding new etiologies of mental retardation and hypotonia: X marks the spot. *Developmental Medicine and Child Neurology*, 50(2), pp.104–111.
- Romano *et al.*, 2016. Genes and sex hormones interaction in neurodevelopmental disorders. *Neuroscience and Biobehavioral Reviews*, 67, pp.9–24.

- Ronemus *et al.*, 2014. The role of de novo mutations in the genetics of autism spectrum disorders. *Nature Reviews. Genetics*, 15(2), pp.133–141.
- Rooman, Van Driessche & Du Caju, 2002. Growth and ovarian function in girls with 48,XXXX karyotype--patient report and review of the literature. *Journal of pediatric endocrinology & metabolism: JPEM*, 15(7), pp.1051–1055.
- Ropers, 2010. Genetics of early onset cognitive impairment. *Annual Review of Genomics and Human Genetics*, 11, pp.161–187.
- Ropers & Hamel, 2005. X-linked mental retardation. *Nature Reviews. Genetics*, 6(1), pp.46–57.
- Rosario *et al.*, 2004. Age-related testosterone depletion and the development of Alzheimer disease. *JAMA*, 292(12), pp.1431–1432.
- Ross *et al.*, 2015. Behavioral phenotypes in males with XYY and possible role of increased NLGN4Y expression in autism features. *Genes, brain, and behavior*, 14(2), pp.137–144.
- Roulet, Lai & Foster, 2013. In utero exposure to valproic acid and autism--a current review of clinical and animal studies. *Neurotoxicology and Teratology*, 36, pp.47–56.
- Rubenstein, Wiggins & Lee, 2015. A Review of the Differences in Developmental, Psychiatric, and Medical Endophenotypes Between Males and Females with Autism Spectrum Disorder. *Journal of Developmental and Physical Disabilities*, 27(1), pp.119–139.
- Ruizeveld de Winter *et al.*, 1994. Androgen receptor status in localized and locally progressive hormone refractory human prostate cancer. *The American Journal of Pathology*, 144(4), pp.735–746.
- Russo *et al.*, 2009. Novel mutations in the CDKL5 gene, predicted effects and associated phenotypes. *Neurogenetics*, 10(3), pp.241–250.
- Rutter, Caspi & Moffitt, 2003. Using sex differences in psychopathology to study causal mechanisms: unifying issues and research strategies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 44(8), pp.1092–1115.
- Saez, 1994. Leydig cells: endocrine, paracrine, and autocrine regulation. *Endocrine Reviews*, 15(5), pp.574–626.
- Sampson *et al.*, 2001. Identification and characterization of androgen receptor associated coregulators in prostate cancer cells. *Journal of Biological Regulators and Homeostatic Agents*, 15(2), pp.123–129.
- Samsel & Seneff, 2015. Glyphosate, pathways to modern diseases III: Manganese, neurological diseases, and associated pathologies. *Surgical Neurology International*, 6, p.45.
- Sanders *et al.*, 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*, 70(5), pp.863–885.
- Sanders *et al.*, 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485(7397), pp.237–241.
- Sandin *et al.*, 2014. The familial risk of autism. *JAMA*, 311(17), pp.1770–1777.
- Sar & Stumpf, 1977. Androgen concentration in motor neurons of cranial nerves and spinal cord. *Science (New York, N.Y.)*, 197(4298), pp.77–79.
- Sato *et al.*, 2004. Brain masculinization requires androgen receptor function. *Proceedings of the National Academy of Sciences of the United States of America*, 101(6), pp.1673–1678.
- Sato *et al.*, 2012. SHANK1 Deletions in Males with Autism Spectrum Disorder. *American Journal of Human Genetics*, 90(5), pp.879–887.
- Schaaf *et al.*, 2011. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Human Molecular Genetics*, 20(17), pp.3366–3375.

