Université Pierre et Marie Curie

THESE DE DOCTORAT

Spécialité

Neurosciences

Présentée par

Elsa Delaby

# Dissection de l'architecture génétique de l'autisme par analyse des variations du nombre de copies de gènes

Soutenue le 30 Septembre 2014 devant le jury composé de :

| | |
|---|---|
| Pr Philippe Jonveaux | Rapporteur |
| Pr Laurence Faivre | Rapporteur |
| Dr Laurence Colleaux | Examinateur |
| Dr Gilles Fischer | Examinateur |
| Pr Pierre Netter | Examinateur |
| Dr Catalina Betancur | Directeur de thèse |

# Remerciements

Je commence par remercier tout particulièrement Catalina Betancur, qui s'est toujours montrée à l'écoute et très disponible pendant ma thèse. Merci pour l'aide et le temps que vous m'avez consacrés dans la rédaction du manuscrit mais surtout pendant ces années de recherche. J'ai énormément appris à votre contact, autant sur le plan scientifique que personnel. Grâce à vous j'ai fait un énorme pas en avant, digne de la physique quantique.

Je tiens à remercier les membres du jury Philippe Jonveaux, Laurence Faivre, Laurence Colleaux, Gilles Fischer et Pierre Netter qui me font l'honneur et le privilège de juger ce travail de thèse.

Au sein du laboratoire je remercierai tout d'abord Bruno Giros et les membres de l'unité pour leur accueil. En particulier ceux qui ont partagé mon bureau ces derniers mois, Thomas et Laurian(n)e, je vous souhaite le meilleur, votre tour approche… Merci aussi à mes compagnons de rédaction Camille et Daniel parce qu'ensemble on est toujours plus forts ; c'est la dernière ligne droite ! Je tiens à remercier également Stéphanie Daumas pour les pauses déjeuner-maquereaux-sardines, Amélie qui va beaucoup me manquer, Nicolas et son étrange humour, Amaia pour sa gentillesse, la multicolore Barbara, Stéphanie De Gois pour ses coups de fils du soir, mais aussi Sophie, Elise, Odile, Vincent, Gaël, Salah, les deux Véroniques et Franck. Un petit mot pour les nouvelles recrues, Nida et Nina, je vous souhaite bonne route, soyez fortes.

Merci à toi Marion, notre rencontre a été essentielle, tu m'as tellement apporté et tu me manques beaucoup ici, même si je sais que tu es toujours là pour moi. Merci pour ton amitié et ton soutien sans faille et j'espère que tu chanteras Céline Dion après ma soutenance.

Je remercie également Marika, pour ton amitié, tes conseils scientifiques et personnels ; merci de nous avoir souvent nourries avec Marion et merci surtout d'être restée si présente depuis ton départ du laboratoire. Je commence le yoga à la rentrée !

Je remercie Hélène, parce que grâce à toi les départs de Marion et Marika se sont fait en douceur. Je suis tellement contente qu'on s'entende si bien et maintenant je vais avoir le temps de venir voir ta maison et de rencontrer enfin les deux hommes de ta vie !

Je tiens également à remercier avec tout mon amour ma famille et en particulier mes parents pour n'avoir jamais douté de moi et pour m'avoir permis d'aller au bout de mes ambitions. J'espère que vous serez fiers de moi.

Enfin, je remercie de tout mon cœur celui qui a partagé ma vie pendant cette période, merci Clément pour ta patience et ton soutien et je te promets de t'accorder un peu plus de temps désormais.

# Liste d'abréviations

| | |
|---|---|
| aCGH | puce d'hybridation génomique comparative *(array comparative genomic hybridization)* |
| ADI-R | *Autism Diagnostic Interview-Revised* |
| ADOS | *Autism Diagnostic Observation Schedule* |
| AGP | *Autism Genome Project* |
| BAC | *bacterial artificial chromosome* |
| CIM-10 | Classification Internationale des Maladies |
| CNV | variation du nombre de copies *(copy number variation)* |
| DSM-IV | Manuel Diagnostique et Statistique des Maladies Mentales, quatrième édition |
| DSM-5 | Manuel Diagnostique et Statistique des Maladies Mentales, cinquième édition |
| Etude PARIS | *Paris Autism Research International Study* |
| FISH | *fluorescent in situ hybridization* |
| FoSTeS | *Fork Stalling and Template Switching* |
| kb | kilobase |
| Mb | mégabase |
| MLPA | *multiplex ligation-dependent probe amplification* |
| NAHR | *non-allelic homologous recombination* |
| NHEJ | *non-homologous end-joining* |
| pb | paire de bases |
| QI | quotient intellectuel |
| SNP | *single nucleotide polymorphism* |
| SNV | *single nucleotide variant* |
| TDAH | trouble de déficit de l'attention/hyperactivité |
| TED | trouble envahissant du développement |
| TED-NOS | trouble envahissant du développement non spécifié |
| TSA | trouble du spectre autistique |
| VOUS | *variant of unknown significance* |
| WES | *whole exome sequencing* |
| WGS | *whole genome sequencing* |
| YAC | *yeast artificial chromosome* |

# Table des matières

**Introduction** ..... 1

Partie I : Troubles du spectre autistique ..... 3

1. Définition et critères diagnostiques ..... 3
   1.1 Evolution de la définition de l'autisme ..... 3
   1.2 Principales caractéristiques des TSA ..... 4
      1.2.1 Troubles de la communication et des relations sociales ..... 4
      1.2.2 Intérêts restreints et comportements répétitifs ..... 4
      1.2.3 Autres manifestations ..... 5
   1.3 Outils diagnostiques ..... 5

2. Hétérogénéité clinique ..... 6

3. Epidémiologie ..... 7
   3.1 Prévalence ..... 7
   3.2 Sexe ratio ..... 8

4. Evolution et prise en charge ..... 10

Partie II : Contribution des facteurs génétiques à l'étiologie de l'autisme ..... 12

1. Implication des facteurs génétiques dans les TSA ..... 12
   1.1 Etudes familiales et de jumeaux ..... 12
   1.2 Facteurs non génétiques ..... 15

2. Recherche de variants communs dans les TSA ..... 15
   2.1 Etudes de liaison ..... 15
   2.2 Etudes d'association ..... 16

3. Recherche de variants rares dans les TSA ..... 18
   3.1 Variations du nombre de copies ..... 18
      3.1.1 Découverte et définition ..... 19
      3.1.2 Mécanismes de formation des CNVs ..... 20
      3.1.3 Impact des CNVs sur l'expression des gènes ..... 23
      3.1.4 Les CNVs dans l'autisme ..... 25
   3.2 Variations de séquence ..... 29
      3.2.1 Séquençage de gènes candidats ..... 29
      3.2.2 Séquençage de l'exome entier ..... 30
      3.2.3 Séquençage du génome entier ..... 33

4. Causes génétiques connues de TSA ..... 35
   4.1 Remaniements chromosomiques ..... 35
   4.2 Syndromes génétiques ..... 35
      4.2.1 Maladies monogéniques ..... 36
      4.2.2 Syndromes de microdélétion/microduplication ..... 40
   4.3 CNVs rares à pénétrance incomplète et/ou expressivité variable ..... 43

5. Multiples voies biologiques impliquées dans l'autisme ..... 48

**Résultats** ..... 55

Article 1 *Convergence of genes and cellular pathways dysregulated in autism spectrum disorders* .....59

**Discussion** ..... 165

1. Complexité de l'architecture génétique des TSA..... 167
   1.1 Multitude de gènes impliqués ..... 167
   1.2 Modes de transmission rares..... 168
   1.3 Expressivité variable et pénétrance incomplète..... 168
   1.4 Différences de susceptibilité aux altérations génétiques en fonction du sexe.....171

2. Héritabilité manquante ..... 173
   2.1 Variants non détectés et de signification clinique inconnue ..... 173
   2.2 Modèles oligogéniques et polygéniques ..... 174
   2.3 Contribution des variants génétiques communs ..... 174

3. Stratégies pour interpréter les variants génétiques ..... 176
   3.1 Bases de données de patients et de témoins ..... 176
   3.2 Listes de gènes et revue de la littérature..... 177
   3.3 Outils bioinformatiques ..... 179
   3.4 Etudes d'expression ..... 179
   3.5 Analyses fonctionnelles ..... 180

4. Intérêt du diagnostic étiologique chez les patients atteints de TSA..... 181

**Références** ..... 185

**Annexes**..... 201

Liste des publications.....203

Article 3 *A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism*.....205

Article 4 *Complex nature of apparently balanced chromosomal rearrangements in patients with autism spectrum disorder* .....215

# Liste des tableaux

**Tableau 1.** Etudes de jumeaux dans les TSA ..... 13
**Tableau 2.** Principales études de CNVs dans les TSA ..... 26
**Tableau 3.** Syndromes génétiques pour lesquels les TSA sont une manifestation fréquente ..... 37
**Tableau 4.** Gènes synaptiques impliqués dans les TSA et autres troubles neurodéveloppementaux ..... 51

# Liste des figures

**Figure 1.** Théorie de l'esprit : test de Sally et Anne ..... 5
**Figure 2.** Hétérogénéité clinique des TSA ..... 7
**Figure 3.** Effet de la substitution de diagnostic dans le prévalence des TSA aux USA ..... 8
**Figure 4.** Distribution du score de systématisation chez des patients atteints de TSA et des hommes et des femmes 'typiques' ..... 10
**Figure 5.** Risque de récurrence dans les TSA ..... 14
**Figure 6.** Contribution des facteurs génétiques et environnementaux dans les TSA ..... 18
**Figure 7.** Evolution des techniques de détection des CNVs appliquées à la recherche étiologique dans l'autisme ..... 20
**Figure 8.** Polymorphismes du nombre de copies répartis sur l'ensemble du génome humain ..... 21
**Figure 9.** Réarrangements par recombinaison homologue non allélique (NAHR) ..... 22
**Figure 10.** Mécanismes à l'origine des CNVs non récurrents ..... 23
**Figure 11.** Variations du nombre de copies et mécanismes pathogènes ..... 24
**Figure 12.** Modèle prédictif de la probabilité d'haploinsuffisance des gènes ..... 25
**Figure 13.** Enrichissement de CNVs touchant des gènes impliqués dans les TSA ou la déficience intellectuelle chez les patients avec TSA ..... 27
**Figure 14.** Enrichissement des mutations *de novo* chez les sujets atteints de TSA, selon la classe fonctionnelle ..... 31
**Figure 15.** Influence de l'âge du père sur le nombre de mutations *de novo* ..... 32
**Figure 16.** Première étude de séquençage du génome entier dans les TSA ..... 34
**Figure 17.** Vue d'ensemble des étiologies génétiques connues de TSA ..... 36
**Figure 18.** Nombre de gènes impliqués dans l'étiologie des TSA ..... 37
**Figure 19.** Gènes liés au chromosome X impliqués dans la déficience intellectuelle et les TSA ..... 39
**Figure 20.** CNVs récurrents et non récurrents ..... 40
**Figure 21.** Délétions 22q13 de taille variable chez les patients atteints du syndrome de Phelan-McDermid ..... 42
**Figure 22.** Pénétrance incomplète et expressivité variable ..... 43
**Figure 23.** CNVs rares à expressivité variable et/ou pénétrance incomplète sur les chromosomes 15 et 16 ..... 44
**Figure 24.** Distribution du QI chez les porteurs d'une délétion 16p11.2 par rapport aux non porteurs de la famille ..... 45
**Figure 25.** Agrégation des délétions *NRXN1* dans la forme alpha du gène ..... 46
**Figure 26.** Protéines synaptiques impliquées dans les TSA ..... 50

**Figure 27.** Convergence des gènes impliqués par les études d'exome vers les mêmes voies biologiques ..... 53
**Figure 28.** Etiologie génétique des TSA : relation entre la fréquence allélique et la taille de l'effet ..... 167
**Figure 29.** Comparaison de la prévalence de quatre CNVs dans des troubles neuropsychiatriques et chez des témoins ..... 169
**Figure 30.** Modèle oligogénique ..... 169
**Figure 31.** Phénomène d'interférence cellulaire associé aux mutations *PCDH19* ..... 172
**Figure 32.** Modèles polygéniques proposés dans l'étiologie des TSA ..... 175
**Figure 33.** Corrélation entre les habilités sociales des parents et leurs enfants ..... 176
**Figure 34.** Implication du gène *SETD5* dans les TSA ..... 178
**Figure 35.** Génération de cellules souches pluripotentes induites pour étudier la physiopathologie des TSA ..... 180
**Figure 36.** CNVs pathogènes identifiés chez 2 446 patients atteints de TSA de l'*Autism Genome Project* ..... 182

# Introduction

# Partie I : Troubles du spectre autistique

## 1. Définition et critères diagnostiques

Le pédopsychiatre américain Leo Kanner fut le premier à employer le terme 'autisme' en 1943 pour décrire des troubles du comportement qu'il observe dans une cohorte de 11 enfants (Kanner, 1943). Ces enfants présentaient un retrait social, des comportements répétitifs, un désir obsessionnel d'immuabilité et des troubles sévères du langage. Il emprunte ce terme à son confrère suisse Eugen Bleuler, terme jusqu'alors réservé à décrire l'une des manifestations de la schizophrénie. Il est d'ailleurs intéressant de noter qu'avant la publication des travaux de Leo Kanner qui décrit les troubles comportementaux de ces enfants comme un trouble à part entière, l'autisme a longtemps été considéré comme une forme infantile de la schizophrénie. Au même moment, le pédiatre autrichien Hans Asperger rapporte lui aussi quatre jeunes garçons présentant un manque d'empathie, une difficulté à se faire des amis et une forte préoccupation pour des intérêts particuliers (Asperger, 1944). Toutefois, contrairement à ceux décrits par Kanner, ces enfants avaient un niveau intellectuel normal et un langage fonctionnel. Malheureusement, ses travaux étant rédigés en allemand, leur diffusion fut retardée et c'est seulement en 1981 que Lorna Wing les réactualise et fait le lien entre les travaux de Leo Kanner et la 'psychopathie autistique de l'enfance' décrite par Asperger (Wing, 1981). Cette psychiatre britannique réhabilite les travaux d'Asperger et propose l'idée d'une continuité entre l'autisme de Kanner et le syndrome d'Asperger en mettant en avant la triade de caractéristiques autistiques qui comprend des troubles de l'interaction sociale, des troubles de la communication ainsi que la présence de gestes répétitifs et stéréotypés (Wing et Gould, 1979 ; Wing, 1981).

### 1.1 Evolution de la définition de l'autisme

Ces caractéristiques seront par la suite largement reconnues et serviront de critères diagnostiques dans le Manuel diagnostique et statistique des troubles mentaux (*Diagnostic and Statistical Manual of Mental Disorders*, DSM) de l'Association Américaine de Psychiatrie ainsi que dans la Classification internationale des maladies (*International statistical classification of diseases and related health problems*, CIM) publiée par l'Organisation Mondiale de la Santé. Ainsi, dans le DSM-IV et dans la CIM 10, des déficits dans les trois domaines décrits par Lorna Wing caractérisent l'autisme, le syndrome d'Asperger et le trouble envahissant du développement non spécifié (TED-NOS), avec un âge d'apparition des troubles avant trois ans. Les manifestations cliniques de ces troubles du comportement étant similaires bien que de sévérité variable, ils ont été regroupés sous le terme de 'troubles envahissants du développement'. La classification utilisée ultérieurement reposait sur des différences subtiles liées à la gravité des symptômes observés ainsi qu'à l'âge d'apparition des troubles. L'autisme au sens strict est considéré comme le trouble le plus sévère des troubles envahissants du développement. Le syndrome d'Asperger diffère par une acquisition non retardée du langage et un développement cognitif en apparence normal. Un individu est diagnostiqué avec un TED-NOS (autisme atypique) lorsque les troubles et déficits comportementaux observés chez lui ne remplissent pas tous les critères d'âge ou de symptomatologie de l'autisme.

Aujourd'hui, le terme 'troubles du spectre autistique' (TSA) est plutôt utilisé et remplace les troubles envahissants du développement. L'autisme, le syndrome d'Asperger et le TED-NOS sont maintenant regroupés en une seule catégorie dans la cinquième édition du DSM (DSM-5) publiée en 2013 (American Psychiatric Association, 2013), se distinguant par leur sévérité et leur nature. De plus, l'habilité à créer des interactions sociales étant intimement liée aux capacités de communication verbale et non verbale, ces deux critères ont été regroupés en un seul. Les TSA forment donc une catégorie unique définie par deux composantes majeures : les troubles de la communication et des relations sociales ainsi que la présence

d'intérêts restreints et de comportements répétés. Ces deux caractéristiques se rapprochent finalement de la première description faite par Leo Kanner pour qui les caractéristiques de l'autisme pouvaient être résumées par un isolement extrême (*aloneness*) et un désir obsessionnel de préserver l'immuabilité (*sameness*).

## 1.2 Principales caractéristiques des TSA

### 1.2.1 Troubles de la communication et des relations sociales

Les enfants atteints de TSA présentent des déficits du langage et dans un tiers des cas, celui-ci ne se développe pas du tout ou seulement quelques mots (Snow et coll., 2009). Pour ceux qui parviennent à développer un langage (habituellement les patients avec autisme dit de 'haut niveau' ou syndrome d'Asperger), ce dernier est souvent atypique. Ces patients ont du mal à comprendre une conversation et à entrer dans un dialogue. Ils ont une prosodie particulière (voix monotone, débit et rythme inhabituels), se trompent dans l'emploi des pronoms (parlent d'eux-mêmes à la deuxième personne, par exemple 'tu' à la place de 'je') et en général n'utilisent pas les mots abstraits. Les enfants atteints d'autisme ont également tendance à répéter tout ou une partie des phrases de leurs interlocuteurs en guise de réponse verbale. Ce phénomène appelé écholalie, involontaire et quasi automatique, fait partie des stéréotypies fréquemment observées chez les enfants avec TSA. En outre, même si certains patients ont un langage, l'utilisation des éléments de communication non verbale (gestes des mains, expressions faciales) est altérée.

L'enfant atteint d'autisme montre des défauts dans la mise en place des relations sociales. Dès les premières années de la vie, il fuit le contact visuel, a des difficultés à sourire et présente également des défauts d'empathie.

***Théorie de l'esprit***

La théorie de l'esprit est la capacité à comprendre les états mentaux qui sous-tendent les comportements des autres. Cette capacité est sous-jacente à l'empathie, à l'expression des sentiments et est donc essentielle à l'établissement de relations sociales entre les individus. Le test de Sally et Anne **(Figure 1)** est l'un des tests utilisés en psychologie du développement pour évaluer l'aptitude d'une personne à comprendre qu'autrui possède des états mentaux différents des siens (Wimmer et Perner, 1983). L'indicateur de réussite du test est de pouvoir concevoir que Sally a son propre point de vue même si ce point de vue diffère de la réalité. Ce test a été réalisé pour la première fois dans une cohorte de patients avec autisme en 1985 (Baron-Cohen et coll., 1985). Soixante et un enfants ont été testés dont 20 avec un diagnostic de TSA, 14 présentant une trisomie 21 et 27 enfants non atteints. Les résultats de cette étude montrent que 80 % des enfants atteints de TSA échouent au test alors que les enfants atteints de trisomie 21 et les enfants non atteints réussissent en grande majorité.

### 1.2.2 Intérêts restreints et comportements répétitifs

Les intérêts restreints et les comportements stéréotypés englobent une très large catégorie de comportements pouvant varier en fréquence et en sévérité. Chez les patients atteints de TSA, on note souvent une passion démesurée pour certains sujets particuliers (les dinosaures ou les dessins animés par exemple). Ce centre d'intérêt devient alors le seul moyen de communiquer, peu importe si l'autre est intéressé ou non. Le détournement des objets de leur fonction première est également fréquemment observé ainsi qu'une obsession pour les objets qui tournent (machine à laver ou roue d'un vélo par exemple).

Pour pallier à l'anxiété liée au monde qui les entoure, les patients atteints de TSA instaurent une routine et des rituels qui vont les rassurer. Souvent, ils ne supportent pas les surprises ou les

changements de dernière minute qui les paralysent et provoquent chez eux une anxiété extrême (changer de chemin pour aller à l'école, visite surprise à la maison ou encore rendez-vous médical non planifié). Le besoin de ritualiser peut aussi se traduire par des mouvements répétés, par exemple des stéréotypies des mains ou des doigts ou par la répétition de comportements (ouvrir et fermer une porte, allumer et éteindre la lumière).

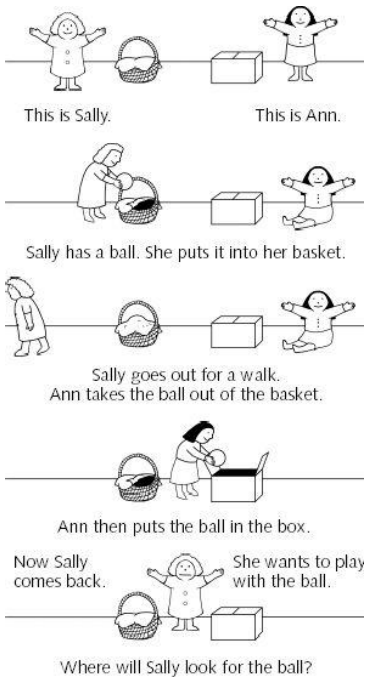



**Figure 1. Théorie de l'esprit : test de Sally et Anne**
Sally a une balle et la range dans son panier. Puis, Sally s'en va et pendant son absence, Anne déplace la balle de Sally dans une boîte. Sally revient et on demande alors à l'enfant où Sally va chercher sa balle. L'enfant doit indiquer le panier de Sally. Cette réponse est le point de vue de Sally, mais pas celui de l'enfant qui a vu la scène.

### 1.2.3 Autres manifestations

Les patients atteints de TSA souffrent fréquemment d'anomalies de la perception, en particulier auditive. En effet, les enfants atteints d'autisme semblent souvent indifférents au monde sonore, notamment, certains ne répondent pas à leur nom. Paradoxalement, certains bruits particuliers peuvent déclencher chez eux des réactions de peur exagérée ou de forte anxiété. Des manifestations d'hyposensibilité ou d'hypersensibilité à d'autres stimuli sensoriels (une insensibilité à la douleur, au froid ou à l'inverse un rejet de certains textiles, textures d'aliments ou lumières trop vives) peuvent également exister. De plus, des problèmes de motricité fine et globale sont parfois observés, avec un manque de coordination, une démarche particulière (souvent rigide) et des difficultés de dextérité fine. En outre, les TSA s'accompagnent souvent de troubles du sommeil parfois durables et intenses. Il est également très fréquent que les patients atteints de TSA présentent des troubles gastro-intestinaux (diarrhées et/ou constipations chroniques, reflux gastro-œsophagien).

## 1.3 Outils diagnostiques

Les parents sont souvent les premiers à déceler les signes évocateurs des TSA chez leur enfant, le plus souvent entre 18 et 36 mois. Il n'existe aucun marqueur biologique permettant de diagnostiquer un TSA. Le diagnostic est principalement basé sur les critères du DSM et de la CIM après un interrogatoire sur le développement de l'enfant et l'observation de son comportement. Pour évaluer de manière plus homogène le degré d'atteinte des patients, des outils diagnostiques standardisés ont été développés. Ces

outils, basés sur les critères du DSM et de la CIM, ne sont pas nécessaires à l'établissement du diagnostic mais sont principalement utilisés en recherche. Les deux tests les plus couramment utilisés sont l'ADI-R (*Autism Diagnostic Interview-Revised*) (Lord et coll., 1994) et l'ADOS (*Autism Diagnostic Observation Schedule*) (Lord et coll., 2000).

L'***ADI-R*** est un entretien comportant 93 questions qui concernent les capacités de l'enfant dans les domaines de l'interaction sociale, de la communication et du comportement. Il est mené par un examinateur expérimenté sur le plan clinique qui interroge les parents (ou une personne qui s'occupe de l'enfant) sur l'histoire du développement et le comportement actuel de l'enfant. L'ADI-R est construit en lien avec les critères du DSM-IV et permet de confirmer un diagnostic d'autisme. Il s'adresse à des enfants dont l'âge mental est supérieur à 1 an et demi.

L'***ADOS*** est une échelle d'observation semi-structurée qui permet d'évaluer les comportements sociaux, la communication, l'interaction sociale réciproque et les comportements face aux jeux faisant appel à l'imagination. L'ADOS inclut une série d'activités standardisées d'une durée de 30 à 45 minutes qui permettent au clinicien d'évaluer un individu avec autisme de l'âge préscolaire jusqu'à des adultes verbaux. En effet, l'ADOS comporte quatre modules, chaque module possédant son propre protocole avec des activités différentes pour les enfants, les adolescents ou les adultes. Le choix du module se fait en fonction de l'âge chronologique et du niveau de langage expressif.

Les informations de l'ADI-R utilisées en complémentarité de l'ADOS permettent d'évaluer de manière plus fiable et plus précise le diagnostic de TSA que s'ils sont utilisés séparément (Risi et coll., 2006 ; Kim et Lord, 2012). En effet, l'ADI-R donne accès à l'histoire développementale du patient, à la description de son fonctionnement et à une évaluation de son comportement grâce aux éléments rapportés par les parents. L'ADOS donne un résumé des observations standardisées du clinicien qui évalue l'enfant sur son comportement dans des contextes particuliers liés aux rapports sociaux et à la communication.

## 2. Hétérogénéité clinique

Les manifestations cliniques et comportementales des TSA varient très largement d'un individu à un autre. Par exemple, il a été observé un âge variable d'apparition des troubles, des niveaux de langage très hétérogènes mais également différents degrés des atteintes comportementales. En plus des caractéristiques principales des TSA décrites dans la **section 1.2**, d'autres manifestations cliniques et comportementales peuvent être associées, à des fréquences variables (**Figure 2**).

Certains patients atteints de TSA peuvent présenter des dysmorphies faciales ou d'autres malformations congénitales ou encore des anomalies neuromusculaires ou métaboliques. Dans ce cas, l'autisme est dit **syndromique**. Le terme autisme syndromique fait donc référence à la présentation clinique du patient et non pas au fait qu'une maladie ou un syndrome génétique ait été identifié chez lui. En effet, un syndrome génétique peut être découvert chez un patient ayant une présentation clinique non syndromique.

Les TSA sont souvent associés à d'autres troubles du développement. La **déficience intellectuelle**, tout comme l'autisme, est une pathologie neurodéveloppementale très hétérogène sur le plan clinique et étiologique. La prévalence dans la population générale est d'environ 2 à 3 % (Leonard et Wen, 2002). Elle est caractérisée par un quotient intellectuel (QI) inférieur à 70 et classée en quatre niveaux de sévérité (légère, modérée, sévère et profonde) en fonction de la valeur globale du QI. Selon les différentes études réalisées, en moyenne 50 à 70 % des patients avec TSA présentent une déficience intellectuelle (Williams et coll., 2006 ; Fombonne, 2009 ; Fombonne et coll., 2011). L'association est variable selon les cohortes étudiées et il a été observé que la prévalence de la déficience intellectuelle est plus élevée dans les cohortes de patients avec un TSA sévère en particulier chez les individus qui présentent un autisme syndromique.

De nombreuses études ont également montré une association entre l'**épilepsie** et les TSA (Tuchman et Rapin, 2002 ; Canitano, 2007 ; Berg et Plioplys, 2012). L'épilepsie est l'une des pathologies neurologiques les plus fréquentes, avec une prévalence d'environ 1,5 % dans la population générale. On estime qu'un tiers des patients atteints de TSA souffrent également d'épilepsie (Mouridsen et coll., 1999, 2011). La fréquence varie selon les cohortes étudiées et notamment en fonction de l'âge des patients, du sexe et du niveau d'atteinte cognitive.

On note également une association entre les TSA et d'**autres troubles psychiatriques** tels que la dépression, l'anxiété, le trouble du déficit de l'attention/hyperactivité (TDAH) (Goldstein et Schwebach, 2004), le trouble bipolaire ou le trouble obsessionnel compulsif (Gillott et coll., 2001 ; Simonoff et coll., 2008 ; Kanne et coll., 2009). Plus rarement, les individus avec TSA peuvent présenter un diagnostic additionnel de schizophrénie. D'ailleurs, certaines caractéristiques comportementales des TSA sont proches des symptômes négatifs observés dans la schizophrénie, en particulier l'absence d'expression d'émotions et le repli social.

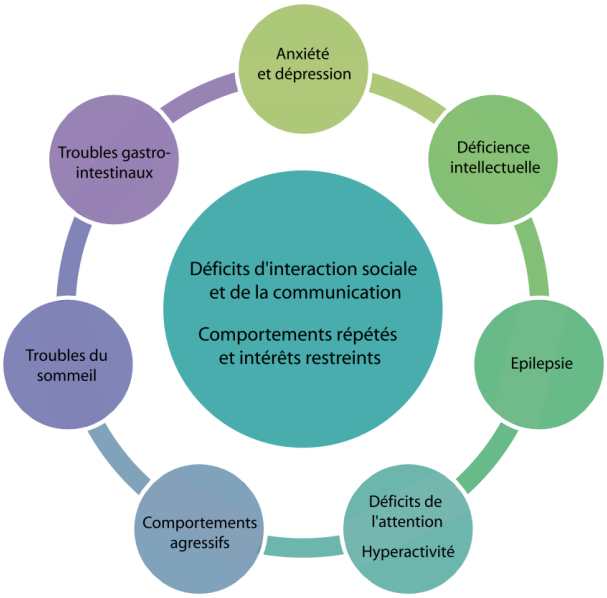



**Figure 2. Hétérogénéité clinique des TSA**
Les caractéristiques principales des TSA sont représentées dans le cercle interne, entourées des autres manifestations communément observées chez les patients atteints d'autisme (figure adaptée de Spooren et coll., 2012).

# 3. Epidémiologie

## 3.1 Prévalence

La prévalence correspond au nombre de cas atteints d'une pathologie donnée rapporté à la population générale. Elle représente donc le nombre de cas à un moment donné, que le diagnostic soit ancien ou récent. Elle ne doit pas être confondue avec l'incidence, qui elle, représente le nombre de nouveaux cas par an. Ces deux indicateurs de fréquence sont généralement utilisés en épidémiologie.

***Augmentation de la prévalence de l'autisme***

D'après les dernières estimations du *Center for Disease Control and Prevention* (CDC) obtenues sur une large cohorte d'enfants âgés de 8 ans en 2010, la prévalence des TSA serait de 1 enfant sur 68 aux Etats-Unis (Autism and Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, 2014). Ce chiffre est plus de 30 fois supérieur à la première estimation de la prévalence de l'autisme obtenue lors d'une étude réalisée en 1966 (Lotter, 1966). Cette étude avait calculé la prévalence

de l'autisme au sens strict (et non de l'ensemble des TSA) en considérant toute la population âgée de 8 à 10 ans dans le comté du Middlesex (ancien comté d'Angleterre) soit près de 78 000 enfants. Après évaluation par des questionnaires comportementaux, la prévalence avait été estimée à environ 4,5/10 000 (soit 0,04 %). Les études parues par la suite, prenant en compte l'ensemble des TSA, ont rapporté un nombre toujours croissant de la prévalence, pour atteindre aujourd'hui cette valeur d'environ 1,5 %.

La prévalence des TSA semble donc avoir considérablement augmenté ces dernières années et pendant un temps, la notion d'une 'épidémie' d'autisme a été évoquée. Divers facteurs environnementaux comme la pollution, les pesticides ou autres toxines ont été incriminés malgré le manque d'évidence scientifique. Il semblerait que l'augmentation de la prévalence soit en réalité liée à d'autres facteurs comme **l'élargissement des critères diagnostiques**. Une **meilleure accessibilité aux services de santé et d'éducation** (notamment à des tests diagnostiques objectifs qui permettent un diagnostic précoce) et une **meilleure information** des professionnels de la santé et de la population générale pourraient également contribuer à expliquer cette augmentation du nombre de cas. Enfin, le phénomène de **substitution de diagnostic** a également été proposé pour expliquer l'augmentation de la prévalence des TSA. En effet, de nombreux enfants ayant reçu un diagnostic initial de déficience intellectuelle ont par la suite été diagnostiqués avec autisme après avoir été réévalués (Shattuck, 2006 ; King et Bearman, 2009). On remarque d'ailleurs que l'augmentation de la prévalence des TSA entre 1994 et 2003 est directement liée à une diminution de la prévalence de la déficience intellectuelle sur cette même période (**Figure 3**) (Shattuck, 2006). De plus, la substitution peut également être expliquée par le fait qu'une meilleure prise en charge et un meilleur accès thérapeutique sont proposés aux patients atteints d'autisme par rapport à ceux atteints de déficience intellectuelle. Ainsi, les parents 'préféreront' que leur enfant soit diagnostiqué avec autisme plutôt qu'avec déficience intellectuelle afin de bénéficier de meilleures opportunités notamment sur le plan éducatif.

Si la prévalence de l'autisme avait effectivement augmenté ces dernières années, la prévalence des TSA dans une cohorte d'adultes devrait être plus faible comparée à la prévalence actuelle. Or, selon une étude épidémiologique réalisée en Angleterre et basée sur une cohorte de patients de 16 ans et plus, la prévalence estimée est similaire à celle calculée dans les cohortes d'enfants (environ 1 %), suggérant que la prévalence des TSA est stable (Brugha et coll., 2011).

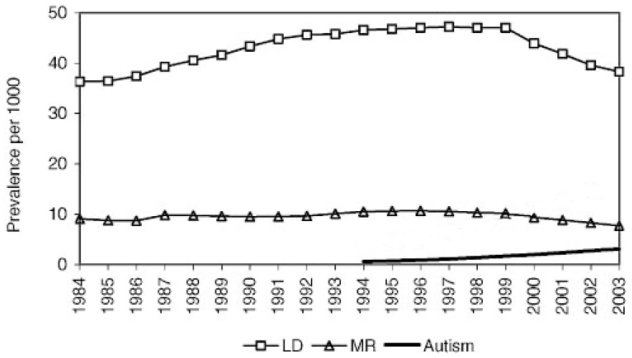



**Figure 3. Effet de la substitution de diagnostic dans la prévalence des TSA aux USA**
L'augmentation de la prévalence des TSA (*Autism*) entre les années 1994 et 2003 est directement corrélée à la diminution de celle de la déficience intellectuelle (MR, *mental retardation*) et des troubles d'aprentissage (LD, *learning disability*) sur la même période (figure adaptée de Shattuck, 2006).

## 3.2 Sexe ratio

Les études épidémiologiques réalisées dans les TSA ont toujours reporté un nombre plus important de garçons atteints comparé aux filles. Le sexe ratio généralement admis est de 4 garçons pour une fille mais

peut varier d'une cohorte à une autre (Fombonne, 2009 ; Fombonne et coll., 2011 ; Centers for Disease Control (CDC), 2012). En particulier, il a été observé que le niveau de déficience intellectuelle influençait grandement le sexe ratio. En effet, le déséquilibre garçons-filles s'accentue dans les cohortes de patients aux capacités cognitives normales voire supérieures pour atteindre jusqu'à 10:1 dans des groupes d'individus atteints du syndrome d'Asperger (Gillberg et coll., 2006). Par contre, le sexe ratio a tendance à s'équilibrer dans les cohortes d'individus plus sévèrement atteints (Rivet et Matson, 2011).

Plusieurs hypothèses pouvant expliquer ce fort déséquilibre ont été avancées (pour revue voir Werling et Geschwind, 2013) :

- Les **chromosomes sexuels** ont été proposés comme un des responsables du déséquilibre du sexe ratio observé dans les TSA. En effet, de nombreux gènes localisés sur le chromosome X ont été impliqués dans l'étiologie des TSA (Betancur, 2011). Le fait que les filles possèdent deux copies du chromosome X les protège contre les altérations génétiques récessives liées à l'X. De plus, il a été montré que les individus porteurs d'aneuploïdies des chromosomes sexuels (syndrome de Klinefelter et syndrome XYY) présentaient un risque accru de développer des TSA (Bishop et coll., 2011 ; Ross et coll., 2012 ; Zafeiriou et coll., 2013). Cependant, même s'il est clair que les chromosomes sexuels jouent un rôle, ils ne peuvent pas expliquer à eux seuls un tel déséquilibre.
- L'hypothèse du **cerveau hypermasculin** (*extreme male brain theory*), aussi appelée théorie d'empathisation-systématisation, proposée par Simon Baron-Cohen, est basée sur l'existence d'un dimorphisme sexuel entre les cerveaux masculins et féminins (Baron-Cohen, 2002). Selon cette hypothèse, les hommes sont plus performants pour analyser les variables d'un système, en déduire les règles sous-jacentes et ainsi prédire et contrôler ce système (systématisation). En contrepartie, leurs capacités à identifier les pensées et les sentiments des autres personnes pour y réagir de manière appropriée (empathisation) seraient moins développées que chez les femmes. Les individus atteints de TSA ont un déficit dans les processus d'empathie mais ne sont pas affectés (voire sont meilleurs) dans les tâches qui requièrent la systématisation, notamment les tâches visuo-spatiales. Ainsi, les personnes atteintes d'autisme auraient un cerveau de type masculin poussé à l'extrême. Par conséquent, les garçons seraient plus susceptibles d'être atteints d'autisme puisque naturellement plus enclins à comprendre et interagir en utilisant la systématisation. Dans une étude, le score de systématisation (SQ pour '*systemizing quotient*') a été calculé chez trois groupes d'individus : un groupe atteint de TSA, un groupe de femmes 'typiques' et un groupe d'hommes 'typiques'. Les résultats montrent que les patients autistes présentent une distribution modale centrée sur un score compris entre 76 et 90 alors que chez les hommes typiques, la distribution est centrée sur l'intervalle 61-75, et sur l'intervalle 31-45 chez les femmes (**Figure 4**). La courbe des hommes est donc par nature plus proche de celle des patients avec TSA.
- Un autre modèle, complémentaire au précédent, propose l'existence de **facteurs protecteurs présents spécifiquement chez les filles**, qui présenteraient donc un seuil de tolérance plus élevé que les garçons par rapport à la pathologie (Robinson et coll., 2013 ; Jacquemont et coll., 2014). D'après cette hypothèse, les filles seraient donc moins susceptibles aux atteintes génétiques et présenteraient une 'charge étiologique' plus forte pour manifester un phénotype d'autisme par rapport aux garçons.
- Enfin, l'existence d'un **biais dans les outils diagnostiques** pourrait également expliquer le déséquilibre filles-garçons dans les TSA. Il a été suggéré que les outils diagnostiques utilisés seraient moins performants pour détecter les filles autistes, notamment celles atteintes du syndrome d'Asperger ou autisme de 'haut niveau'. En effet, les filles auraient plus tendance à intérioriser les manifestations cliniques (anxiété, dépression) par rapport aux garçons, elles montreraient donc moins facilement des comportements agressifs ou violents, et seraient plus rarement présentées en consultation. De plus, du fait de leurs meilleures aptitudes pour l'expression verbale et les interactions sociales, les filles

pourraient montrer un niveau de fonctionnement plus préservé dans ces domaines lorsqu'elles sont atteintes.

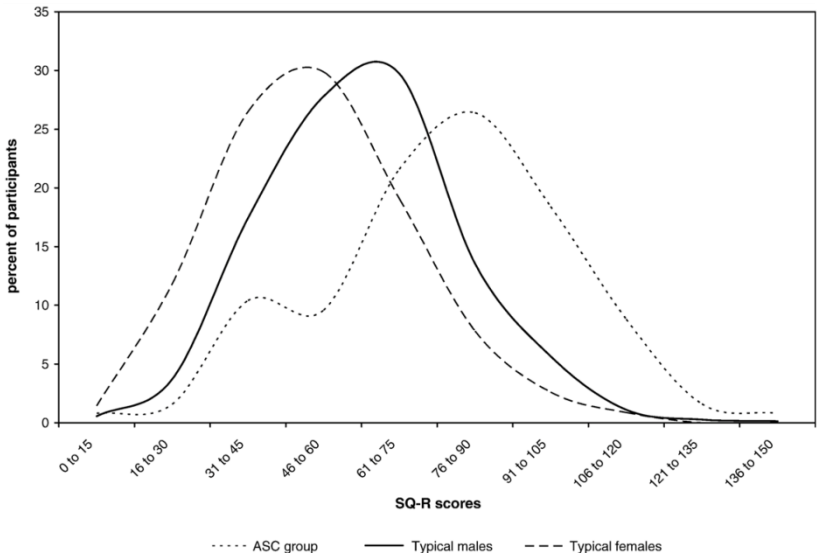



**Figure 4. Distribution du score de systématisation chez des patients atteints de TSA et des hommes et des femmes 'typiques'**

La courbe du groupe des hommes typiques est plus proche de celle des patients atteints de TSA par rapport à celle des femmes (figure tirée de Wheelwright et coll., 2006).