- Schaafsma. & Pfaff, 2014. Etiologies underlying sex differences in Autism Spectrum Disorders. *Frontiers in Neuroendocrinology*, 35(3), pp.255–271.
- Scheiffele *et al.*, 2000. Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell*, 101(6), pp.657–669.
- Schmidt *et al.*, 2011. Tumor suppressor down-regulated in renal cell carcinoma 1 (DRR1) is a stress-induced actin bundling factor that modulates synaptic efficacy and cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 108(41), pp.17213–17218.
- Schoenmakers *et al.*, 2000. Differences in DNA binding characteristics of the androgen and glucocorticoid receptors can determine hormone-specific responses. *The Journal of Biological Chemistry*, 275(16), pp.12290–12297.
- Schulze, 1984. Sertoli cells and Leydig cells in man. *Advances in Anatomy, Embryology, and Cell Biology*, 88, pp.1–104.
- Schwarz *et al.*, 2011. Sex-specific serum biomarker patterns in adults with Asperger's syndrome. *Molecular Psychiatry*, 16(12), pp.1213–1220.
- Sealey *et al.*, 2016. Environmental factors in the development of autism spectrum disorders. *Environment International*, 88, pp.288–298.
- Sebat *et al.*, 2007. Strong association of de novo copy number mutations with autism. *Science (New York, N.Y.)*, 316(5823), pp.445–449.
- Selevan, Kimmel & Mendola, 2000. Identifying critical windows of exposure for children's health. *Environmental Health Perspectives*, 108 Suppl 3, pp.451–455.
- Serajee & Mahbubul Huq, 2009. Association of Y chromosome haplotypes with autism. *Journal of Child Neurology*, 24(10), pp.1258–1261.
- Shaffer *et al.*, 2004. Structural basis of androgen receptor binding to selective androgen response elements. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), pp.4758–4763.
- Shaw, Li & Tomljenovic, 2014. Are there negative CNS impacts of aluminum adjuvants used in vaccines and immunotherapy? *Immunotherapy*, 6(10), pp.1055–1071.
- Shmueli *et al.*, 2001. DCX in PC12 cells: CREB-mediated transcription and neurite outgrowth. *Human Molecular Genetics*, 10(10), pp.1061–1070.
- Sholl & Kim, 1990. Androgen receptors are differentially distributed between right and left cerebral hemispheres of the fetal male rhesus monkey. *Brain Research*, 516(1), pp.122–126.
- Shyr *et al.*, 2010. Tumor suppressor PAX6 functions as androgen receptor co-repressor to inhibit prostate cancer growth. *The Prostate*, 70(2), pp.190–199.
- Simental *et al.*, 1991. Transcriptional activation and nuclear targeting signals of the human androgen receptor. *The Journal of Biological Chemistry*, 266(1), pp.510–518.
- Simerly *et al.*, 1990. Distribution of androgen and estrogen receptor mRNA-containing cells in the rat brain: an in situ hybridization study. *The Journal of Comparative Neurology*, 294(1), pp.76–95.
- Simoncini & Genazzani, 2003. Non-genomic actions of sex steroid hormones. *European Journal of Endocrinology / European Federation of Endocrine Societies*, 148(3), pp.281–292.
- Singh *et al.*, 2006. Testosterone inhibits adipogenic differentiation in 3T3-L1 cells: nuclear translocation of androgen receptor complex with beta-catenin and T-cell factor 4 may bypass canonical Wnt signaling to down-regulate adipogenic transcription factors. *Endocrinology*, 147(1), pp.141–154.
- Skaletsky *et al.*, 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423(6942), pp.825–837.

- Skuse, 1999. Genomic imprinting of the X chromosome: a novel mechanism for the evolution of sexual dimorphism. *The Journal of Laboratory and Clinical Medicine*, 133(1), pp.23–32.
- Skuse, 2000. Imprinting, the X-chromosome, and the male brain: explaining sex differences in the liability to autism. *Pediatric Research*, 47(1), pp.9–16.
- Skuse *et al.*, 1997. Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature*, 387(6634), pp.705–708.
- Smith & Dahodwala, 2014. Sex differences in Parkinson's disease and other movement disorders. *Experimental Neurology*, 259, pp.44–56.
- Snijders Blok *et al.*, 2015. Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. *American Journal of Human Genetics*, 97(2), pp.343–352.
- Sobreira *et al.*, 2015. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Human Mutation*, 36(10), pp.928–930.
- Song *et al.*, 2003. Beta-catenin binds to the activation function 2 region of the androgen receptor and modulates the effects of the N-terminal domain and TIF2 on ligand-dependent transcription. *Molecular and Cellular Biology*, 23(5), pp.1674–1687.
- Sowalsky *et al.*, 2015. Whole transcriptome sequencing reveals extensive unspliced mRNA in metastatic castration-resistant prostate cancer. *Molecular cancer research: MCR*, 13(1), pp.98–106.
- Speiser, & White, 2003. Congenital adrenal hyperplasia. *The New England Journal of Medicine*, 349(8), pp.776–788.
- Srour *et al.*, 2010. Mutations in DCC cause congenital mirror movements. *Science (New York, N.Y.)*, 328(5978), p.592.
- Stanbrough *et al.*, 2001. Prostatic intraepithelial neoplasia in mice expressing an androgen receptor transgene in prostate epithelium. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19), pp.10823–10828.
- Steffenburg *et al.*, 1989. A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 30(3), pp.405–416.
- Steinberg *et al.*, 2012. Identification of rare X-linked neuroligin variants by massively parallel sequencing in males with autism spectrum disorder. *Molecular Autism*, 3(1), p.8.
- Stevens *et al.*, 2013. Autism characteristics in children with fetal alcohol spectrum disorders. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 19(6), pp.579–587.
- Stevenson *et al.*, 2003. Genetic syndromes among individuals with mental retardation. *American Journal of Medical Genetics. Part A*, 123A(1), pp.29–32.
- Stevenson *et al.*, 2005. Renpenning syndrome comes into focus. *American Journal of Medical Genetics. Part A*, 134(4), pp.415–421.
- Strömmland *et al.*, 1994. Autism in thalidomide embryopathy: a population study. *Developmental Medicine and Child Neurology*, 36(4), pp.351–356.
- Südhof, 2008. Neuroligins and neuexins link synaptic function to cognitive disease. *Nature*, 455(7215), pp.903–911.
- Sudol, McDonald & Farooq, 2012. Molecular insights into the WW domain of the Golabi-Ito-Hall syndrome protein PQBP1. *FEBS letters*, 586(17), pp.2795–2799.
- Surani, Barton & Norris, 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, 308(5959), pp.548–550.

- Surén *et al.*, 2013. Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. *JAMA*, 309(6), pp.570–577.
- Suzuki *et al.*, 2012. Activity-dependent proteolytic cleavage of neuroligin-1. *Neuron*, 76(2), pp.410–422.
- Sybert & Hall, 1979. Inheritance of tuberous sclerosis. *Lancet (London, England)*, 1(8119), p.783.
- Szafranski *et al.*, 2015. Neurodevelopmental and neurobehavioral characteristics in males and females with CDKL5 duplications. *European journal of human genetics: EJHG*, 23(7), pp.915–921.
- Szatmari *et al.*, 2007. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genetics*, 39(3), pp.319–328.
- Tabuchi *et al.*, 2007. A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science (New York, N.Y.)*, 318(5847), pp.71–76.
- Tadmouri *et al.*, 2009. Consanguinity and reproductive health among Arabs. *Reproductive Health*, 6, p.17.
- Takahashi *et al.*, 2010. Polyglutamine tract-binding protein-1 binds to U5-15kD via a continuous 23-residue segment of the C-terminal domain. *Biochimica Et Biophysica Acta*, 1804(7), pp.1500–1507.
- Takayama *et al.*, 2008. FOXP1 is an androgen-responsive transcription factor that negatively regulates androgen receptor signaling in prostate cancer cells. *Biochemical and Biophysical Research Communications*, 374(2), pp.388–393.
- Takayama *et al.*, 2011. Integration of cap analysis of gene expression and chromatin immunoprecipitation analysis on array reveals genome-wide androgen receptor signaling in prostate cancer cells. *Oncogene*, 30(5), pp.619–630.
- Talebizadeh *et al.*, 2004. Do known mutations in neuroligin genes (NLGN3 and NLGN4) cause autism? *Journal of Autism and Developmental Disorders*, 34(6), pp.735–736.
- Talebizadeh *et al.*, 2006. Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism. *Journal of Medical Genetics*, 43(5), p.e21.
- Tan & Pu, 2003. A pilot study on the effects of testosterone in hypogonadal aging male patients with Alzheimer's disease. *The Aging Male: The Official Journal of the International Society for the Study of the Aging Male*, 6(1), pp.13–17.
- Tan *et al.*, 2015. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica*, 36(1), pp.3–23.
- Tapia *et al.*, 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. *The Journal of Biological Chemistry*, 285(25), pp.19391–19401.
- Tartaglia *et al.*, 2010. A review of trisomy X (47,XXX). *Orphanet Journal of Rare Diseases*, 5, p.8.
- Tartaglia *et al.*, 2012. Attention-deficit hyperactivity disorder symptoms in children and adolescents with sex chromosome aneuploidy: XXY, XXX, XYY, and XXYY. *Journal of developmental and behavioral pediatrics: JDBP*, 33(4), pp.309–318.
- Tassone *et al.*, 2011. MAOA, DBH, and SLC6A4 variants in CHARGE: a case-control study of autism spectrum disorders. *Autism Research: Official Journal of the International Society for Autism Research*, 4(4), pp.250–261.
- Taylor, Cook & Counsell, 2007. Heterogeneity in male to female risk for Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 78(8), pp.905–906.