## 4. Evolution et prise en charge

Aujourd'hui, il n'existe pas de traitement pour l'autisme. Cependant, il a été montré qu'une prise en charge précoce et adaptée permet aux enfants atteints d'améliorer leurs capacités à interagir avec le monde extérieur et à mieux s'y adapter. Le nombre de méthodes de prise en charge des patients avec TSA a beaucoup augmenté ces dernières années. Elles sont essentiellement basées sur le comportement, avec comme objectif premier de développer le langage, les compétences cognitives, sensorielles et motrices et surtout d'apprendre à interagir avec les autres et à acquérir de l'autonomie. Parmi ces programmes, on retrouve la méthode de l'analyse du comportement appliquée (*Applied Behavior Analysis*, ABA) (Lovaas, 1987), le programme TEACCH (*Treatment and Education of Autistic and related Communication handicapped CHildren*) plutôt ciblé sur le développement de l'autonomie (Mesibov et Shea, 2010) ou encore la méthode PECS (*Picture Exchange Communication System*) qui propose des techniques alternatives et augmentatives de la communication (Bondy et Frost, 2001). Dans tous les cas, la littérature scientifique insiste sur la nécessité d'une intervention éducative intensive, précoce et personnalisée.

Des traitements médicamenteux ont également montré des bénéfices pour certaines manifestations associées aux TSA. La mélatonine, une neurohormone impliquée dans le contrôle des cycles du sommeil et du rythme circadien, aide à réguler les troubles du sommeil fréquemment observés chez les individus atteints de TSA (Harrington et coll., 2006 ; Malow et coll., 2012). D'autres traitements pharmacologiques de type neuroleptiques, anxiolitiques ou psychostimulants sont aussi utilisés pour traiter les troubles psychiatriques associés aux TSA tels que l'anxiété, l'hyperactivité, l'agression ou encore les comportements répétés (Kolevzon, 2013). Par ailleurs, de récentes études ont montré que l'administration intranasale d'ocytocine (hormone impliquée dans la reconnaissance sociale, l'anxiété ou encore les comportements maternels) améliore les comportements sociaux chez des patients avec TSA (Andari et coll., 2010 ; Guastella et coll., 2010), mais des recherches additionnelles sont nécessaires afin de prouver l'efficacité de ce traitement.

La plupart des patients atteints d'autisme sont vus lorsqu'ils sont enfants ou jeunes adultes. Par conséquent, peu d'éléments sont connus quant à l'évolution des manifestations comportementales au cours du temps. Cependant, certains exemples de patients adultes montrent que les individus atteints de TSA peuvent parfois réussir à s'intégrer à la société, suivre un parcours scolaire, obtenir un travail et même exceller dans un domaine particulier, notamment pour les autistes dits 'de haut niveau'. A l'inverse, une étude épidémiologique réalisée sur une cohorte d'adultes a montré que la prévalence des TSA était significativement augmentée chez ceux n'ayant pas reçu d'éducation contrairement à ceux ayant suivi des études universitaires (Brugha et coll., 2011). De plus, le diagnostic de TSA était significativement associé au fait d'être célibataire et socio-économiquement dépendant. Les techniques de prise en charge actuelles, la meilleure accessibilité aux services de santé et d'éducation ainsi que la meilleure compréhension de la pathologie vont certainement améliorer la qualité de vie des futures générations de patients atteints de TSA.

# Partie II : Contribution des facteurs génétiques à l'étiologie de l'autisme

## 1. Implication des facteurs génétiques dans les TSA

Bien que Leo Kanner et Hans Asperger aient suggéré dès leurs premières descriptions le caractère inné des comportements anormaux qu'ils observaient chez leurs jeunes patients, il faudra attendre plus de vingt ans avant que soit reconnu le rôle des facteurs génétiques dans l'étiologie de l'autisme. Deux grands types d'études ont contribué à démontrer la composante génétique de l'autisme : les études familiales et les études de jumeaux. En outre, la description de syndromes génétiques rares mais fortement associés aux TSA ainsi que de remaniements chromosomiques identifiés chez des patients ont également mis en avant la contribution des facteurs génétiques dans l'autisme.

### 1.1 Etudes familiales et de jumeaux

L'héritabilité est définie comme la part de variation phénotypique pouvant être expliquée par des facteurs génétiques, en opposition aux facteurs dits 'non génétiques', tels que les effets environnementaux, épigénétiques ou encore stochastiques.

Les **études de jumeaux** permettent d'estimer la contribution des facteurs génétiques et non génétiques à l'étiologie d'une pathologie. Elles sont basées sur la comparaison du taux de concordance entre les jumeaux monozygotes (qui ont l'intégralité de leur génome en commun) et les jumeaux dizygotes (qui ne partagent que la moitié de leurs gènes). Ainsi la discordance observée chez les jumeaux monozygotes correspond à la part attribuable aux facteurs non génétiques (influences épigénétiques, environnementales, stochastiques, ou erreurs de mesure). Dès la première étude de jumeaux dans l'autisme réalisée sur 21 paires (11 monozygotes et 10 dizygotes), il a été observé un **taux de concordance** beaucoup plus élevé chez les jumeaux monozygotes (36 %) comparé aux jumeaux dizygotes (0 %) (Folstein et Rutter, 1977). D'autres études, elles aussi réalisées sur de petits groupes de jumeaux, ont confirmé ces résultats, avec des taux de concordance variables allant de 60 à 96 % chez les jumeaux monozygotes contre 0 à 24 % chez les dizygotes (Ritvo et coll., 1985 ; Steffenburg et coll., 1989 ; Bailey et coll., 1995). Ces premières études ont ainsi révolutionné la vision de l'époque sur l'étiologie de l'autisme en impliquant une base génétique pour la première fois. Plus tard, des études réalisées sur des échantillons de taille plus grande allant jusqu'à 277 paires de jumeaux avec TSA, ont reporté des taux de concordance compris entre 39 et 88 % chez les monozygotes et entre 15 et 31 % pour les dizygotes (Rosenberg et coll., 2009 ; Lichtenstein et coll., 2010 ; Hallmayer et coll., 2011). Le **Tableau 1** présente les principales études de jumeaux réalisées dans les TSA et les estimations de l'héritabilité des TSA. Dans leur étude, Hallmayer et coll. (2011) ont utilisé un modèle incluant des effets liés aux facteurs environnementaux partagés, non pris en compte par les études précédentes et ont calculé une héritabilité plus faible que précédemment estimée (37 % contre 70-93 %), suggérant des effets 'environnementaux' plus importants dans le risque de développer un TSA. Cependant, les intervalles de confiance de ces estimations sont très larges (8-84 %), chevauchant les estimations des précédentes études.

Les études familiales comparent la fréquence d'une pathologie chez les apparentés de premier degré d'un individu atteint par rapport à celle de la population générale. Ce type d'études a montré l'existence d'une forte **agrégation familiale** dans les TSA et d'un **risque de récurrence élevé** pour la fratrie d'un individu atteint de TSA. Les premières études ne considéraient que l'autisme au sens strict et ont estimé des taux de récurrence de l'ordre de 3 à 6 % (August et coll., 1981 ; Tsai et coll., 1981 ; Baird et August,

1985 ; Ritvo et coll., 1989), ce qui correspondait à un risque 50 à 100 fois supérieur par rapport à la prévalence dans la population générale estimée à l'époque. Une étude récente prenant en compte les TSA dans leur globalité a estimé le taux de récurrence à 18,7 % dans une cohorte de 600 enfants ayant au moins un frère ou une sœur plus âgé(é) atteint(e) de TSA, par rapport à la prévalence dans la population générale qui est d'environ 1 % (Ozonoff et coll., 2011). Une autre étude, réalisée sur plus de 5 000 familles avec un enfant atteint de TSA et au moins un frère ou une sœur, a montré que le risque de récurrence des caractéristiques autistiques chez les frères et sœurs d'enfants atteints est deux fois plus important que celui des demi-frères et demi-sœurs (environ 10 et 5 % respectivement), ce qui est un argument supplémentaire en faveur d'un composante génétique forte dans les TSA (Constantino et coll., 2013).

**Tableau 1. Etudes de jumeaux dans les TSA**

| Etude | Diagnostic | Nombre de jumeaux | Concordance MZ DZ | Estimation de l'héritabilité |
|---|---|---|---|---|
| Folstein et Rutter (1977) | Critères de Kanner | 11 MZ, 10 DZ | Autisme : MZ 36 %, DZ 0 % BAP : MZ 82 %, DZ 10 % | |
| Ritvo et coll. (1985) | DSM-III | 23 MZ, 17 DZ | Autisme : MZ 96 %, DZ 24 % | |
| Steffenburg et coll. (1989) | DSM-III-R | 11 MZ, 10 DZ et une paire de triplets identiques | Autisme : MZ 91 %, DZ 0 % BAP : MZ 91%, DZ 30 % | |
| Bailey et coll. (1995) | ICD-10 | 25 MZ, 20 DZ et deux paires de triplets | Autisme : MZ 60 %, DZ, 0 % BAP : MZ 92 %, DZ 10 % | 91-93 % |
| Taniai et coll. (2008) | CARS | 19 MZ, 26 DZ | TSA : MZ 95 %, DZ 31 % | 73 % pour les garçons, 87% pour les filles |
| Rosenberg et coll. (2009) | IAN; diagnostic fourni par les familles | 67 MZ, 210 DZ | TSA : MZ 88 %, DZ 31 % | |
| Lichtenstein et coll. (2010) | A-TAC | 29 MZ, 88 DZ | TSA : MZ 39 %, DZ 15 % | 80 % |
| Hallmayer et coll. (2011) | ADI-R et ADOS | 54 MZ, 138 DZ | Autisme : MZ 58-60 %, DZ 21-27 %<br>TSA : MZ 50-77 %, DZ 31-36 % | 37 % autisme, 38 % TSA |

ADI-R, *Autism Diagnostic Interview-Revised* ; ADOS, *Autism Diagnostic Observation Schedule* ; A-TAC, *Autism, tics, ADHD and other comorbidities inventory* ; BAP, *Broader Autism Phenotype*, phénotype élargi de l'autisme ; CARS, *Childhood Autism Rating Scale* ; CIM-10, Classification Internationale des Maladies, 10ème édition ; DSM-III-R, Manuel Diagnostique et Statistique des Maladies Mentales, troisième édition, révisé ; DZ, dizygote; IAN, *Interactive Autism Network* ; MZ, monozygote (tableau adapté de Gaugler et coll. 2014).

Les études familiales ont également montré que les apparentés de premier degré de patients atteints de TSA présentaient une fréquence élevée de traits autistiques, similaires à ceux présents chez les individus atteints mais de sévérité moindre. Ces phénotypes plus légers et qui suggèrent un effet familial ont été regroupés sous le terme de '**phénotype élargi de l'autisme**' (*broader autism phenotype*) (Bolton et coll., 1994 ; Constantino et Todd, 2005 ; Losh et coll., 2008). Une récente étude réalisée sur une cohorte incluant près de 20 000 paires de jumeaux suédois recrutés dans la population générale a montré une forte corrélation entre l'autisme au sens strict et les traits autistiques, avec des estimations de l'héritabilité similaires, suggérant que ces deux catégories de troubles présentent des bases génétiques communes (Lundstrom et coll., 2012). L'autisme pourrait donc correspondre à l'extrême variation de traits phénotypiques s'inscrivant dans un continuum au sein de la population générale.

Plus récemment, la plus grande étude familiale dans les TSA a été réalisée sur une cohorte incluant plus de deux millions d'enfants suédois dont environ 14 500 avec un diagnostic de TSA (Sandin et coll.,

2014). Afin d'augmenter la puissance statistique, cette étude a inclus les frères et sœurs, les demi-frères et demi-sœurs ainsi que les jumeaux. Les auteurs ont estimé les influences génétiques additives et non additives ainsi que les influences environnementales partagées et non partagées. D'après leurs résultats, ils ont retenu un modèle incorporant une covariance due à des facteurs génétiques additifs et à l'environnement non partagé. Le risque relatif de récurrence (RRR) chez les jumeaux monozygotes est de 153 contre 8 chez les dizygotes. Le RRR chez les frères et sœurs d'enfants atteints est similaire à celui des jumeaux dizygotes (10) et environ 3 fois supérieur à celui des demi-frères et demi-sœurs d'enfants atteints **(Figure 5)**. Le RRR est d'environ 2 pour les cousins, ce qui confirme le fait que le risque de récurrence est corrélé au degré de parenté avec l'individu atteint.

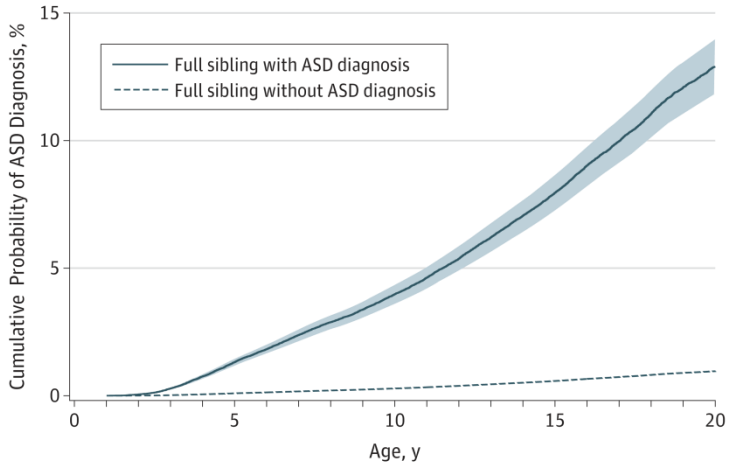



**Figure 5. Risque de récurrence dans les TSA**

La courbe en trait continu représente la probabilité d'être diagnostiqué avec un TSA pour un frère ou une sœur d'un enfant atteint. Celle-ci est plus de 10 fois supérieure à celle d'un enfant n'ayant pas de frère ou de sœur atteint (courbe en pointillé). La partie grisée représente l'intervalle de confiance à 95 % (figure tirée de Sandin et coll., 2014).

Les auteurs ont estimé l'héritabilité des TSA à environ 50 %, suggérant une part de variabilité liée aux influences 'environnementales' plus importante que précédemment supposée. Cependant, cette conclusion est à tempérer puisqu'elle est largement dépendante des modèles statistiques sous-jacents pour lesquels aucun consensus n'a été admis à l'heure actuelle. En outre, la discordance entre les jumeaux monozygotes est ici résumée par le terme 'influences environnementales non partagées' qui n'implique pas uniquement les facteurs environnementaux extérieurs. En effet, cette discordance peut également correspondre à des effets stochastiques, à des erreurs de mesure ou à une expression variable interindividuelle au cours du développement.

Malgré les variations dues aux différences de la taille des cohortes et des critères diagnostiques utilisés, les études familiales ont toujours montré un risque de récurrence très élevé dans la fratrie alors que ce risque est certainement sous-estimé par l'existence du phénomène d''arrêt de reproduction' (*reproductive stoppage*) observé dans les TSA. En effet, il a été observé que dans les familles avec un premier enfant atteint de TSA, la probabilité d'avoir un deuxième enfant est réduite comparée à celle d'un couple sans enfant atteint (Hoffmann et coll., 2014). De même, les taux de concordance calculés entre jumeaux monozygotes, bien que variables, sont toujours largement supérieurs à ceux observés chez les dizygotes. L'ensemble de ces études montre l'existence d'une composante génétique forte dans l'étiologie des TSA. Cependant, l'héritabilité dans l'autisme serait plus faible que ce qu'il avait été estimé avant, laissant une part non négligeable à l'influence de facteurs dits 'non génétiques'.

### 1.2 Facteurs non génétiques

Longtemps, des causes environnementales jouant un rôle étiologique ont été suspectées. Il a effectivement été montré que certains **facteurs environnementaux, en particulier liés à la grossesse**, constituent des facteurs de risque importants. Par exemple, la consommation d'alcool par la mère pendant sa grossesse peut entraîner un syndrome d'alcoolisation fœtale, qui peut être associé à un TSA chez l'enfant (Stevens et coll., 2013). De même, l'exposition à l'antiépileptique acide valproïque pendant la grossesse (Christensen et coll., 2013), une hypothyroïdie de la mère non traitée (Andersen et coll., 2014) ou encore des infections maternelles virales telles que la rubéole (Chess, 1971 ; Chess et coll., 1978) ou le cytomégalovirus (Yamashita et coll., 2003 ; Sweeten et coll., 2004) ont également été impliquées dans l'étiologie de l'autisme. D'autres facteurs pré-, peri- et post-nataux comme la prématurité, la souffrance fœtale ou l'hyperbilirubinémie augmentent également le risque de développer un TSA (Gardener et coll., 2011 ; Guinchat et coll., 2012). Cependant, ces causes environnementales avérées n'expliquent qu'une très faible fraction de cas.

Parmi les autres influences environnementales étudiées dans la population comme l'exposition à des toxiques, à des pesticides ou à la pollution, la plupart sont mal comprises pour le moment et aucune association claire n'a pu être mise en évidence. Cependant, la part de la variance phénotypique non attribuée aux causes génétiques n'est pas limitée aux facteurs dits 'environnementaux' mais comprend également les facteurs épigénétiques, les effets stochastiques ou les effets de mesure. L'étude des facteurs épigénétiques potentiellement impliqués dans l'autisme est un domaine en pleine expansion dans lequel peu d'évidences convaincantes ont été apportées pour le moment mais qui pourrait révéler de nouvelles pistes étiologiques. Par exemple, il a été observé des différences de méthylation entre les génomes de jumeaux discordants pour les TSA, indiquant que ce phénomène pourrait jouer un rôle dans l'étiologie de l'autisme (Wong et coll., 2014).

Si la part de l'héritabilité attribuable aux facteurs non génétiques représente effectivement 50 %, la très grande majorité d'entre eux reste à mettre en évidence.

## 2. Recherche de variants communs dans les TSA

La recherche d'une étiologie génétique de l'autisme s'est focalisée, dans un premier temps, sur la recherche de facteurs communs entre les individus atteints. En effet, à l'époque, l'autisme était considéré comme une entité unique ayant une cause unique. En outre, puisque les études familiales dans les TSA avaient montré l'existence d'un phénotype élargi de l'autisme, il a été supposé l'existence de variants communs chez les apparentés des individus atteints, lesquels ne seraient pas suffisants à l'expression du phénotype complet de l'autisme mais seraient responsables des traits autistiques chez eux. A cette époque, les outils d'analyse les plus adaptés à la recherche de variants communs dans les maladies à hérédité complexe comme les troubles psychiatriques étaient les études de liaison et d'association. Pendant deux décennies, le domaine de la génétique de l'autisme s'est donc attelé à la recherche de variants communs à tous les individus atteints.

### 2.1 Etudes de liaison

Les analyses de liaison permettent de mettre en évidence des régions chromosomiques liées à un phénotype particulier. Elles sont basées sur le calcul d'un logarithme des probabilités, le LOD (*logarithm of odds*) score, qui permet d'établir le degré de liaison génétique entre deux loci en se basant sur les probabilités de transmission des allèles à la descendance. La méthode d'analyse de liaison non paramétrique permet de chercher des facteurs liés sans modèle de transmission *a priori*. Dans l'analyse de paires de germains atteints, on calcule le pourcentage d'allèles parentaux partagés par les enfants

atteints. Si la proportion des allèles partagés à un marqueur donné diffère de la proportion attendue, les germains atteints ont donc plus souvent reçu les mêmes allèles de leurs parents que ne le voudrait le hasard, ce qui indique une liaison génétique entre le marqueur et le phénotype.

En 1998, la première étude de liaison réalisée dans l'autisme sur une centaine de familles à cas multiples (ayant au moins deux enfants atteints) a montré l'existence d'une liaison avec six régions chromosomiques différentes et notamment avec une région d'environ 10 Mb en 7q (International Molecular Genetic Study of Autism Consortium, 1998). De nombreuses études de ce type ont ensuite été réalisées dans les TSA dans des échantillons de familles à cas multiples de taille variable mais toujours faible. La plupart ont mis en évidence différentes régions génétiques liées significativement à l'autisme mais aucune région n'a été répliquée de manière fiable. En 2010, la plus grande étude de liaison réalisée sur un échantillon de 1 181 familles n'a identifié aucune région significativement liée aux TSA (Szatmari et coll., 2007). En particulier, cette étude n'a répliqué aucun des résultats obtenus auparavant, suggérant que les régions précédemment incriminées correspondaient en réalité à des faux-positifs. Rétrospectivement, ces échecs étaient dus en grande partie à la taille très insuffisante des échantillons testés et surtout à la forte hétérogénéité génétique des TSA, qui avait été largement sous-estimée.

## 2.2 Etudes d'association

Les études d'association consistent à comparer la fréquence d'allèles dans une population d'individus atteints par rapport à une population d'individus témoins dans le but de mettre en évidence des allèles de prédisposition à la pathologie. Elles permettent donc de déterminer une association entre un haplotype ou un marqueur donné et le fait de développer une maladie et sont plus puissantes que les études de liaison dans la recherche de variants à effet faible. Il existe deux types d'études d'association : les études de population (cas/témoins) et les études familiales (par exemple, le test du déséquilibre de transmission).

Les premières études d'association ont été réalisées avec l'approche 'gène candidat' dont le principe est de tester un nombre souvent restreint de variants génétiques localisés dans des loci intéressants de par leur fonction ou leur localisation. Ainsi, dans les années 90 et jusqu'au début des années 2000, de multiples études de ce type ont été réalisées dans les TSA et d'autres troubles neuropsychiatriques, en analysant des gènes candidats fonctionnels ou ceux mis en évidence par des études de liaison ou par des découvertes cytogénétiques. Cependant, aucun des résultats n'a été répliqué de façon convaincante, en partie à cause de la taille très limitée des échantillons utilisés dans ces premières études (souvent sur une centaine de familles) dans les TSA mais aussi dans d'autres pathologies à hérédité complexe (Hirschhorn et coll., 2002).

Grâce aux avancées technologiques et notamment à l'essor des puces à ADN, des études d'association 'génome entier' (*genome-wide association studies*, GWAS) ont été entreprises ces dernières années. Leur avantage est qu'elles ne nécessitent pas la sélection de gènes ou loci candidats *a priori*. Quatre grandes études d'association ont été publiées avec des cohortes allant d'environ 1 500 à plus de 2 700 familles (Wang et coll., 2009 ; Weiss et coll., 2009 ; Anney et coll., 2010 ; Anney et coll., 2012) en utilisant des puces SNP (*single nucleotide polymorphism*). En effet, la découverte de ces variations nucléotidiques, très nombreuses et très variables entre les individus, les a rapidement placées comme marqueurs de choix pour les études d'association. Deux loci différents du chromosome 5 ont d'abord été impliqués: une région non génique en 5p14.1 entre les gènes *CDH9* et *CDH10* (codant pour des cadhérines neuronales) (Wang et coll., 2009) et une région comprenant le gène *SEMA5A* (*semaphorin 5*) au locus 5p15.2 (Weiss et coll., 2009). En augmentant la résolution à un million de marqueurs, Anney et coll. (2010) ont par la suite montré une association significative avec un groupe de SNPs situés dans un intron du gène *MACROD2* (*MACRO domain containing 2*) localisé en 20p12.2, non répliquée en doublant la taille de leur échantillon

(Anney et coll., 2012). Par ailleurs, cette étude n'a répliqué aucune des associations précédemment rapportées. En conclusion, malgré l'utilisation de milliers voire un million de marqueurs, aucune région n'a pu être associée aux TSA de façon reproductible.

Selon une méta-analyse qui reprend tous les résultats des études d'association génome entier dans les TSA (à l'exception de l'étude d'Anney et coll. 2012), les variants communs à effet fort ou modéré (risque relatif, *odds ratio*, OR > 1,5) n'existent pas dans les TSA sinon ils auraient été identifiés avec la taille des échantillons étudiés (99 % de puissance statistique pour trouver un variant associé à un OR > 1,4 avec 2 500 familles) (Devlin et coll., 2011). Par contre, selon cette même analyse, l'existence de facteurs de risque communs à faible effet (risque relatif < 1,2) semble probable mais la taille des cohortes à réunir pour les mettre en évidence a largement été sous-estimée dans les TSA, d'au moins un ordre de grandeur. Pour espérer pouvoir trouver des variants communs il faudra donc réunir des échantillons d'au moins 15 000 à 20 000 patients et un nombre équivalent de témoins, tout en sachant que si des allèles de susceptibilité sont ainsi détectés, ils seront associés à un risque faible. Ces résultats ont été confortés par une autre analyse qui a estimé la part de la variance dans les TSA attribuable à des facteurs communs à 40 % pour les familles à cas isolés et 60 % pour les familles à cas multiples (Klei et coll., 2012). Les auteurs concluent qu'une myriade de variants communs à effet (très) faible agissant de façon additive, contribue significativement au risque de TSA mais que leur identification nécessite de réunir des cohortes très importantes de patients.

D'ailleurs, dans d'autres troubles neurodéveloppementaux et notamment dans la schizophrénie, des études d'association génome entier ont permis de mettre en évidence des facteurs de risque avec des évidences statistiques convaincantes et qui ont été répliqués dans des cohortes indépendantes. A la différence des études réalisées dans les TSA, les cohortes étudiées étaient de très grande taille avec plus de 15 000 patients et 20 000 témoins (Purcell et coll., 2009 ; Shi et coll., 2009 ; Stefansson et coll., 2009 ; Schizophrenia Psychiatric Genome-Wide Association Study Consortium, 2011). Une dizaine de régions génomiques montrent des évidences fiables d'association avec la schizophrénie, incluant la région du complexe majeur d'histocompatibilité en 6p22.1 (Stefansson et coll., 2008 ; Purcell et coll., 2009 ; Shi et coll., 2009), la région 11q24.2 en amont du gène *NRGN* (codant pour une neurogranine) (Stefansson et coll., 2009), ou encore le microARN *MIR137* (Schizophrenia Psychiatric Genome-Wide Association Study Consortium, 2011 ; Ripke et coll., 2013). Cependant, les variants mis en évidence sont associés à des risques relatifs très faibles de l'ordre de 1,09-1,25. Une méta-analyse de ces données a estimé que le nombre de variants fréquents impliqués dans la schizophrénie est compris entre 6 000 et 10 000 SNPs, qui expliqueraient collectivement environ 50 % de la variance dans la susceptibilité à la schizophrénie (Ripke et coll., 2013). Il est intéressant de noter que certains SNPs conférant un risque dans la schizophrénie contribuent à la variance dans la susceptibilité à d'autres troubles neuropsychiatriques, incluant les TSA (Lee et coll., 2013). Bien que faible, la co-héritabilité de ces deux troubles est significative, suggérant qu'ils partagent des bases génétiques communes.

Malgré des résultats négatifs, les études d'association dans les TSA ont permis de montrer que l'effet associé à des variants communs avait été largement surestimé et que les variants communs à effet fort n'existent pas, réorientant les recherches vers les variants rares à effet fort. En effet, la pression de sélection négative des facteurs génétiques impliqués dans l'autisme n'avait pas été considérée. En outre, la taille des échantillons utilisés jusqu'à maintenant n'a pas apporté la puissance statistique nécessaire à l'identification de variants communs qui seraient en fait associés à un risque (très) faible. Ces études ont également permis la prise de conscience de la communauté quant à la forte hétérogénéité génétique associée aux TSA qui, bien que suspectée, n'avait pas été prise en compte par la majorité jusqu'alors.

Plus récemment, une analyse statistique novatrice basée sur une très large cohorte épidémiologique suédoise dans les TSA a estimé les contributions relatives des variants rares, communs, hérités et *de novo* dans la susceptibilité aux TSA. Cette étude conclut que la variance dans la susceptibilité aux TSA est

principalement liée à des facteurs génétiques et que les variants communs représenteraient l'essentiel du risque génétique associé aux TSA (Gaugler et coll., 2014) (**Figure 6**). S'il est vrai que les variants communs jouent un rôle substantiel dans l'étiologie des TSA, aucun d'entre eux n'a encore été mis en évidence. Ceci implique que le risque associé à ces variants communs de faible effet correspond probablement à un effet dû à la présence de nombre d'entre eux avec potentiellement des influences environnementales additionnelles. On peut alors se demander quelle serait la pertinence de la mise en évidence, chez un individu donné, d'un risque associé à des milliers de variants à effet faible voire extrêmement faible, en termes de diagnostic, de prise en charge, de modélisation *in vivo* ou *in vitro* et de perspective thérapeutique.

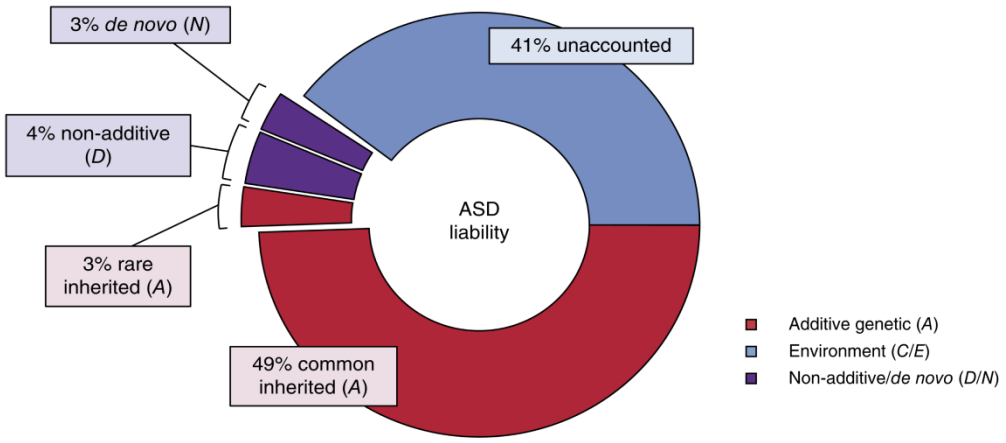



**Figure 6. Contribution des facteurs génétiques et environnementaux dans les TSA**
Les facteurs génétiques incluent les effets additifs (*A*), les effets non additifs (*D* ; dominance, récessivité, épistasie) et les effets *de novo* (*N*). Selon le modèle statistique développé par les auteurs, la grande majorité de la variance de l'autisme liée aux facteurs génétiques est expliquée par les facteurs communs hérités (49 %) alors que les facteurs rares (*de novo* et hérités) ne contribueraient que faiblement à l'étiologie des TSA (figure tirée de Gaugler et coll., 2014).

## 3. Recherche de variants rares dans les TSA

Pendant que de nombreux groupes continuaient d'explorer la piste des variants communs, d'autres s'intéressaient aux variants rares en se basant sur l'existence de nombreux syndromes génétiques rares associés aux TSA ainsi que sur l'identification de remaniements chromosomiques chez des individus atteints d'autisme. Ces évidences, et notamment les multiples anomalies ou mutations *de novo*, cohérentes avec un mécanisme de sélection négative contre des variants génétiques délétères à effet fort, orientaient déjà la recherche de l'étiologie de l'autisme vers les variants rares.

### 3.1 Variations du nombre de copies

De nombreuses délétions et duplications de taille variable touchant tous les chromosomes et détectées par caryotype ou par FISH avaient été rapportées chez des individus atteints de TSA (pour revue voir Veenstra-Vanderweele et coll., 2004). Cependant, ces descriptions représentaient des cas souvent isolés et apparaissaient à l'époque comme marginales. Dès 2007, la résolution offerte par l'utilisation des puces à ADN permet de détecter des délétions et duplications passées inaperçues jusqu'alors par les techniques cytogénétiques traditionnelles et ce, sur l'ensemble du génome. Ces réarrangements **submicroscopiques**, appelés variations du nombre de copies (*copy number variant*, CNV) ont ouvert une nouvelle voie de recherche de l'étiologie dans les TSA qui a permis la mise en évidence de nombreux gènes et loci rares impliqués dans les TSA.

### 3.1.1 Découverte et définition

Un CNV est un segment d'ADN de plus de 1 kb[1] dont le nombre varie par rapport à la diploïdie normale du génome humain. Une perte de matériel génétique à un locus donné est appelée délétion, un gain de matériel génétique correspond à une duplication. Les CNVs peuvent être hérités ou *de novo* (non retrouvés chez les parents). Dans le cas d'une apparition *de novo*, l'altération génétique est survenue soit au niveau de l'un des deux gamètes parentaux, soit elle est spécifique au zygote et est apparue au cours du développement très précoce de l'œuf fécondé.

Les délétions et les duplications du génome ont été décrites dès l'apparition du caryotype dans les années 1960. Bien que cette technique permette une analyse globale du génome, sa faible résolution en fait une technique limitée puisque seules les anomalies de très grande taille (souvent compris entre 3 et 10 Mb) peuvent être détectées. Dans les années 1990, la technique de FISH permet l'étude de régions de plus petite taille de l'ordre de centaines de kilobases, en utilisant des chromosomes artificiels de levure (*yeast artificial chromosome*, YAC) et des chromosomes artificiels de bactéries (*bacterial artificial chromosome*, BAC). Cependant, cette technique n'est utilisée que sur des régions ciblées et ne permet pas l'étude du génome entier. Une autre technique appelée MLPA (*multiplex ligation-dependent probe amplification*), développée au début des années 2000 et basée sur la PCR quantitative multiple permet de tester une quarantaine de régions génomiques simultanément et a été largement utilisée dans la détection de syndromes de microdélétion et microduplication. C'est finalement l'avancée technologique et l'avènement des puces à ADN qui a rendu possible l'étude globale du génome à haute résolution et la mise en lumière de l'abondance et de la large distribution des CNVs (**Figure 7**).

Il existe deux types de micropuces : les puces d'hybridation génomique comparative (*array comparative genomic hybridization*, aCGH) et les micropuces SNP. Les premières consistent en la comparaison d'un ADN test avec un ADN de référence, sur le principe de l'hybridation compétitive. Chacun des ADNs est marqué avec un fluorochrome différent et hybridé sur une même puce sur laquelle des sondes d'ADN représentatives de l'ensemble du génome sont fixées. Le nombre de copies de l'ADN test pour chacune des sondes est calculé par les différences relatives de fluorescence par rapport à l'ADN de référence (qui est en général un mélange d'ADNs de témoins). Dans le cas des micropuces SNP, un seul ADN est hybridé à la puce sur laquelle sont fixées des sondes oligonucléotidiques. Ces puces permettent non seulement de détecter les CNVs d'un individu mais également de génotyper l'ADN test sur des milliers de SNP simultanément et sont donc utilisées pour les études d'association génome entier.

C'est seulement en 2004 que la notion de CNV fut définie pour la première fois. Deux groupes décrivent simultanément l'existence de délétions et de duplications submicroscopiques dans les génomes d'individus de la population générale (Iafrate et coll., 2004 ; Sebat et coll., 2004). Ces variations du nombre de copies ont d'abord été étudiées en tant qu'acteurs majeurs du polymorphisme humain. En effet, ces études rapportent la présence de CNVs chez tous les individus, touchant tous les chromosomes, certains étant très fréquents et d'autres plus rares. Jusqu'alors, les SNPs étaient considérés comme principaux responsables de la variabilité interindividuelle. Grâce à cette découverte, on place alors la variabilité entre les individus au niveau structurel de façon prépondérante par rapport aux variations de séquence. La différence interindividuelle expliquée par les CNVs serait comprise entre 4 et 24 Mb (Tuzun et coll., 2005 ; Redon et coll., 2006) alors que les SNPs ne représentent qu'environ 2,5 Mb de variabilité génomique (Wang et coll., 1998 ; Altshuler et coll., 2000 ; Sachidanandam et coll., 2001). De plus, ces études ont permis de mettre en évidence la structure dynamique du génome.

[1] La taille a longtemps fait partie de la définition du CNV mais cette notion est subjective puisqu'elle est directement liée à la résolution des techniques employées pour les détecter. Ainsi, la limite inférieure de 1 kb habituellement utilisée est maintenant fréquemment remplacée par 50 pb et en-dessous. Cette limite inférieure est surtout utilisée pour discriminer les CNVs de réarrangements plus petits tels que les petites insertions/délétions, appelées indels.

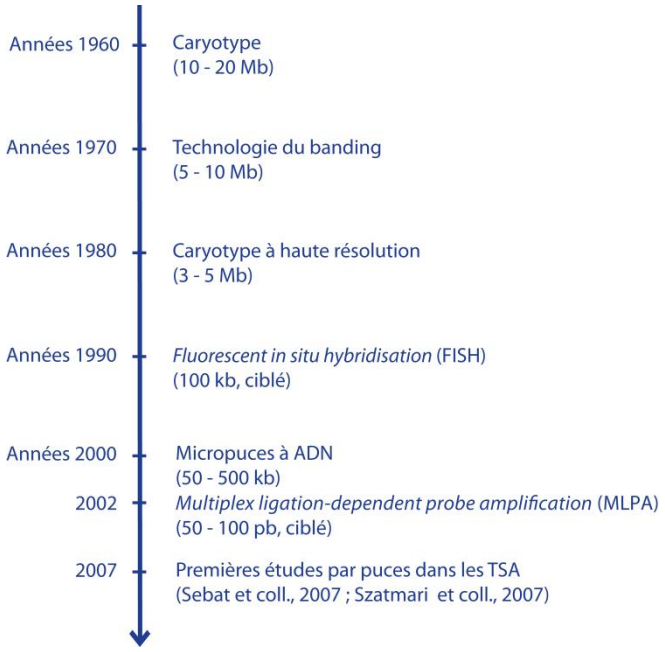



**Figure 7. Evolution des techniques de détection des CNVs appliquées à la recherche étiologique dans l'autisme**
L'utilisation des micropuces à ADN à haute résolution permet aujourd'hui d'étudier des remaniements de petite taille répartis sur l'ensemble du génome.

Iafrate et coll. (2004) ont étudié le génome de 55 individus sans lien de parenté par la technique de puce d'hybridation génomique comparative et ont ainsi identifié 255 loci montrant des déséquilibres génomiques, répartis sur l'ensemble du génome, certains d'entre eux étant récurrents et fréquents alors que d'autres étaient uniques à un individu (**Figure 8**). En utilisant la même technique, Sebat et coll. (2004) ont analysé 20 individus et ont détecté 221 CNVs différents, dont 76 uniques à un individu et environ 50 % récurrents. Ces deux études soulignent la large distribution des CNVs le long du génome mais aussi la présence de régions particulièrement sujettes à l'apparition de réarrangements qui sont par ailleurs riches en séquences homologues appelées duplications segmentales.

### 3.1.2 Mécanismes de formation des CNVs

Les CNVs sont répartis sur l'ensemble du génome bien qu'il existe des régions pour lesquelles leur apparition semble plus fréquente (*hotspots*). Ils peuvent être récurrents (c'est-à-dire de taille similaire et avec des points de cassure fixes) ou non récurrents selon le mécanisme à l'origine de leur formation, qui va directement dépendre de l'architecture génomique de la région remaniée. Trois principaux mécanismes ont été proposés pour expliquer la majorité des réarrangements observés dans le génome humain : la recombinaison homologue non allélique (*non-allelic homologous recombination*, NAHR) et les mécanismes de *non-homologous end-joining* (NHEJ) et du *Fork Stalling and Template Switching* (FoSTeS).