- Thorne-Lyman & Fawzi, 2012. Vitamin D during pregnancy and maternal, neonatal and infant health outcomes: a systematic review and meta-analysis. *Paediatric and Perinatal Epidemiology*, 26 Suppl 1, pp.75–90.
- Toma *et al.*, 2013. Neurotransmitter systems and neurotrophic factors in autism: association study of 37 genes suggests involvement of DDC. *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry*, 14(7), pp.516–527.
- Tordjman *et al.*, 1997. Androgenic activity in autism. *The American Journal of Psychiatry*, 154(11), pp.1626–1627.
- Toro *et al.*, 2010. Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. *Trends in genetics: TIG*, 26(8), pp.363–372.
- Tostain, Rossi & Martin, 2004. [Physiology of androgens in adult men]. *Progrès En Urologie: Journal De l'Association Française D'urologie Et De La Société Française D'urologie*, 14(5), pp.639–660.
- Treubert & Brümmendorf, 1998. Functional cooperation of beta1-integrins and members of the Ig superfamily in neurite outgrowth induction. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 18(5), pp.1795–1805.
- Trubanova *et al.*, 2014. Under-identification of ASD in females: A case series illustrating the unique presentation of ASD in young adult females.
- Truss & Beato, 1993. Steroid hormone receptors: interaction with deoxyribonucleic acid and transcription factors. *Endocrine Reviews*, 14(4), pp.459–479.
- Turner *et al.*, 2002. Variable expression of mental retardation, autism, seizures, and dystonic hand movements in two families with an identical ARX gene mutation. *American Journal of Medical Genetics*, 112(4), pp.405–411.
- Tyagi *et al.*, 2000. Dynamics of intracellular movement and nucleocytoplasmic recycling of the ligand-activated androgen receptor in living cells. *Molecular Endocrinology (Baltimore, Md.)*, 14(8), pp.1162–1174.
- Unger *et al.*, 2007. Filamin A mutation is one cause of FG syndrome. *American Journal of Medical Genetics. Part A*, 143A(16), pp.1876–1879.
- Vaarala *et al.*, 2012. Identification of androgen-regulated genes in human prostate. *Molecular Medicine Reports*, 6(3), pp.466–472.
- van der Steen, Tindall & Huang, 2013. Posttranslational modification of the androgen receptor in prostate cancer. *International Journal of Molecular Sciences*, 14(7), pp.14833–14859.
- van der Zwaag *et al.*, 2009. Gene-network analysis identifies susceptibility genes related to glycobiology in autism. *PloS One*, 4(5), p.e5324.
- van den Beld *et al.*, 2000. Measures of bioavailable serum testosterone and estradiol and their relationships with muscle strength, bone density, and body composition in elderly men. *The Journal of Clinical Endocrinology and Metabolism*, 85(9), pp.3276–3282.
- van Laar *et al.*, 1989. The human androgen receptor is a 110 kDa protein. *Molecular and Cellular Endocrinology*, 63(1–2), pp.39–44.
- van Rijn & Swaab, 2015. Executive dysfunction and the relation with behavioral problems in children with 47,XXY and 47,XXX. *Genes, Brain, and Behavior*, 14(2), pp.200–208.
- van Rijn *et al.*, 2008. Social behavior and autism traits in a sex chromosomal disorder: Klinefelter (47XXY) syndrome. *Journal of Autism and Developmental Disorders*, 38(9), pp.1634–1641.
- Varoqueaux *et al.*, 2006. Neuroligins determine synapse maturation and function. *Neuron*, 51(6), pp.741–754.

- Veenstra-VanderWeele & Cook, 2004. Molecular genetics of autism spectrum disorder. *Molecular Psychiatry*, 9(9), pp.819–832.
- Velasco *et al.*, 2004. Identification and validation of novel androgen-regulated genes in prostate cancer. *Endocrinology*, 145(8), pp.3913–3924.
- Veldscholte *et al.*, 1992. The androgen receptor in LNCaP cells contains a mutation in the ligand binding domain which affects steroid binding characteristics and response to antiandrogens. *The Journal of Steroid Biochemistry and Molecular Biology*, 41(3–8), pp.665–669.
- Veltman & Brunner, 2012. De novo mutations in human genetic disease. *Nature Reviews. Genetics*, 13(8), pp.565–575.
- Venkatesh *et al.*, 2015. Neuronal Activity Promotes Glioma Growth through Neuroligin-3 Secretion. *Cell*, 161(4), pp.803–816.
- Verkerk *et al.*, 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, 65(5), pp.905–914.
- Verma *et al.*, 2014. Sexual dimorphic effect in the genetic association of monoamine oxidase A (MAOA) markers with autism spectrum disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 50, pp.11–20.
- Verrijdt *et al.*, 2000. Change of specificity mutations in androgen-selective enhancers. Evidence for a role of differential DNA binding by the androgen receptor. *The Journal of Biological Chemistry*, 275(16), pp.12298–12305.
- Vierhapper, Nowotny & Waldhäusl, 2000. Production rates of testosterone in patients with Cushing's syndrome. *Metabolism: Clinical and Experimental*, 49(2), pp.229–231.
- Villard, Lacombe & Fontés, 1996. A point mutation in the XNP gene, associated with an ATR-X phenotype without alpha-thalassemia. *European journal of human genetics: EJHG*, 4(6), pp.316–320.
- Vincent *et al.*, 2004. Mutation screening of X-chromosomal neuroligin genes: no mutations in 196 autism probands. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 129B(1), pp.82–84.
- Vissers, Gilissen & Veltman, 2016. Genetic studies in intellectual disability and related disorders. *Nature Reviews. Genetics*, 17(1), pp.9–18.
- Vissers *et al.*, 2010. A de novo paradigm for mental retardation. *Nature Genetics*, 42(12), pp.1109–1112.
- Viveiros *et al.*, 2015. Screening for fragile X syndrome in males from specialized institutions in the northeast region of Brazil. *Genetics and molecular research: GMR*, 14(2), pp.6897–6905.
- Vohra *et al.*, 2007. Reduced endothelin converting enzyme-1 and endothelin-3 mRNA in the developing bowel of male mice may increase expressivity and penetrance of Hirschsprung disease-like distal intestinal aganglionosis. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 236(1), pp.106–117.
- Volaki *et al.*, 2013. Mutation screening in the Greek population and evaluation of NLGN3 and NLGN4X genes causal factors for autism. *Psychiatric Genetics*, 23(5), pp.198–203.
- Vorstman *et al.*, 2006. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Molecular Psychiatry*, 11(1), pp.1, 18–28.
- Waghray *et al.*, 2001. Identification of androgen-regulated genes in the prostate cancer cell line LNCaP by serial analysis of gene expression and proteomic analysis. *Proteomics*, 1(10), pp.1327–1338.



- Walcott & Merry, 2002. Ligand promotes intranuclear inclusions in a novel cell model of spinal and bulbar muscular atrophy. *The Journal of Biological Chemistry*, 277(52), pp.50855–50859.
- Wald & Sneddon, 1991. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet (London, England)*, 338(8760), pp.131–137.
- Wang, Protheroe, *et al.*, 2009. Müllerian inhibiting substance contributes to sex-linked biases in the brain and behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17), pp.7203–7208.
- Wang, Wang, *et al.*, 2009. Both endogenous and exogenous testosterone decrease myocardial STAT3 activation and SOCS3 expression after acute ischemia and reperfusion. *Surgery*, 146(2), pp.138–144.
- Wang, Zhang, *et al.*, 2009. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*, 459(7246), pp.528–533.
- Wang *et al.*, 1996. Testosterone replacement therapy improves mood in hypogonadal men--a clinical research center study. *The Journal of Clinical Endocrinology and Metabolism*, 81(10), pp.3578–3583.
- Wang *et al.*, 1997. Novel point mutation within intron 10 of FMR-1 gene causing fragile X syndrome. *Human Mutation*, 10(5), pp.393–399.
- Wang *et al.*, 2013. PQBP1, a factor linked to intellectual disability, affects alternative splicing associated with neurite outgrowth. *Genes & Development*, 27(6), pp.615–626.
- Wang *et al.*, 2014. Comparative RNA-seq analysis reveals potential mechanisms mediating the conversion to androgen independence in an LNCaP progression cell model. *Cancer Letters*, 342(1), pp.130–138.
- Waragai *et al.*, 1999. PQBP-1, a novel polyglutamine tract-binding protein, inhibits transcription activation by Brn-2 and affects cell survival. *Human Molecular Genetics*, 8(6), pp.977–987.
- Weeber *et al.*, 2002. Reelin and ApoE receptors cooperate to enhance hippocampal synaptic plasticity and learning. *The Journal of Biological Chemistry*, 277(42), pp.39944–39952.
- Weiner, 1980. Possible role of androgen receptors in amyotrophic lateral sclerosis. A hypothesis. *Archives of Neurology*, 37(3), pp.129–131.
- Weiss *et al.*, 2008. Association between microdeletion and microduplication at 16p11.2 and autism. *The New England Journal of Medicine*, 358(7), pp.667–675.
- Weiss *et al.*, 2009. A genome-wide linkage and association scan reveals novel loci for autism. *Nature*, 461(7265), pp.802–808.
- Wengler *et al.*, 1995. Nonrandom inactivation of the X chromosome in early lineage hematopoietic cells in carriers of Wiskott-Aldrich syndrome. *Blood*, 85(9), pp.2471–2477.
- Werling & Geschwind, 2013. Sex differences in autism spectrum disorders. *Current Opinion in Neurology*, 26(2), pp.146–153.
- Wermter *et al.*, 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: The Official Publication of the International Society of Psychiatric Genetics*, 147B(4), pp.535–537.
- Wheelwright *et al.*, 2006. Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Research*, 1079(1), pp.47–56.
- White *et al.*, 2010. Cyclin-dependent kinase-like 5 (CDKL5) mutation screening in Rett syndrome and related disorders. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, 13(2), pp.168–178.