***Les CNVs récurrents***
La recombinaison homologue fait partie des mécanismes de réparation de l'ADN. Elle permet de recréer une copie identique à la séquence endommagée en utilisant la séquence homologue présente à la même position génomique sur la chromatide sœur ou les chromatides du chromosome homologue (autrement dit au même allèle). Elle nécessite donc la présence de séquences identiques ou présentant un très grand

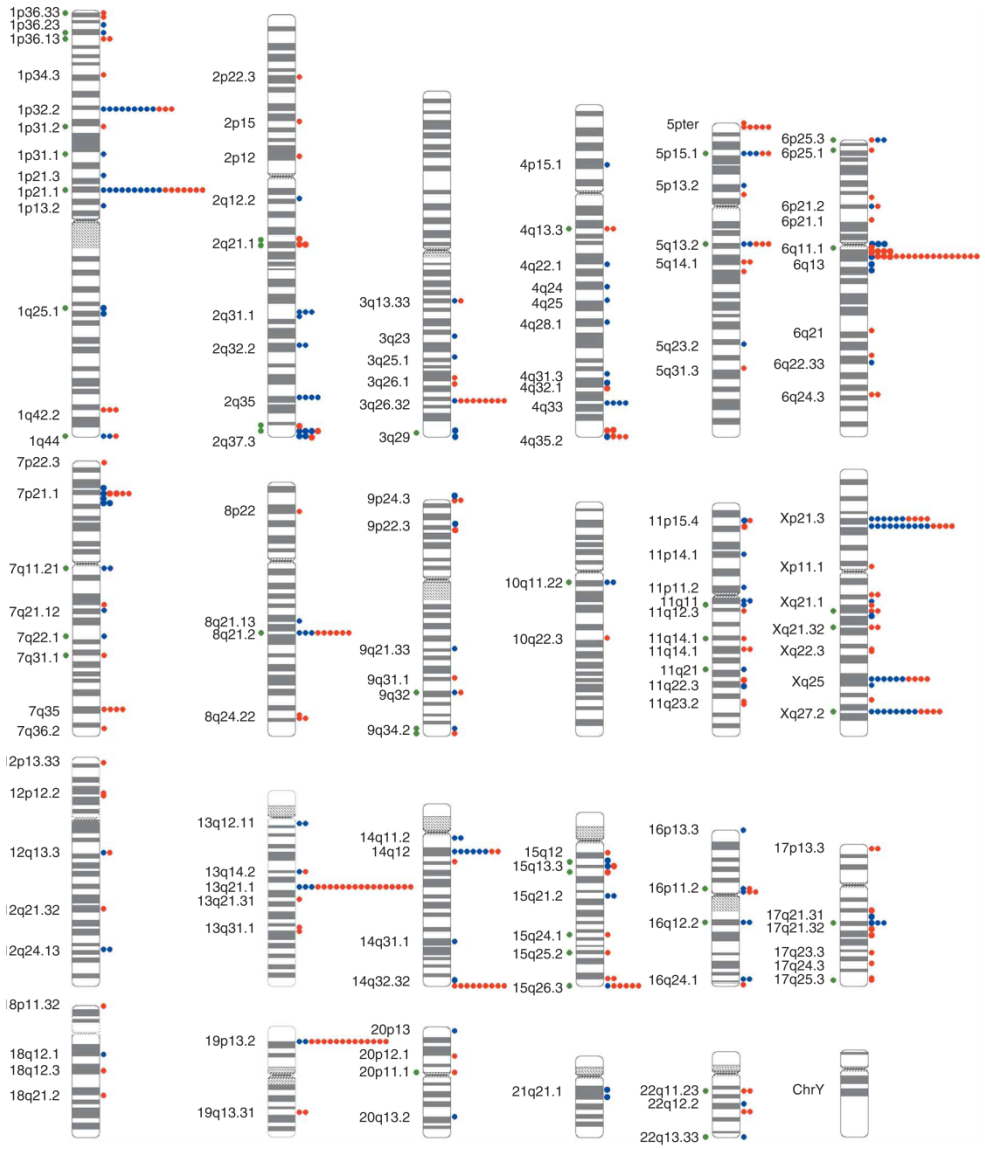



**Figure 8. Polymorphismes du nombre de copies répartis sur l'ensemble du génome humain**

Les cercles à droite de chaque chromosome représentent les délétions (en rouge) et les duplications (en bleu) pour chacun des clones détectés chez les individus témoins. Les cercles verts à gauche des chromosomes indiquent les régions correspondant à des duplications segmentales connues ou à des régions de lacunes dans le génome (*gaps*), qui coïncident avec les clones (figure tirée de Iafrate et coll., 2004).

degré d'identité. Les CNVs sont, dans la majorité des cas, formés lors d'une recombinaison homologue non allélique (NAHR), due à la présence de régions d'homologie flanquant le réarrangement. Ces régions d'homologie, appelées duplications segmentales (ou *low-copy repeats*, LCR) correspondent à des séquences répétées dans le génome (au moins deux copies par génome haploïde) qui favorisent l'apparition de remaniements inter et intra-chromosomiques. Pour des raisons pratiques, on les définit souvent par leur taille (> 1 kb) et par leur degré d'identité (> 95 %). Ainsi, dans les régions génomiques riches en duplications segmentales, des alignements entre des séquences non alléliques appelés mésappariements pourront être créés. Ces derniers peuvent avoir lieu à plusieurs niveaux : entre chromosomes homologues, entre chromatides sœurs ou encore dans une même chromatide. Selon le sens d'orientation des régions homologues flanquantes et selon le niveau du réarrangement, la recombinaison pourra entraîner l'apparition d'une délétion, d'une duplication, d'une inversion ou même de chromosomes isodicentriques ou acentriques (**Figure 9**).

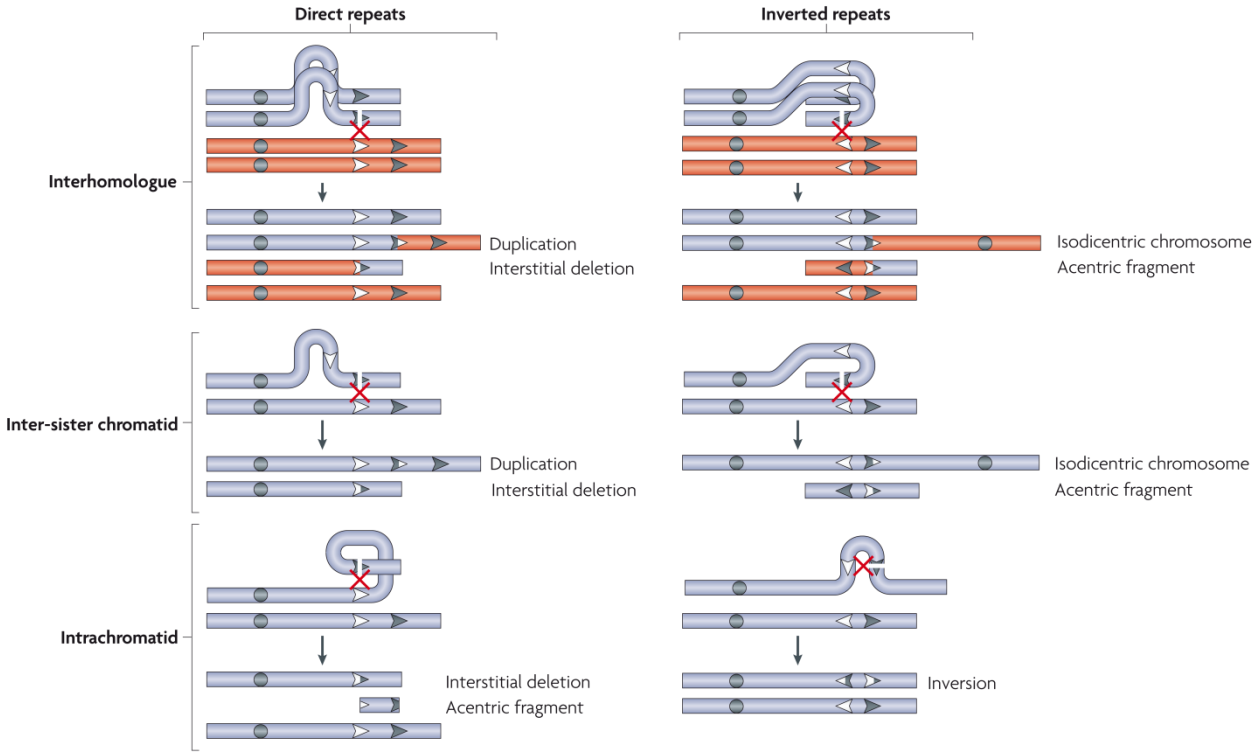



**Figure 9. Réarrangements par recombinaison homologue non allélique (NAHR)**

Une recombinaison (croix rouge) entre séquences d'ADN présentant un degré d'homologie élevé à des positions non alléliques peut générer une délétion, une duplication, une inversion ou l'apparition d'un chromosome isodicentrique. Six cas sont représentés selon le niveau du remaniement (entre chromosomes homologues, entre chromatides sœurs ou dans une même chromatide) et l'orientation des séquences répétées (directe ou inversée). Les chromosomes homologues sont représentés en bleu et rouge ; les chromatides sœurs sont indiquées dans une même couleur (bleu). Les flèches blanches et noires correspondent à des duplications segmentales (figure tirée de Sasaki et coll., 2010).

Il est intéressant de noter que théoriquement, le nombre de délétions générées par NAHR entre deux duplications segmentales données sera toujours supérieur au nombre de duplications. En effet, les remaniements interchromosomiques ou entre chromatides sœurs peuvent donner soit des délétions soit des duplications alors que si la recombinaison a lieu dans une même chromatide, seule une délétion pourra être créée.

***Les CNVs de taille variable***

A l'inverse des CNV récurrents, les non récurrents peuvent apparaître n'importe où, puisque les mécanismes à l'origine de leur formation ne requièrent pas la présence de séquences répétées. Dans la plupart des cas, il s'agit des mécanismes de *non-homologous end-joining* (NHEJ) et du *Fork Stalling and Template Switching* (FoSTeS) (**Figure 10**).

Le NHEJ est l'un des deux mécanismes utilisés par les cellules pour réparer une cassure d'ADN double brin. La réparation a lieu en quatre étapes distinctes : la détection du point de cassure, la création d'un pont moléculaire rapprochant les extrémités, l'édition des régions terminales pour les rendre compatibles et liables et la ligation. Ce mécanisme présente une signature particulière appelée 'cicatrice génomique' liée à l'édition des extrémités avant la ligation (clivage ou ajout de plusieurs nucléotides aux points de cassure). Parfois, il arrive qu'une séquence d'ADN ectopique (élément libre de type transposon par exemple) soit ajoutée aux points de cassure avant la ligation. Par ailleurs, bien que la présence de régions d'homologie ne soit pas nécessaire à ce mécanisme, il a été montré qu'il pouvait tout de même être stimulé par la présence d'éléments d'ADN répétés et de duplications segmentales.

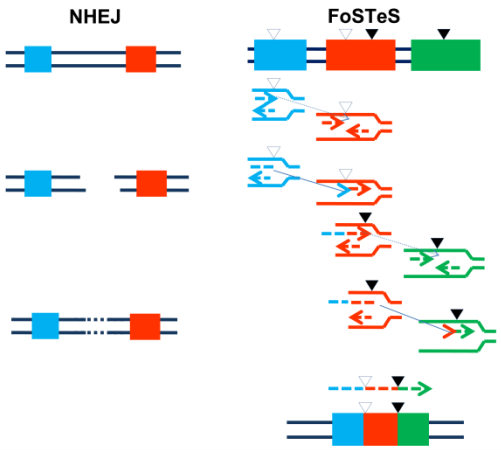



**Figure 10. Mécanismes à l'origine des CNVs non récurrents**
Le mécanisme de *non-homologous end-joining* (NHEJ) cause la délétion de la région présente entre les points de cassure (avec parfois l'ajout de séquences ectopiques), ainsi qu'un cicatrice génomique ; le mécanisme de *Fork Stalling and Template Switching* (FoSTeS), qui se produit lors de la réplication de l'ADN, entraîne des réarrangements plus complexes (figure tirée de Gu et coll., 2008).

Contrairement aux mécanismes de NAHR ou de NHEJ, le FoSTeS a lieu lors de la réplication de l'ADN. Au cours de la synthèse, il arrive parfois que la fourche créée se bloque à une position particulière. Le brin en élongation va alors de désengager de la fourche et s'apparier à un autre substrat, dans une autre fourche de réplication située à proximité physique de la première. La réplication reprend et le brin naissant continue son élongation, entraînant la juxtaposition de séquences génomiques normalement non adjacentes. Si l'échange de substrat a lieu avec une fourche située en amont, une duplication sera créée alors que si la fourche est située en aval, ce sera une délétion. Le mécanisme peut être répété en série (FoSTeS x 2, FoSTeS x 3, et ainsi de suite) ayant pour conséquence la formation de réarrangements complexes avec une succession de régions délétées/normales ou dupliquées/normales. Le FoSTeS ne nécessite pas de séquences d'homologie, cependant il est fortement favorisé par l'existence de microhomologies (de 2 à 5 paires de bases) entre les brins substrats des différentes fourches impliquées.

Beaucoup d'autres mécanismes peuvent engendrer des remaniements génomiques, cependant, les trois cités précédemment permettent d'expliquer l'apparition de la majorité des CNVs chez l'humain.

### 3.1.3 Impact des CNVs sur l'expression des gènes

Les CNVs peuvent n'avoir aucun rôle fonctionnel ou au contraire entraîner l'apparition d'un phénotype s'ils touchent partiellement la séquence codante d'un gène ou la totalité d'un gène sensible au dosage, s'ils dérégulent le niveau d'expression du gène en altérant un élément régulateur, s'ils démasquent un allèle récessif ou s'ils créent de nouveaux gènes de fusion (Weischenfeldt et coll., 2013) (**Figure 11**). Même s'ils ont un effet fonctionnel, ceci n'implique pas que le phénotype auquel ils contribuent a une conséquence négative pour l'individu. En effet, il existe des polymorphismes qui peuvent influencer certaines caractéristiques physiques des individus ou des différences interindividuelles (par exemple, les récepteurs olfactifs).

Il a été montré que l'effet d'un CNV sur l'expression d'un gène et donc sur le niveau d'ARN messager (ARNm) correspondant n'était pas toujours corrélé (Weischenfeldt et coll., 2013). En effet, il est rare qu'une délétion hétérozygote entraîne une diminution de moitié de la quantité d'ARNm et qu'une duplication entraîne son augmentation d'un ratio 3:2. De plus, certains CNVs ne changent pas du tout l'expression du gène qu'ils touchent voire vont entraîner l'effet inverse du nombre de copies géniques au

niveau ARNm. Par exemple, la délétion d'un élément régulateur de type répresseur (*silencer*) ou insulateur pourra entraîner l'augmentation de l'expression du gène ou même son expression ectopique.

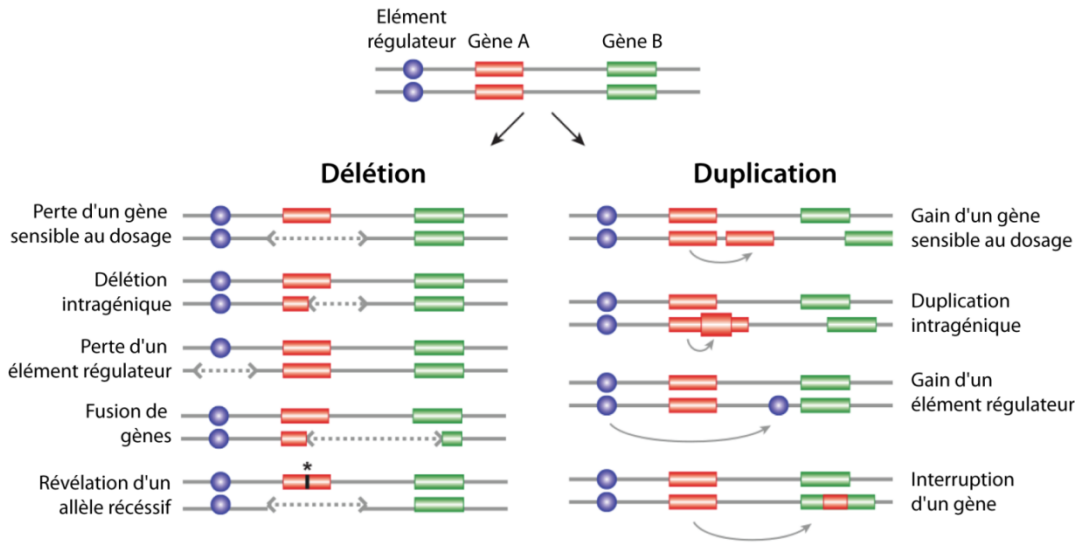



**Figure 11. Variations du nombre de copies et mécanismes pathogènes**
A un locus donné, une perte de matériel génétique est appelée une délétion alors qu'un gain d'une copie correspond à une duplication. Les CNVs peuvent affecter un élément fonctionnel (gène ou élément régulateur) à travers divers mécanismes.

***Gènes sensibles au dosage et haploinsuffisants***

Les CNVs altèrent le dosage génique, soit en l'augmentant soit en le diminuant. Pour certains gènes, l'existence d'une copie supplémentaire ou d'une copie manquante n'aura aucun effet sur leur expression ou sur la fonction de la protéine traduite. A l'inverse, pour certains autres gènes dits 'dosage-sensitifs', une déviation de la diploïdie ne sera pas tolérée. En particulier, certains gènes sont haploinsuffisants et seront sensibles à toute perte hétérozygote. Notre connaissance actuelle de quels gènes sont sensibles ou non au dosage génique reste limitée bien qu'une étude récente ait développé un algorithme permettant de prédire la probabilité d'haploinsuffisance d'un gène donné (Huang et coll., 2010). Les auteurs se sont basés sur les évidences d'études de génétique humaine ayant impliqué des gènes comme haploinsuffisants pour les comparer avec des gènes altérés chez des témoins et donc haplosuffisants (**Figure 12**). Cette étude a montré que les gènes haploinsuffisants sont typiquement plus longs et montrent un plus haut degré de conservation. En outre, ils ont souvent un niveau d'expression plus important que les autres gènes pendant le développement précoce. En se basant sur ces différences, les auteurs ont créé un modèle prédictif qui leur a permis d'annoter environ 12 000 gènes. Plus le pourcentage d'haploinsuffisance d'un gène est faible, plus il est probable qu'il soit haploinsuffisant. En prenant la valeur de 15 % comme limite supérieure stricte, environ 1 800 gènes parmi ceux qui ont été évalués sont haploinsuffisants (soit 15 % des gènes évalués).

La majorité des gènes de notre génome sont haplosuffisants et non sensibles au dosage, ce qui implique que la plupart des CNVs sont neutres vis-à-vis de la sélection naturelle et retrouvés chez de multiples individus. Cependant, certains CNVs sont rares, suggérant pour une fraction d'entre eux, un possible effet délétère associé à une sélection négative. Parmi les CNVs rares (< 1 % dans la population) certains pourraient donc être associés à une pathologie.

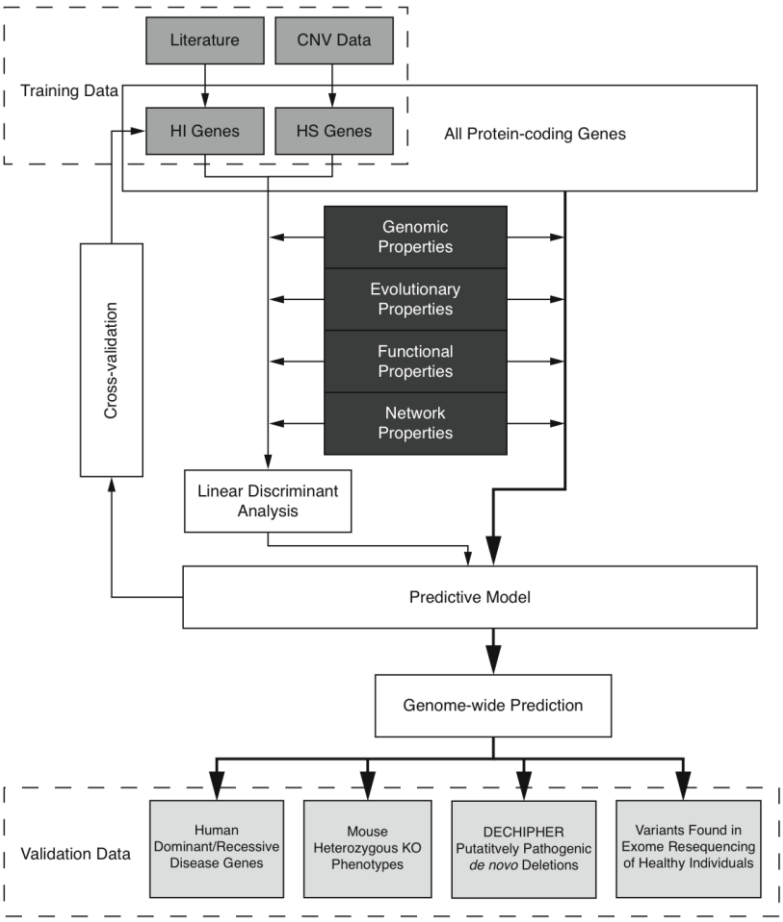



**Figure 12. Modèle prédictif de la probabilité d'haploinsuffisance des gènes**
Une liste de 301 gènes haploinsuffisants (HI) d'après des évidences de la littérature et de 1 079 gènes haplosuffisants (HS) altérés chez au moins deux individus témoins, a permis de déterminer les propriétés des deux types de gènes. Ces caractéristiques génomiques, de conservation, fonctionnelles ou d'interaction en réseau ont ensuite été utilisées pour prédire si un gène donné est HI ou HS. Les prédictions ont ensuite été validées dans 4 groupes de données : une liste de gènes dominants et récessifs associés à des pathologies chez l'humain, une liste de gènes associés à un phénotype chez des souris hétérozygotes, une liste de 487 délétions pathogéniques *de novo* rapportées dans DECIPHER et une liste de 349 gènes pour lesquels des variants perte de fonction ont été détectés chez des individus témoins (figure tirée de Huang et coll., 2010).

### 3.1.4 Les CNVs dans l'autisme

Bien que de nombreux syndromes de microdélétion ou microduplication, pour la plupart *de novo*, avaient déjà été impliqués dans l'étiologie des TSA, c'est en 2007 que la première étude **pangénomique** de recherche de CNVs dans l'autisme a été publiée (Sebat et coll., 2007). Dans cette étude, réalisée sur 264 familles avec autisme non syndromique (118 à cas uniques et 47 à cas multiples) et une centaine de familles témoins, la présence de CNVs *de novo* est significativement associée aux TSA. Les puces à ADN utilisées dans cette étude, avec 85 000 sondes réparties sur l'ensemble du génome, offraient une résolution moyenne de 35 kb. S'ensuivirent plusieurs études sur de grandes cohortes allant de 427 à 1 124 familles (Szatmari et coll., 2007 ; Marshall et coll., 2008 ; Pinto et coll., 2010 ; Levy et coll., 2011 ; Sanders et coll., 2011), utilisant des puces jusqu'à 2 millions de SNPs, qui ont permis de mieux comprendre la contribution de ces variations à l'étiologie de l'autisme (**Tableau 2**).

**Tableau 2. Principales études de CNVs dans les TSA**

| Etude | Sujets | Micropuce | Analyse |
|---|---|---|---|
| Sebat et coll. (2007) | 264 familles (118 à cas sporadiques et 47 à cas multiples), incluant 195 cas<br>99 familles témoins, incluant 196 sujets | Puces à oligonucléotides ROMA | CNVs *de novo* uniquement |
| Szatmari et coll. (2007) | 1 181 familles (*Autism Genome Project*) | Affymetrix 10K | |
| Marshall et coll. (2008) | 427 familles (237 à cas sporadiques et 189 à cas multiples), incluant 427 cas<br>500 témoins et 1 152 témoins additionnels | Affymetrix 500K | |
| Pinto et coll. (2010) | 876 familles (393 à cas sporadiques et 348 à cas multiples, *Autism Genome Project*), incluant 996 cas et 1 752 parents<br>1 287 témoins | Illumina 1M | CNVs rares (<1 %) et >30 kb |
| Sanders et coll. (2011) | 1 124 familles à cas sporadiques (*Simons Simplex Collection*), incluant 1 124 cas, 2 248 parents et 872 frères et sœurs non atteints | Illumina 1M | CNVs rares (<1 %) et >20 sondes consécutives |
| Levy et coll. (2011) | 887 familles à cas sporadiques (*Simons Simplex Collection*) incluant 858 cas, 1 774 parents et 863 frères et sœurs non atteints | Nimblegen 2.1M | CNVs rares (<1 %) et >20 sondes consécutives |

***Différences observées entre patients et témoins***

Plusieurs de ces études ont montré un **enrichissement en CNVs *de novo*** chez les patients avec TSA comparé aux témoins, confirmant les résultats rapportés par Sebat et coll. (Marshall et coll., 2008 ; Pinto et coll., 2010 ; Sanders et coll., 2011). En outre, la proportion de CNVs *de novo* est considérablement augmentée chez les patients présentant des formes syndromiques de TSA (Jacquemont et coll., 2006). Par contre, il n'a pas été observé d'augmentation du nombre global de CNVs ni de la taille moyenne des CNVs entre les individus atteints et les témoins (Marshall et coll., 2008 ; Pinto et coll., 2010). On note toutefois que les CNVs de taille supérieure à 1 Mb sont plus fréquents chez les individus atteints et que la taille d'un CNV est corrélée avec son risque de pathogénicité. En effet, il a été observé que les CNVs de très grande taille (> 5 Mb) sont absents chez les témoins et le plus souvent pathogènes chez les patients.

L'ensemble des CNVs ne diffère pas par le nombre ou la taille entre patients et témoins mais par leur nature. En effet, le nombre de **CNVs géniques et notamment exoniques est enrichi** chez les patients. De plus, les CNVs chez les individus atteints sont enrichis en gènes déjà impliqués dans les TSA ou la déficience intellectuelle (Pinto et coll., 2010) (**Figure 13**).

Il a également été montré que parmi les CNVs *de novo*, le nombre de délétions est plus important chez les patients que chez les témoins, contrairement aux duplications. En effet, dans la toute première étude, Sebat et coll. (2007) avaient rapporté que sur les 15 CNVs *de novo* détectés chez les patients, 12 étaient des délétions (80 %) alors que les deux CNVs *de novo* des témoins étaient des duplications. Ces résultats ont été confirmés dans les études de cohortes plus larges (Marshall et coll., 2008 ; Pinto et coll., 2010 ; Levy et coll., 2011 ; Sanders et coll., 2011), soulignant l'**effet plus délétère des délétions**.

***Familles à cas isolés versus familles à cas multiples***

Longtemps, les défauts génétiques à l'origine des TSA chez les familles à cas isolés et les familles à cas multiples étaient considérés comme différents, notamment vis-à-vis du mode de transmission. Conceptuellement, il semble logique que si plusieurs enfants d'une même famille sont atteints, une cause héritée soit plus probable que dans une famille où un seul enfant est atteint. Les premières études ont effectivement rapporté un nombre plus important de CNVs *de novo* dans les familles à cas isolés, de 7-10 % contre 2-3 % chez les familles ayant plusieurs enfants atteints (Sebat et coll., 2007 ; Marshall et coll., 2008). Cependant, ces résultats n'ont pas été répliqués dans une étude plus large, dans laquelle la

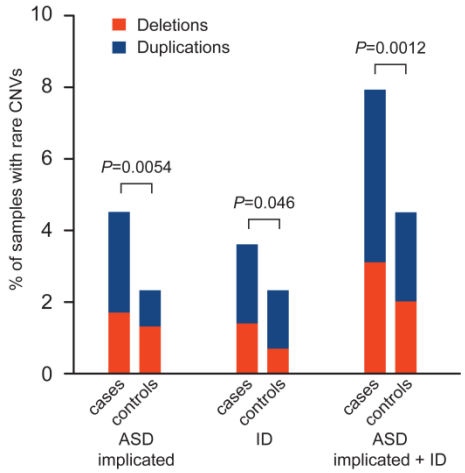



**Figure 13. Enrichissement de CNVs touchant des gènes impliqués dans les TSA ou la déficience intellectuelle chez les patients avec TSA**
Proportion de patients et de témoins présentant des CNVs rares (fréquence < 1 % dans l'ensemble de la population) et de taille supérieure à 30 kb touchant des gènes ou loci déjà impliqués dans les TSA et la déficience intellectuelle (*ASD implicated,* 36 gènes et 10 loci), impliqués uniquement dans la déficience intellectuelle (*ID,* 110 gènes et 17 loci) ou touchant les gènes et loci de ces deux listes combinées. Pour être comptabilisé, le CNV doit être exonique ou chevaucher au moins 50 % de la région dans le cas des syndromes de microdélétion/microduplication (figure adaptée de Pinto et coll., 2010).

proportion de CNVs *de novo* chez les familles à cas sporadiques (n = 393) était similaire à celui observé chez les familles à cas multiples (n = 348), respectivement 5,6 et 5,5 % (Pinto et coll., 2010). Dans le but de mettre en évidence les différences de structure génétique pouvant exister entre les deux types de familles, un projet appelé '*Simons Simplex Collection*' a été initié par une fondation de recherche pour l'autisme (*Simons Foundation Autism Research Initiative*, SFARI) afin de réunir une cohorte de familles composée uniquement de cas sporadiques (Fischbach et Lord, 2010), avec prélèvement des frères et sœurs non atteints. Deux grandes études de CNVs utilisant des familles de cette cohorte ont été réalisées et ont montré une proportion de CNVs *de novo* chez leurs patients (5,6-7,9 %) équivalente à ce qui a été observé dans les familles à cas multiples de Pinto et coll. (Levy et coll., 2011 ; Sanders et coll., 2011). De plus, les résultats récemment obtenus par *l'Autism Genome Project* sur l'ensemble de leur cohorte (n = 2 446), ont confirmé des taux de CNVs *de novo* identiques entre les deux types de familles (Pinto et coll., 2014). L'ensemble de ces résultats suggère que la contribution des CNVs *de novo* dans l'étiologie des familles à cas isolés ou des familles à cas multiples est similaire. Néanmoins, des études supplémentaires sur des échantillons de familles de plus grande taille sont nécessaires afin de comprendre le rôle des CNVs et des autres mutations *de novo* dans les familles à cas sporadiques et multiples.

***Individus atteints versus frères et sœurs non atteints***
Les études réalisées sur la cohorte *Simons*, de par leur conception novatrice où les frères et sœurs non atteints sont utilisés comme témoins des patients, s'affranchissent des biais dus aux différences ethniques ou de génotypage entre les cohortes de patients et témoins (Levy et coll., 2011 ; Sanders et coll., 2011). Elles ont montré que la proportion de CNVs rares *de novo* est augmentée chez les patients par rapport aux frères et sœurs non atteints (5,8 contre 1,7 % (Sanders et coll., 2011) et 7,9 contre 2 % (Levy et coll., 2011)). Ceci implique que la proportion de CNVs rares *de novo* conférant un risque chez les patients correspondrait à environ 70 % de ceux détectés. En outre, ces études présentent un intérêt majeur dans l'étude des CNVs hérités. Sanders et coll. ont montré que la proportion de frères et sœurs non atteints présentant des CNVs hérités était identique à celle des patients et ce, même en considérant uniquement les plus rares, uniquement les géniques ou uniquement les délétions. Cependant, certains CNVs hérités

très rares et spécifiquement trouvés chez les patients, chevauchent des régions associées à plusieurs CNVs *de novo* récurrents déjà décrits dans les TSA (par exemple 1q21.1, 15q13.3, 16p11.2) (Sanders et coll., 2011). De plus, Levy et coll. (2011) ont observé un nombre plus important d'événements transmis 'ultra rares' (CNVs touchant au moins un gène dans une seule famille) chez les patients comparés aux frères et sœurs non atteints, en particulier quand l'enfant atteint est un garçon, suggérant que les filles sont moins susceptibles à ce type de CNVs. Par ailleurs, ces CNVs hérités et ultra rares sont plus fréquemment des duplications, contrairement à ce qui a été observé pour les CNVs *de novo*, soulignant à nouveau l'effet plus délétère des délétions.

***CNVs hérités***

Deux types de variations héritées contribuent à l'étiologie des TSA et doivent être prises en compte séparément puisqu'elles sont conceptuellement très différentes. Les premières sont les variants hérités qui contribuent à des **formes récessives d'autisme**. Dans ce cas, les parents sont porteurs de la mutation à l'état hétérozygote et c'est le fait d'être homozygote ou hétérozygote composé qui entraînera l'apparition de la pathologie chez l'enfant. L'association des TSA avec de nombreux syndromes génétiques récessifs rares a été décrite très tôt dans la recherche d'étiologie dans l'autisme (pour revue voir Betancur, 2011). L'analyse par micropuces à ADN a été appliquée à une cohorte de 104 familles non syndromiques ayant au moins un enfant atteint de TSA dont 88 montrant une consanguinité parentale, dans le but d'identifier d'autres causes héritées d'autisme (Morrow et coll., 2008). La recherche de délétions dans des régions de perte d'hétérozygotie a mis en évidence plusieurs nouveaux gènes candidats comme les gènes *PCDH10* (codant pour une cadhérine), *RNF8* (qui joue un rôle dans les processus d'ubiquitinylation) et *SCN7A* (canal sodique de la même famille que *SCN1A et SCN2A*, impliqués dans les TSA). Des évidences supplémentaires sont nécessaires afin de valider l'implication de ces gènes dans les TSA. En outre, il a été observé qu'il existe un excès de grandes régions d'homozygotie chez les patients avec TSA de *l'Autism Genome Project* par rapport aux témoins (Casey et coll., 2012). Ces résultats ont été répliqués dans étude réalisée sur un échantillon de 2 108 familles de la cohorte *Simons*, confirmant la forte contribution des loci autosomiques récessifs dans l'étiologie des formes de TSA associées à une déficience intellectuelle (Gamsiz et coll., 2013).

D'autres variations héritées rares suivent un mode de transmission **dominant à pénétrance incomplète et/ou expressivité variable**. Il s'agit par exemple de certains CNVs qui, bien que transmis de parents (apparemment) sains, sont responsables du phénotype chez l'enfant. Des exemples incluent les délétions 15q13.3, les délétions et duplications proximales 16p11.2 et les délétions 16p13.11. Ces CNVs sont associés à des phénotypes peu spécifiques et à un large spectre de troubles neurodéveloppementaux mais sont rarement détectés chez des témoins (voir **section II.4**).

Les études de CNVs dans l'autisme ont permis de réaliser de multiples diagnostics étiologiques et également d'impliquer de nouveaux gènes et loci dans l'étiologie des TSA. Le premier gène impliqué par l'utilisation des micropuces dans les TSA est le gène *NRXN1* (*neurexin 1*), codant une molécule d'adhésion cellulaire (Szatmari et coll., 2007). L'analyse des CNVs sur plus de 1 000 familles à cas multiples de l'*Autism Genome Project*, a permis d'identifier une délétion *de novo* de 300 kb en 2p16 chez deux sœurs atteintes, due à un mosaïcisme germinal chez le père. Cette délétion implique uniquement le gène *NRXN1*, qui était un candidat fonctionnel de par son rôle dans la synaptogenèse et son interaction avec les neuroligines, déjà impliquées dans les TSA. L'implication de ce gène dans les TSA mais aussi dans d'autres troubles neurodéveloppementaux a ensuite été confirmée dans des cohortes indépendantes (Ching et coll., 2010 ; Pinto et coll., 2010). Un autre exemple est le gène *SHANK2* pour lequel deux patients porteurs de délétions exoniques *de novo* ont été identifiés par Pinto et coll. (2010). Ce gène appartient à la même famille que le gène *SHANK3*, précédemment impliqué dans les TSA, et code également une protéine

d'échafaudage synaptique. Cette étude a également démontré le rôle des gènes *SYNGAP1*, déjà impliqué dans la déficience intellectuelle et *PTCHD1* jamais impliqué avant, dans l'étiologie des TSA.

Les études de CNVs pangénomiques dans l'autisme et la déficience intellectuelle ont dévoilé l'existence d'une multitude de nouveaux syndromes de microdélétion et microduplication. Le grand nombre de patients étudiés a permis aux différents groupes de répliquer les nouvelles régions mises en cause. En particulier, des CNVs récurrents dans la région 16p11.2 ont été retrouvés dans la plupart de ces études, faisant de ce réarrangement l'un des plus fréquemment associés aux TSA (environ 1 %) et à d'autres troubles neuropsychiatriques (Marshall et coll., 2008 ; Weiss et coll., 2008 ; Fernandez et coll., 2010 ; Shinawi et coll., 2010). De plus, Sanders et coll. (2011) rapportent plusieurs duplications de la région délétée dans le syndrome de Williams-Beuren (7q11.23), qui avaient déjà été décrites dans des cas isolés de TSA (Somerville et coll., 2005 ; Berg et coll., 2007 ; Depienne et coll., 2007 ; Van der Aa et coll., 2009).

Il est intéressant de noter que l'analyse des gènes identifiés par ces études a mis en évidence l'existence de réseaux fonctionnels communs. Parmi les gènes affectés par des CNVs rares chez leurs patients, Pinto et coll. (2010) ont montré un enrichissement des gènes impliqués dans le développement et la fonction des synapses, les gènes de la voie de signalisation Ras et en particulier les Rho GTPAses impliquées dans la régulation des dendrites et la plasticité des épines dendritiques. Une nouvelle méthode d'analyse de voies biologiques (*network-based analysis of genetic associations*, NETBAG) appliquée aux gènes impliqués par les CNVs rares *de novo* identifiés par Levy et coll. (2011), a montré que, malgré la diversité des réseaux touchés, il existe une convergence fonctionnelle, en particulier vers les voies impliquées dans la synaptogenèse, le guidage axonal et la motilité neuronale (Gilman et coll., 2011).

En conclusion, l'ensemble de ces résultats montre le rôle considérable des CNVs rares *de novo* mais aussi hérités dans l'étiologie des TSA. D'ailleurs, la recherche de CNVs s'est rapidement imposée comme la méthode de choix dans l'exploration étiologique chez les patients avec TSA ou déficience intellectuelle et est maintenant réalisée en routine par les laboratoires cliniques (Miller et coll., 2010). De plus, les études de CNVs nous ont permis d'étendre considérablement nos connaissances sur l'architecture étiologique des TSA en identifiant de nombreux gènes causaux et une multitude de gènes et loci candidats, ainsi que des voies biologiques communes à ces différents gènes.

## 3.2 Variations de séquence

### 3.2.1 Séquençage de gènes candidats

Le séquençage traditionnel a été largement utilisé dans les TSA pour identifier des mutations dans des gènes déjà impliqués dans d'autres troubles neurodéveloppementaux (notamment la déficience intellectuelle) ou dans des gènes candidats positionnels ou fonctionnels. Le premier exemple d'un gène candidat séquencé dans les TSA et impliqué dans l'étiologie est celui des gènes *NLGN3* et *NLGN4X* (*neuroligin 3* et *neuroligin 4X*). Des délétions *de novo* de la région Xp22.3 avaient été décrites chez trois filles atteintes d'autisme (Thomas et coll., 1999). Cette région contenait le gène *NLGN3*, candidat par sa fonction. En effet, les neuroligines sont des protéines appartenant à la famille des molécules d'adhésion cellulaire, spécifiquement localisées à la synapse et qui jouent un rôle dans le maintien et la fonction des synapses excitatrices et inhibitrices. La famille des neuroligines comprend 5 gènes dont deux sont localisés sur le chromosome X : *NLGN3* et *NLGN4X*. Le séquençage de ces deux gènes dans une cohorte de 158 familles avec TSA a permis d'identifier une mutation de décalage du cadre de lecture dans *NLGN4X* et une mutation faux-sens dans *NLGN3* au sein de deux familles présentant chacune deux frères atteints (l'un avec autisme et l'autre avec syndrome d'Asperger dans les deux cas) (Jamain et coll., 2003). Les mutations des neuroligines sont très rares mais ces résultats ont toutefois été répliqués dans d'autres

cohortes, notamment pour la *NLGN4X* pour laquelle une mutation non-sens a été décrite dans une famille multigénérationelle avec treize individus atteints de sexe masculin présentant soit une déficience intellectuelle (n = 10), une déficience intellectuelle associée à des TSA (n = 2) ou encore un TSA sans déficience intellectuelle (n = 1) (Laumonnier et coll., 2004). Dans le cas de la *NLGN3*, il aura fallu attendre huit ans après la description de la première famille par Jamain et coll. (2003) pour qu'une délétion exonique *de novo* de ce gène soit décrite chez un garçon atteint d'autisme (Sanders et coll., 2011).

### 3.2.2 Séquençage de l'exome entier

Les techniques de séquençage ont beaucoup évolué ces dernières années et aujourd'hui, le séquençage classique (technique de Sanger) a été remplacé par les nouvelles technologies dites de nouvelle génération. Elles ont révolutionné la génétique humaine puisqu'elles permettent notamment de séquencer l'exome entier des individus de façon rapide et à un coût modéré. L'exome correspond à l'ensemble des exons (régions codantes des gènes) d'un individu et représente environ 1 % du génome entier. Il a été estimé qu'environ 85 % des mutations impliquées dans une pathologie affectent les régions génomiques codantes (Choi et coll., 2009). Cette stratégie a largement montré son efficacité pour identifier de nouveaux gènes causaux dans de nombreuses maladies rares tout comme dans des pathologies plus fréquentes comme la déficience intellectuelle (Rabbani et coll., 2014).

La toute première étude de séquençage de l'exome entier dans la déficience intellectuelle a été réalisée en 2010 sur une petite cohorte de 10 patients (Vissers et coll., 2010). Intelligemment, les auteurs ont séquencé des trios parents-enfants, leur permettant ainsi de déterminer si les variants détectés étaient *de novo* ou hérités. Grace à cette approche, les auteurs ont pu détecter et valider six mutations *de novo* non synonymes chez six patients, correspondant probablement à la cause de la déficience intellectuelle chez ces individus. Puis, deux autres études (de Ligt et coll., 2012 ; Rauch et coll., 2012) réalisées sur des cohortes de 100 et 51 trios, ont confirmé le fort rendement étiologique de cette approche puisque la cause de la déficience intellectuelle a été identifiée dans environ 16 et 30 % des cas, respectivement. Des mutations *de novo* ont été rapportées dans des gènes déjà impliqués dans la déficience intellectuelle tels que *GRIN2A*, *SYNGAP1* ou *SCN2A* mais aussi dans plusieurs gènes qui n'avaient jamais été impliqués dans une pathologie. Pour certains de ces gènes candidats (par exemple *GATAD2B*, *SETD5*, *CHD2* et *DEAF1*), des études postérieures ont confirmé leur implication dans la déficience intellectuelle et/ou dans les TSA.

***Etudes de l'exome entier dans les TSA***

L'analyse de l'exome entier a été appliquée aux TSA pour la première fois en 2011 dans une étude pilote réalisée sur 20 trios (O'Roak et coll., 2011). Tout comme la première étude d'exome dans la déficience intellectuelle, cette étude a montré des résultats prometteurs avec l'identification de 4 mutations *de novo* potentiellement pathogènes sur 20 patients étudiés (20 %) affectant les gènes *FOXP1*, *GRIN2B*, *SCN1A* et *LAMC3*. A l'exception de *LAMC3*, ces gènes ont été répliqués et impliqués dans l'étiologie des TSA. Quatre grandes études ont ensuite été publiées avec des cohortes plus grandes allant de 175 à 343 familles (Iossifov et coll., 2012 ; Neale et coll., 2012b ; O'Roak et coll., 2012b ; Sanders et coll., 2012a).

Toutes ont montré qu'il n'y a pas d'augmentation significative du nombre de mutations *de novo* chez les patients comparé aux frères et sœurs ou aux témoins. Le taux de mutation *de novo* est d'environ un SNV (*single nucleotide variant*) par exome, ce qui correspond à $2 \times 10^{-8}$ par base et par génération (Gratten et coll., 2013). Cependant, si les mutations *de novo* sont séparées selon leur classe fonctionnelle[2], un enrichissement est observé pour les mutations avec un effet potentiellement délétère (Iossifov et coll., 2012 ; Sanders et coll., 2012a) **(Figure 14)**.

[2] Les mutations sont appelées synonymes si elles ne changent pas la séquence protéique. Parmi les non synonymes, on trouve les faux-sens (changement d'un acide aminé par un autre), les non-sens (introduction d'un codon stop prématuré) et

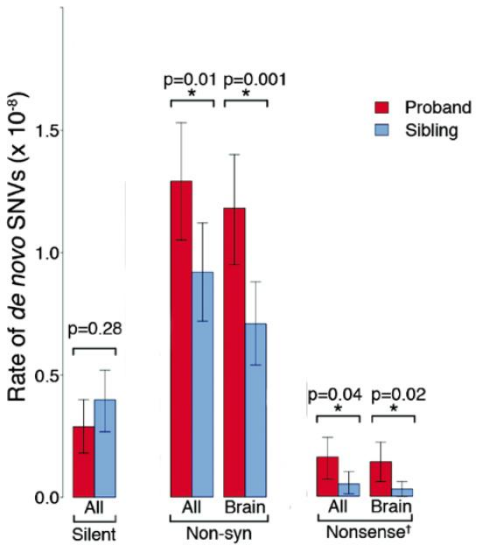



**Figure 14. Enrichissement des mutations *de novo* chez les sujets atteints de TSA, selon la classe fonctionnelle**

Le taux de mutations *de novo* est indiqué pour 200 patients (en rouge) et pour les frères et sœurs non atteints (en bleu). 'All' se réfère à toutes les mutations touchant un gène RefSeq, 'Brain' correspond aux mutations présentes dans des gènes exprimés dans le cerveau. 'Silent' se réfère aux mutations synonymes, 'Non-syn' fait référence à toutes les mutations non synonymes (faux-sens, non-sens ou d'épissage) et 'Nonsense' prend en compte les mutations non-sens et également les mutations qui affectent l'épissage (figure tirée de Sanders et coll., 2012).

Les premières études se sont largement focalisées sur les mutations *de novo* et en particulier sur les mutations perte de fonction (non-sens, de décalage du cadre de lecture et d'épissage) qui sont en principe les plus délétères. En effet, l'interprétation de la masse de données générées par cette technique étant très complexe, il est plus simple de se focaliser sur ce type de mutations, potentiellement les plus délétères et qui sont en outre plus rares. Cependant, il est clair que parmi les mutations faux-sens, beaucoup jouent également un rôle dans la pathologie. Pourtant, le signal associé au risque qu'elles représentent est fortement dilué par les variants faux-sens qui n'ont aucun effet fonctionnel. D'ailleurs, contrairement à ce qui a été observé pour les mutations perte de fonction, aucun enrichissement en mutations faux-sens n'a été montré chez les patients atteints de TSA par rapport à leurs frères et sœurs non atteints (Iossifov et coll., 2012).

***Influence de l'âge du père dans le taux de mutations de novo***

Toutes les études d'exome dans les TSA ont rapporté une corrélation positive entre le nombre de mutations *de novo* observées chez les patients et l'âge paternel. Ce résultat est cohérent avec l'observation du risque accru de développer un TSA chez les enfants de pères âgés (Hultman et coll., 2011). Cette observation a également été confirmée dans une étude de 78 trios atteints principalement de TSA ou de schizophrénie (Kong et coll., 2012) (**Figure 15**).

L'augmentation du nombre de mutations *de novo* avec l'âge du père serait liée à la nature du processus de spermatogenèse qui implique un plus grand nombre de divisions cellulaires, contrairement à l'ovogenèse (Crow, 2000).

---

les décalages du cadre de lecture (*frameshift*, dues à l'insertion ou à la délétion d'un nombre de paires de bases non multiple de 3). Les mutations non-sens, de décalage du cadre de lecture et celles qui affectent des sites d'épissage sont considérées comme les plus délétères et regroupées sous le terme 'perte de fonction'. D'autres variants appelés indels correspondent à des insertions ou à des délétions de quelques paires de base et ont été peu étudiées pour le moment du fait des difficultés techniques liées à leur détection.

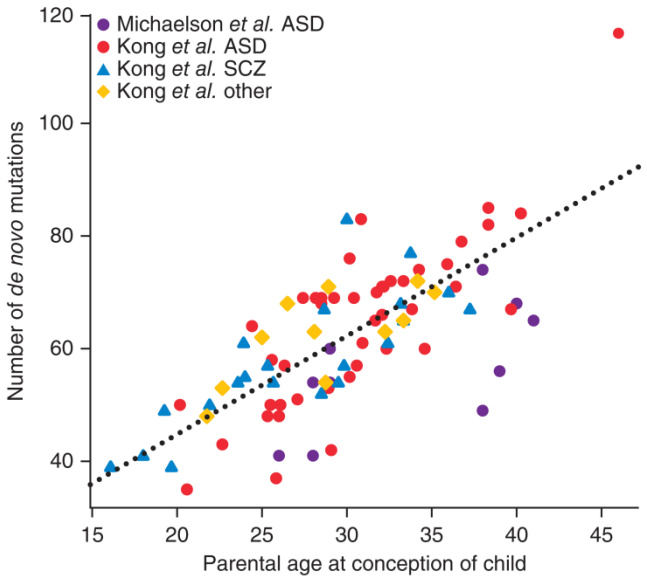



**Figure 15. Influence de l'âge du père sur le nombre de mutations *de novo***
Le nombre de mutations *de novo* est indiqué en fonction de l'âge du père au moment de la conception de l'enfant pour 78 trios rapportés par Kong et coll. (2012) : 44 atteints de TSA, 21 de schizophrénie et 13 inclus pour d'autres raisons (dont la construction de familles multigénérationnelles) ; les données de 10 trios avec TSA décrits par Michaelson et coll. (2012) sont également représentées (figure tirée de Gratten et coll., 2014).

***Gènes impliqués de façon récurrente***

Le taux de mutations *de novo* étant faible (environ 1 mutation par exome et par génération), il semble statistiquement peu probable d'observer plusieurs mutations *de novo* dans un même gène chez plusieurs patients. En se basant sur ce fait et en prenant en compte différents paramètres tels que la taille de l'échantillon, la taille des gènes et leur contenu en GC, Sanders et coll. (2012) ont estimé que la découverte de deux mutations perte de fonction dans un même gène pouvait suffire à l'impliquer dans les TSA de façon statistiquement significative. En combinant les différentes analyses d'exome réalisées et les études de séquençage ciblé de certains candidats dans des cohortes de réplication, douze gènes ont été touchés par au moins deux mutations perte de fonction, incluant *ARID1B*, déjà impliqué dans la déficience intellectuelle et les TSA ainsi que *CHD8*, *DYRK1A*, *KATNAL2*, *SCN2A, TBR1, GRIN2B, ADNP*, *POGZ, ANK2, RIMS1* et *CUL3* (Iossifov et coll., 2012 ; Kong et coll., 2012 ; Neale et coll., 2012b ; O'Roak et coll., 2012a ; O'Roak et coll., 2012b ; Sanders et coll., 2012a ; Willsey et coll., 2013 ; Ronemus et coll., 2014). Cette récurrence constitue un argument fort pour considérer ces gènes comme candidats mais des évidences additionnelles sont nécessaires pour les impliquer dans l'étiologie des TSA. D'ailleurs, depuis la publication de ces études, d'autres ont confirmé l'implication de plusieurs de ces gènes dans l'étiologie des TSA (*CHD8*, *DYRK1A*, *SCN2A*, *GRIN2B* et *ADNP*).

***Mutations héritées et formes récessives de TSA***

Après les premières analyses d'exomes entiers focalisées sur les variants *de novo*, deux études ont utilisé cette technique pour mettre en évidence des causes de TSA récessives ou liées à l'X (Lim et coll., 2013 ; Yu et coll., 2013). Yu et coll. (2013) ont séquencé l'exome de familles consanguines et se sont focalisés sur les régions d'homozygotie, réduisant ainsi largement l'analyse des variants détectés chez leurs familles. Plusieurs mutations homozygotes ont été identifiées dans des gènes responsables de syndromes génétiques connus, chez des patients ne présentant pas le phénotype typique associé à ces pathologies. Chez trois familles, des mutations faux-sens homozygotes touchant les gènes *POMGNT1* (dystrophie musculaire congénitale), *AMT* (encéphalopathie sévère avec hyperglycémie) et *VPS13B* (syndrome de Cohen) ont été identifiées. Ces maladies génétiques récessives rares sont habituellement associées à des

présentations cliniques très sévères non observées chez les patients de cette étude. Les auteurs postulent que les mutations qu'ils observent chez ces individus correspondent à des allèles hypomorphes qui n'abolissent pas complètement la fonction de la protéine mais diminuent son activité. Cette étude montre à nouveau la contribution des mutations héritées récessives dans l'étiologie des TSA et souligne la difficulté de diagnostiquer certains patients chez qui la présence d'allèles hypomorphes de gènes impliqués dans des syndromes engendre des présentations atypiques.

Une autre étude, cas-témoins cette fois, s'est intéressée à la contribution à l'étiologie des TSA des mutations perte de fonction à l'état homozygote ou hétérozygote composé (c'est-à-dire quand les deux allèles d'un même gène sont touchés par des mutations différentes) (Lim et coll., 2013). Les variants de ce type, aussi appelés '*knockouts*' puisqu'ils abolissent complètement la fonction de la protéine, sont rares chez les individus normaux (5 à 20 en moyenne). Ils concernent la plupart du temps un nombre limité de gènes et sont par exemple retrouvés dans les gènes codant les récepteurs olfactifs (MacArthur et coll., 2012). En se focalisant sur les *knockouts* rares (dont la fréquence est inférieure à 5 %), les auteurs ont montré un enrichissement significatif des cas présentant au moins un *knockout* rare par rapport aux témoins (6 % et 3 %, respectivement). De plus, le nombre de variants hémizygotes sur le chromosome X est 1,5 fois plus important chez les garçons atteints de TSA comparé aux témoins, confirmant la contribution des gènes de l'X à l'étiologie des TSA chez les sujets masculins.

En conclusion, il est aujourd'hui clair que les deux types de variations rares *de novo* et héritées contribuent à l'étiologie de l'autisme. D'ailleurs de plus en plus d'études utilisent des modèles dans lesquels les deux types de variations sont pris en compte. En particulier, une nouvelle analyse statistique appelée TADA (*Transmission And De novo Association*) intègre les données des variants *de novo* et hérités en utilisant les données des études d'exome familiales mais aussi cas-témoins (He et coll., 2013). Cette méthode est basée sur un model statistique qui prend en compte pour chacun des variants les fréquences alléliques et la pénétrance gène-spécifique et a montré une puissance statistique nettement supérieure aux autres modèles prédictifs utilisés pour identifier des gènes contribuant à la vulnérabilité aux TSA.

Au total, les grandes études de séquençage de l'exome entier réalisées dans les TSA ont montré un fort enrichissement des mutations *de novo*, en particulier celles associées à une perte de fonction, chez les patients comparés aux frères et sœurs non atteints ou aux témoins. Les variants hérités, plus difficiles à interpréter, semblent également jouer un rôle, en particulier dans les formes récessives de TSA. Ces études ont également permis de réaliser de multiples diagnostics étiologiques et de mettre en évidence de nouveaux gènes impliqués dans les TSA et la déficience intellectuelle. Par ailleurs, elles ont confirmé l'hétérogénéité génétique extrême associée aux TSA puisque peu de gènes ont été trouvés mutés chez plusieurs patients indépendants pour le moment. L'interprétation des SNVs est complexe, notamment pour les variants faux-sens, les indels et les variants rares hérités qui n'ont été que très peu explorés jusqu'à maintenant. Cependant, de plus en plus d'exomes de patients sont disponibles ce qui va nous permettre d'augmenter nos connaissances et de pouvoir mieux interpréter les différents variants détectés. En effet, un consortium international ayant pour projet le séquençage d'exomes dans l'autisme (*Autism Sequencing Consortium*, ASC) a prévu de réunir 10 000 exomes de patients très prochainement (Buxbaum et coll., 2012).

### 3.2.3 Séquençage du génome entier

Le séquençage du génome entier présente encore un coût très élevé et n'est pas pour le moment réalisé en routine. La première étude de ce type réalisée dans les TSA a été publiée en 2013 (Jiang et coll., 2013). Le génome complet de 32 trios a été analysé (**Figure 16**). Environ 65 SNVs et 5 indels *de novo* par individu (sujets atteints et parents) ont été détectés. Comme dans les études d'exomes, le taux de mutations *de novo* est corrélé avec l'âge paternel. Pour faciliter l'interprétation, les trios ont été sélectionnés de telle

façon qu'aucun membre ne soit porteur d'un CNV pathogène connu. Des mutations *de novo* ont été identifiées chez 6 patients (19 %), contribuant potentiellement au phénotype chez eux. Par exemple, une petite délétion *de novo* de 32 pb causant un décalage du cadre de lecture a été identifiée dans le gène *SCN2A* chez une patiente avec autisme et déficience intellectuelle sévère. Ce gène a été précédemment impliqué dans les TSA et la déficience intellectuelle et bien qu'une sœur atteinte ne porte pas la mutation, on peut tout de même imaginer qu'elle est responsable des troubles de cette patiente. D'autres mutations héritées (liées à l'X ou autosomiques) potentiellement délétères ont été détectées dans 10 familles et incluent une mutation faux-sens du gène *AFF2* (responsable du syndrome FRAXE), sur le chromosome X, prédite néfaste et transmise d'une mère à son fils atteint. Pour d'autres variants touchant des gènes qui n'ont pas encore été impliqués ni dans les TSA ni dans la déficience intellectuelle ou touchant des gènes candidats, des preuves additionnelles sont nécessaires pour conclure sur leur contribution au phénotype.

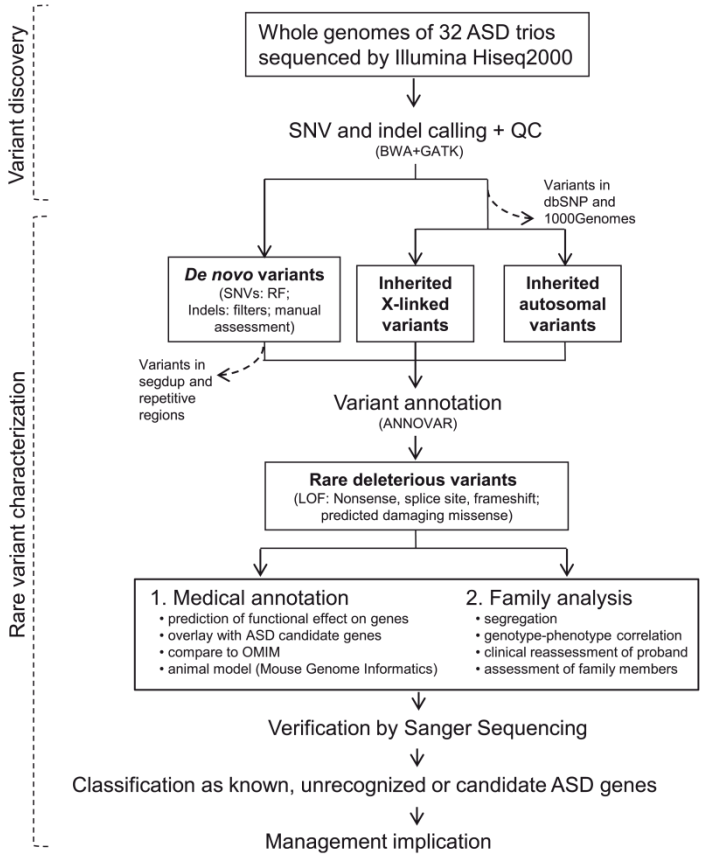



**Figure 16. Première étude de séquençage du génome entier dans les TSA**
Le génome de 32 trios a été séquencé et les variants *de novo*, hérités liés à l'X ou sur les autosomes ont ensuite été annotés pour sélectionner les mutations rares potentiellement délétères (non-sens, d'épissage, de décalage du cadre de lecture et faux-sens prédites comme néfastes). Selon le statut des gènes touchés et l'analyse des données familiales, les variants d'intérêt ont été validés et classés en 3 catégories : variants touchant des gènes impliqués dans les TSA et/ou la déficience intellectuelle, touchant des gènes non impliqués ou touchant des gènes candidats (figure tirée de Jiang et coll., 2013).

Bien que cette technique reste marginale pour le moment dans les TSA, elle promet d'apporter de nouvelles ressources d'autant qu'elle offre une meilleure couverture que le séquençage d'exome entier, même pour les régions codantes. Elle permettra aussi d'identifier des mutations dans les régions régulatrices et les autres régions non codantes qui ont été jusqu'alors très peu explorées. Par contre, elle ne permet pas encore de détecter les CNVs de façon fiable.

Devant une telle hétérogénéité étiologique, il est clair que la recherche de variants rares que ce soit au niveau structurel ou au niveau de la séquence a porté ses fruits dans l'identification de nombreux gènes et loci impliqués dans les TSA. Cependant, bien que beaucoup ait été accompli dans la compréhension des facteurs étiologiques de l'autisme, il nous reste beaucoup à découvrir. En particulier, si des variants fréquents existent, ils n'ont pas encore été mis en évidence et d'autres approches vont devoir être développées pour les détecter.

## 4. Causes génétiques connues de TSA

Grâce aux dernières années de recherche de variants rares, de multiples causes de TSA ont été découvertes. Plus d'une centaine de gènes et loci jouant un rôle étiologique ont été impliqués dans les TSA (Betancur, 2011) et des centaines d'autres présentent des évidences fortes mais restent candidats pour le moment. Toutes les causes identifiées de TSA sont également des causes de déficience intellectuelle, ce qui démontre le fort chevauchement étiologique existant entre ces deux pathologies. Les causes impliquées dans les TSA incluent des réarrangements chromosomiques ainsi que de nombreux syndromes génétiques (monogéniques ou non). La **Figure 17** offre une vue d'ensemble des causes génétiques connues associées aux TSA à ce jour, avec des exemples de chacune des catégories.

### 4.1 Remaniements chromosomiques

Des anomalies chromosomiques visibles au caryotype ont été reportées chez environ 2-6 % des individus atteints de TSA (Wassink et coll., 2001 ; Kielinen et coll., 2004 ; Veenstra-Vanderweele et coll., 2004 ; Shen et coll., 2010). Ce taux varie selon la composition des cohortes étudiées et notamment selon la proportion de patients présentant une déficience intellectuelle ou une forme syndromique de TSA. Des anomalies de nombre des chromosomes (aneuploïdies) ainsi que des remaniements impliquant un ou plusieurs chromosomes comme des délétions/duplications de grande taille visibles au caryotype (> 5 Mb), des translocations, des chromosomes en anneaux et des inversions ont été décrites. Ces réarrangements de grande taille sont le plus souvent trouvés chez des patients avec des présentations cliniques syndromiques.

La trisomie 21 (syndrome de Down) ainsi que les aneuploïdies des chromosomes sexuels comme le syndrome de Turner (45,X), le syndrome de Klinefelter (47,XXY), le syndrome 47,XYY et le syndrome 48,XXYY sont associés à un risque élevé de TSA (Bishop et coll., 2011 ; Ross et coll., 2012 ; Zafeiriou et coll., 2013).

En outre, des réarrangements chromosomiques apparemment équilibrés tels que des translocations ou des inversions peuvent être associés aux TSA. En effet, ces réarrangements, en apparence équilibrés, peuvent parfois présenter en réalité des anomalies cryptiques non détectées qui sont la cause des manifestations cliniques du patient (Feenstra et coll., 2011). Les remaniements réellement équilibrés peuvent exercer un effet délétère s'ils interrompent un gène haploinsuffisant, ont des effets de position ou influencent l'empreinte parentale (Talkowski et coll., 2012).

### 4.2 Syndromes génétiques

De multiples syndromes génétiques correspondant à des maladies monogéniques ou à des syndromes de microdélétion/microduplication ont été décrits chez des patients atteints d'autisme. Pour certains d'entres eux, les TSA font partie des manifestations cliniques principales (**Tableau 3**). Tous ces syndromes ont également été décrits dans la déficience intellectuelle, indiquant un fort chevauchement étiologique entre ces deux troubles neurodéveloppementaux. Parfois, d'autres troubles neuropsychiatriques comme l'épilepsie ou plus rarement la schizophrénie peuvent faire partie des caractéristiques phénotypiques de

certains de ces syndromes, indiquant que des processus neurobiologiques communs sous-tendent ces différentes manifestations. Le but de cette section n'est pas de faire la liste exhaustive des nombreux syndromes génétiques rares décrits chez des patients atteints d'autisme mais de citer quelques exemples pour lesquels les TSA sont fréquemment associés ou qui permettent d'expliquer le plus de cas.

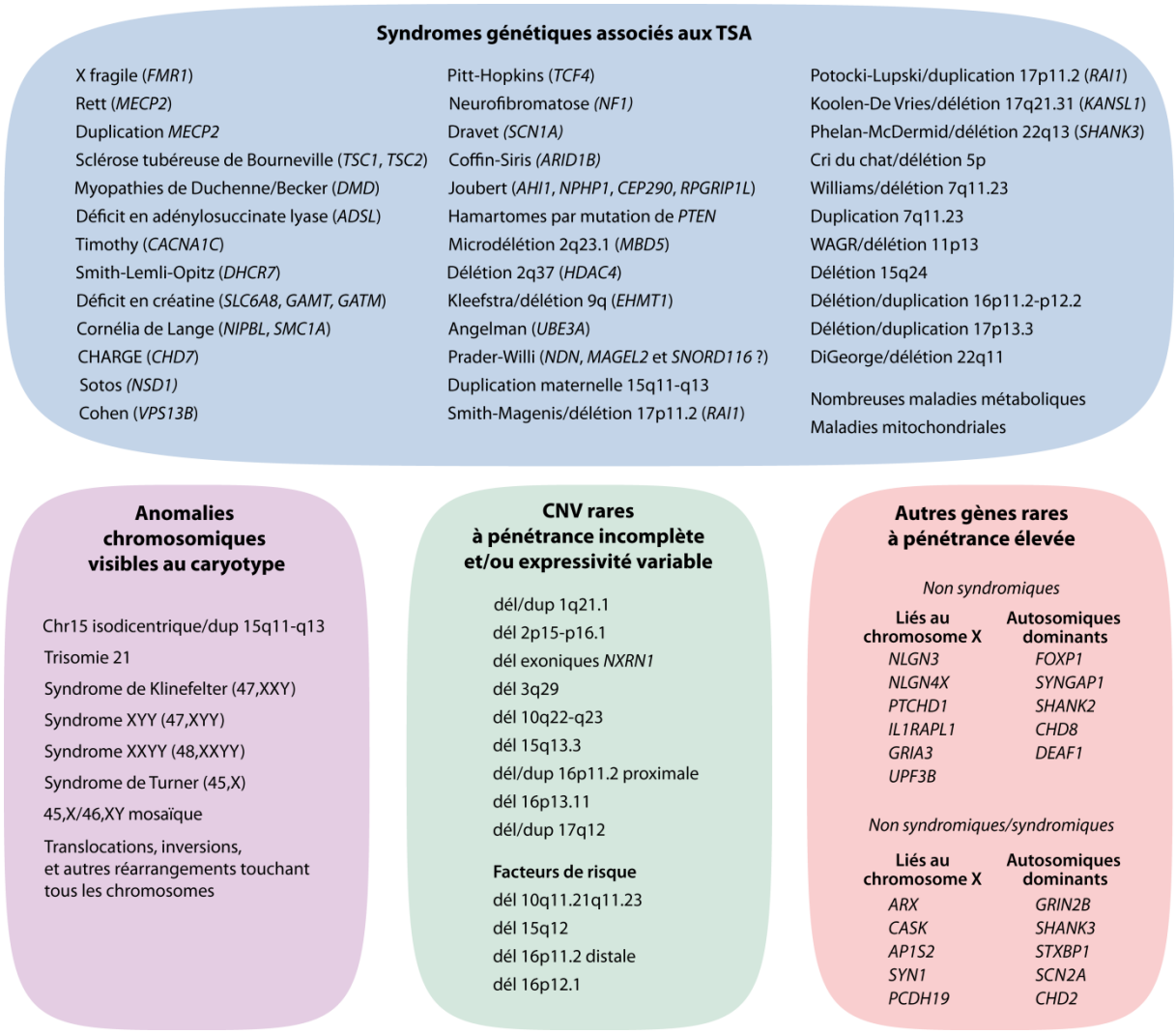



**Figure 17. Vue d'ensemble des étiologies génétiques connues de TSA**

Les variants génétiques rares associés aux TSA ont été regroupés en quatre catégories avec des exemples représentatifs. Dans leur ensemble, ces anomalies génétiques rares expliquent la pathologie chez environ 25 % des patients.

### 4.2.1 Maladies monogéniques

De nombreuses pathologies monogéniques rares ont été impliquées dans l'étiologie des TSA (pour revues voir Betancur, 2011 ; Betancur et Coleman, 2013). A ce jour, le nombre de gènes impliqués dans l'autisme est de 148 et ne cesse d'augmenter (**Figure 18**).

**Tableau 3. Syndromes génétiques pour lesquels les TSA sont une manifestation fréquente**

| Maladie génétique | Gène | Fréquence | Références |
|---|---|---|---|
| Syndrome de l'X Fragile | *FMR1* | TSA : 67 % garçons et 23 % filles | (Clifford et coll., 2007) |
| Syndrome de Rett | *MECP2* | TSA/traits autistiques fréquents | (Neul, 2012) |
| Syndrome de duplication *MECP2* | *MECP2* | 76 % autisme/traits autistiques | (Ramocki et coll., 2010) |
| Sclérose tubéreuse de Bourneville | *TSC1, TSC2* | 40-50 % TSA | (Jeste et coll., 2008 ; Numis et coll., 2011) |
| Syndrome de duplication 15q11–q13 | *UBE3A* | 92 % TSA (dont 81 % autisme) | (Hogart et coll., 2010) |
| Syndrome d'Angelman (délétion 15q11-q13, mutations) | *UBE3A* | 61 % TSA | (Bonati et coll., 2007) |
| Déficit en adénylosuccinate lyase | *ADSL* | ~50 % autisme/traits autistiques | (Spiegel et coll., 2006) |
| Syndrome de Smith-Magenis (délétion 17p11.2, mutations) | *RAI1* | 90 % TSA | (Laje et coll., 2010) |
| Syndrome de Potocki-Lupski (duplication 17p11.2) | *RAI1* | 67 % TSA | (Treadwell-Deering et coll., 2010) |
| Syndrome de Smith-Lemli-Opitz | *DHCR7* | 71 % TSA ; 53 % autisme | (Tierney et coll., 2001 ; Sikora et coll., 2006) |
| Syndrome de Timothy | *CACNA1C* | 80 % TSA | (Splawski et coll., 2004) |
| Syndrome de Phelan-McDermid (délétion 22q13) | *SHANK3* | 84 % TSA (dont 75 % autisme) | (Soorya et coll., 2013) |
| Syndrome CHARGE | *CHD7* | 68 % TSA/traits autistiques (dont 48 % autisme) | (Johansson et coll., 2006) |
| Syndrome de dysplasie corticale avec épilepsie focale | *CNTNAP2* | 67 % TSA | (Strauss et coll., 2006) |
| Syndrome de Cohen | *VPS13B* | 49 % autisme | (Howlin et coll., 2005) |
| Syndrome de Cornelia de Lange | *NIPBL, SMC1A* | 43-65 % autisme/TSA | (Moss et coll., 2012 ; Nakanishi et coll., 2012) |
| Maladie de Steinert | *DMPK* | 49 % TSA | (Ekstrom et coll., 2008) |
| Syndrome de déficit en créatine cérébrale | *SLC6A8* | TSA/traits autistiques fréquents | (Longo et coll., 2011) |
| Syndrome de Williams | ? | 50 % TSA | (Klein-Tasman et coll., 2009) |

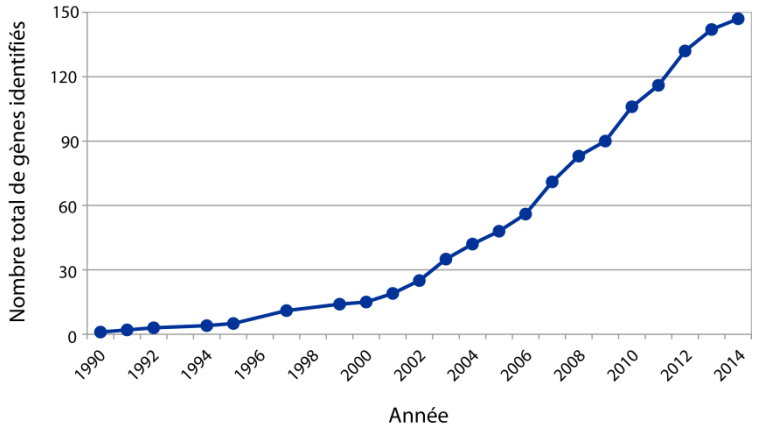



**Figure 18. Nombre de gènes impliqués dans l'étiologie des TSA**

Grâce à la recherche de variants rares, le nombre de gènes impliqués dans l'étiologie des TSA n'a cessé d'augmenter ces dernières années pour atteindre 148 en 2014 (figure faite à partir d'une version mise à jour des gènes impliqués dans les TSA publiée dans Betancur, 2011).

Parmi les plus fréquemment décrites, on trouve notamment le syndrome de l'X fragile, le syndrome de Rett et la sclérose tubéreuse de Bourneville.

- Le ***syndrome de l'X fragile*** est l'anomalie génétique la plus souvent retrouvée chez les sujets atteints de TSA, présente chez environ 2 % des cas (Kielinen et coll., 2004 ; Reddy, 2005 ; Hagerman et coll., 2010). Le syndrome de l'X fragile représente également la cause monogénique la plus fréquente de déficience intellectuelle. Elle est due à l'expansion anormale d'une répétition de triplets dans la région régulatrice du gène *FMR1*, situé sur le chromosome X. Le trinucléotide CGG normalement présent à environ 30 copies sera, par erreur, répété plus de 200 fois. Cette augmentation du triplet CGG va entraîner une hyperméthylation de la région 5' du gène provoquant la répression de son expression. Quand le nombre de copies atteint 200 et plus la mutation est complète, s'il se situe entre 55 et 200 l'individu est porteur de la pré-mutation. Cette pré-mutation est très instable et pourra se déséquilibrer à la génération suivante, on parle de mutation dynamique. La transmission est de type dominante liée à l'X, ce qui implique que tous les garçons porteurs de la mutation complète sont atteints mais également les filles, bien qu'elles présentent un phénotype plus variable et souvent moins sévère que les garçons. Soixante-sept pour cent des garçons porteurs de la mutation complète présentent un TSA et 23 % des filles (Clifford et coll., 2007). La pré-mutation est également associée à un risque accru de développer un autisme (Clifford et coll., 2007).
- Le ***syndrome de Rett*** est également fréquemment associé aux TSA. Une mutation ou une délétion du gène *MECP2* (Xq28) est responsable de l'apparition du syndrome qui touche principalement les filles (Neul, 2012). Il est souvent létal chez les garçons bien que quelques cas rares aient été décrits. Le syndrome de duplication *MECP2* est fortement associé aux TSA, principalement chez des garçons, avec 76 % des sujets ayant un autisme ou des traits autistiques (Ramocki et coll., 2010).
- La ***sclérose tubéreuse de Bourneville*** est une pathologie neurologique grave associée à des lésions tubéreuses. Elle est causée par des délétions ou des mutations des gènes *TSC1* et *TSC2*, qui sont des gènes suppresseurs de tumeurs (Curatolo et coll., 2008). La prévalence de cette maladie dans la population générale ou dans des cohortes d'individus avec TSA est faible. Cependant, des TSA ont été rapportés chez environ 40-50 % des patients atteints de la sclérose tubéreuse de Bourneville (Jeste et coll., 2008 ; Numis et coll., 2011).

En outre, de nombreux gènes localisés sur le chromosome X ont été impliqués dans l'étiologie des TSA (Betancur, 2011) (**Figure 19**). Parmi eux, certains sont responsables de formes syndromiques comme par exemple les gènes *DMD* (dystrophies musculaires de Duchenne/Becker) et *ATRX* (syndrome d'α-thalassémie associée à une déficience intellectuelle) ; d'autres ont été impliqués chez des patients ayant des présentations non syndromiques comme les gènes *NLGN3* et *NLGN4X*.

De nombreux autres syndromes génétiques monogéniques peuvent inclure les TSA parmi leurs manifestations cliniques (pour revues voir Betancur et Coleman, 2013 ; Zafeiriou et coll., 2013). Certains peuvent être regroupés dans des sous-groupes cliniques et étiologiques, par exemple :

- Les **encéphalopathies épileptiques précoces** sont un groupe d'épilepsies très sévères caractérisées par l'apparition de crises réfractaires très tôt dans l'enfance et évoluant souvent vers une déficience intellectuelle profonde ou une régression sévère. Plusieurs gènes responsables d'encéphalopathies épileptiques comme les gènes *SCN1A* (syndrome de Dravet), *SCN2A*, *CDKL5*, *STXBP1* et *CHD2* ont également été impliqués dans les TSA et la déficience intellectuelle, soulignant le chevauchement étiologique existant entre ces trois troubles neurodéveloppementaux (O'Roak et coll., 2011 ; Neale et coll., 2012b ; Sanders et coll., 2012a ; Carvill et coll., 2013 ; Pinto et coll., 2014).

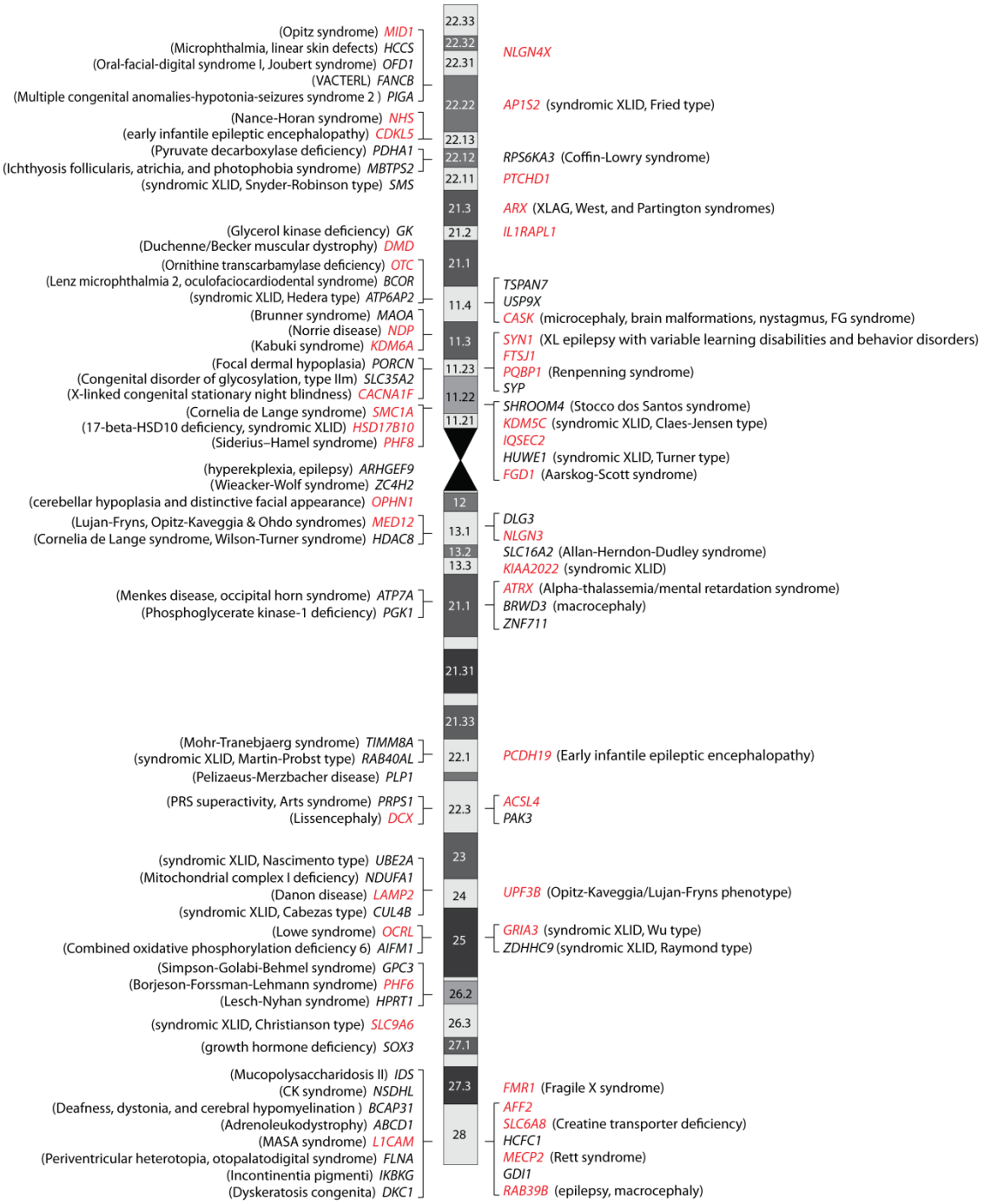



**Figure 19. Gènes liés au chromosome X impliqués dans la déficience intellectuelle et les TSA**

A ce jour, 93 gènes liés à l'X ont été impliqués dans la déficience intellectuelle dont 43 sont également impliqués dans les TSA (indiqués en rouge). Les gènes situés à gauche de la représentation du chromosome sont associés à des formes syndromiques et les gènes à droite à des formes non syndromiques. La distinction entre gènes syndromiques et non syndromiques n'est pas stricte puisque des mutations dans plusieurs gènes à droite ont été rapportées dans des familles avec des formes syndromiques de TSA ou de déficience intellectuelle. Les phénotypes associés sont indiqués entre parenthèses (version mise à jour d'une figure publiée dans Betancur, 2011).

- Les **syndromes métaboliques** sont des maladies très rares, également appelées 'erreurs innées du métabolisme', liées à un dysfonctionnement d'une protéine impliquée dans le métabolisme cellulaire. Mêmes si ces maladies sont extrêmement rares et ne concernent qu'une faible proportion de cas, il est important de les diagnostiquer puisque certaines sont potentiellement traitables. Par exemple, la phénylcétonurie, maladie récessive due à des mutations du gène *PAH* codant pour la phénylalanine

hydroxylase, peut être prévenue par la mise en place précoce d'un régime alimentaire adapté. Autrefois une cause relativement fréquente d'autisme, elle n'est retrouvée aujourd'hui que dans les pays sous-développés qui n'ont pas de programmes de dépistage néonatal. D'autres maladies métaboliques impliquées dans les TSA incluent le syndrome de Smith-Lemli-Opitz dû à des mutations ou délétions récessives du gène *DHCR7* impliqué dans la synthèse du cholestérol (Sikora et coll., 2006) ; des déficits du métabolisme des purines et pyrimidines liées à des mutations récessives des gènes *ADSL* (Spiegel et coll., 2006) et *DPYD* (van Kuilenburg et coll., 2009) ; les syndromes de déficit en créatine cérébrale dont les gènes responsables sont *SLC6A8*, *GAMT* et *GATM* (Longo et coll., 2011) ainsi que les mucopolysaccharidoses de type III (maladie de Sanfilippo) causées par des pertes homozygotes des gènes *SGSH*, *NAGLU*, *HGSNAT* et *GNS* (Heron et coll., 2011).