- Williams & Casanova, 2011. Above genetics: lessons from cerebral development in autism. *Translational Neuroscience*, 2(2), pp.106–120.
- Williams *et al.*, 2010. Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *American Journal of Human Genetics*, 87(2), pp.219–228.
- Wilson & McPhaul, M.J., 1994. A and B forms of the androgen receptor are present in human genital skin fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America*, 91(4), pp.1234–1238.
- Wilson, 2001. Androgens, androgen receptors, and male gender role behavior. *Hormones and Behavior*, 40(2), pp.358–366.
- Wilson, 2011. Analysis of interdomain interactions of the androgen receptor. *Methods in Molecular Biology (Clifton, N.J.)*, 776, pp.113–129.
- Windham *et al.*, 2006. Autism spectrum disorders in relation to distribution of hazardous air pollutants in the san francisco bay area. *Environmental Health Perspectives*, 114(9), pp.1438–1444.
- Winters & Dalkin, 2004. Neuroendocrine Control of Testicular Function. In S. J. W. MD, ed. *Male Hypogonadism*. Contemporary Endocrinology. Humana Press, pp. 1–21. Available at: [http://link.springer.com.gate2.inist.fr/chapter/10.1007/978-1-59259-727-7\\_1](http://link.springer.com.gate2.inist.fr/chapter/10.1007/978-1-59259-727-7_1)
- Wittmann & McLennan, 2013a. Anti-Müllerian hormone may regulate the number of calbindin-positive neurons in the sexually dimorphic nucleus of the preoptic area of male mice. *Biology of Sex Differences*, 4(1), p.18.
- Wittmann & McLennan, 2013b. The bed nucleus of the stria terminalis has developmental and adult forms in mice, with the male bias in the developmental form being dependent on testicular AMH. *Hormones and Behavior*, 64(4), pp.605–610.
- Wong *et al.*, 1993. Steroid requirement for androgen receptor dimerization and DNA binding. Modulation by intramolecular interactions between the NH<sub>2</sub>-terminal and steroid-binding domains. *The Journal of Biological Chemistry*, 268(25), pp.19004–19012.
- Wu *et al.*, 2009a. Estrogen masculinizes neural pathways and sex-specific behaviors. *Cell*, 139(1), pp.61–72.
- Wu *et al.*, 2009b. Regulation of monoamine oxidase A by the SRY gene on the Y chromosome. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 23(11), pp.4029–4038.
- Wudy *et al.*, 1995. 17 alpha-hydroxyprogesterone, 4-androstenedione, and testosterone profiled by routine stable isotope dilution/gas chromatography-mass spectrometry in plasma of children. *Pediatric Research*, 38(1), pp.76–80.
- Wudy *et al.*, 1999. Profiling steroid hormones in amniotic fluid of midpregnancy by routine stable isotope dilution/gas chromatography-mass spectrometry: reference values and concentrations in fetuses at risk for 21-hydroxylase deficiency. *The Journal of Clinical Endocrinology and Metabolism*, 84(8), pp.2724–2728.
- Wyce *et al.*, 2010. Research Resource: The androgen receptor modulates expression of genes with critical roles in muscle development and function. *Molecular Endocrinology (Baltimore, Md.)*, 24(8), pp.1665–1674.
- Xiong *et al.*, 2015. Genotype/phenotype analysis in Chinese laminin- $\alpha$ 2 deficient congenital muscular dystrophy patients. *Clinical Genetics*, 87(3), pp.233–243.
- Xu, Ionita-Laza, *et al.*, 2012. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature Genetics*, 44(12), pp.1365–1369.