- Les **dystrophies musculaires** sont un groupe de myopathies génétiques qui peuvent être associées aux TSA. On trouve dans cette catégorie les dystrophies musculaires de Duchenne/Becker, causées par des anomalies dans le gène *DMD* localisé sur le chromosome X. Hormis le phénotype musculaire, les individus atteints peuvent également présenter un TSA et/ou une déficience intellectuelle (Young et coll., 2008 ; Hinton et coll., 2009). Il est intéressant de noter que plusieurs cas de patients atteints de TSA porteurs de duplications ou de délétions intragéniques du gène *DMD* ont été décrits sans aucun phénotype musculaire (Pinto et coll., 2010 ; Pagnamenta et coll., 2011 ; Pinto et coll., 2014). Ces cas pourraient correspondre à des formes moins sévères de type Becker, à expression plus tardive ou à des cas atypiques sans aucune atteinte musculaire. Les TSA sont également une manifestation fréquente des patients atteints de la maladie de Steinert, une autre forme de dystrophie musculaire causée par une expansion de triplet dans la région 3' du gène *DMPK* (Ekstrom et coll., 2008).
- Les **maladies mitochondriales** correspondent à des anomalies de la chaîne respiratoire mitochondriale et peuvent parfois être associées aux TSA. Ces troubles touchent des gènes de l'ADN mitochondrial ou plus souvent des gènes nucléaires codant pour des protéines mitochondriales (Haas, 2010).

### 4.2.2 Syndromes de microdélétion/microduplication

De nombreux syndromes de microdélétion ou microduplication ont été identifiés chez des patients avec TSA. Ils correspondent à des CNVs qui peuvent être récurrents ou de taille variable selon le mécanisme d'apparition (**Figure 20**).

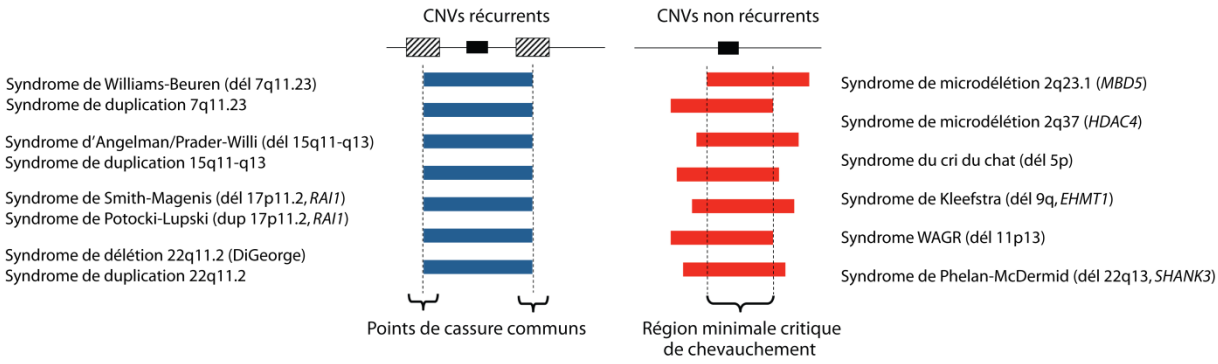



**Figure 20. CNVs récurrents et non récurrents**

La ligne horizontale noire représente la région génomique impliquée dans le réarrangement, le rectangle noir représente un gène localisé dans cette région et les rectangles rayés correspondent à des duplications segmentales. Les lignes épaisses bleues ou rouges correspondent aux CNVs. Les CNVs récurrents ont une taille et des points de cassure similaires ; les CNVs non récurrents ont des tailles variables mais peuvent présenter une région minimale commune. Des exemples de chaque type de CNV, impliqués dans l'étiologie des TSA, sont présentés de part et d'autre du schéma (figure adaptée de Gu et coll., 2008).

***Syndromes associés à des CNVs de taille variable***

Parmi les syndromes de microdélétion/microduplication de taille variable associés aux TSA, on trouve notamment plusieurs délétions subtélomériques comme le syndrome de Phelan-McDermid, dû à une délétion terminale du chromosome 22q13.33 et dont le gène responsable est *SHANK3*. D'autres exemples incluent le syndrome de microdélétion 2q37 (délétion terminale 2q, gène *HDAC4*), le syndrome de Kleefstra (délétion terminale 9q, gène *EHMT1*) et le syndrome de Jacobsen (délétion 11q terminale, gène(s) impliqué(s) non connu).

Dans le cas des CNVs de taille variable impliqués dans un syndrome, la description d'un nombre suffisant de patients et notamment la description de réarrangements atypiques et/ou de petite taille, permet parfois de définir une région critique. Dans de rares cas, la région ainsi délimitée ne contient qu'un seul gène, ce qui permet d'attribuer les manifestations associées au syndrome aux altérations de ce dernier. La détection de mutations d'un gène de la région chez plusieurs patients et absentes de témoins peut également apporter la preuve de sa contribution au phénotype. Dans les cas où la région minimale critique contient plusieurs gènes, la présence d'un point de cassure d'une translocation interrompant l'un de ces gènes chez un individu ayant une présentation clinique similaire peut apporter une évidence, bien que ce type de réarrangement complexe ne puisse pas impliquer un gène directement. En effet, les translocations présentent toujours deux points de cassure et même lorsque la région impliquée dans le deuxième point de cassure est non génique, on ne peut exclure sa contribution au phénotype. Par exemple, dans le syndrome de microdélétion 22q13 (syndrome de Phelan-McDermid), de nombreuses délétions de tailles très variables avaient été décrites. Une translocation *de novo* interrompant le gène *SHANK3* avait été rapportée chez un garçon présentant toutes les caractéristiques cliniques du syndrome de microdélétion 22q13 (Bonaglia et coll., 2001). Ce gène était l'un des meilleurs candidats fonctionnels dans la région puisqu'il code une protéine d'échafaudage synaptique qui interagit avec les neuroligines, déjà impliquées dans l'étiologie des TSA. Puis, la région minimale délétée a été réduite à environ 90 kb (Bonaglia et coll., 2006), contenant uniquement 3 gènes, dont le gène *SHANK3*. Le séquençage de ce gène dans une cohorte de patients atteints de TSA a permis la découverte d'une mutation *de novo* causant un décalage du cadre de lecture chez deux frères atteints, impliquant définitivement *SHANK3* comme responsable des manifestations principales du syndrome et notamment des altérations neurocognitives (retard sévère ou absence de langage, déficience intellectuelle et TSA) (Durand et coll., 2007). Depuis, plusieurs délétions interstitielles n'impliquant que le gène *SHANK3* ont également été décrites chez des patients atteints du syndrome de Phelan-McDermid (Bonaglia et coll., 2011) (**Figure 21**).

Dans d'autres cas, la région critique est trop large pour impliquer un gène en particulier. Il est possible que nombre de ces syndromes soient causés par l'altération simultanée de plusieurs gènes chacun contribuant indépendamment au phénotype avec des effets additifs ou synergiques, dans ce cas, ils sont appelés syndromes de gènes contigus.

***Syndromes associés à des CNVs récurrents***

Les CNVs récurrents sont formés dans des régions génomiques riches en duplications segmentales, qui favorisent le mécanisme de recombinaison homologue non allélique à leur origine. Les duplications segmentales correspondent aux points de cassure récurrents des réarrangements (*breakpoints*, BP) qui ont ainsi pu être caractérisés moléculairement et phénotypiquement pour définir des syndromes de microdélétion et microduplication.

Sur le bras long du chromosome 15, différents groupes de duplications segmentales entraînent la formation de différentes délétions et duplications récurrentes associées aux TSA. La duplication interstitielle de la région chromosomique 15q11-q13 entre les BP2-BP3 ou BP1-BP3 (la région critique se situant entre le BP2 et le BP3) est l'un des réarrangements les plus fréquemment associés à l'autisme (Depienne et coll., 2009b ; Hogart et coll., 2010). Ces duplications peuvent être interstitielles ou

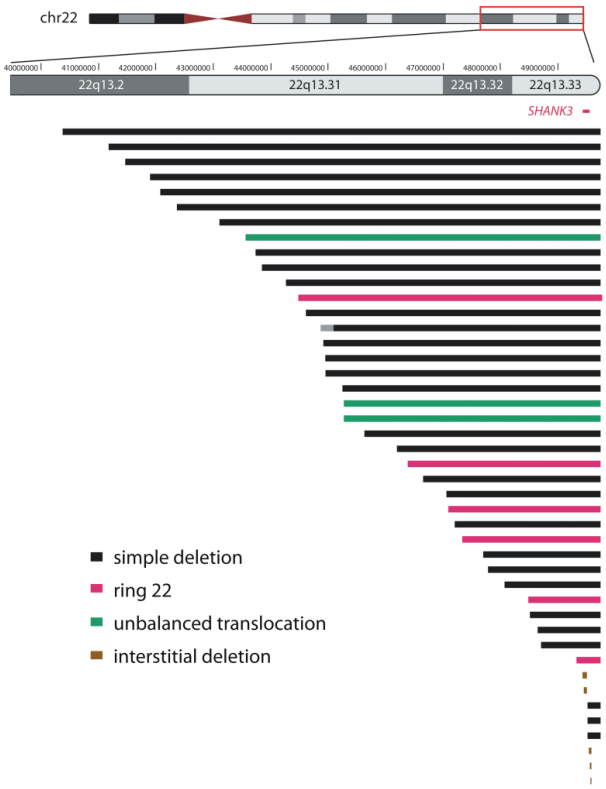



**Figure 21. Délétions 22q13 de taille variable chez les patients atteints du syndrome de Phelan-McDermid**
Les délétions peuvent être simples (terminales ou interstitielles) ou correspondre à un remaniement complexe (chromosome en anneau ou translocation). Le seul gène contenu dans la région de délétion commune est *SHANK3* (figure adaptée de Bonaglia et coll., 2011).

correspondre à la présence d'un chromosome 15 isodicentrique et sont la plupart du temps d'origine maternelle. Ces dernières sont caractérisées par une pénétrance élevée et une forte association avec les TSA, contrairement aux duplications interstitielles paternelles qui sont plus rarement associées à un phénotype anormal (Depienne et coll., 2009b) et peuvent parfois être détectées chez des individus non atteints (Veltman et coll., 2005). Le phénotype inclut également un retard de développement, un retard de langage, une déficience intellectuelle plus ou moins sévère, des difficultés de coordination motrice ainsi que quelques dysmorphies. Par ailleurs, des triplications de cette région ont également été décrites, associées à un phénotype plus sévère peu importe l'origine parentale du réarrangement (Hogart et coll., 2010). Les délétions de cette région sont impliquées dans les syndromes d'Angelman et Prader-Willi. Ces deux syndromes associés aux TSA sont dus à des altérations des gènes soumis à empreinte dans la région. Le syndrome d'Angelman est causé par une déficience de la copie maternelle du gène *UBE3A* (mutation, délétion, disomie uniparentale paternelle ou défaut d'empreinte). Dans le cas du syndrome de Prader-Willi, c'est la copie paternelle de la région qui est impliquée cependant, le gène (ou les gènes) responsable n'a pas encore été clairement identifié.

Parmi les CNVs récurrents pour lesquels les TSA sont une manifestation fréquente, on trouve les délétions et duplications 17p11.2, incluant le gène *RAI1*. Les délétions causent le syndrome de Smith-Magenis alors que les duplications sont responsables du syndrome de Potocki-Lupski. Ces deux syndromes sont très rares, cependant, les individus atteints présentent des TSA dans environ 90 % et 70 % des cas, respectivement (Laje et coll., 2010 ; Treadwell-Deering et coll., 2010). De même, des délétions au locus 7q11.23 sont responsables du syndrome de Williams-Beuren pour lequel environ 50 % des patients ont un TSA (Klein-Tasman et coll., 2009 ; Tordjman et coll., 2012). La duplication en miroir de cette région

est également pathogène et environ 40 % des individus atteints présentent un autisme (Depienne et coll., 2007 ; Van der Aa et coll., 2009 ; Sanders et coll., 2011). Dans la région 7q11.23, le ou les gènes impliqués dans le phénotype neurocognitif ne sont pas connus.

## 4.3 CNVs rares à pénétrance incomplète et/ou expressivité variable

Bien que de nombreux CNVs récurrents pathogènes aient été décrits précédemment, le développement de puces à ADN de résolution croissante a permis de détecter de nombreux **nouveaux syndromes de microdélétion/microduplication**. L'étude de la localisation des duplications segmentales tout le long du génome a été utilisée comme stratégie pour détecter de nouveaux remaniements (Sharp et coll., 2006). Plusieurs CNVs récurrents ont ainsi été dévoilés et correspondent à de nouveaux syndromes dits '*genotype-first*' puisqu'ils sont caractérisés d'abord par l'anomalie génétique commune aux patients et non par un ensemble de manifestations cliniques similaires (Mefford, 2009). En outre, un même CNV peut être impliqué dans une large gamme de phénotypes neurologiques et psychiatriques qui n'incluent pas seulement les TSA, la déficience intellectuelle et l'épilepsie mais aussi la schizophrénie et les troubles du déficit de l'attention/hyperactivité ; certains peuvent également être détectés chez des individus témoins ou transmis par des parents apparemment sains.

La pénétrance représente la probabilité pour un individu porteur d'un génotype donné, d'exprimer le phénotype associé. En pathologie, si la pénétrance est complète, tous les individus porteurs du variant causal exprimeront le phénotype anormal associé. L'expressivité correspond au degré de variation du phénotype associé à une mutation donnée. En effet, dans certains cas, une même anomalie génétique peut entraîner différents tableaux cliniques chez les différents individus porteurs, on dit alors que l'expressivité est variable. Ces deux phénomènes sont très souvent liés en pratique et peuvent correspondre aux mêmes mécanismes biologiques sous-jacents (**Figure 22**).

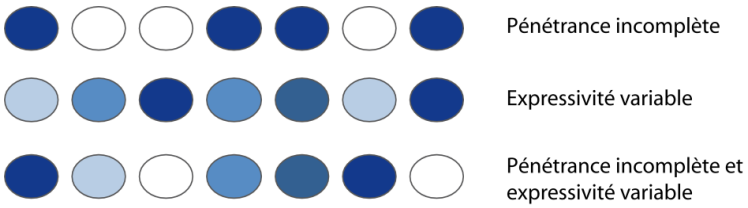



**Figure 22. Pénétrance incomplète et expressivité variable**
Chaque cercle représente un individu. En cas de pénétrance incomplète, certains individus portant un CNV délétère n'expriment pas le phénotype pathologique associé (en bleu). Des degrés de sévérité variables ou des phénotypes différents peuvent être causés par un même CNV, on parle alors d'expressivité variable. Ces deux phénomènes peuvent être observés pour un même CNV.

Ce phénomène a été observé dès la description des premiers syndromes de microdélétion et microduplication. En effet, les délétions associées au **syndrome de délétion 22q11** (syndrome de DiGeorge ou syndrome vélo-cardio-facial) ont été décrites dès le début des années 1980 (de la Chapelle et coll., 1981). Ces délétions ont été rapportées chez des individus atteints de déficience intellectuelle, de TSA, de schizophrénie (Fine et coll., 2005 ; Vorstman et coll., 2006 ; Antshel et coll., 2007 ; Niklasson et coll., 2009). En outre, ces délétions sont également associées à la présence de nombreuses caractéristiques syndromiques à des fréquences variables telles qu'une insuffisance vélopharyngienne, des malformations faciales, des fentes palatines, des anomalies cardiaques et/ou rénales. Deux tailles de délétions sont typiquement observées, l'une de 3 Mb et l'autre de 1,5 Mb ; le gène responsable des troubles neuropsychiatriques n'a pas encore été identifié. Par contre, il a été montré que le gène *TBX1*

(qui code pour un facteur de transcription) était responsable des anomalies cardiaques associées à ce syndrome.

Plus récemment, des microdélétions et microduplications réciproques d'environ 600 kb associées à une expressivité très variable ont été décrites dans la région **16p11.2 proximale** (BP4-BP5) (**Figure 23**). Les **délétions** ont été rapportées dans des cohortes de patients atteints de déficience intellectuelle et de TSA (Weiss et coll., 2008 ; Rosenfeld et coll., 2010 ; Shinawi et coll., 2010). Cette délétion récurrente est l'un des CNVs les plus fréquemment associés aux TSA puisqu'on la détecte chez environ 0.6 % des patients (Kumar et coll., 2008 ; Marshall et coll., 2008 ; Weiss et coll., 2008 ; Hanson et coll., 2014).

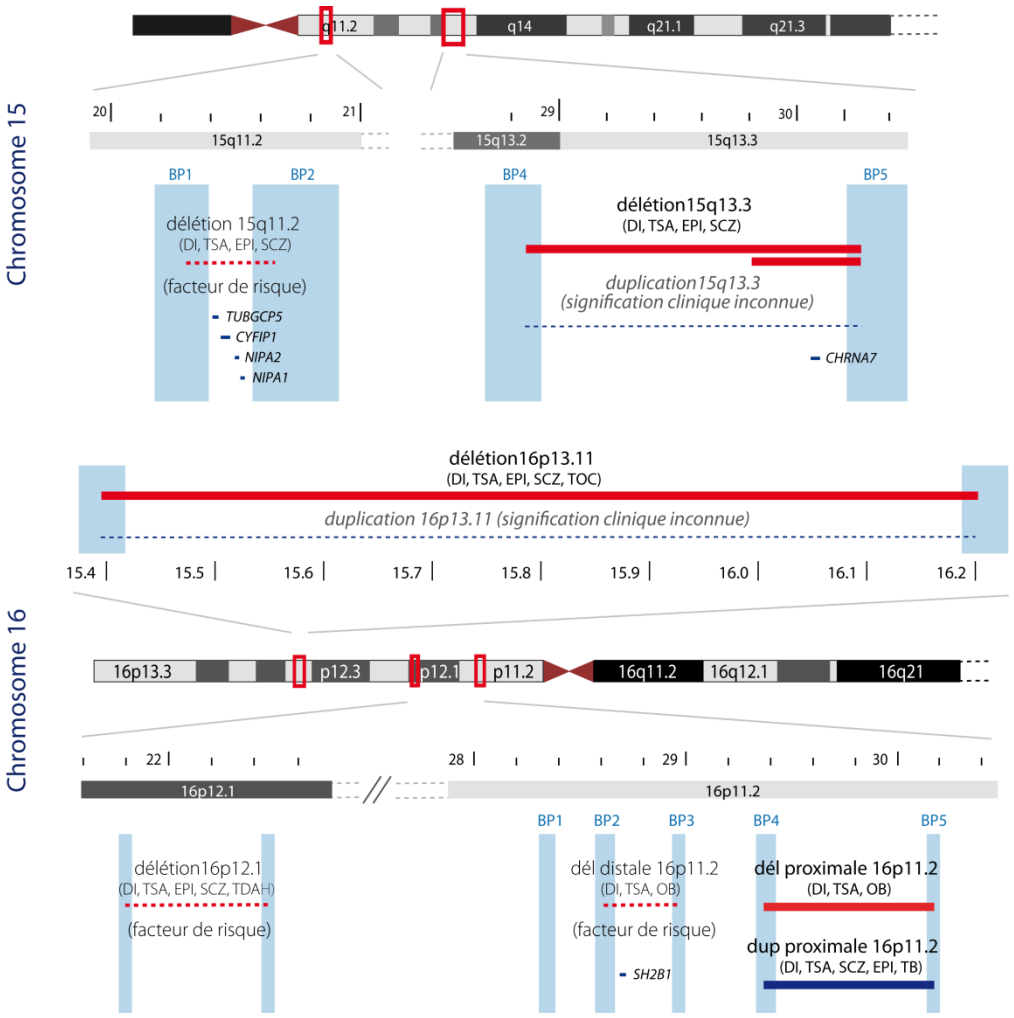



**Figure 23. CNVs rares à expressivité variable et/ou pénétrance incomplète sur les chromosomes 15 et 16**

Les régions des chromosomes 15 et 16 sujettes à des remaniements sont représentées ; les lignes horizontales indiquent les régions délétées (en rouge) ou dupliquées (en bleu) ; les points de cassure récurrents (BP) sont indiqués en bleu clair. Les CNVs pathogènes sont indiqués en noir alors que ceux considérés comme facteurs de risque sont indiqués en pointillés avec les légendes en gris et ceux dont la signification clinique est inconnue en gris italique. Les phénotypes associés à chacun des CNVs sont notés entre parenthèses. Les distances sont en Mb. Les quatre gènes de la région 15q11.2 (BP1-BP2) sont indiqués ainsi que les gènes *CHRNA7* et *SH2B1*, candidats des régions 15q13.3 et 16p11.2 distale, respectivement. DI, déficience intellectuelle ; EPI, épilepsie ; OB, obésité ; TB, trouble bipolaire ; TOC, troubles obsessionnels compulsifs ; TSA, troubles du spectre autistique ; SCZ, schizophrénie.

Une récente étude a réalisé une caractérisation phénotypique détaillée de 85 individus porteurs de délétions 16p11.2 proximales, leurs parents et frères/sœurs non porteurs (Hanson et coll., 2014). Les résultats ont confirmé la présence de troubles neuropsychiatriques chez 93 % des porteurs contre 21 %

chez les non porteurs. Parmi les phénotypes les plus fortement associés on retrouve les troubles du langage et de la coordination (chez plus de 50 % des porteurs), les TSA (26 %), des troubles du déficit de l'attention/hyperactivité (19 %) et la déficience intellectuelle (10 %). En effet, la distribution des scores de QI chez les individus porteurs est décalée de presque 2 SD (*standard deviation*, écart-type) par rapport aux non-porteurs de la famille (**Figure 24**). La plupart des délétions identifiées sont *de novo* (79 %), suggérant qu'elles pourraient avoir un effet délétère sur la reproduction. Les **duplications** de cette région sont quant à elles retrouvées dans des cohortes de patients atteints de TSA, de déficience intellectuelle, de schizophrénie, d'épilepsie et de troubles bipolaires (McCarthy et coll., 2009 ; Fernandez et coll., 2010 ; Shinawi et coll., 2010 ; Cooper et coll., 2011 ; Reinthaler et coll., 2014). Par ailleurs, les anomalies de dosage génique dans cette région sont corrélées à l'indice de masse corporelle chez les individus porteurs. De façon intéressante, les délétions ont été identifiées chez des patients obèses (Walters et coll., 2010) alors que les duplications sont associées à un indice de masse corporelle faible (Jacquemont et coll., 2011). Les délétions et duplications de cette région ont également des effets inverses sur le périmètre crânien puisqu'elles sont associées à une macrocéphalie et à une microcéphalie, respectivement (Shinawi et coll., 2010).

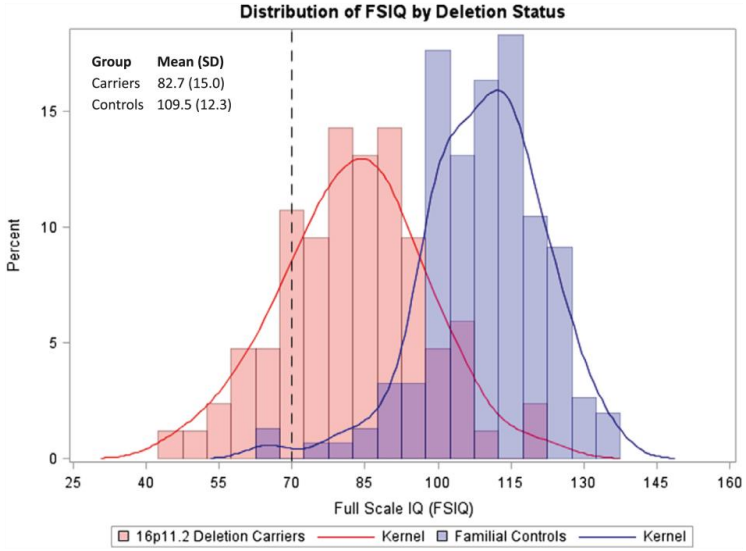



**Figure 24. Distribution du QI chez les porteurs d'une délétion 16p11.2 par rapport aux non porteurs de la famille**
La distribution des scores de QI chez les porteurs (*carriers*, en rouge) présente un décalage de 1,8 SD par rapport à celles des non porteurs de la famille (*familial controls*, en bleu) (figure tirée de Hanson et coll. 2014).

Une pénétrance incomplète et une expressivité variable sont également observées pour les **délétions 15q13.3** (BP4-BP5, distales aux délétions impliquées dans les syndromes d'Angelman et Prader-Willi, **Figure 23**), qui ont d'abord été décrites chez des patients avec déficience intellectuelle et épilepsie (Sharp et coll., 2008), puis ont été impliquées dans les TSA et la schizophrénie (Ben-Shachar et coll., 2009 ; Miller et coll., 2009 ; Masurel-Paulet et coll., 2010 ; Pinto et coll., 2010 ; Costain et coll., 2013). Une étude récente, rapportant toutes les délétions de cette région décrites à ce jour (195 délétions typiques BP4-BP5, 8 délétions BP3-BP5 et 43 délétions plus petites incluant *CHRNA7*), a montré qu'elles sont héritées la plupart du temps (85 %) et sont le plus souvent d'origine maternelle (Lowther et coll., 2014). Ces délétions sont très pénétrantes (environ 80 % des porteurs présentent un phénotype neuropsychiatrique) et absentes chez les témoins (non retrouvées chez 23 838 témoins adultes). Le gène *CHRNA7,* qui code pour un canal ionique ligand-dépendant impliqué dans la transmission neuronale, a été proposé comme candidat pour l'épilepsie et les manifestations neuropsychiatriques associées à ces délétions. La **duplication** en miroir de cette région est de signification clinique inconnue puisqu'elle a été rapportée à

des fréquences similaires entre des cohortes d'individus atteints de déficience intellectuelle ou de TSA et de témoins (van Bon et coll., 2009 ; Moreno-De-Luca et coll., 2013). Par contre, un enrichissement significatif de ces duplications a été rapporté dans des cohortes de patients atteints de troubles du déficit de l'attention/hyperactivité, par rapport aux témoins (Williams et coll., 2012). L'ensemble de ces résultats, négatifs et positifs, nécessite d'être confirmé dans des cohortes indépendantes de grande taille.

Les **délétions et duplications 16p13.11** sont également associées à un large spectre de manifestations phénotypiques et à une pénétrance incomplète (**Figure 23**). La délétion a été décrite chez des patients avec TSA, déficience intellectuelle, épilepsie, schizophrénie et troubles obsessionnels compulsifs, parfois héritées de parents apparemment sains (Ullmann et coll., 2007 ; Hannes et coll., 2009 ; Heinzen et coll., 2010 ; Ingason et coll., 2011b ; McGrath et coll., 2014) alors que la duplication a été trouvée chez des patients avec TSA, déficience intellectuelle et schizophrénie (Ullmann et coll., 2007 ; Ingason et coll., 2011b). Les CNVs de cette région ont également été identifiés chez des individus témoins ; cependant, la délétion est significativement enrichie dans les cohortes de patients (Moreno-De-Luca et coll., 2013) ce qui n'est pas le cas de la duplication qui reste de signification clinique inconnue pour le moment. Certaines études suggèrent qu'elle pourrait tout de même jouer un rôle de facteur de risque dans les TSA et d'autres troubles neurodéveloppementaux (Ramalingam et coll., 2011 ; Tropeano et coll., 2013).

Les **délétions exoniques du gène *NRXN1***, non récurrentes, présentent aussi une pénétrance incomplète et une expressivité variable. Elles ont été décrites dans la déficience intellectuelle et les TSA mais également chez des patients atteints d'épilepsie, de schizophrénie ou de troubles du déficit de l'attention/hyperactivité (Szatmari et coll., 2007 ; Ching et coll., 2010 ; Pinto et coll., 2010 ; Schaaf et coll., 2012 ; Bena et coll., 2013). Ces délétions sont de tailles variables et semblent affecter préférentiellement la forme longue (α) du gène (**Figure 25**).

D'autres CNVs récurrents, rares, pathogènes et associés à un spectre phénotypique varié incluent les **délétions et duplications 1q21.1** (Szatmari et coll., 2007 ; Mefford et coll., 2008 ; Pinto et coll., 2010), les **délétions 3q29** (Willatt et coll., 2005 ; Ballif et coll., 2008 ; Quintero-Rivera et coll., 2010) et les **délétions et duplications 17q12** (Moreno-De-Luca et coll., 2010).

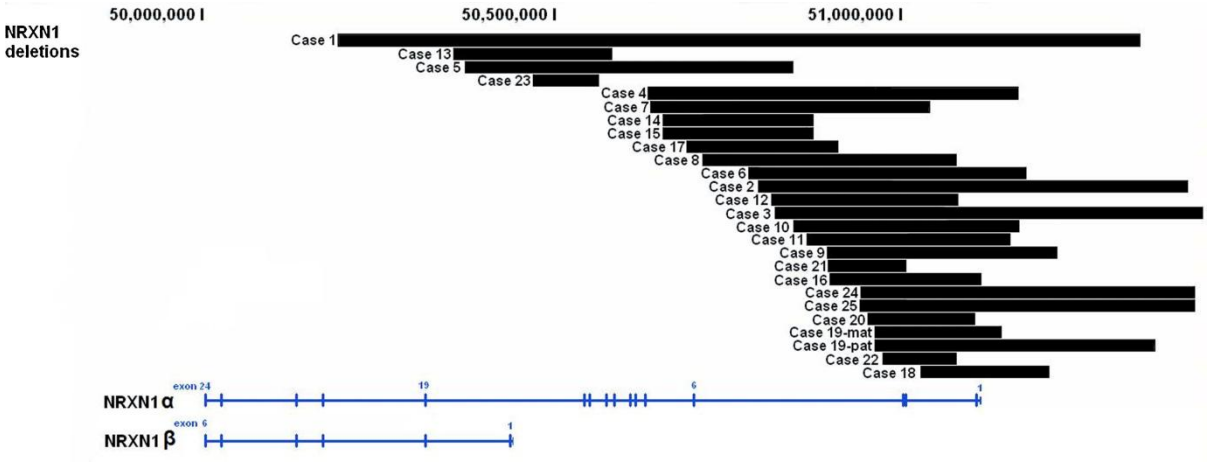



**Figure 25. Agrégation des délétions *NRXN1* dans la forme alpha du gène**

Des délétions exoniques du gène *NRXN1* détectées chez 25 individus atteints de troubles neuropsychiatriques incluant les TSA, la déficience intellectuelle, l'épilepsie et le retard de langage sont représentées par les lignes horizontales noires. Les deux formes du gène sont indiquées en bleu. Les délétions sont regroupées dans les premiers exons de la forme longue (α) du gène (tirée de Bena et coll. 2013).

D'autres remaniements génomiques identifiés ces dernières années, caractérisés par une forte variabilité phénotypique et une pénétrance incomplète, semblent avoir un effet plus modeste dans la susceptibilité aux TSA. Ces délétions et duplications sont le plus souvent héritées et sont détectées à des

fréquences non-négligeables chez les témoins. Cependant, certaines d'entre elles apparaissent significativement enrichies dans des cohortes de patients et pourraient donc exercer un rôle de **facteur de risque**. C'est le cas des délétions proximales 15q11.2, des délétions distales 16p11.2 et des délétions 16p12.1 et 10q11.21-q11.23 :

- Les **délétions et duplications proximales 15q11.2** (BP1-BP2) ont été décrites très tôt chez des patients atteints de TSA et leur statut pathogénique a longtemps été débattu. La région impliquée, adjacente à la région des syndromes d'Angelman et Prader-Willi, contient quatre gènes (*NIPA1*, *NIPA2*, *CYFIP1* et *TUBGCP5*) (**Figure 23**). Bien que retrouvées chez des individus témoins et héritées de parents apparemment sains, les délétions semblent enrichies dans de grandes cohortes de patients avec épilepsie, schizophrénie et d'autres troubles neurodéveloppementaux variés (International Schizophrenia Consortium, 2008 ; de Kovel et coll., 2010 ; Burnside et coll., 2011 ; Abdelmoity et coll., 2012). Par contre, les duplications ont été observées à des fréquences similaires entre patients et témoins, suggérant qu'elles ne constituent pas un facteur de risque (Cooper et coll., 2011). Une étude récente a estimé la pénétrance ainsi que le risque associés à ces CNVs dans un groupe de 2 525 familles avec TSA (Chaste et coll., 2014). Un CNV impliquant cette région a été identifié chez 85 individus de 43 familles (27 délétions et 58 duplications). Aucune différence significative entre la fréquence des délétions et des duplications entre les enfants atteints et leurs frères et sœurs non atteints n'a été trouvée. Dans cet échantillon, les délétions ne sont pas significativement associées aux TSA ; par contre, un déséquilibre de transmission a été observé pour les duplications qui sont plus fréquentes que dans les cohortes précédemment étudiées. En outre, la pénétrance estimée de ces CNVs est très faible (plus faible pour les délétions que pour les duplications, respectivement 0,013 et 0,018) et le risque associé est très modeste (OR = 1,3 pour les délétions et 1,8 pour les duplications). Les auteurs concluent qu'en cas de conseil génétique, les CNVs de cette région doivent être considérés comme 'de signification clinique inconnue' (*variants of unknown significance*, VOUS) et que des échantillons de taille beaucoup plus importante doivent être analysés pour confirmer leurs résultats (environ 20 000 patients et témoins pour la délétion et 3 500 patients et témoins pour la duplication, plus fréquente).
- Des **délétions 16p11.2 distales** (BP2-BP3) d'environ 220 kb et contenant le gène *SH2B1* ont été décrites chez des patients avec TSA, obésité sévère à début précoce, déficience intellectuelle et schizophrénie (Weiss et coll., 2008 ; Bachmann-Gagescu et coll., 2010 ; Bochukova et coll., 2010 ; Barge-Schaapveld et coll., 2011 ; Tabet et coll., 2012 ; Guha et coll., 2013), ainsi que d'autres phénotypes incluant des anomalies rénales et urinaires ainsi que dans la maladie de Hirschsprung (Sampson et coll., 2010) (**Figure 23**). L'haploinsuffisance du gène *SH2B1* a clairement été impliquée dans le phénotype d'obésité sévère précoce puisque des mutations de ce gène ont été trouvées chez plusieurs individus présentant une obésité parfois associée à une hyperphagie, une résistance à l'insuline et/ou une petite taille (Doche et coll., 2012). De façon intéressante, ces individus présentaient des problèmes de comportements incluant un isolement social, un retard de langage et des comportements agressifs, suggérant une possible implication de ce gène dans les phénotypes neurologiques observés chez les individus porteurs de la délétion récurrente. Une méta-analyse réalisée sur une très large cohorte de patients présentant un retard de développement, une déficience intellectuelle, un TSA et/ou d'autres troubles neurodéveloppementaux (environ 30 000 cas et plus de 10 000 témoins) a montré un enrichissement de ces délétions chez les cas comparés aux témoins (Moreno-De-Luca et coll., 2013). Cet enrichissement n'a pas été observé dans l'analyse combinée d'environ 4 000 patients atteints de TSA issus de plusieurs cohortes. Ce résultat a été interprété par les auteurs comme un manque de puissance statistique dû à une taille d'échantillon plus petite. Des études additionnelles dans des cohortes de plus grande taille sont donc nécessaires pour mieux comprendre l'implication de ces délétions dans la susceptibilité aux TSA.

- Les **microdélétions 16p12.1** ont également été associées à divers troubles neurodéveloppementaux incluant la schizophrénie, les TSA et la déficience intellectuelle (International Schizophrenia Consortium, 2008 ; Marshall et coll., 2008 ; Girirajan et coll., 2010) (**Figure 23**). Tout comme les délétions 16p11.2 distales, leur enrichissement a été montré dans une très large cohorte de patients avec retard de développement et déficience intellectuelle mais pas dans les TSA (Moreno-De-Luca et coll., 2013). Elles sont le plus souvent héritées et les individus qui les portent peuvent présenter un second CNV de grande taille (Girirajan et coll., 2010). Les auteurs ont proposé que ces délétions joueraient un rôle de facteur de risque, en association à un premier événement (modèle oligogénique 'second-hit').
- Les **délétions 10q11.21-q11.23** ont été récemment reportées chez des patients avec des phénotypes variables dont la seule caractéristique commune était de présenter une déficience intellectuelle et un retard de développement (Stankiewicz et coll., 2012). La taille de ces délétions est variable mais la plupart du temps les points de cassure sont localisés dans des duplications segmentales. Un TSA a été décrit chez 27 % (4/15) des cas. Ces délétions sont souvent héritées de parents apparemment sains mais montrent un enrichissement significatif chez les cas et pourraient donc jouer un rôle dans la vulnérabilité aux TSA et à d'autres troubles neurodéveloppementaux. Deux patients de la cohorte de l'*Autism Genome Project* sont porteurs de cette délétion ; l'une étant *de novo*, l'autre est héritée d'un père apparemment sain (Pinto et coll., 2014).

Les duplications en miroir de ces quatre CNVs n'ont pas montré d'enrichissement significatif chez les cas par rapport aux témoins et sont donc de signification clinique inconnue (Stankiewicz et coll., 2012 ; Moreno-De-Luca et coll., 2013). Les délétions jouant un rôle de facteur de risque dans les TSA agissent certainement en interaction avec d'autres facteurs, non identifiés pour le moment, pour provoquer l'apparition de ces troubles. Le risque associé à ces réarrangements est modeste voire faible et des cohortes de taille beaucoup plus importante sont nécessaires afin de tester leur implication réelle dans l'étiologie des TSA et des autres troubles neurodéveloppementaux.

En conclusion, l'ensemble de ces résultats montre la forte contribution des variants rares à l'étiologie génétique des TSA. Bien que de nombreuses causes soient *de novo*, des formes héritées autosomiques récessives ou liées à l'X ont également été décrites. Plus récemment, un rôle important des variations autosomiques dominantes à pénétrance incomplète et expressivité variable a émergé, ouvrant une nouvelle voie dans la recherche de facteurs impliqués dans les TSA. L'hétérogénéité génétique des TSA est extrême puisque plus d'une centaine de gènes et loci ont déjà été impliqués. Longtemps considérés comme une entité unique, les TSA apparaissent aujourd'hui comme la manifestation comportementale de nombreux défauts génétiques et génomiques plus ou moins fréquents et qui permettent collectivement d'expliquer une importante fraction des cas. Malgré la remarquable avancée dans nos connaissances des facteurs impliqués dans les TSA, il nous reste encore beaucoup à découvrir et l'étude d'un nombre très important de cas et de témoins pourra nous aider à mieux comprendre le rôle de certains variants candidats identifiés. Ces résultats ont également mis en évidence des bases génétiques communes entre les TSA et d'autres troubles neurodéveloppementaux incluant la déficience intellectuelle et l'épilepsie mais aussi parfois la schizophrénie ou les troubles du déficit de l'attention/hyperactivité. Ceci suggère que ces différents troubles partagent des mécanismes physiopathologiques communs altérant des voies biologiques impliquées dans la mise en place et la régulation des circuits cérébraux.

## 5. Multiples voies biologiques impliquées dans l'autisme

Plus de cent gènes ont été mis en cause dans les TSA et des centaines d'autres sont candidats. Ces gènes codent des protéines impliquées dans un large éventail de fonctions biologiques incluant des molécules

d'adhésion cellulaire, des récepteurs, des canaux ioniques, des protéines du cytosquelette, des protéines de transport vésiculaire, des molécules de signalisation intracellulaire, des enzymes, des facteurs de transcription/traduction, et des protéines de liaison à l'ADN. Cependant, malgré le nombre de protéines impliquées dans l'étiologie de l'autisme, une convergence vers certaines voies biologiques communes a rapidement été observée. Historiquement, l'une des premières voies biologiques impliquées dans les TSA a été celle de la synapse. A plusieurs reprises, des gènes codant des protéines synaptiques et mutés dans l'autisme ont été identifiés comme les acteurs principaux de réseaux fonctionnels hautement connectés.

***Protéines synaptiques***

La formation de circuits neuronaux fonctionnels permettant l'intégration et le traitement des informations dans le cerveau dépend de la mise en place des synapses qui ont donc été particulièrement étudiées dans la neurobiologie des troubles neurodéveloppementaux. La voie synaptique englobe des processus cellulaires et moléculaires divers tels que la migration neuronale, la croissance axonale, la stabilité synaptique et la dynamique de épines dendritiques. Notamment, plusieurs gènes codant pour des protéines d'adhésion cellulaire synaptiques ont été impliqués dans l'étiologie des TSA (pour revues voir Sudhof, 2008 ; Betancur et coll., 2009). Parmi eux, les exemples classiquement présentés sont les gènes codant pour des **neuroligines**, protéines d'adhésion post-synaptiques ayant un rôle dans la formation, le maintien et la fonction des synapses ; leurs partenaires pré-synaptiques les **neurexines** qui jouent un rôle dans le fonctionnement des synapses et la libération de neurotransmetteurs et les **protéines d'échafaudage post-synaptiques SHANK** qui sont des partenaires indirects des neuroligines (**Figure 26**). Comme mentionné précédemment, des variants rares pathogènes ont été trouvés chez des patients atteints de TSA et de déficience intellectuelle dans les gènes *NLGN3*, *NLGN4X*, *NRXN1*, *SHANK2* et *SHANK3*. D'autres gènes codant des protéines d'adhésion cellulaire localisées à la synapse et impliqués dans les TSA incluent *PCDH19* (Hynes et coll., 2010 ; Marini et coll., 2010), *L1CAM* (Simonati et coll., 2006) et *CNTNAP2* (Strauss et coll., 2006 ; Zweier et coll., 2009). Bien que l'attention se soit longtemps focalisée sur les gènes codant des molécules d'adhésion cellulaire synaptiques, de nombreux autres gènes codant des protéines synaptiques ont été impliqués dans les TSA, la déficience intellectuelle et d'autres troubles neurodéveloppementaux (**Tableau 4**). Il est intéressant de noter qu'un enrichissement en gènes codant des protéines de la densité post-synaptique a été observé parmi les gènes impliqués dans la déficience intellectuelle liée à l'X (Laumonnier et coll., 2007). Une analyse protéomique ayant isolé la densité post-synaptique à partir d'échantillons de néocortex humain a montré qu'il existerait environ 1 400 protéines dans cette structure (Bayes et coll., 2011) ; l'étude des gènes codant ces protéines peut permettre d'impliquer de nouveaux gènes dans les TSA, la déficience intellectuelle et les autres troubles neuropsychiatriques. D'ailleurs, une analyse basée sur cette liste a montré un enrichissement en gènes de la densité post-synaptique parmi les gènes touchés par des CNVs chez des patients avec TSA (Pinto et coll., 2014).

La découverte de mutations rares touchant de nombreux gènes jouant un rôle dans les processus épigénétiques et en particulier dans le remodelage de la chromatine a également permis de mettre en évidence l'implication de cette voie dans les TSA et la déficience intellectuelle (pour revues voir Ronan et coll., 2013 ; Zhubi et coll., 2014).

***Protéines de modification de l'ADN/remodelage de la chromatine***

Un nombre croissant de gènes codant pour des protéines impliquées dans la modification de l'ADN ou le remodelage de la chromatine ont été impliqués dans les TSA. Parmi eux, le plus décrit est le gène ***MECP2*** (*methyl-CpG-binding protein 2*), qui code pour une protéine de liaison à l'ADN méthylé jouant un rôle dans la répression de la transcription. Les mutations et délétions de ce gène sont responsables du syndrome de Rett alors qu'une surexpression du gène cause le syndrome de duplication *MECP2*. Ces deux syndromes ont été décrits chez des patients atteints de TSA (Ramocki et coll., 2010 ; Schaefer et Mendelsohn, 2013).

Des gènes codant pour des protéines impliquées dans les modifications des histones ont également été impliqués dans l'étiologie des TSA. C'est par exemple le cas du gène ***EHMT1*** (*euchromatic histone-lysine N-methyltransferase 1*) qui code pour une méthyltransférase d'histones. La méthylation de l'ADN est une marque de l'hétérochromatine associée à une diminution de la transcription. Le syndrome de Kleefstra (délétion subtélomérique 9q) est dû à des délétions ou des mutations du gène *EHMT1* et les TSA font souvent partie des manifestations observées chez ces patients (Kleefstra et coll., 2009). D'autres gènes codant des protéines modifiant les histones ont été impliqués dans les TSA. A titre d'exemples, on peut citer le gène ***HDAC4*** (codant une désacétylase d'histone), impliqué dans le syndrome de microdélétion 2q37 (Williams et coll., 2010 ; Pinto et coll., 2014) ; le gène ***CREBBP*** (codant une acétylase d'histone) responsable du syndrome de Rubinstein-Taybi (Schorry et coll., 2008) et le gène ***KDM5C*** (précédemment appelé *JARID1C* et qui code une déméthylase d'histone) (Adegbola et coll., 2008). Plusieurs gènes appartenant à la famille CHD (*chromodomain helicase DNA binding*) jouent également un rôle dans l'étiologie des TSA : ***CHD7*** est responsable du syndrome CHARGE associé aux TSA (Johansson et coll., 2006), ***CHD8*** a été mis en évidence par les études d'exome dans les TSA (O'Roak et coll., 2012a) et confirmé comme étant à l'origine d'une forme de TSA avec des caractéristiques phénotypiques communes (Bernier et coll., 2014) ; et ***CHD2*** qui a d'abord été impliqué dans les encéphalopathies épileptiques et plus récemment dans les TSA (Carvill et coll., 2013 ; Pinto et coll., 2014).

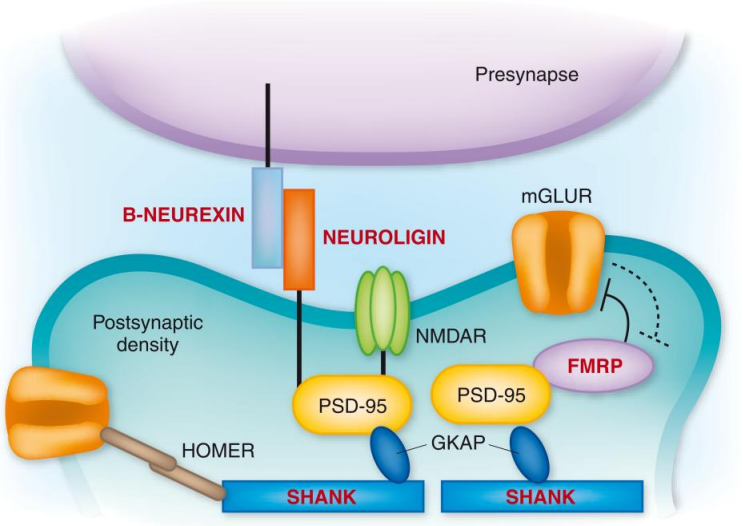



**Figure 26. Protéines synaptiques impliquées dans les TSA**

Les molécules d'adhésion cellulaires neuronales de la famille des neurexines interagissent avec les neuroligines post-synaptiques qui elles sont liées aux protéines PSD-95 dans la densité post-synaptique. Les protéines SHANK forment une matrice qui sert d'échafaudage aux molécules de la densité post-synaptiques incluant HOMER et les protéines associées aux guanylate kinases (GKAP), liées aux PSD-95. La densité post-synaptique est caractérisée par la présence des récepteurs au NMDA (*N*-methyl-D-aspartate, NMDARs) et au glutamate (mGluRs). Les mGluRs activent la synthèse de certaines protéines ainsi que certaines voies de signalisation et sont négativement régulés par la protéine FMRP (*fragile X mental retardation protein*). Les protéines impliquées dans les TSA sont indiquées en rouge (figure tirée de State, 2010).

Des analyses de voies biologiques (*pathway analyses*) des gènes touchés par des CNVs ou des SNVs dans les TSA ont confirmé que beaucoup de ces gènes étaient connectés en un nombre limité de réseaux fonctionnels. De façon remarquable, bien que les approches utilisées par les différentes analyses soient très hétérogènes, en particulier au niveau des gènes sur lesquelles elles sont basées et des méthodes analytiques, toutes ont montré une convergence, en particulier vers la voie synaptique (Pinto et coll., 2010 ; Gilman et coll., 2011 ; O'Roak et coll., 2012b ; Pinto et coll., 2014). En utilisant une liste de gènes touchés par des mutations *de novo* (perte de fonction ou faux-sens prédites délétères) détectées par

**Tableau 4. Gènes synaptiques impliqués dans les TSA et autres troubles neurodéveloppementaux**

| Gène | Localisation | Fonction | Pathologie | Evidence | Références |
|---|---|---|---|---|---|
| **Adhésion cellulaire et échafaudage** | | | | | |
| *NRXN1* | 2p16.3 | Adhésion cellulaire, formation des synapses | TSA, DI, schizophrénie, TDAH, syndrome de Tourette et autres troubles neurodéveloppementaux et psychiatriques (dominant); syndrome de Pitt-Hopkins-like-2 (récessif) | Délétions, mutations, translocations | (Szatmari et coll., 2007 ; Rujescu et coll., 2009 ; Zweier et coll., 2009 ; Ching et coll., 2010 ; Pinto et coll., 2010) |
| *CNTNAP2* | 7q35-q36.1 | Membre de la famille des neurexines | Syndrome de dysplasie corticale avec épilepsie focale, associé à DI et TSA (récessif)[a] | Mutations récessives | (Strauss et coll., 2006 ; Zweier et coll., 2009) |
| *NLGN4X* | Xp22.32-p22.31 | Adhésion cellulaire, formation des synapses | TSA et/ou DI non syndromique liée à l'X | Mutations, délétions | (Jamain et coll., 2003 ; Laumonnier et coll., 2004) |
| *NLGN3* | Xq13.1 | Adhésion cellulaire, formation des synapses inhibitrices | TSA et DI non syndromique liée à l'X | Mutations, délétion | (Jamain et coll., 2003 ; Sanders et coll., 2011) |
| *PCDH19* | Xq22.1 | Adhésion cellulaire, formation des synapses | Epilepsie liée à l'X spécifique des filles, avec ou sans déficits cognitifs et TSA ; encéphalopathie épileptique précoce | Mutations, délétions | (Camacho et coll., 2012 ; Depienne et LeGuern, 2012) |
| *L1CAM* | Xq28 | Adhésion cellulaire, formation et développement des synapses | Déficience intellectuelle syndromique liée à l'X avec hydrocéphalie, pouces adductus, paraplégie spastique, hypoplasie du corps calleux et TSA | Délétions, mutations | (Saghatelyan et coll., 2004 ; Simonati et coll., 2006 ; Schafer et Altevogt, 2010) |
| *SHANK3* | 22q13.33 | Echafaudage synaptique | Syndrome de Phelan-McDermid, associé à DI, absence ou un retard de langage, TSA, hypotonie ; des mutations ont également été rapportées dans deux familles avec schizophrénie et DI | Délétions, mutations | (Durand et coll., 2007 ; Moessner et coll., 2007 ; Gauthier et coll., 2010b ; Hamdan et coll., 2011b ; Gong et coll., 2012) |
| *SHANK2* | 11q13.3 | Echafaudage synaptique | DI et TSA | Délétions, mutations | (Berkel et coll., 2010 ; Pinto et coll., 2010 ; Leblond et coll., 2012) |
| *CASK* | Xp11.4 | Protéine adaptatrice/d'échafaudage ; plasticité synaptique | DI liée à l'X avec microcéphalie et hypoplasie ponto-cérébelleuse, TSA | Mutations, délétions | (Najm et coll., 2008 ; Tarpey et coll., 2009 ; Hackett et coll., 2010 ; Lin et coll., 2013) |
| **Récepteurs et canaux ioniques** | | | | | |
| *GRIN2B* | 12p13.1 | Canalopathie (canal ionique ligand-dépendant, récepteur ionotropique glutamatergique, NMDA) | DI, TSA et épilepsie | Mutations, translocations | (Endele et coll., 2010 ; O'Roak et coll., 2012a) |
| *GRIA3* | Xq25 | Canalopathie (canal ionique ligand-dépendant, récepteur ionotropique glutamatergique, AMPA) | DI et TSA | Délétions, duplications, mutations, translocations | (Chiyonobu et coll., 2007 ; Wu et coll., 2007) |
| *CHRNA7* | 15q13.3 | Canalopathie (canal ionique ligand-dépendant, récepteur nicotinique de l'acétylcholine) | Syndrome de microdélétion 15q13.3, associé à DI, épilepsie, TSA, schizophrénie et autres troubles neuropsychiatriques | Délétions | (Sharp et coll., 2008 ; Shinawi et coll., 2009 ; Levinson et coll., 2011 ; Hoppman-Chaney et coll., 2013) |
| *SCN1A* | 2q24.3 | Canalopathie (canal sodique voltage-dépendant) | Epilepsie myoclonique juvénile sévère (syndrome de Dravet), DI et TSA | Délétions, mutations | (Marini et coll., 2009 ; O'Roak et coll., 2011 ; Carvill et coll., 2013) |
| *SCN2A* | 2q24.3 | Canalopathie (canal sodique voltage-dépendant) | Encéphalopathie épileptique précoce ; convulsions bénignes infantiles familiales, DI et TSA | Délétions, mutations | (Neale et coll., 2012a ; Sanders et coll., 2012b ; Carvill et coll., 2013) |
| *CACNA1C* | 12p13.33 | Canalopathie (canal calcique voltage-dépendant) | Syndrome de Timothy, associé à DI et TSA | Mutations gain de fonction | (Splawski et coll., 2004) |
| **Transduction, transcription, traduction** | | | | | |
| *PTEN* | 10q23.31 | Voie de signalisation AKT-mTOR (tyrosine phosphatase), développement synaptique, plasticité synaptique activité-dépendante | Syndrome d'harmatomes par mutation de *PTEN* avec macrocéphalie, associé à DI et TSA | Délétions, mutations | (Butler et coll., 2005 ; Buxbaum et coll., 2007 ; McBride et coll., 2010) |
| *TSC1*<br>*TSC2* | 9q34.13,<br>16p13.3 | Voie de signalisation inhibitrice mTOR (protéine suppresseur de tumeurs) | Sclérose tubéreuse de Bourneville, associée à DI, TSA et épilepsie | Délétions, mutations | (Numis et coll., 2011) |

| Gène | Localisation | Fonction | Pathologie | Evidence | Références |
|---|---|---|---|---|---|
| *MECP2* | Xq28 | Facteur de transcription (liaison à l'ADN méthylé) | Haploinsuffisance : syndrome de Rett chez les filles ; encéphalopathie congénitale ou DI non syndromique chez les garçons, un cas unique décrit avec schizophrénie ; syndrome de duplication *MECP2*, principalement chez les garçons, DI et TSA | Délétions, mutations, duplications | (Cohen et coll., 2002 ; Carney et coll., 2003 ; Ramocki et coll., 2010 ; Na et coll., 2013) |
| *FMR1* | Xq27.3 | Répresseur traductionnel de nombreuses protéines synaptiques | Syndrome de l'X fragile, associé à DI, TSA, TDAH et épilepsie | Expansion (CGG)n, mutations | (Sidorov et coll., 2013) |
| *UBE3A* | 15q11.2 | Ubiquitine ligase, marque les protéines pour dégradation ; formation des synapses | Syndrome d'Angelman, associé à DI, TSA, absence de langage et épilepsie (haploinsuffisance) ; syndrome de duplication 15q11-q13, associé à DI, TSA, schizophrénie, troubles du langage et épilepsie | Délétions, mutations, duplications | (Ingason et coll., 2011a ; Mabb et coll., 2011) |
| *CREBBP* | 16p13.3 | Modulateur transcriptionnel (acétyltransférase d'histones) | Syndrome de Rubinstein-Taybi, associé à DI et TSA | Mutations, délétions | (Schorry et coll., 2008) |
| *EHMT1* | 9q34.3 | Régulation transcriptionnelle (méthyltransférase d'histones), développement des synapses | Syndrome de Kleefstra (délétion subtélomérique 9q), associé à DI, TSA et schizophrénie | Délétions, mutations | (Kirov et coll., 2012 ; Willemsen et coll., 2012 ; Balemans et coll., 2013) |
| *ARID1B* | 6q25.3 | Remodelage de la chromatine neurone-spécifique, plasticité synaptique | DI, défauts de langage, TSA, anomalies du corps calleux et syndrome de Coffin-Siris | Mutations, délétions | (Halgren et coll., 2011 ; Santen et coll., 2012 ; Wieczorek et coll., 2013) |
| *SMARCA2* | 9p24.3 | Remodelage de la chromatine neurone-spécifique (ATPase/hélicase), plasticité synaptique | Syndrome de Nicolaides-Baraitser, syndrome de Coffin-Siris, DI et TSA | Mutations, délétions | (Sousa et coll., 2009 ; Van Houdt et coll., 2012) |
| *SRCAP* | 16p11.2 | Remodelage de la chromatine neurone-spécifique (ATPase), plasticité synaptique | Syndrome de Floating-Harbor, DI variable, TDAH et TSA | Mutations | (White et coll., 2010 ; Hood et coll., 2012) |
| *HDAC4* | 2q37.3 | Régulation transcriptionnelle (désacétylase d'histones), plasticité synaptique | Syndrome de brachydactylie avec déficience intellectuelle (syndrome de délétion 2q37), associé à DI et TSA | Délétions, mutations | (Williams et coll., 2010 ; Kim et coll., 2012 ; Sando et coll., 2012) |
| *STXBP1* | 9q34.11 | Fusion des vésicules synaptiques (protéine SNARE) | Encéphalopathie épileptique précoce, DI non syndromique sans épilepsie, TSA et malformations cérébrales | Mutations, délétions | (Hamdan et coll., 2011c ; Milh et coll., 2011 ; Neale et coll., 2012a ; Barcia et coll., 2014) |
| *SYN1* | Xp11.23 | Phosphoprotéine de vésicule synaptique qui module la libération de neurotransmetteurs, le trafic vésiculaire synaptique et synaptogenèse | Epilepsie liée à l'X avec déficits d'apprentissage variables et troubles comportementaux | Mutations, délétions | (Garcia et coll., 2004 ; Boido et coll., 2010 ; Cesca et coll., 2010 ; Fassio et coll., 2011) |
| *IL1RAPL1* | Xp21.2-p21.3 | Formation des synapses | DI et TSA liés à l'X, avec ou sans traits dysmorphiques | Mutations, délétions | (Piton et coll., 2008 ; Pinto et coll., 2010 ; Valnegri et coll., 2011) |
| *NF1* | 17q11.2 | Voie de signalisation MAPK (protéine activatrice de RAS-GTPases) | Neurofibromatose de type 1, associée à DI et TSA | Délétions, mutations | (Garg et coll., 2013) |
| *SYNGAP1* | 6p21.32 | Voie de signalisation RAS, régulation du trafic de récepteurs glutamatergiques et de l'efficacité synaptique, maturation des synapses excitatrices | DI non syndromique, TSA et épilepsie ; encéphalopathie épileptique | Délétions, mutations | (Pinto et coll., 2010 ; Hamdan et coll., 2011a ; Clement et coll., 2012 ; Berryer et coll., 2013 ; Carvill et coll., 2013) |
| *RAB39B* | Xq28 | Formation et maintien des synapses | DI liée à l'X avec TSA, épilepsie et macrocéphalie | Mutations | (Giannandrea et coll., 2010) |
| *CDKL5* | Xp22.13 | Développement et plasticité synaptiques | Encéphalopathie épileptique précoce, DI et TSA | Mutations, délétions | (Archer et coll., 2006 ; Russo et coll., 2009 ; Su et Tsai, 2011 ; Fehr et coll., 2013) |
| *MEF2C* | 5q14.3 | Régulation de la formation et de la fonction des synapses | Syndrome de microdélétion 5q14.3, associé à DI, épilepsie et TSA | Délétions, mutations | (Barbosa et coll., 2008 ; Novara et coll., 2010 ; Zweier et coll., 2010) |

Abréviations : DI, déficience intellectuelle ; *MAGUK, membrane-associated guanylate kinase ;* mTOR, *mammalian target of rapamycin* ; SNARE, *soluble N-ethylmaleimide-sensitive-factor attachment protein receptor* ; TDAH, troubles du déficit de l'attention/hyperactivité.

[a]Des délétions ou des remaniements chromosomiques hétérozygotes de *CNTNAP2* ont été reportés chez des patients atteints de TSA, déficience intellectuelle, schizophrénie et troubles bipolaires et également chez des témoins ; leur signification clinique est inconnue

quatre études d'exome dans les TSA et deux dans la déficience intellectuelle, Krumm et coll. (2014) ont montré que ces gènes sont impliqués dans trois voies biologiques majeures : le remodelage de la chromatine et la fonction synaptique ainsi que la voie de signalisation *Wnt* (Krumm et coll., 2014) (**Figure 27**). De même, les analyses de voies biologiques réalisées dans la dernière étude de CNVs de l'*Autism Genome Project*, basées sur les gènes touchés par les CNVs *de novo* détectés dans une cohorte de 2 446 patients atteints de TSA, ont souligné la convergence de ces gènes vers des réseaux impliqués dans le développement neuronal, la fonction synaptique et la régulation de la chromatine (Pinto et coll., 2014).

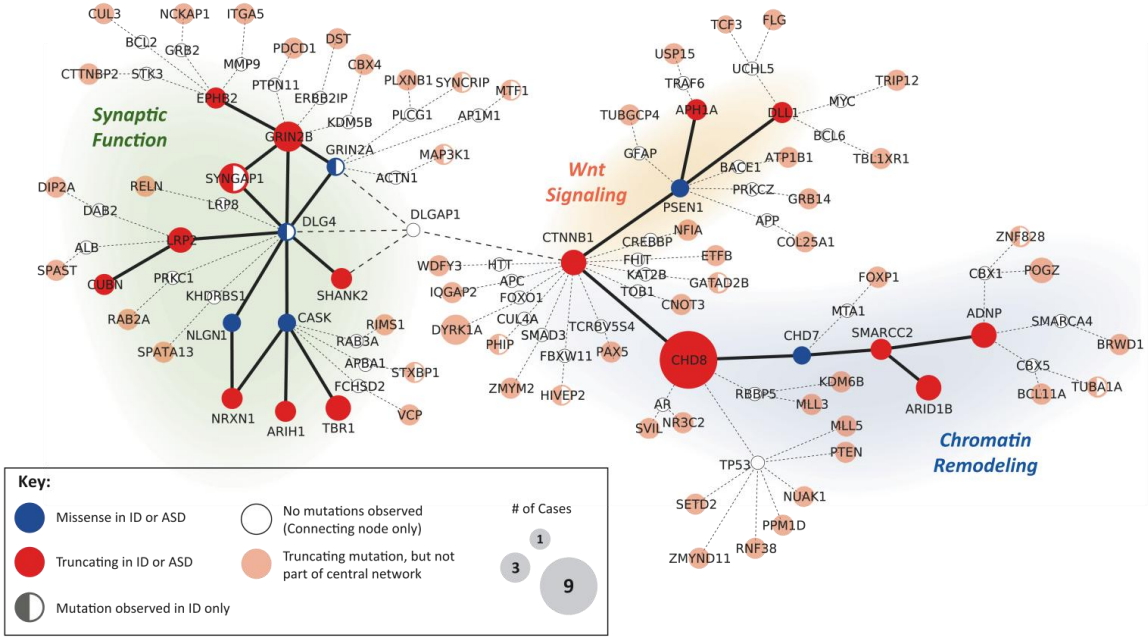



**Figure 27. Convergence des gènes impliqués par les études d'exome vers les mêmes voies biologiques**
Les protéines dans lesquelles des mutations *de novo* potentiellement délétères ont été trouvées chez des patients de 4 études d'exome dans les TSA et 2 études d'exome dans la déficience intellectuelle sont représentées dans un même réseau fonctionnel. Les mutations perte de fonction sont représentées en rouge, les faux-sens prédites délétères en bleu (figure tirée de Krumm et coll., 2014).

Un enrichissement significatif des cibles de la protéine FRMP (*Fragile X mental retardation protein*) a été montré parmi les gènes touchés par des mutations *de novo* chez des patients avec TSA, non retrouvé en analysant les mutations *de novo* chez les frères et sœurs non-atteints (Iossifov et coll., 2012). De plus, deux études d'exome dans la schizophrénie ont également montré un enrichissement des gènes codant des cibles de la protéine FRMP parmi ceux touchés par des mutations *de novo* chez leurs patients, soulignant l'existence de processus physiopathologiques communs à ces troubles neurodéveloppementaux (Fromer et coll., 2014 ; Purcell et coll., 2014). La protéine FMRP, codée par le gène *FMR1* responsable du syndrome de l'X fragile, se lie aux ARNs messagers pour moduler leur traduction et possède des centaines de cibles qui n'ont pas toutes été identifiées à ce jour (Darnell et coll., 2011 ; Ascano et coll., 2012). Plusieurs de ces cibles appartiennent aux voies de signalisation RAS/MAPK et mTOR, dans lesquelles plusieurs gènes impliqués dans les TSA jouent un rôle (*PTEN*, *TSC1* et *TSC2* appartiennent à la voie de signalisation mTOR ; *NF1*, *SPRED1* et *SYNGAP1* à la voie des kinases RAS/MAP). Un enrichissement en délétions et duplications touchant des cibles de la protéine FMRP a également été observé chez les cas de la cohorte de l'*Autism Genome Project* (Pinto et coll., 2014). Cet enrichissement persiste après avoir soustrait de l'analyse les CNVs interprétés comme pathogènes, suggérant l'implication d'autres gènes codant des cibles de FMRP dans l'étiologie des TSA.

Observer les différents gènes mis en évidence par les études de CNVs et d'exomes dans le contexte d'un réseau fonctionnel peut orienter vers des voies biologiques communes et de nouveaux gènes candidats qui appartiennent à ces voies biologiques. De plus, le fait que les gènes impliqués convergent vers un nombre restreint de voies biologiques permet d'envisager des **approches thérapeutiques communes** à différentes causes de TSA et/ou de déficience intellectuelle.

L'analyse des voies biologiques impliquées dans les TSA peut être réalisée à un niveau plus intégré, en utilisant des données transcriptomiques par exemple. Une analyse de ce type a été réalisée en utilisant des données d'expression de micropuces à ARN de la base de données *Brainspan* qui contient des données transcriptomiques de plusieurs régions du cerveau au cours du développement (Willsey et coll., 2013). Les auteurs ont construit un réseau de co-expression afin de définir dans quelles régions du cerveau et à quels stades de développement l'expression des gènes impliqués dans les TSA est cruciale. La liste des gènes qu'ils ont utilisée contient les gènes touchés de façon récurrente par des mutations *de novo* perte de fonction et 144 autres gènes candidats tirés des études d'exome dans les TSA. Les résultats de leur analyse ont montré une convergence vers les régions temporales du cerveau dans la période de développement fœtal entre 15-20 semaines. De façon intéressante, une étude parue simultanément a trouvé des résultats similaires malgré la différence des approches analytiques utilisées (Parikshak et coll., 2013). Pour cette analyse, une liste comprenant un plus grand nombre de gènes candidats pour les TSA et des données de séquençage d'ARN de *Brainspan* ont été utilisés. Les résultats montrent que les gènes potentiellement impliqués dans les TSA sont exprimés dans les couches superficielles du cortex, pendant la période fœtale précoce. Ces deux études soulignent l'importance du **contexte spatial et temporel spécifique** de l'expression des gènes impliqués dans les TSA et ce type d'analyse promet d'enrichir considérablement nos connaissances des processus physiopathologiques impliqués dans les TSA et les autres troubles neurodéveloppementaux.

# Résultats

Notre équipe étudie une cohorte de patients recrutés dans la cadre d'un consortium international, l'étude PARIS (*Paris Autism Research International Study*), qui regroupe plus de 700 familles avec un ou plusieurs enfants atteints de TSA idiopathique ; les patients pour lesquels nous avons déjà identifié une étiologie sont exclus des analyses.

Depuis plusieurs années, notre équipe participe au consortium international *Autism Genome Project* (AGP), constitué de nombreux groupes travaillant sur la génétique de l'autisme aux Etats-Unis, au Canada et en Europe. Le projet principal de ce consortium a été de réaliser une analyse pangénomique par micropuces à haute résolution sur près de 3 000 familles. Chaque équipe du consortium fournit les informations cliniques et les échantillons d'ADN de ses propres familles ; le génotypage par micropuces et l'analyse des CNVs sont centralisés. Les résultats sont ensuite stockés dans une base de données qui est accessible à tous les membres. Le génotypage par puce des familles de l'AGP à été réalisé en deux parties (stades 1 et 2), chacune incluant la moitié des familles. Grâce à la participation de notre équipe à ce projet collaboratif, 282 familles de la cohorte PARIS ont été analysées par micropuces. Les analyses des CNVs réalisées par l'AGP ont donné lieu à la publication de deux articles, correspondant aux deux stades de l'étude ; le premier a été publié en 2010 (Pinto et coll., *Nature*, 2010) et le second, auquel j'ai participé, a été publié en 2014 (Pinto et coll., *Am J Hum Genet*, 2014).

Mon travail de thèse a consisté dans un premier temps à analyser les CNVs des familles de l'étude PARIS, obtenus à partir des micropuces lors du stade 2 de l'étude de l'AGP, afin d'identifier et de valider des CNVs pathogènes chez nos patients. Puis, j'ai étendu l'analyse à l'ensemble des familles du consortium et j'ai procédé à la ré-analyse des données combinées des stades 1 et 2 afin d'identifier des CNVs pathogènes (**Article 1, Convergence des gènes et des voies biologiques dérégulés dans les TSA,** Pinto et coll., *Am J Hum Genet*, 2014). Les analyses ont confirmé la forte contribution des CNVs rares à l'étiologie de l'autisme, qu'ils soient *de novo* ou hérités et ont permis d'identifier de nombreux CNVs pathogènes. De plus, cette étude a permis d'impliquer pour la première fois dans les TSA des gènes responsables d'autres pathologies neurodéveloppementales comme les gènes *CHD2, GDI1* et *HDAC4*, ainsi que des gènes nouveaux comme le gène *SETD5* et de mettre en évidence de nombreux gènes candidats pour des analyses futures. Enfin, des analyses de voies biologiques ont montré que malgré la très forte hétérogénéité génétique, les gènes impliqués convergent vers un nombre limité de réseaux fonctionnels.

En parallèle de cette étude, j'ai pu appliquer mes connaissances dans l'interprétation des CNVs à deux autres projets collaboratifs sur les patients de la cohorte PARIS afin d'explorer le rendement d'une évaluation clinique et d'un bilan étiologique très complet dans les TSA. Deux manuscrits sont en cours de rédaction.

Par ailleurs, notre équipe a participé, en tant que cohorte de réplication, à une étude dont le but était d'étudier *TMLHE*, gène lié à l'X jouant un rôle dans la biosynthèse de la carnitine, en tant que facteur de risque dans les TSA. Bien que les résultats n'aient pas montré d'évidence claire en faveur de l'implication de ce gène dans les TSA, les données suggèrent qu'une dérégulation du métabolisme de la carnitine pourrait jouer un rôle de facteur de risque dans la vulnérabilité aux TSA (**Annexe, Article 2**). J'ai également participé à une étude de notre équipe par micropuces à ADN de 18 patients atteints d'autisme et porteurs d'un remaniement chromosomique apparemment équilibré, dans le but d'identifier une anomalie cryptique associée au réarrangement qui pourrait être responsable du phénotype. Cette analyse a mis en évidence la présence d'une anomalie cliniquement significative chez 4 patients (22 %) liée ou non au remaniement initial. Cet article est en cours de soumission pour publication (**Annexe, Article 3**).

Article 1

# *Convergence of genes and cellular pathways dysregulated in autism spectrum disorders*

Pinto D, **Delaby E**, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, Xu X, Ziman R, Wang Z, Vorstman JA, Thompson A, Regan R, Pilorge M, Pellecchia G, Pagnamenta AT, Oliveira B, Marshall CR, Magalhaes TR, Lowe JK, Howe JL, Griswold AJ, Gilbert J, Duketis E, Dombroski BA, De Jonge MV, Cuccaro M, Crawford EL, Correia CT, Conroy J, Conceição IC, Chiocchetti AG, Casey JP, Cai G, Cabrol C, Bolshakova N, Bacchelli E, Anney R, Gallinger S, Cotterchio M, Casey G, Zwaigenbaum L, Wittemeyer K, Wing K, Wallace S, van Engeland H, Tryfon A, Thomson S, Soorya L, Rogé B, Roberts W, Poustka F, Mouga S, Minshew N, McInnes LA, McGrew SG, Lord C, Leboyer M, Le Couteur AS, Kolevzon A, Jiménez González P, Jacob S, Holt R, Guter S, Green J, Green A, Gillberg C, Fernandez BA, Duque F, Delorme R, Dawson G, Chaste P, Café C, Brennan S, Bourgeron T, Bolton PF, Bölte S, Bernier R, Baird G, Bailey AJ, Anagnostou E, Almeida J, Wijsman EM, Vieland VJ, Vicente AM, Schellenberg GD, Pericak-Vance M, Paterson AD, Parr JR, Oliveira G, Nurnberger JI, Monaco AP, Maestrini E, Klauck SM, Hakonarson H, Haines JL, Geschwind DH, Freitag CM, Folstein SE, Ennis S, Coon H, Battaglia A, Szatmari P, Sutcliffe JS, Hallmayer J, Gill M, Cook EH, Buxbaum JD, Devlin B, Gallagher L, Betancur C*, Scherer SW*

*Am J Hum Genet* (2014) 94(5), 677-694

*These authors contributed equally to this work

**Article 1. Convergence des gènes et des voies biologiques dérégulés dans les TSA**

La forte contribution des CNVs rares dans l'étiologie des TSA a été montrée et des centaines de gènes et loci ont été impliqués. Néanmoins, nos connaissances de l'architecture génétique des TSA restent très incomplètes puisque les causes connues n'expliquent l'étiologie que chez environ 25 % des cas, indiquant qu'il reste encore de nombreux gènes et loci à identifier. Dans cette étude, le consortium international *Autism Genome Project* (AGP), a réalisé une analyse par micropuces Illumina à très haute résolution sur l'échantillon de familles le plus grand jamais réuni pour ce type d'analyse dans les TSA (n = 2 845 familles ayant un enfant ou plusieurs enfants atteints de TSA). Les analyses se sont concentrées sur les CNVs rares, dans le but d'identifier de nouveaux gènes impliqués dans les TSA, de réaliser des diagnostics étiologiques et de mieux comprendre les mécanismes physiopathologiques qui sous-tendent l'apparition de cette pathologie.

Le projet a été divisé en deux stades, chacun incluant environ la moitié des familles. Les résultats du premier stade ont été publiés en 2010 (Pinto et coll., 2010). Dans ce second stade, le reste des familles a été analysé et l'ensemble des données des deux stades a été combiné pour être ré-analysé (n = 2 446 après contrôles de qualité). Grâce aux données de CNVs de 4 768 témoins analysés de la même façon que les familles, les analyses se sont concentrées sur les CNVs rares (<1 % dans la population étudiée). De plus, afin de minimiser le taux de faux-positifs, les analyses ont été restreintes aux CNVs de taille supérieure à 30 kb. Les résultats ont confirmé l'enrichissement en CNVs rares géniques chez les cas comparés aux témoins, notamment pour les CNVs exoniques qui chevauchent des gènes et loci déjà impliqués dans les TSA et/ou la déficience intellectuelle. Des CNVs pathogènes ont été identifiés chez 82 patients, ce qui correspond à un rendement étiologique de 3,3 % pour cette cohorte. En outre, l'analyse des CNVs a permis de mettre en évidence plusieurs gènes impliqués dans les TSA pour la première fois et de nombreux gènes candidats. Parmi les gènes touchés par les CNVs, un enrichissement a été observé chez les cas comparés aux témoins pour les cibles de la protéine FMRP et les gènes de la densité post-synaptique. Un enrichissement limité aux délétions a également été observé pour plusieurs autres listes de gènes candidats : les gènes avec une probabilité élevée d'être haploinsuffisants, les gènes associés à des phénotypes neurologiques chez l'homme ou chez la souris ainsi que les gènes exprimés dans le cerveau. Enfin, des analyses de voies biologiques basées sur les CNVs *de novo* ont montré une convergence des gènes impliqués vers la fonction synaptique, la régulation de la transcription ainsi que le remodelage de la chromatine.

En utilisant les données des micropuces obtenues pour les familles du stade 2 de la cohorte PARIS, mon travail a consisté à analyser les milliers de CNVs détectés dans le but d'identifier ceux qui pourraient être impliqués dans l'étiologie de nos patients. Pour sélectionner les CNVs potentiellement pathogènes, j'ai utilisé différents critères comme le fait qu'un CNV soit *de novo*, lié au chromosome X et maternellement transmis à un garçon, qu'il soit de grande taille, qu'il touche un gène ou loci déjà impliqué dans les TSA et/ou la déficience intellectuelle, qu'il soit récurrent chez plusieurs patients, qu'il soit rare voire absent chez les témoins ou encore qu'il touche un gène candidat par sa fonction ou qu'il soit impliqué dans une voie neurobiologique déjà mise en cause dans l'autisme. J'ai ensuite validé ces CNVs potentiellement pathogènes par PCR quantitative et étudié la ségrégation dans la famille. J'ai également validé des CNVs potentiellement pathogéniques identifiées chez les patients d'autres pays qui ne pouvaient pas réaliser ces analyses eux-mêmes. Parmi les CNVs validés, nous avons identifié plusieurs CNVs rares *de novo* ou liés à l'X touchant des gènes candidats par leur fonction et qui pourraient jouer un rôle dans l'étiologie de l'autisme. J'ai ensuite étendu la recherche de CNVs pathogènes à toutes les familles du consortium, pour les deux stades, ce qui correspond à environ 130 000 CNVs chez 2 446 patients. En appliquant ces mêmes critères, nous avons identifié de nombreux CNVs rares, dont certains chevauchant des gènes déjà impliqués dans les TSA et/ou la déficience intellectuelle (la liste de gènes et

loci impliqués a été créée par notre équipe) (**Tableaux S6A-D et Tableau S7A**). Nous avons analysé individuellement chacun de ces CNVs pour les séparer en différentes catégories : pathogènes, de signification clinique inconnue et bénins (**Figure 1 et Figure S1**). Puis, nous avons analysé les autres CNVs rares, notamment ceux qui étaient de grande taille, ceux de taille inférieure à 30 kb touchant un gène d'intérêt et validés expérimentalement et ceux qui étaient *de novo* ; et ainsi identifié plusieurs CNVs pathogènes additionnels. Au total, nous avons identifié 84 CNVs pathogènes chez 82 patients, correspondant à 38 causes différentes et soulignant l'extrême hétérogénéité étiologique des TSA (**Figure 2 et Tableau S7B**). Parmi les CNVs délétères identifiés, plusieurs correspondent à des syndromes génétiques bien définis (syndrome de Phelan-McDermid, syndrome de Kleefstra et syndrome de Smith-Magenis, par exemple), d'autres correspondent à des CNVs récurrents associés aux TSA et à d'autres troubles neurodéveloppementaux (délétions et duplications 16p11.2, duplications 15q11-q13 et délétions 15q13.3, par exemple). Le rendement étiologique faible (3,3 %) s'explique d'une part par le fait que la cohorte d'AGP soit composée en majorité de patients non syndromiques, dont 54 % ne présentent pas de retard intellectuel ; et d'autre part par le fait que plusieurs sites (dont l'étude PARIS) avaient déjà recherché les principaux syndromes de microdélétion/microduplication chez leurs patients par FISH ou MLPA et avaient exclus ceux pour lesquels une cause avait été identifiée.

Parmi les CNVs pathogènes identifiés, nous avons observé qu'un nombre substantiel d'entre eux étaient hérités (36 %), en particulier dans le groupe des CNVs à pénétrance incomplète et expressivité variable, soulignant la contribution des variants rares hérités au risque de développer un TSA. En analysant le sexe des patients porteurs de CNVs pathogènes, nous avons remarqué que la proportion de filles est significativement augmentée parmi les porteurs de CNVs à forte pénétrance alors qu'aucune différence n'est observée pour les CNVs associés à une pénétrance incomplète et/ou une expressivité variable (**Figure 2**). Ces résultats soutiennent l'hypothèse selon laquelle il existerait des facteurs protecteurs spécifiques chez les filles, qui devraient donc présenter des altérations génétiques plus délétères pour être atteintes. De plus, les données cliniques des patients porteurs de CNVs pathogènes que nous avons récoltées des différentes équipes participantes (**Tableau S8**) ont montré que 83 % avaient un TSA non syndromique et que plus de la moitié d'entre eux ne présentaient pas de déficience intellectuelle (**Figure 2**). Bien qu'il soit communément admis que les patients pour lesquels une anomalie génétique a été identifiée sont plus fréquemment atteints de déficience intellectuelle syndromique comparés aux patients pour lesquels une étiologie n'a pas encore été identifiée, nos résultats montrent que certains CNVs pathogènes peuvent causer un TSA sans déficience intellectuelle et sans phénotype syndromique.

Nous avons également présenté quelques exemples détaillés de gènes touchés par des CNVs rares chez les patients. Ainsi, dans la section ***Highlighted genes*** (**Figures S6-S18, pages 12-34**), nous présentons plusieurs cas de CNVs pathogènes touchant des gènes récemment impliqués dans d'autres troubles neurodéveloppementaux et pour lesquels les évidences rassemblées nous permettent de les impliquer pour la première fois dans les TSA (*CHD2*, *GDI1* et *HDAC4*), ainsi que des arguments suggérant l'implication de plusieurs autres gènes dans l'étiologie des TSA. Parmi ces gènes, certains montrent des évidences convaincantes de leur implication dans l'étiologie des TSA et de la déficience intellectuelle, comme le gène *SETD5* et plusieurs gènes sont candidats incluant *SYNCRIP*, *LSAMP*, *TRIP12* et *MIR137*.

Par ailleurs, nous avons réalisé une méta-analyse de 17 gènes et loci touchés par des CNVs rares dans trois grandes cohortes de patients atteints de TSA dont la cohorte AGP (n = 5 106), montrant un enrichissement significatif de plusieurs de ces remaniements chez les cas comparés aux témoins (**Tableau S9**). Les délétions de NRXN1 et de 16p11.2 sont les plus fréquentes, retrouvées chez 0,3 % des patients ; la duplication 15q11-q13 est quant à elle détectée chez 0,25 % des cas. Enfin, nous avons participé à la réalisation des tableaux contenant tous les CNVs rares *de novo* (**Tableau S14**), les anomalies

chromosomiques chez les patients (**Tableau S1C**) et chez les parents et témoins (**Tableau S17B**) ainsi que tous les CNVs validés expérimentalement (**Tableau S17C**).