- Xu, Li, *et al.*, 2012. AutismKB: an evidence-based knowledgebase of autism genetics. *Nucleic Acids Research*, 40(Database issue), pp.D1016-1022.
- Xu *et al.*, 2001. Quantitative expression profile of androgen-regulated genes in prostate cancer cells and identification of prostate-specific genes. *International Journal of Cancer*, 92(3), pp.322–328.
- Xu *et al.*, 2014. Variations analysis of NLGN3 and NLGN4X gene in Chinese autism patients. *Molecular Biology Reports*, 41(6), pp.4133–4140.
- Yamana, Labrie & Luu-The, 2010. Human type 3 5 $\alpha$ -reductase is expressed in peripheral tissues at higher levels than types 1 and 2 and its activity is potently inhibited by finasteride and dutasteride. *Hormone Molecular Biology and Clinical Investigation*, 2(3), pp.293–299.
- Yan *et al.*, 2008. Analysis of the neuroligin 4Y gene in patients with autism. *Psychiatric Genetics*, 18(4), pp.204–207.
- Yanagi *et al.*, 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, p.724072.
- Yanase *et al.*, 2003. Aromatase in bone: roles of Vitamin D3 and androgens. *The Journal of Steroid Biochemistry and Molecular Biology*, 86(3–5), pp.393–397.
- Yang & Gill, 2007. A review of gene linkage, association and expression studies in autism and an assessment of convergent evidence. *International Journal of Developmental Neuroscience: The Official Journal of the International Society for Developmental Neuroscience*, 25(2), pp.69–85.
- Yang *et al.*, 2002. Linking beta-catenin to androgen-signaling pathway. *The Journal of Biological Chemistry*, 277(13), pp.11336–11344.
- Yang *et al.*, 2011. Inhibition of androgen receptor activity by histone deacetylase 4 through receptor SUMOylation. *Oncogene*, 30(19), pp.2207–2218.
- Yassa, 2014. Autism: a form of lead and mercury toxicity. *Environmental Toxicology and Pharmacology*, 38(3), pp.1016–1024.
- Yasuda *et al.*, 2011. Gene expression analysis in lymphoblasts derived from patients with autism spectrum disorder. *Molecular Autism*, 2(1), p.9.
- Yeargin-Allsopp *et al.*, 2003. Prevalence of autism in a US metropolitan area. *JAMA*, 289(1), pp.49–55.
- Yeh *et al.*, 2002. Generation and characterization of androgen receptor knockout (ARKO) mice: an in vivo model for the study of androgen functions in selective tissues. *Proceedings of the National Academy of Sciences of the United States of America*, 99(21), pp.13498–13503.
- Ylisaukko-oja *et al.*, 2005. Analysis of four neuroligin genes as candidates for autism. *European journal of human genetics: EJHG*, 13(12), pp.1285–1292.
- Yrigollen *et al.*, 2008. Genes controlling affiliative behavior as candidate genes for autism. *Biological Psychiatry*, 63(10), pp.911–916.
- Yu *et al.*, 2013. Using whole-exome sequencing to identify inherited causes of autism. *Neuron*, 77(2), pp.259–273.
- Yu *et al.*, 2016. Lack of Association Between Polymorphisms in Dopa Decarboxylase and Dopamine Receptor-1 Genes With Childhood Autism in Chinese Han Population. *Journal of Child Neurology*, 31(5), pp.560–564.
- Zhang *et al.*, 1994. Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. *Nature Genetics*, 7(4), pp.531–535.