Cette étude par micropuces à haute résolution, la plus vaste réalisée à ce jour dans un échantillon de patients atteints de TSA, souligne le rôle majeur des CNVs rares dans l'étiologie de l'autisme qu'ils soient *de novo* ou hérités et associés à une pénétrance incomplète et/ou une expressivité variable. L'ensemble des résultats confirme l'extrême hétérogénéité génétique associée aux TSA et la puissance de l'analyse des CNVs dans la recherche étiologique et de nouveaux gènes et voies biologiques impliqués dans les TSA.

**The American Journal of Human Genetics, Volume *94***
**Supplemental Data**

# Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders

**Dalila Pinto, Elsa Delaby, Daniele Merico, Mafalda Barbosa, Alison Merikangas, Lambertus Klei, Bhooma Thiruvahindrapuram, Xiao Xu, Robert Ziman, Zhuozhi Wang, Jacob A.S. Vorstman, Ann Thompson, Regina Regan, Marion Pilorge, Giovanna Pellecchia, Alistair T. Pagnamenta, Bárbara Oliveira, Christian R. Marshall, Tiago R. Magalhaes, Jennifer K. Lowe, Jennifer L. Howe, Anthony J. Griswold, John Gilbert, Eftichia Duketis, Beth A. Dombroski, Maretha V. De Jonge, Michael Cuccaro, Emily L. Crawford, Catarina T. Correia, Judith Conroy, Inês C. Conceição, Andreas G. Chiocchetti, Jillian P. Casey, Guiqing Cai, Christelle Cabrol, Nadia Bolshakova, Elena Bacchelli, Richard Anney, Steven Gallinger, Michelle Cotterchio, Graham Casey, Lonnie Zwaigenbaum, Kerstin Wittemeyer, Kirsty Wing, Simon Wallace, Herman van Engeland, Ana Tryfon, Susanne Thomson, Latha Soorya, Bernadette Rogé, Wendy Roberts, Fritz Poustka, Susana Mouga, Nancy Minshew, L. Alison McInnes, Susan G. McGrew, Catherine Lord, Marion Leboyer, Ann S. Le Couteur, Alexander Kolevzon, Patricia Jiménez González, Suma Jacob, Richard Holt, Stephen Guter, Jonathan Green, Andrew Green, Christopher Gillberg, Bridget A. Fernandez, Frederico Duque, Richard Delorme, Geraldine Dawson, Pauline Chaste, Cátia Café, Sean Brennan, Thomas Bourgeron, Patrick F. Bolton, Sven Bölte, Raphael Bernier, Gillian Baird, Anthony J. Bailey, Evdokia Anagnostou, Joana Almeida, Ellen M. Wijsman, Veronica J. Vieland, Astrid M. Vicente, Gerard D. Schellenberg, Margaret Pericak-Vance, Andrew D. Paterson, Jeremy R. Parr, Guiomar Oliveira, John I. Nurnberger, Anthony P. Monaco, Elena Maestrini, Sabine M. Klauck, Hakon Hakonarson, Jonathan L. Haines, Daniel H. Geschwind, Christine M. Freitag, Susan E. Folstein, Sean Ennis, Hilary Coon, Agatino Battaglia, Peter Szatmari, James S. Sutcliffe, Joachim Hallmayer, Michael Gill, Edwin H. Cook, Joseph D. Buxbaum, Bernie Devlin, Louise Gallagher, Catalina Betancur, and Stephen W. Scherer**

# CONTENTS

## SUPPLEMENTAL FIGURES


Figure S1. CNV burden in genes and loci implicated in ASD and ID in affected and control subjects of all ancestries ..... 3
Figure S2. Projection of the glutamergic (A) and cholinergic (B) synapse subclusters enriched in cases over controls, onto the KEGG pathways ..... 4
Figure S3. Candidacy tier for *de novo* CNV genes selected by NETBAG ..... 5
Figure S4. Expression profile for 113 *de novo* CNV genes selected by NETBAG in the neocortex across 12 developmental stages ..... 8
Figure S5. Overlap among gene lists pre-DAPPLE analysis ..... 10


## HIGHLIGHTED GENES


Figure S6. CNV and SNV identified in *CHD2* in chromosome 15q26.1 ..... 12
Figure S7. CNV and SNV identified in *HDAC4* in chromosome 2q37.3 ..... 14
Figure S8. Recurrent duplications at Xq28 including *GDI1* ..... 16
Figure S9. CNV and SNV identified in *SETD5* in chromosome 3p25.3 ..... 18
Figure S10. CNV identified in *LSAMP* in chromosome 3q13.31 ..... 21
Figure S11. CNV identified in *SH3KBP1* in chromosome Xp22.12 ..... 22
Figure S12. CNV and SNV identified in *TRIP12* in chromosome 2q36.3 ..... 23
Figure S13. CNV and SNV identified in *SYNCRIP* in chromosome 6q14.3 ..... 24
Figure S14. CNV identified in *DTNA* in chromosome 18q12.1 ..... 26
Figure S15. CNV overlapping *MIR137* in chromosome 1p21.3 ..... 28
Figure S16. CNV identified in *PIK3CB* in chromosome 3q22.3 ..... 30
Figure S17. CNV and SNV identified in *HDAC9* in chromosome 7p21.1 ..... 32
Figure S18. CNV identified in the distal 16p11.2 region containing *SH2B1* ..... 33


## SUPPLEMENTAL TABLES


Table S1A. Autism strict and spectrum classifications ..... 35
Table S1B. Quality control – Family and control sample breakdown ..... 36
Table S1C. Quality control – Chromosome abnormalities detected in probands ..... 38
Table S2A. Sample characteristics ..... 39
Table S2B. Sample characteristics (*continued*) ..... 40
Tables S3A-S3C. CNV burden ..... 41
  Table S3A. Platform comparison ..... 41
  Table S3B. Stage 1 (Pinto et al.) versus Stage 2 (new cases) versus Combined (all 2147 European cases) ..... 42
  Table S3C. Characteristics of rare CNVs in 2,147 European ASD probands and 2,640 European controls ..... 43
Table S4. Rare *de novo* CNVs in probands confirmed experimentally ..... 44
Table S5A. Parent of origin for rare *de novo* validated CNVs in probands ..... 46
Tables S5B-S5H. Parent of origin of *de novo* CNVs – breakdown by family type and CNV characteristics ..... 48
Tables S6A-S6D. List of genes and loci implicated in ASD and ID ..... 50
Table S7A. CNVs overlapping ASD or ID genes and loci in affected and control subjects (all ancestries) ..... 58
Table S7B. Pathogenic CNVs in affected subjects (all ancestries) ..... 61
Table S8. Phenotypes in ASD subjects with pathogenic CNVs or with selected CNVs of uncertain significance ..... 63
Table S9. Meta-analysis of loci and genes affected by rare CNVs in large ASD cohorts ..... 64
Table S10. FMRP targets affected by deletions in probands and not yet implicated in ASD or ID ..... 65
Tables S11A-S11E. Multigene analyses – various models ..... 66
Tables S12A-S12D. GO terms, pathways, and MPO enrichment in affecteds versus control subjects ..... 71
Table S13A. Gene Ontology terms and pathways used to generate a list of neurodevelopmental functions ..... 71
Table S13B. Effect size for neurobiological-related clusters ..... 72
Table S13C. Neuronal synapse main cluster ..... 73
Table S13D. Genes and subjects represented in the enriched cholinergic and glutamergic synapse subclusters ..... 74
Tables S14A-S14E. Characterization of genes selected by NETBAG ..... 75
Table S15. Functional-group enrichment for DAPPLE results ..... 75
Table S16. List of 97 high-confidence CNV/SNV genes ..... 75
Table S17A. Listing of CNV calls in affected subjects ..... 75
Table S17B. Chromosome abnormalities in parents and control subjects ..... 75
Table S17C. Experimentally validated CNVs ..... 75

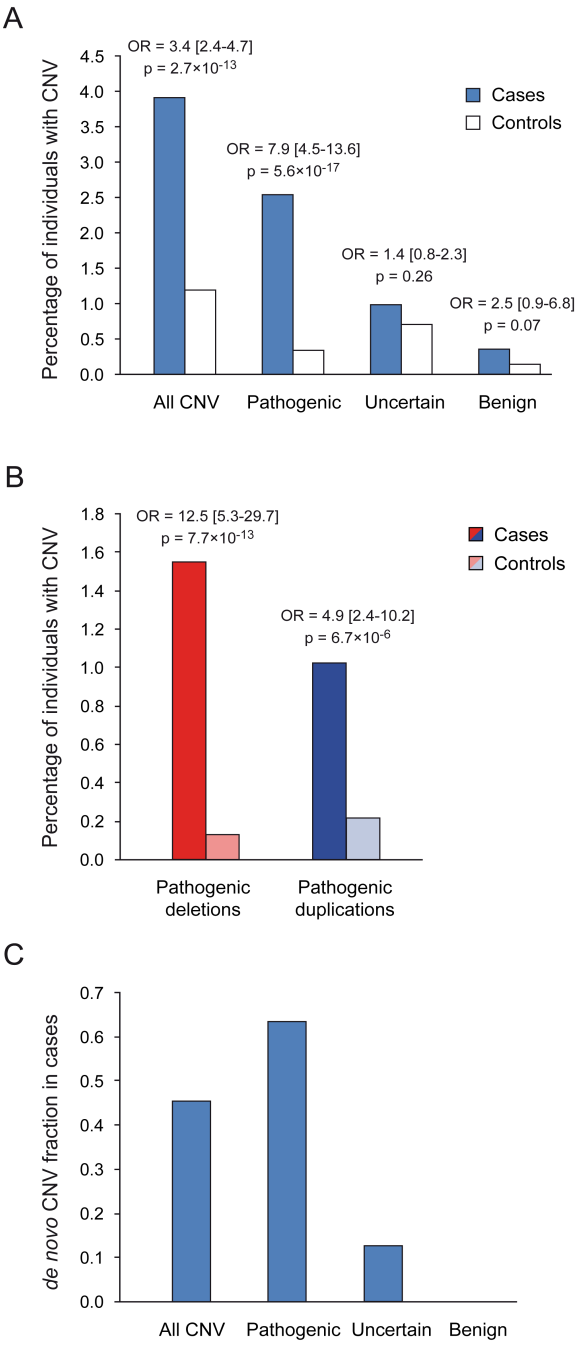

## SUPPLEMENTAL DATAFILES


Table S8. Phenotypes in ASD subjects with pathogenic CNV or with selected CNV of uncertain significance ..... 76
Tables S12A-S12D. GO terms, pathways, and MPO enrichment in affecteds versus control subjects ..... 76
Tables S14A-S14E. Characterization of genes selected by NETBAG ..... 76
Table S15. Functional-group enrichment for DAPPLE results ..... 76
Table S17A. Listing of CNV calls in affected subjects ..... 76
Table S17B. Chromosome abnormalities in parents and control subjects ..... 76
Table S17C. Experimentally validated CNVs ..... 76

**SUPPLEMENTAL ACKNOWLEDGEMENTS** ..... 76

**SUPPLEMENTAL REFERENCES** ..... 77

## SUPPLEMENTAL FIGURES

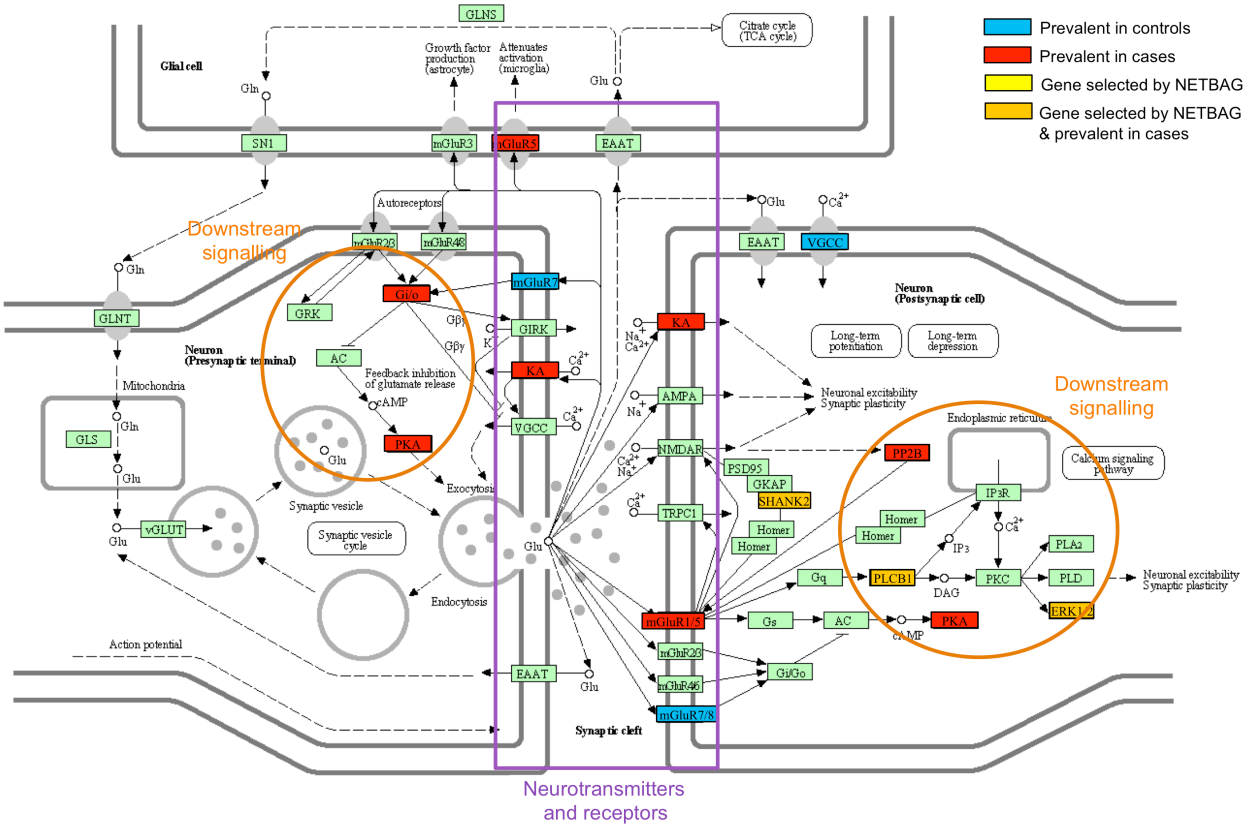



### Figure S1. CNV burden in genes and loci implicated in ASD and ID in affected and control subjects of all ancestries

CNV data from 2,446 cases and 4,768 controls irrespective of ancestry were analyzed for overlaps with a list of genes and loci causally implicated in ASD/ID (**Tables S6A-S6D**). Overlap results include 299 non-European cases and 1,843 non-Europeans controls. Only CNVs affecting autosomal dominant and X-linked dominant genes/loci in both genders (132 genes, 56 loci) as well as X-linked recessive genes/loci in males (52 genes, 2 loci) were considered ('all CNV'). Exonic CNVs ≥30 kb affecting an ASD/ID gene and CNVs overlapping at least 50% of the target loci were selected for further analysis. After curation, CNVs were divided into three categories: pathogenic, uncertain clinical significance or benign. **(A)** Percentage of individuals with CNVs overlapping genes and loci implicated in ASD/ID ('all CNV'), pathogenic, uncertain or benign CNVs and odds ratio (OR) in cases and controls. **(B)** Percentage of individuals with pathogenic deletions or duplications and OR in cases and controls. **(C)** *De novo* CNV fraction in each category in cases.

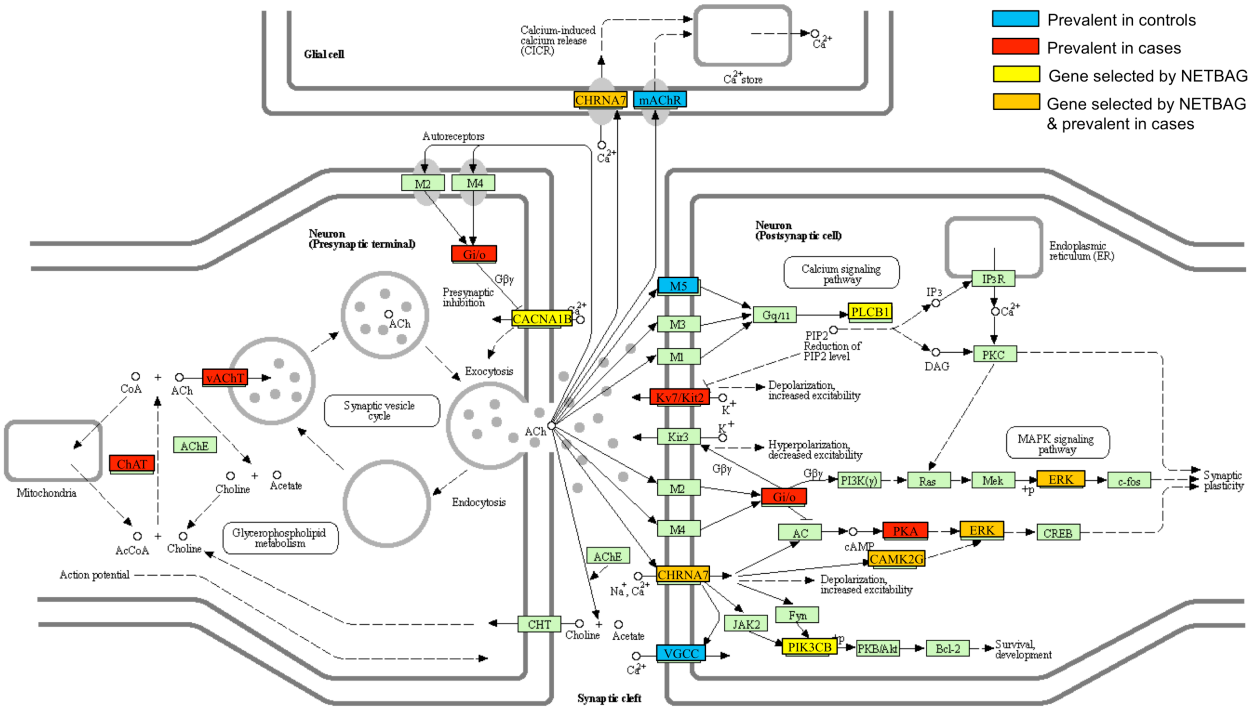
## Figure S2. Projection of the glutamergic (A) and cholinergic (B) synapse subclusters enriched in cases over controls, onto the KEGG pathways

### A. Glutamergic pathway

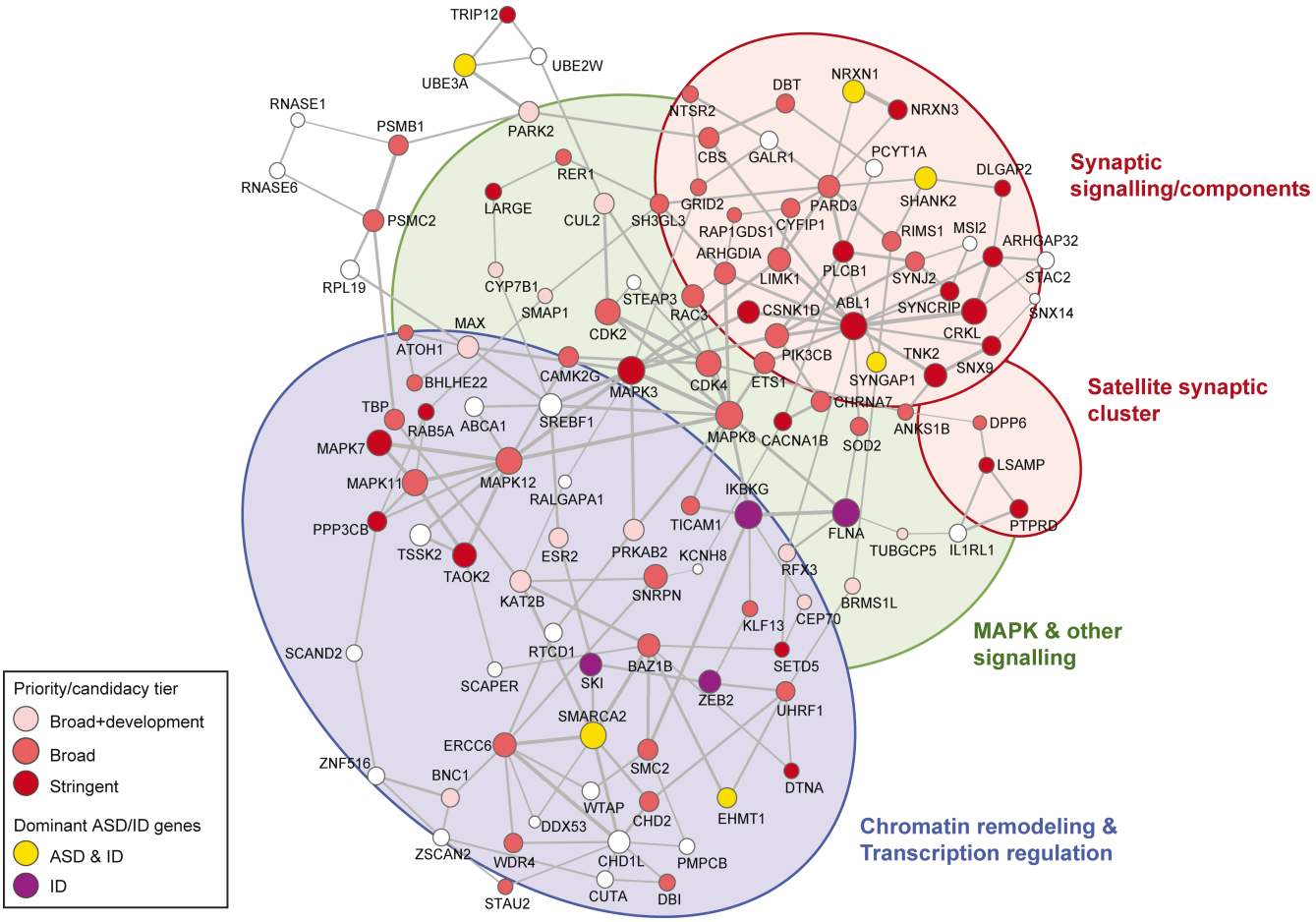



19 cases and 2 controls in the glutamatergic pathway: 11 deleted genes in 19 cases (average size= 635 kb, median size= 265 kb; average number of genes= 16, median number of genes= 3): *GNG13* (1), *MAPK3* (5), *GNG2* (2), *PRKACB* (1), *SHANK2* (3), *PPP3CB* (1), *SHANK3* (3), *GRIK2* (1), *GRM5* (1), *PLCB1* (1), *SHANK1* (1); 2 deleted genes in 2 controls (average/median size= 85 kb; average/median number of genes= 1): *CACNA1C*, *GRM7*. Genes selected by NETBAG (main **Figure 4B**) are highlighted with yellow or orange shaded boxes.

### B. Cholinergic pathway




17 cases and 2 controls in the cholinergic pathway. Genes selected by NETBAG (main **Figure 4B**) are highlighted with yellow or orange shaded boxes




## Figure S3. Candidacy tier for *de novo* CNV genes selected by NETBAG

Genes implicated in dominant forms of ASD and ID or in ID are highlighted in yellow and purple, respectively. Genes implicated in autosomal recessive ID (*ERCC6*, *CBS*, *DBT*, *PLCB1* and *LARGE*) were not given a specific color for ease of comprehension of the figure. Multiple lines of evidence from the literature and dedicated databases (DECIPHER, ISCA and ECARUCA) for the *CHD2* gene (recently involved in other neurodevelopmental disorders), and the candidate genes *SETD5*, *LSAMP*, *TRIP12*, *SYNCRIP*, *DTNA*, and *PIK3CB*, are given in the section '**Higlighted genes**'.

**NETBAG analysis of *de novo* CNV genes:** For this analysis, recurrent events were counted once (or combined in one region) and events that did not intersect any genes were removed; the resulting final set used in the analyses consisted of 75 CNV regions intersecting 874 unique RefSeq genes. Derived gene clusters were scored using the de-weighted method, where the contribution of each individual gene to the overall cluster score is given by the weighted sum of its connections (edges) to the other cluster genes,[1] and significance was obtained by generating random events with the same gene count as observed in the *de novo* CNV dataset with 1,000 randomizations (main **Figure 4B**; **Tables S14A-S14E**). The highest scoring cluster obtained using a searching procedure of up to two genes per CNV (global p-value 0.02) is shown. The list of up to 2 genes selected per CNV is given in **Table S14A**. The majority of the genes identified by NETBAG (92.0%, 104/113) are highly expressed in brain. When using a search procedure of up to one gene per CNV, a list of 69 genes was obtained (global p-value <0.01); this list is given in **Table S14A**. Analysis of established annotation resources, such as PubMed, OMIM, GeneReviews, EntrezGene and iHOP, suggested that a significant fraction of genes in the identified network either play a well-defined functional role in the brain or have been previously implicated in neurological and psychiatric disorders. An overview of functional information about each of the genes forming the cluster can be found in **Table S14B**.

**Prenatal, postnatal or unbiased brain expression of genes selected by NETBAG** (additional information relevant to main **Figure 4B**, based on NETBAG data). Nodes were colored according to whether a gene's brain expression is prenatally or postnatally biased, or has no biased expression in an analysis of 12 developmental stages of the BrainSpan RNA-seq dataset. Genes were considered biased if their mean expression in prenatal developmental stages was increased more than two-fold compared to their post-natal expression, or vice versa. Genes with related function according to Gene Ontology (GO)[2] and KEGG pathways[3] were indicated by shaded areas; a complete listing of over-represented GO terms and pathways is given in **Table S14D**. By further examining these genes (**Table S14C**), we found that 77% (87/113) of them have a strong link with neuronal and brain-specific genes. We explored whether affected genes were biased to certain developmental stages by interrogating the temporal expression of genes selected by NETBAG. We found no overall significant

bias to prenatally expressed genes, but we did observe prenatal bias for a cluster of genes participating in chromatin remodeling and transcription regulation (main **Figure 5B**, **Table S14E**). The prenatally-biased expression profile was characterized by high expression during development and sharp decrease after birth. We further compared expression levels (**Table S14B**) and found that most of the genes are highly expressed in the neocortex. Similar results were obtained when looking at LoF *de novo* single SNVs in 122 genes assembled from four ASD exome-sequencing studies.[4-7]

**Candidacy tier of ASD candidate genes**

To identify potential new ASD targets and gain mechanistic insights, we expanded our analysis to several gene lists described below. These gene lists were used to build a candidacy tier, by identifying which genes are more likely to be implicated in ASD based on membership in the gene lists. These gene lists were used for the analyses presented in main **Figure 3**, **Figure S3** and **Table S14C**. Our analyses focused on exonic events overlapping or disrupting exons, and deletions and duplications were analysed separately. P-values associated with odds ratios were calculated using Fisher exact tests.

1) ***Fragile X mental retardation protein (FMRP) targets*****:** the gene encoding FMRP, *FMR1*, responsible for fragile X syndrome, is mutated in ~2% of ASD cases.[8] Significant overlap between FMRP-targets and ASD candidate genes was reported recently,[9] and Iossifov et al.[4] had previously shown that rare *de novo* LoF SNVs identified in their exome sequencing study were enriched in this class of genes. We used a set of FMRP-RNA interactors identified experimentally (n=842), using crosslinking and immunoprecipitation (CLIP) experiments in mouse.[9] We note that a later paper[10] detected FMPR targets computationally, using human RNA sequence motif analysis (n=939 top genes as provided by the authors), but given the limited overlap between the genes in the two papers (~20%), we decided to use for our enrichment analyses only the list of targets that were experimentally identified by Darnell et al.[9].

2) ***Post-synaptic density genes*****:** The full list of experimentally identified human post-synaptic density genes was collected from Bayes et al. (n=1,453).[11]

3) ***Haploinsufficiency index*****:** We used the predicted haploinsufficiency index (HI) by Huang et al.[12] In particular, we used the imputed predictions as these have better gene coverage, and showed a similar performance compared to the original list of predicted genes pre-imputation when looking at dominant and recessive ASD and ID genes (data not shown). We defined four gene-sets: all genes with any value of HI (17,081 genes), the top 55% HI genes (pHI ≥0.15, 8,862 genes), top 26% (pHI ≥0.35, 4,136 genes), and top 14% (pHI ≥0.55, 2,214 genes). We focused our analysis on deletions because deletions, not duplications, are associated with HI, and show that HI is an important contributor to ASD etiology (duplication breakpoints can also disrupt genes or their regulatory elements and lead to haploinsufficiency).

4) ***Neurodevelopmental/neuropsychiatric phenotypes*****:** Genes associated with neurodevelopmental/neuropsychiatric phenotypes were mined from HPO (Human Phenotype Ontology)[13] for human and MGI/MPO for mouse (Mouse Genome Informatics/Mammalian Phenotype Ontology, The Jackson Laboratory, Bar Harbor, Maine, http://www.informatics.jax.org; data retrieved on November, 2012).[14]
   a) For human, we selected all genes annotated for "Behavioural/Psychiatric abnormality" (HP:0000708) and/or "Cognitive impairment" (HP:0100543), as well as children terms in the ontology. We only considered genes with HPO phenotype annotation derived from OMIM, as for these we could more reliably infer the mode of inheritance based on the first digit of the OMIM ID (1 = autosomal dominant, 2 = autosomal recessive, 3 = X-linked).
   b) For mouse, we selected all genes annotated for "nervous system phenotype" (MP:0003631) and "behavior/neurological phenotype" (MP:0005386), as well as children terms in the ontology; eQTL and complex phenotypes were removed. Autosomal dominant mode of inheritance was inferred by parsing MGI mutation allele tables and retaining only gene-phenotype annotations supported by a heterozygous mutated allele; genes on the human X chromosome were removed to obtain the autosomal dominant subset.

5) ***Neurodevelopmental function*****:** Two neurodevelopmental lists were generated to help evaluate the candidacy of the genes selected by NETBAG: "NeuroF" was purely based on the manual curation of Gene Ontology terms and pathways, and "NeuroF_EM" was based on gene-sets found associated by the logistic regression test in the enrichment map (EM) (results shown in main **Figure 4A**) and clustered as: (a) Cell projection/Neural development/Axonogenesis, or (b) Neuronal synapse. The Gene Ontology terms and pathways used to generate NeuroF are listed in **Table S13A**.

6) ***Developmental genes*****:** Genes were collected from all gene-sets found associated by the logistic regression test and clustered as Development/Cell Proliferation/Cell Motility in the enrichment map ("Dev_EM").

7) ***Brain-expressed genes*****:** Three gene lists were constructed based on the human BrainSpan RNA-seq dataset:
   a) *ExprBspan_BrainAny_log2rpkm3.0*: genes with RPKM ≥3.0 in at least five BrainSpan data-points (a data-point corresponds to any unique combination of donor subject x brain region x developmental time-point).
   b) *ExprBspan_BrainAny_log2rpkm4.5*: genes with RPKM ≥4.5 in at least five BrainSpan data-points.
   c) *ExprBspan_NotBrainExprF*: genes RPKM ≥1.0 not satisfying the above definition, yet annotated for at least one Gene Ontology term or pathway; these are similar in number to ExprBspan_BrainAny_log2rpkm4.5 and therefore were used as a negative control list, representing functionally characterized genes not expressed or expressed at very low levels in brain.

   Brain-expressed gene lists were also defined based on the Novartis expression atlas (HG-U133A Affymetrix arrays), and used only for the definition of candidate gene lists:

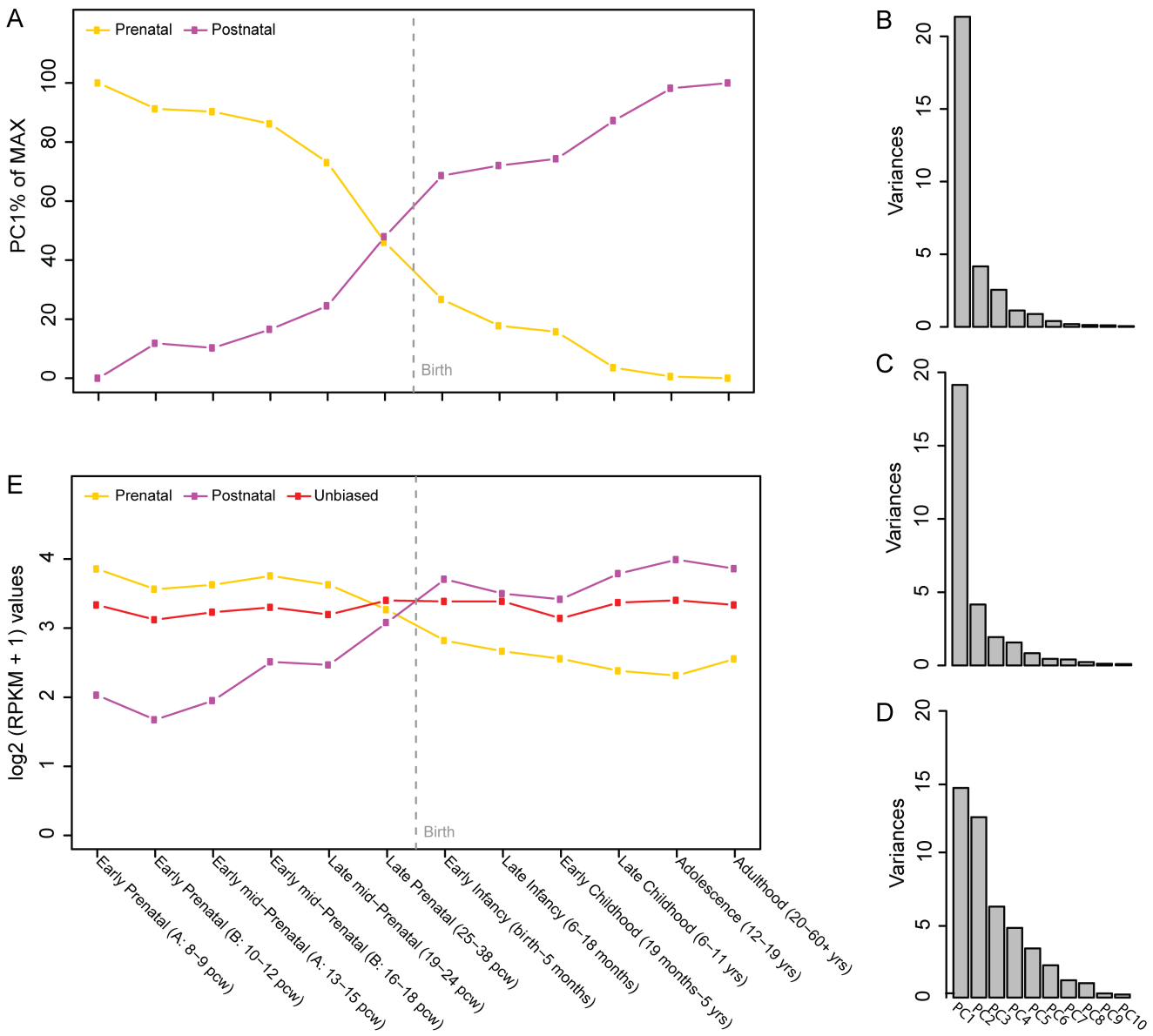
a) *Novartis Brain-expressed genes*: for a gene to be considered brain expressed it needed to have the robust multi-array average (RMA) gene expression values in whole-brain or fetal brain: (i) greater than the array median expression value (considering all tissues), and (ii) greater than the median expression in non-nervous tissues.
b) *Novartis Brain-specific genes*: for a gene to be considered brain-specific it needed to have the RMA gene expression values in whole brain or fetal brain: (i) greater than the array median expression value (considering all tissues), and (ii) greater than the double of the median expression in non-nervous tissues.

Three lists were compiled based on the results from the 1-gene hit odds ratio analysis of single gene-sets and their combinations (main **Figures 3A-B**):

1) Stringent (1,088 genes): clearly linked to synaptic components and/or autosomal dominant neurodevelopmental/neuropsychiatric phenotypes.
2) Broad (5,342 genes): linked to synaptic components, or neurodevelopmental/neuropsychiatric phenotypes, or brain expression.
3) Broad development (7,158 genes): extended by gene-sets corresponding to broader developmental functions found by the logistic regression gene-set test and enrichment map clusters.

In particular, the following definitions were used:

1) Stringent, union of:
   a) FMRP targets from Darnell et al.[9]
   b) Genes found in at least two of these lists:
      i) NeuroPheno Dom
      ii) Intersection of: (a) Novartis Brain-expressed and (b) HI ≥ 0.55
      iii) Postsynaptic density genes, full dataset from Bayes et al.[11]
      iv) FMRP targets from Ascano et al.[10]
2) Broad, union of:
   a) Stringent candidates (as defined above, in Stringent)
   b) Novartis Brain-specific
   c) Intersection of: (i) ExprBspan_BrainAny_log2rpkm4.5 and (ii) HI ≥ 0.35
   d) Intersection of: (i) NeuroF and (ii) HI ≥ 0.35
   e) Intersection of: (i) ExprBspan_BrainAny_log2rpkm3.0 and (ii) NeuroPheno All
3) Broad development, union of:
   a) Broad candidates (as defined above, in Broad)
   b) Intersection of: (i) NeuroF_EM and (ii) NeuroPheno All
   c) Development cluster in the enrichement map (**Figure 4A**, main text)




## Figure S4. Expression profile for 113 *de novo* CNV genes selected by NETBAG in the neocortex across 12 developmental stages

The time course of BrainSpan RNA-seq gene expression data was quantile normalized and log2 transformed. **(A)** The first principle components (PC1, as percentage of the maximum, %/MAX) for the 113 CNV genes are shown in 2 expression groups (prenatally biased, postnatally biased). The amount of variance explained by each principle component is shown for **(B)** Prentatally-biased genes, **(C)** Postnatally-biased genes, and **(D)** Unbiased genes. **(E)** The group centroids (median) of the 3 expression groups (Prenatally biased, Postnatally biased, Unbiased) are plotted at each developmental stage with the Y axis showing log2 transformed expression profiles log2(RPKM + 1). PCW, post-conceptional weeks. Panels B-D compared the variance explained by PC1 for the 3 expression groups. PC1 for Unbiased genes is not representative, i.e., it does not explain more variance than other PCs and it is unable to profile the major direction of the expression group. Therefore we did not include the unbiased group in panel A.

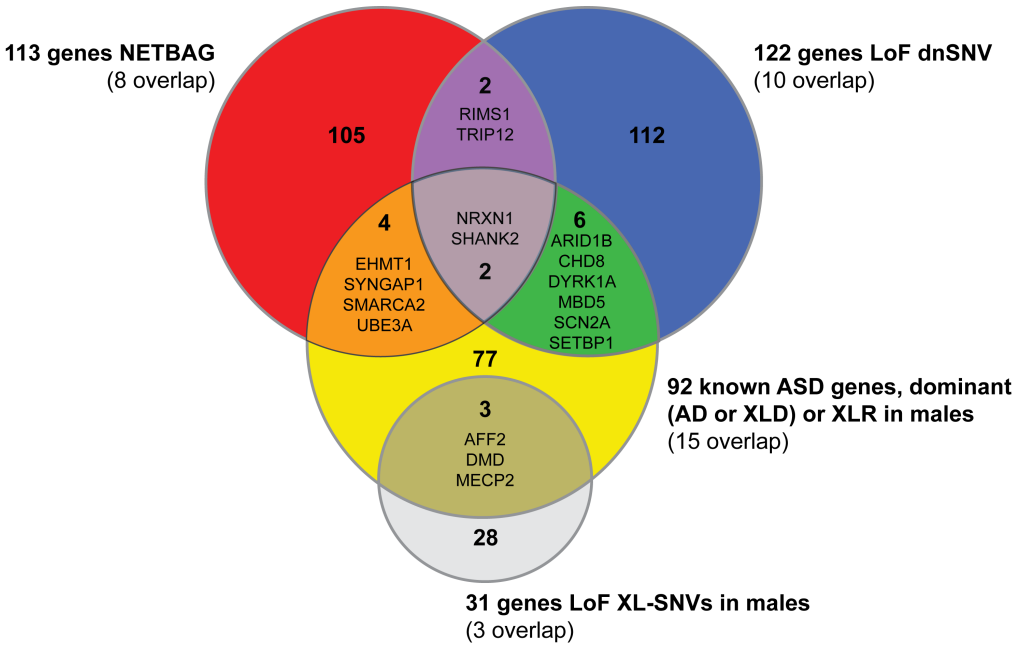
**RNA-Seq data preprocessing and normalization:** The human brain developmental transcriptome data, consisting of a total of 12 developmental stages (from 8-9 post-conceptional weeks [PCW] to 20-60 years) in 524 samples, was downloaded from BrainSpan resource (http://brainspan.org/; Allen Institute for Brain Science, 2012; analyses presented here used the October 2013 version).[15,16] In this dataset, gene expression profiles measured by RNA-sequencing experiments were represented as reads per kilobase of transcript per million mapped reads (RPKM),[17] which were pre-summarized to gene levels based on a composite model defining a gene as the union of all exonic nucleotides across all transcripts. A sample level filtering procedure was implemented after comparison of the distribution of each sample to the entire distribution (all 524 samples); the top 2% outliers were removed based on the sum of chi-squares with visual assistance on boxplot. Subsequently, a gene level filter was applied so that only genes with a RPKM value > 0 in

more than 80% of the samples in any brain region of any developmental stage were kept. To ensure quality of the data, the samples were further filtered by RNA integrity number (RIN) > 8. As a result, 42,965 gene entries and 416 samples passed QC and entered the downstream analysis where quantile normalization and log2 transformation procedures were performed.