- Zhang *et al.*, 2011. Androgen receptor variants occur frequently in castration resistant prostate cancer metastases. *PloS One*, 6(11), p.e27970.
- Zhang *et al.*, 2013. Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. *Molecular Biology and Evolution*, 30(12), pp.2588–2601.
- Zhao *et al.*, 2007. A unified genetic theory for sporadic and inherited autism. *Proceedings of the National Academy of Sciences of the United States of America*, 104(31), pp.12831–12836.
- Zhou *et al.*, 1994. A ligand-dependent bipartite nuclear targeting signal in the human androgen receptor. Requirement for the DNA-binding domain and modulation by NH<sub>2</sub>-terminal and carboxyl-terminal sequences. *The Journal of Biological Chemistry*, 269(18), pp.13115–13123.
- Zhou *et al.*, 1995. Specificity of ligand-dependent androgen receptor stabilization: receptor domain interactions influence ligand dissociation and receptor stability. *Molecular Endocrinology (Baltimore, Md.)*, 9(2), pp.208–218.
- Zhou, Kemppainen & Wilson, 1995. Identification of three proline-directed phosphorylation sites in the human androgen receptor. *Molecular Endocrinology (Baltimore, Md.)*, 9(5), pp.605–615.
- Zuloaga *et al.*, 2008. The role of androgen receptors in the masculinization of brain and behavior: what we've learned from the testicular feminization mutation. *Hormones and Behavior*, 53(5), pp.613–626.
- Zup, Edwards & McCarthy, 2014. Sex- and age-dependent effects of androgens on glutamate-induced cell death and intracellular calcium regulation in the developing hippocampus. *Neuroscience*, 281C, pp.77–87.
- Zweier *et al.*, 2007. Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *American Journal of Human Genetics*, 80(5), pp.994–1001.
- Zweier *et al.*, 2010. Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. *Human Mutation*, 31(6), pp.722–733.

# Etude des mécanismes à l'origine de l'excès de garçons dans la déficience intellectuelle et les troubles du spectre autistique

## Résumé

La déficience intellectuelle (DI) et les troubles du spectre autistique (ASD) sont deux troubles neurodéveloppementaux (NDD) présentant de nombreux chevauchements génétiques et phénotypiques ainsi qu'un biais de sexe important, avec plus de garçons atteints (1,4x plus pour la DI et 4x plus pour l'ASD). Au sein de notre laboratoire, le taux de diagnostic des patients souffrant de DI et/ou d'ASD est significativement plus élevé chez les filles que chez les garçons. De façon surprenante, nous n'avons pas observé de différence significative entre filles et garçons au niveau de la proportion de mutations pathogènes sur le chromosome X (5,3% *versus* 7,6%), confirmant ainsi que les mutations causales totalement pénétrantes sur ce chromosome ne peuvent pas expliquer la totalité de l'excès de garçons atteints de DI ou d'ASD. Nous avons donc choisi d'étudier une autre des hypothèses, plus environnementale, qui pourrait rendre le cerveau masculin plus susceptible au développement de NDD : le rôle des androgènes au cours du développement du cerveau. J'ai étudié l'effet de ces hormones masculines dans des précurseurs neuronaux humains (hNSCs) et observé que les androgènes augmentent la prolifération des hNSCs et les protègent contre la mort cellulaire en conditions stressantes. J'ai également mis en évidence que les androgènes, via leur récepteur (le récepteur aux androgènes), régulent une centaine de gènes dans les hNSCs avec, parmi eux, un enrichissement en gènes connus pour être différenciellement exprimés chez les individus avec ASD (dont NRCAM et FAM107A). La régulation de ces gènes par les androgènes pendant le développement du cerveau pourrait ainsi participer à la sensibilité accrue du cerveau masculin, exposé à d'autres facteurs génétiques et environnementaux, à développer une NDD.

**Mots-clés :** Déficience intellectuelle, Troubles du spectre autistique, Biais de sexe, Androgènes, Précurseurs neuronaux humains

## Résumé en anglais

Intellectual disability (ID) and autism spectrum disorders (ASD) are two common neurodevelopmental disorders (NDD) with many genetic and phenotypic overlaps. Another common feature is the existence of a gender bias, very strong for ASD (4 males for 1 female) and notable for ID (1.4:1). In our team, the diagnostic yield of patients affected by ID with or without ASD is significantly higher in girls than in boys. Surprisingly, we did not observe a significant difference between girls and boys in the proportion of pathogenic mutations on the X chromosome (5.3% versus 7.6%), confirming that rare and fully penetrant mutations on this chromosome can not explain the totality this male bias observed in ID and ASD. We have therefore chosen to study another hypothesis, more environmental, which could make the male brain more susceptible to develop NDD: the role of androgens during brain development. I studied the effect of these male hormones in human neuronal precursors (hNSCs) and observed that androgens increase the proliferation of hNSCs and protect them against cell death under stressful conditions. I also showed that androgens, via their receptor (the androgen receptor), regulate a hundred genes in hNSCs with, among them, an enrichment of genes known to be differentially expressed in individuals with ASD (e.g NRCAM and FAM107A). The regulation of these genes by androgens during brain development could thus contribute to the increased sensitivity of the male brain, exposed to other genetic and environmental factors, to develop an NDD.

**Keywords:** Intellectual deficiency, Autism spectrum disorders, Sex bias, Androgens, Human neuronal precursors