**Gene expression level classification:** We then sought to classify genes by their expression levels across multiple brain regions and developmental stages. The genes with lower 20% expression level in all 8 brain regions (neocortex, amygdala, hippocampus, striatum, diencephalon, upper rhombic lip, cerebellar cortex, and dorsolateral prefrontal cortex) were classified as "non-expressed" genes. The remaining 80% genes were classified into 3 expressional categories specifically for two brain regions, the neocortex (NCX) and dorsolateral prefrontal cortex (DFC): 1) lowly expressed (bottom 25% averaging across NCX or DFC samples); 2) highly expressed (top 25% averaging across NCX or DFC samples); and 3) the remainder of genes were considered moderately expressed for NCX or DFC regions.

**Identification of differentially expressed genes:** We analysed the 113 genes resulting from the NETBAG functional network analysis of 102 *de novo* CNVs, and determined their prenatal-biased or postnatal-biased status by a nonparametric algorithm, similar to the method used in Xu et al.[18] Specifically, for each of the 8 brain regions, we used the median of samples to represent the expression level of a certain given gene under a specific developmental stage. Subsequently, two groups were defined: prenatal (periods 2A, 2B, 3A, 3B, 4, 5) and postnatal (periods 6 through 11). Wilcoxon rank sum test with a p value cut-off of 0.05 and group median fold change cut-off of 1 were applied to determine differentially expressed genes between prenatal and postnatal groups. Genes with significantly higher expression in prenatal or postnatal group were considered prenatally or postnatally biased genes, while the remaining genes were classified as unbiased. Interestingly, we observed that some genes (e.g. *KCNH8*, *CRKL*) presented a peculiar profile: a "U shape", i.e., high expression in the first developmental stages followed by a decrease in expression and later, towards the last developmental stages, a new increase in expression. If there was a predominance of high expression in the left or in the right branch of the "U" curve, the genes were classified as prenatally-biased or postnatally-biased, respectively. The inverse could be observed for genes with a profile resembling an inverted U (e.g. *RAC3*).

**Profile of gene expression across 12 developmental stages:** Expression profile across 12 developmental stages of the 113 genes resulting from the NETBAG functional network categorized into 3 groups (prenatal, postnatal and unbiased) was plotted for 8 brain regions after Principle Component Analysis was performed with a custom R script. The first principle component of each expressional group was plotted in **Figure S4** for the neocortex region. Of note, the chromatin/transcription module in **Figures 4B** and **S4** showed a predominance of genes with a prenatally-biased expression profile, suggesting that besides the function of the altered gene, the timing of the effect of the genetic perturbation may also be of critical importance in determining disease risk.




## Figure S5. Overlap among gene lists pre-DAPPLE analysis

For the protein-protein interaction (PPI) network analyses presented in **Figures 5A and S5**, we used the Disease Association Protein-Protein Link Evaluator (DAPPLE)[19] under the assumption that causal genetic variants affect a limited number of mechanisms that are detectable by PPIs. DAPPLE takes a list of genes or loci as input and searches for significant physical connectivity among proteins encoded by those genes based on protein-protein interactions available in the InWeb database.[20] DAPPLE will then assess the statistical significance of several network connectivity parameters as well as of the connectivity of individual proteins to other seed proteins using a within-degree node-label permutation method. The Venn diagram shows the overlap between four lists used as input in DAPPLE (i.e. pre-DAPPLE):

1) 113 genes affected by *de novo* CNVs in AGP probands selected by NETBAG.
2) 122 genes with *de novo* LoF SNVs from four published whole exome sequencing studies comprising 965 ASD cases.[4-7] LoF variants included nonsense, frameshift, and splice site-disrupting mutations. One of the 122 genes is X-linked (*RPS6KA3*), and the remainder are autosomal. Three genes were not present in the InWeb database and were thus not included in the analyses (*DHRS4L1*, *PCDHA13*, and *RAD21L1*).
3) 92 genes known to be involved in ASD (list updated from Betancur, 2011)[21] (**Table S6A**); only autosomal dominant (AD) or X-linked (XL) genes were included (XL dominant, XLD, and XL recessive, XLR), since heterozygous CNVs affecting genes involved in autosomal recessive disorders are not deleterious *per se*.
4) 31 X-linked genes with hemizygous LoF SNVs in male ASD cases and not observed in male controls.[22]

Of the 336 genes analysed (339 unique genes pre-DAPPLE minus 3 not present in the Inweb database), 17 genes overlap between two or more lists. Specifically, the *de novo* CNV genes selected by NETBAG overlap with the list of genes with *de novo* LoF SNVs and with the list of genes known to be involved in ASD; two of the genes (*NRXN1*, *SHANK2*) were shared by the three groups. There is no overlap between the 113 genes selected by NETBAG and the X-linked genes with LoF SNVs. Likewise, the Venn diagram post-DAPPLE depicted in main **Figure 5A** shows the overlap among the 151 genes selected by DAPPLE. It resembles the pre-DAPPLE Venn diagram since the relationship between the four gene lists was maintained, including the sharing of two genes (*NRXN1*, *SHANK2*) by the three main lists of genes. Over-representation analysis identified convergent functional themes related to neuronal development and axon guidance, signalling pathways, and chromatin/transcription regulation. Moreover, 90.7% (137/151) of the gene products reside within the same interconnected cluster.

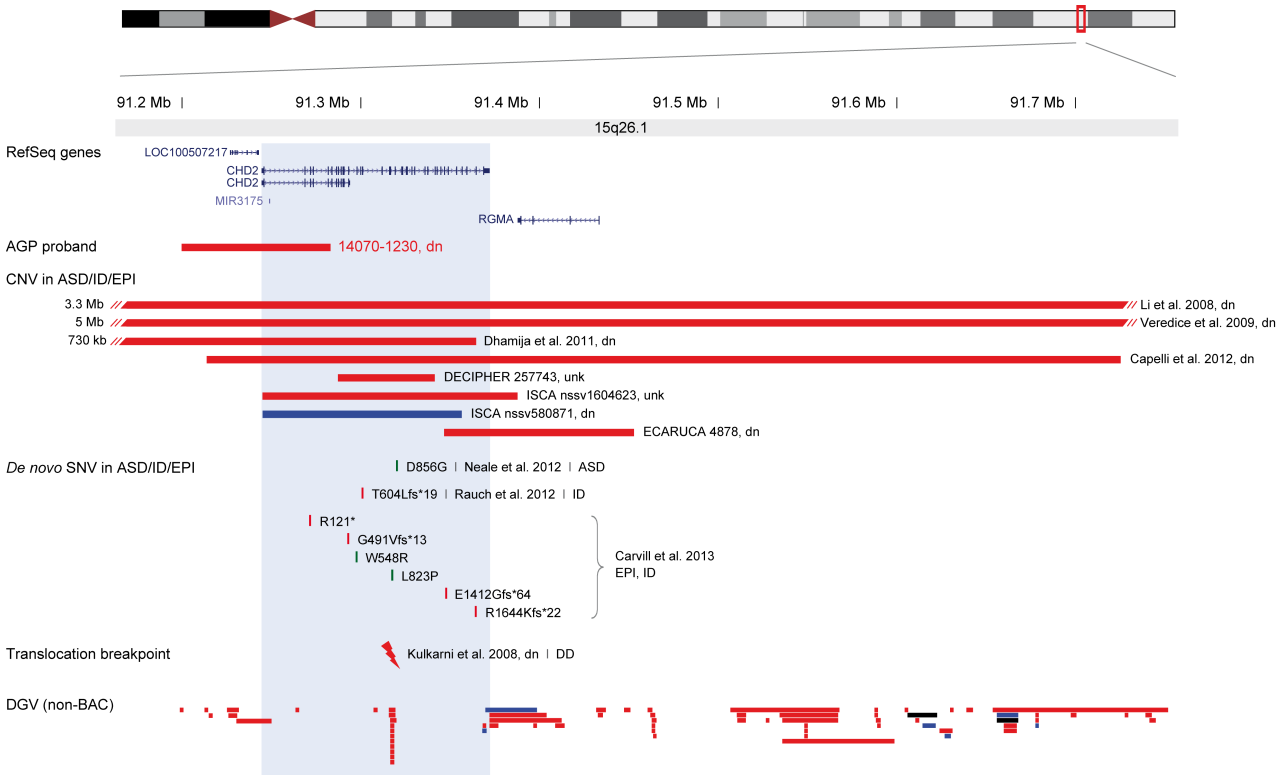
We then further analysed the 113 genes selected by NETBAG and the DAPPLE output network of 151 proteins in terms of brain expression during development and in terms of functional groups (**Table S14E** and **S15**, respectively). Regarding the 113 gene list, we observed that the functional group composed of chromatin remodelers and transcription factors is significantly enriched for prenatal-biased genes and that the functional group composed of signalling genes is enriched for postnatal-biased genes. Similarly, for the DAPPLE output network, the chromatin remodelers/transcription factors functional group is significantly enriched for prenatal-biased genes, with a smaller p-value than the one estimated for the same functional group on the 113 genes list. Interestingly, in this case the functional group composed of neuronal genes is significantly enriched for postnatal-biased genes.

**Phylop and GERP scores:** Basewise conservation scores (PhyloP) across 45 vertebrate genomes (including the human genome) were downloaded from the UCSC website (http://genome.ucsc.edu/)[23] and processed with reference to the UCSC gene track (known genes). Using only the exome, the average base score was calculated. When different variants were present in a gene, all isoforms were aggregated and the average score was used. Likewise, the Genomic Evolutionary Rate Profiling (GERP++) score was downloaded from

Sidow's Lab website (http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html). GERP score was calculated by taking the average basewise score of a gene. Both Phylop and GERP++ scores of genes were ranked. The ranking percentage of the scores (e.g. X% meaning X% of genes in the entire genome are ranked behind the given gene in terms of GERP or Phylop scores) are calculated to characterize each of the genes. Genes in the DAPPLE network (main **Figure 5B**) are among the top 75% more conserved genes compared to the genome average (based on GERP scores and PhyloP), and are typically long (median 103 kb *vs*. 10 kb for the genes in whole genome). Functionally significant and highly conserved genes tend to be more central in physical protein-protein and regulatory networks.[24]

# HIGHLIGHTED GENES

## Genes involved recently in other disorders




## Figure S6. CNV and SNV identified in *CHD2* (chromodomain helicase DNA binding protein 2) in chromosome 15q26.1

Chr15:91,165,000-91,750,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; DD, developmental delay; EPI, epilepsy; ID, intellectual disability; unk, unknown inheritance.

### 1) AGP

- Proband 14070-1230, male, *CHD2* exonic deletion (chr15:91200007-91283004), *de novo*. The deletion also involves *LOC100507217* and *MIR3175*. The affected younger brother also carries the deletion, absent from both parents. SNP genotyping indicated that the deletion occurred in the paternal chromosome, suggesting paternal germline mosaicism. Phenotype: index case, 14 yo, autism (meets criteria both on ADI-R and ADOS), walked at 14 mo, first words 24 mo, phrases 26 mo, currently verbal, ID (Griffiths: language DQ 68, performance DQ 46, global DQ 53); normal growth, normal neurological exam; minor dysmorphic features: micrognatia, protruding ears; no epilepsy; brain MRI: altered angular gyrus (normal variant, unknown pathological significance). Brother, 12 yo, mild autism on ADOS, walked at 15 mo, first words 24 mo, phrases 36 mo, verbal, ID (Griffiths: language DQ 65, performance DQ 61, global DQ 51); normal growth, normal neurological exam; minor dysmorphic features similar to his brother (micrognatia, protruding ears); epilepsy, onset 9 y, partial with secondary generalization, difficult to control with carbamazepine; EEG showed paroxistic activity in the left temporal lobe.
- No *CHD2* deletions among 4,768 controls and 4,875 parents.

### 2) Other human genetic evidence

**CNV:**

- Li et al.[25]: Male with t(15;22)(q26.1;q11.2) translocation and a 3.3 Mb deletion encompassing *CHD2* (chr15:89197342-92489641) at the 15q breakpoint, *de novo*. Phenotype: developmental delay, walked at 20 mo, language delay, developmental motor coordination disorder, height and head circumference consistently ≤3$^{rd}$ centile, weight ≤5$^{th}$ centile, delayed bone age (6-mo-old bone age at 18 mo of chronological age, and 18-mo-old bone age at 36 mo chronological age), mild dysmorphic features (anteverted nares, unilateral auricular pit, fetal fingertip pads, low posterior hairline, and back hirsutism), left eye amblyopia corrected surgically at age 2. He had two episodes of febrile seizures. *IGF1R* was translocated to chromosome 22 and showed 50%

reduction of expression, which could be responsible for the growth deficiency. Angelman syndrome was ruled out by methylation analysis.

- Veredice et al.[26]: Female, 5 Mb deletion encompassing *CHD2* (chr15:87796000-92700000), *de novo*. Phenotype: severe intractable myoclonic epilepsy with photosensitivity, with onset at 6 mo, ID, growth delay, peculiar facial features and minor physical anomalies. Born at 37 weeks, weight 3rd centile, length 10th centile, head circumference 2nd centile. At 21 mo her weight was 3rd centile, with length and head circumference <<3rd centile. She had congenital hypothyroidism, bicuspid aortic valve, diffuse hypotonia, ligamentous laxity, and dysmorphic features (upward slanting eyes, epicanthal fold, depressed nasal bridge, full cheeks, prominent lips, downturned corners of the mouth, protruding tongue, large ears with anteverted lobe, single palmar crease, increased 1st–2nd interdigital space and redundant nuchal skin). Her language was limited to a few words and she had mild developmental delay (Griffith's developmental quotient 67). Brain MRI showed cerebellar vermis hypoplasia with mega-cisterna magna.
- Dhamija et al.[27]: Female, deletion encompassing *CHD2*, with a minimum size of 731 kb (chr15:90633409-91364628) and a maximum size of 936 kb (chr15:90530351-91466733), *de novo*. Phenotype: 9 yo girl with developmental delay, ASD, growth delay, and intractable generalized epilepsy, with onset at 3.5 y. Initially the seizures were partial complex but later became generalized; most prominent were absence seizures, occurring many times a day. She had delayed motor development (walked at 24 mo) and delayed language development with echolalia and mild generalized cognitive delay upon formal testing. On examination, she had mild dysmorphic features (widely set eyes, bilateral pits on the helix of the ears, and crowded teeth with prominent incisors), short stature and head circumference <3rd centile. Brain MRI was normal.
- Capelli et al.[28]: Female, 511 kb deletion encompassing 4 genes, *CHD2*, *LOC100507217*, *MIR3175* and *RGMA* (chr15:91213864-91724860), *de novo*. *RGMA* (repulsive guidance molecule, member A) exerts a negative control on axon growth and could also potentially contribute to the phenotype. Phenotype: when examined at the age of 6 y, she presented global developmental delay, epilepsy, autistic behavior, severe speech impairment with minimal use of words, short attention span, facial dysmorphisms (prognathia, wide mouth, short and widely spaced teeth, strabismus), gait ataxia with uplifted arms and hand flapping, slight hypotonia and hyperactive deep tendon reflexes. Weight, height and head circumference were at the 50th, 75th and 10th centiles respectively. Angelman syndrome was suspected but the methylation pattern at *SNRPN* was normal. Seizures started at the age of 24 mo and were partially controlled with valproic acid. (Corresponds to subject 249888 in DECIPHER.)
- DECIPHER[29]: Subject 257743, male, *CHD2* intragenic deletion (chr15:91287269-91341234), inheritance unknown. Phenotype: autism, macrocephaly, proportionate short stature, scoliosis. This individual also has a 16p11.2 deletion syndrome, inheritance unknown.
- ISCA[30]: Subject nssv1604623, unknown gender, deletion encompassing 3 genes, *CHD2*, *MIR3175* and *RGMA* (chr15:91245098-91387818), inheritance unknown. Phenotype: seizures.
- ISCA: Subject nssv580871, unknown gender, intragenic *CHD2* duplication (chr15:91245028-91356680), *de novo*. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ECARUCA[31]: Subject 4878, male, deletion encompassing only 2 genes, *CHD2* and *RGMA* (chr15:91346585-91452586), *de novo*. Phenotype: ID, truncal obesity, large ear lobules, bulbous nasal tip, low posterior/trident hairline, cryptorchid testes, small hands and feet, tapering fingers, hypotonia, seizures, abnormal EEG.

**Translocation breakpoint:**

- Kulkarni et al.[32]: Female proband with a balanced *de novo* translocation t(X;15)(p22.2;q26.1); the 15q26.1 breakpoint disrupts *CHD2*, whereas the Xp22.2 breakpoint lies in a gene-free region. Phenotype: developmental delay, learning problems, scoliosis, hirsutism, high arched palate, normal face (including eyes and ears), syndactyly of the toes, height <30th centile and head circumference <25th centile. No seizures reported at age 17 y.

**SNV:**

- Carvill et al.[33] (targeted sequencing in epileptic encephalopathies): 6 individuals (4 males and 2 females) with *de novo* mutations in *CHD2* (3 frameshift, 1 stop, 2 missense). Phenotype: 2 had myoclonic atonic epilepsy, 3 had epileptic encephalopathy not otherwise specified and 1 had Lennox-Gastaut syndrome. Seizure onset occurred between 1-3 y; myoclonic seizures were present in all six, photosensitivity in three. All had ID (2 moderate, 4 severe), 1 had ASD.
- Neale et al.[5] (exome sequencing in ASD): Subject 10C100480, male, *CHD2* missense variant, *de novo*. The variant is predicted to be either benign by PolyPhen2[34] and PANTHER,[35] or damaging by SIFT,[36] SNAP,[37] and Mutation Taster;[38] the affected residue is highly conserved: GERP 5.48, [39] ConSurf 6/9.[40] Phenotype: ASD (no other information provided).
- Rauch et al.[41] (exome sequencing in nonsyndromic sporadic ID): Subject MS134, female, *CHD2* frameshift mutation, *de novo*. Phenotype: severe nonsyndromic sporadic ID (IQ 50-69), absence seizures (onset at 5 y), Duane anomaly, no ASD.

### 3) Functional evidence

- *CHD2* encodes a member of the chromodomain helicase DNA-binding (CHD) family of proteins, characterized by the presence of chromodomains (chromatin organization modifier) and SNF2-related helicase/ATPase domains. CHD proteins alter gene expression by modification of chromatin structure, playing critical roles during development.

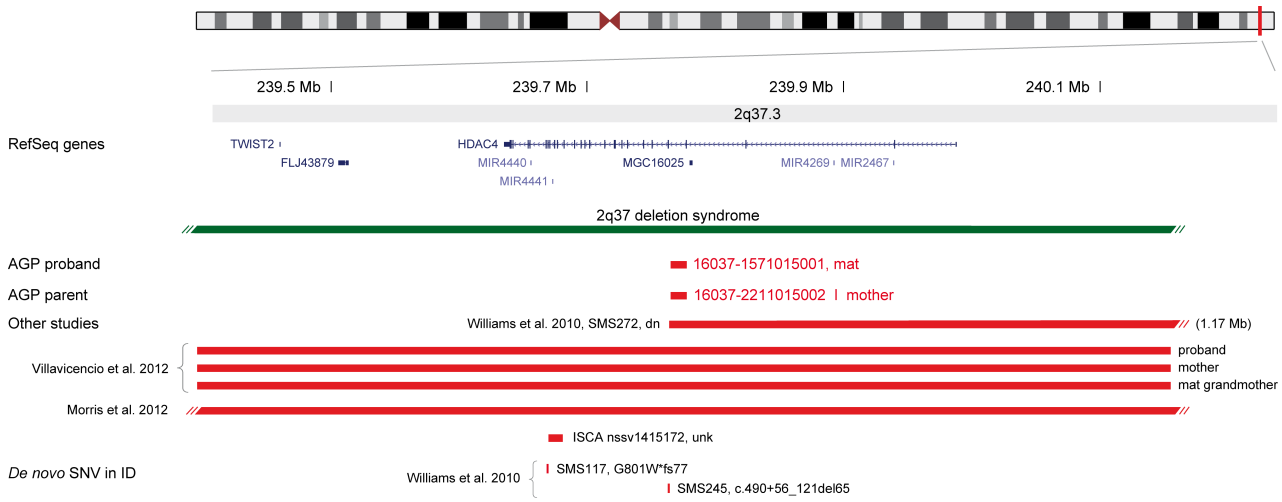
- Expressed in the brain.
- *Chd2* knockout in mice results in embryonic lethality; heterozygous mice have variable postnatal lethality and growth retardation, with reduced body fat, pronounced lordokyphosis, renal disease and anemia in adults.[42,43]

### 4) Other evidence

- % haploinsufficiency (HI) = 15.8 (likely to be haploinsufficient).[12]
- Mutations in *CHD7* cause CHARGE (coloboma, heart defects, atresia of the choanae, retardation of growth and developmental, genital and/or urinary defects and ear abnormalities) syndrome, associated with syndromic ID and sometimes ASD.[44,45]
- Nine *de novo* SNV in *CHD8* reported in subjects with ASD (3 nonsense, 4 frameshift, 1 splice, and 1 single amino acid deletion).[6,46]

### 5) Comment

*CHD2* mutations were implicated in epileptic encephalopathies (Carvill et al. 2013) while this manuscript was in preparation. The deletion in the two siblings with ASD identified in our study, together with the review of other cases, also implicates this gene in ASD. *CHD2* haploinsufficiency is highly penetrant (all cases with known inheritance are *de novo*), associated with ID and variably with epilepsy and ASD. Although the phenotype description is missing in several cases and is incomplete in others, precluding complete analyses, among 18 cases (the missense variant reported by Neale et al. not included), ID is reported in 16, epilepsy in 13 and autism/ASD in 6 cases. Language delay and/or limited speech are described in 5. Eight individuals are reported with dysmorphic features, usually mild. Features reported in 2 or more individuals include: strabismus or amblyopia (n=3), ear pits (n=2), hirsutism (n=2), and low posterior hairline (n=2). At least 4 had growth delay/short stature, and 2 have scoliosis, two phenotypes reported in heterozygous *Chd2* mice.




## Figure S7. CNV and SNV identified in *HDAC4* (histone deacetylase 4) in chromosome 2q37.3

Chr2:239,350,000-240,180,000 (hg18). Abbreviations: dn, *de novo*; ID, intellectual disability; mat, maternal; unk, unknown inheritance

### 1) AGP

- Proband 16037-1571015001, male, 12 kb intragenic deletion of *HDAC4,* removing exon 4 (chr2:239766528-239778481), maternal. Phenotype: autism (ADI-R and ADOS positive), language delay (first words and phrases 42 mo), functional language, WISC-III at 21 y: verbal IQ 72, performance IQ 87, full scale IQ 77. Born at 37 weeks of gestation, no congenital malformations, normal general exam. Had seizures at age 4-5 y, treated with Depakote; treatment stopped for 2-3 y without seizure activity. Unfortunately, we were unable to have the proband re-evaluated by a clinical geneticist to assess the presence of mild dysmorphic features and hand or feet abnormalities. The father has schizophrenia; the mother has depression, no other information is available about her phenotype.
- No *HDAC4* deletions among 4,768 controls and 4,874 parents (1 deletion in the mother of proband 16037-1571015001).

### 2) Other human genetic evidence

- Over 120 individuals with 2q37 terminal deletions (2q37 deletion syndrome, also known as brachydactyly mental retardation syndrome) have been reported. Common features include developmental delay/ID, ASD, hypotonia, mild facial dysmorphism (frontal bossing, round face, depressed nasal bridge, abnormal or prominent ears, deep-set eyes, anteverted nares, and thin upper lip), short stature, obesity and brachymetaphalangy of digits 3-5 (>50%).[47] In 2010, haploinsufficiency of *HDAC4* was shown

to result in brachydactyly mental retardation syndrome. Williams et al.[48] described 5 individuals with 2q37 deletion syndrome in whom the smallest region of overlap contained only *HDAC4*. Sequencing of *HDAC4* identified *de novo* loss-of-function mutations in two additional subjects. Although *HDAC4* mutations may be causative for most of the features of the 2q37 microdeletion syndrome, other genes might also be involved, since individuals with distal deletions not including *HDAC4* have been reported with ID, ASD and seizures. Because no *HDAC4* mutations or single gene deletions had been reported previously in ASD, this gene was not yet considered as a cause of ASD (i.e. we had included it in the list of ID genes, not in the ASD list).

- 2q37 deletions as well as the 2 reported *HDAC4* mutations occur *de novo*, so the maternal transmission observed in the AGP proband would appear a highly unusual finding. However, a recent case report[49] described for the first time a deletion overlapping *HDAC4* inherited from a mildly affected parent (see below).

**CNV:**

As noted above, there are many large 2q37 deletions reported in the literature and in the databases. Here we review only cases with small deletions overlapping *HDAC4* and two recently reported familial cases.

- Villavicencio-Lorini et al.[49]: Report of the first three generation familial case of brachydactyly mental retardation syndrome with an interstitial 2q37.3 microdeletion of 758 kb (chr2:239395957-240154599). Subject 1: female index case, only child, motor development and growth delays during the first year of life, with subsequent catch up. When evaluated at 2 y 8 mo, she had midface hypoplasia, deep set eyes, posteriorly rotated and low-set ears, thin upper lip and pointed chin. Speech development was normal; she exhibited aggressive tantrums and sleep difficulties. Brachydactyly was excluded clinically and radiologically. Subject 2: 45-year-old mother of the index case, history of developmental and growth delays during childhood, but she was later able to attend normal school. She reported reduced spatial orientation and memory deficits. Examination revealed coarse facial appearance with broad and depressed nasal bridge, highly arched eyebrows, deep set eyes and narrow palpebral fissures; growth parameters were normal. Her hands and feet appeared normal, and brachydactyly was excluded radiologically. Subject 3: 68-year-old grandmother of the index case, she is the mother of subject 2, her only child. Family history revealed that her sister's daughter had ID with hydrocephalus and paraplegia of unknown cause. Subject 3 had severe osteoarthritis, dysmorphic facial features similar to those of her daughter with highly arched eyebrows, narrow palpebral fissures and everted full lips, with normal growth parameters. She communicated in simple sentences and her intellectual skills appeared to be lower than normal. Clinically her hands were normal.
- Morris et al.[50]: Report of a familial case of brachydactyly mental retardation syndrome, including a parent with mild symptoms and a child exhibiting a more severe phenotype. Cytogenetic testing showed a cryptic balanced translocation in the mother, t(2;10)(q37;q26), which resulted in a 9.84 Mb deletion on chromosome 2q37.1 (chr2:232810566-242654701) and a 10q26.1 duplication (chr10:131931089–135253240) in her son. *HDAC4* was deleted in the child but present and translocated in his mother. Interestingly, *HDAC4* expression in lymphocytes was 67% in the mother and 23% in the son compared to normal controls. Since the predicted expression after loss of one copy of *HDAC4* is ~50%, these findings suggest that there is an additional unknown mechanism decreasing *HDAC4* dosage, both in the mother and the child.
  Subject 1: male proband, only child; upon examination at the age of 15 y he was overweight (BMI 28), with short stature (–3 SD), facial dysmorphism with round face, prominent forehead, highly arched eyebrows, sparse hair, low-set ears, downslanted palpebral fissures, depressed nasal bridge, thin upper lip, and high palate. He also presented type E brachydactyly of fourth fingers and toes, and syndactyly of the second and third toes. Brain MRI showed abnormal gyration of the frontal lobes. He was schooled in a specialized institute. Subject 2: mother of subject 1, she had one previous termination of pregnancy for occipital meningoencephalocele. Family history showed a maternal aunt and three maternal cousins with ID. She presented with a similar dysmorphic features as her son (round face, prominent forehead, highly arched eyebrows, low-set ears and thin upper lip), short stature, obesity (BMI 45), and brachydactyly of the fourth finger of the right hand, and third and fourth fingers on the left. She had no obvious ID.
- ISCA: Subject nssv1415172, unknown gender, *HDAC4* intragenic deletion of 10 kb (chr2:239670063-239680639), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.

### 3) Functional evidence

- Histones play a critical role in transcriptional regulation, cell cycle progression, and development. Histone acetylation/deacetylation alters chromosome structure and affects transcription factor access to DNA. *HDAC4* encodes a histone deacetylase that shuttles between the nucleus and cytoplasm in response to calcium-regulated signals. HDAC4 has been shown to regulate a transcriptional program essential for synaptic plasticity and memory; by repressing genes encoding synaptic proteins, it affects the strength and architecture of excitatory synapses.[51] Conditional deletion of *Hdac4* in mouse forebrain neurons leads to impairments in spatial learning and memory and long-term synaptic plasticity.[52]
- Highly expressed in the brain.

### 4) Comment

Although many cases of 2q37 deletion syndrome with ASD have been reported in the literature, no deletion involving only *HDAC4* or mutation in this gene had been reported in individuals with ASD. The identification of an intragenic exonic deletion of the *HDAC4* gene in one AGP proband implicates haploinsufficiency of this gene as a cause for autism.




## Figure S8. Recurrent duplications at Xq28 including *GDI1* (GDP dissociation inhibitor 1)

ChrX:153,085,000-153,640,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; EPI, epilepsy; ID, intellectual disability; mat, maternal; unk, unknown inheritance

### 1) AGP

- Proband 20013-1075001, male, Xq28 duplication of 292 kb (chrX:153222048-153514311), *de novo*. The duplication encompasses 21 genes, including 3 genes involved in ID (*FLNA, GDI1, IKBKG*) and is flanked by segmental duplications. Phenotype: sporadic autism (ADI-R and ADOS positive), no language delay, normal IQ (WISC-III at 16 y: verbal IQ 93, performance IQ 94, full scale IQ 93), no dysmorphic features, head circumference -1.4 SD, normal neurological exam, no seizures.
- Proband 14216-3470, male, Xq28 duplication of 211 kb (chrX:153263157-153474401), maternally inherited. The duplication overlaps 19 genes, including 2 genes involved in ID (*GDI1, IKBKG*) and is flanked by segmental duplications. Phenotype: autism (ADI-R and ADOS positive), neurodevelopmental delay with onset at 2 y (first words 18 mo, first phrases 48 mo), mild ID (verbal IQ 61, performance IQ 65, full scale IQ 72), no dysmorphic features, sleep problems, no epilepsy. A brother with developmental delay also carries the duplication. He has a confirmed learning disorder and probably mild intellectual disability (he is an adult now and was not evaluated formally for this study).
- No similar duplications among 2,022 male controls and 2,441 fathers.

### 2) Other human genetic evidence

- Mutations in *GDI1* are a rare cause of nonsyndromic X-linked ID[53,54] and have not yet been reported in ASD.

**CNV:**

- Vandewalle et al.[55]: Four unrelated families with X-linked ID with recurrent 0.3 Mb Xq28 copy number gain (153.20-153.54 Mb) mediated by segmental duplications. Only males are affected, carrier mothers show skewed X chromosome inactivation. The copy-number gain is variable, ranging from 2 to 5 copies, and includes *GDI1, FLNA* and *IKBKG*, involved in ID through mutations. The authors suggest *GDI1* is the most likely candidate gene; it is highly expressed in the brain and its expression in blood is correlated with the severity of the phenotype. *FLNA* duplications have been reported in four males with intestinal dysfunction, without ID[56] and in one male control in the Database of Genomic Variants (DGV); a duplication including *FLNA* and *IKBKG* was reported in a healthy male.[55]
  Family 1: four affected males in two generations, Xq28 copy number gain (chrX:153218000-153535000). Phenotype: non-syndromic moderate ID. Family 2: two affected brothers, Xq28 copy number gain (chrX:153218000-153542000). Phenotype: severe ID, epilepsy in the elder brother. Family 3: three affected males, Xq28 copy number gain (chrX:153218000-153530000). Phenotype: the index case presented with delayed speech, learning disabilities, and mild ID. All affected male subjects had moderate ID, a peculiar face with dysmorphic features, and macrocephaly. Reported previously as case 6 in Madrigal et al.[57] Family 4: male (sporadic case), Xq28 copy number gain (chrX:153218000-153535000). Phenotype: global psychomotor delay, mild dysmorphism.
- Sakai et al.[58]: Subject 11092, male, Xq28 duplication (chrX:153229170-153433332), *de novo*. Phenotype: ASD, history of seizures, normal IQ (verbal IQ 108, performance IQ 125, full scale IQ 109).

- DECIPHER: Subject 262772, male, Xq28 duplication (chrX:153219789-153490319), maternal. No phenotype information; the subject carries two other duplications (chr3:685234-1166340 and chr17:31862-164024), both inherited.
- DECIPHER: Subject 252424, male, Xq28 duplication (chrX:153230084-153475911), *de novo*. No phenotype information; the subject carries an additional CNV (chr1:91820025-92502808 duplication, *de novo*).
- DECIPHER: Subject 253549, male, Xq28 duplication (chrX:153230084-153475911), maternal. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 265379, male, Xq28 duplication (chrX:153230109-153485889), unknown inheritance. Phenotype: ID, hypotonia, sleep apnea. Only CNV reported in the subject.
- ECARUCA: Subject 4882, male, Xq28 duplication (chrX:153051459-153436637), maternal. Phenotype: ID, spasticity/increased tendon reflexes, macrocephaly, constipation, short stature.

  (Three ISCA individuals with duplications of this region —2 with developmental delay, 1 with ASD— were not added because their gender is unknown.)

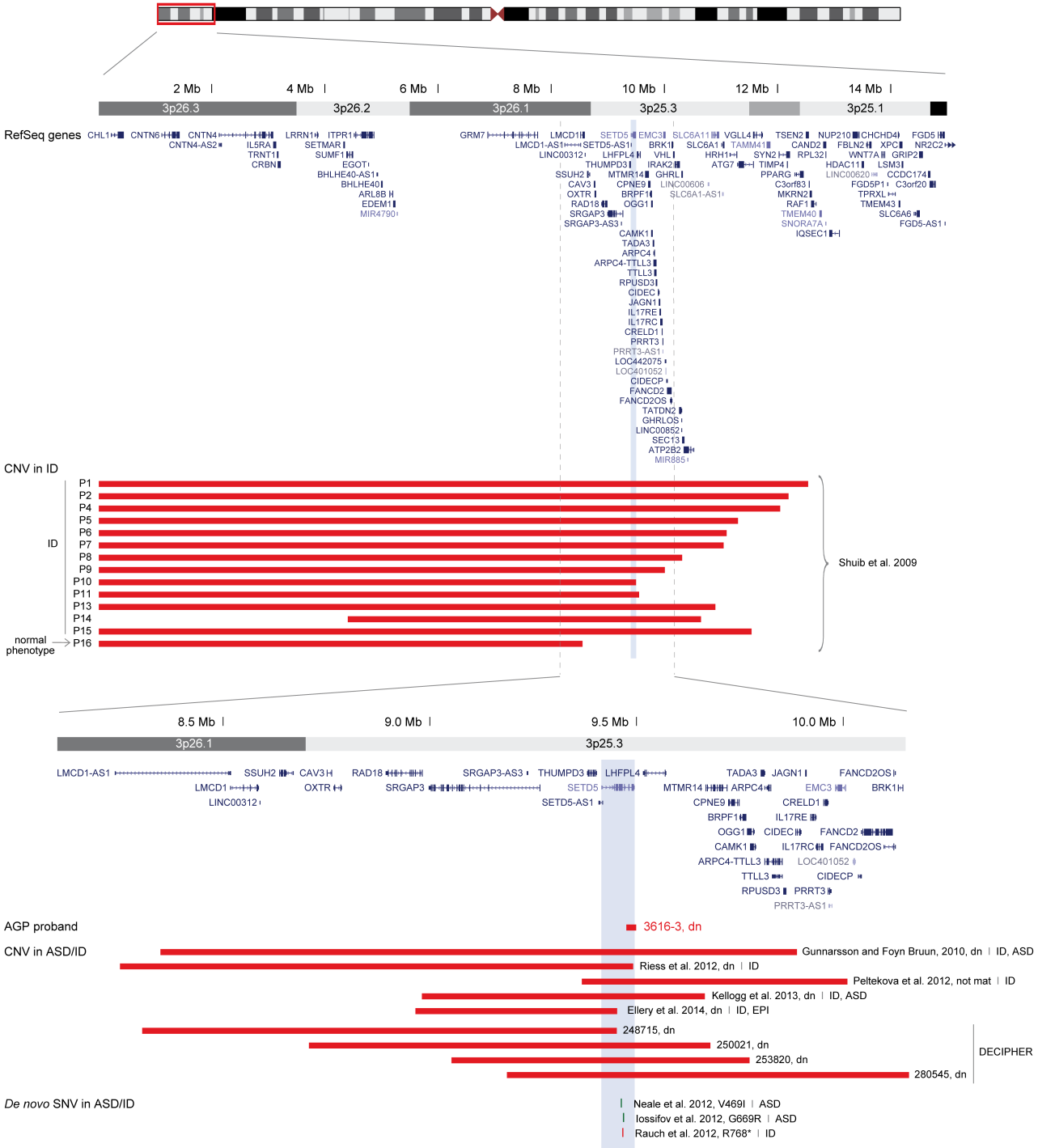
**3) Functional evidence**

- *GDI1* encodes the GDP-dissociation inhibitor alpha (αGDI), which regulates the activity of small GTPases of the Rab family involved in intracellular vesicular trafficking. *Gdi1*-deficient mice exhibit membrane accumulation of Rab GTPases and decrease in the reserve pool of synaptic vesicles in the hippocampus, leading to altered synaptic plasticity and short-term memory deficits.[59]
- *GDI1* is highly expressed in the brain.

- **4) Comment**

  We identified two AGP probands with a recurrent Xq28 duplication corresponding to a genomic disorder recently described in ID. The identification of these two independent ASD cases, together with an ASD subject reported previously[58], implicates duplications of this region in ASD. Among the genes contained in the region, aberrant gene dosage of *GDI1* is likely to be responsible for the neurodevelopmental phenotype of subjects carrying this CNV since mutations in this gene are involved in nonsyndromic X-linked ID[53,54] and the expression of *GDI1* in the blood of individuals with the Xq28 duplication correlated with the severity of the phenotype.[55] Finally, *GDI1* is highly expressed in the brain and its function in intracellular vesicular trafficking together with the synaptic phenotype of the *Gdi1*-deficient mice make this gene particularly interesting with regard to the ASD/ID phenotypes.

## Candidate genes affected by *de novo* CNV in AGP probands




## Figure S9. CNV and SNV identified in *SETD5* (SET domain containing 5) in chromosome 3p25.3

Chr3:1-15,000,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; EPI, epilepsy; ID, intellectual disability; mat, maternal

### 1) AGP

- Proband 3616_3, male, 24 kb deletion including only *SETD5* (chr3:9474320-9498362), *de novo*. Phenotype: Born at term after normal pregnancy and delivery. No feeding or sleep problems; body length and head circumference during the first year of life

were consistently between 0 to –1 SD; walked at 15 mo, language delay (first words 24 mo), functional language, autism (meets criteria for autism on both ADI-R and ADOS), borderline IQ (WISC-III: verbal IQ 64, performance IQ 82, full scale IQ 70). Concerns about ADHD-like behavior in childhood, but did not fulfill criteria for ADHD and there are no concerns about hyperactivity at present. Normal vision and hearing; had sensory-integration training because of over and undersensitivity; asthma and hay fever, no other medical issues, no seizures. At 16 y, normal height and weight (50$^{th}$ centile), no obvious dysmorphic features (no ptosis, normal ears, nose and mouth), normal fingers and toes.

- No *SETD5* deletions among 4,768 controls and 4,875 parents.

### 2) Other human genetic evidence

- 3p deletion syndrome is a rare contiguous gene disorder resulting from terminal and interstitial deletions involving chromosome 3p25-p26. Common features include ID, growth retardation, microcephaly, dysmorphic facial features, and ptosis.[60] Rare individuals with a 3p26-p25 deletion and normal intelligence or only mild abnormalities have been reported, including several cases inherited from seemingly unaffected parents, indicating variable penetrance. There are over 30 individuals with 3p25-p26 deletions reported, the majority with large cytogenetically visible rearrangements, making delineation of the critical region and identification of candidate genes difficult. Shuib et al.[60] reported 14 individuals with cytogenetically visible deletions of 3p25, all with marked ID except one subject (P16) with a normal phenotype, ascertained incidentally because of recurrent miscarriages. Notably, the deletions in subjects with ID all overlapped *SETD5* whereas the subject without ID had the smallest terminal deletion, with a breakpoint proximal to *SETD5* (**Figure S9**). Here we review recently described cases with clinical features of distal 3p deletion syndrome carrying smaller interstitial deletions.

**CNV:**

- Gunnarsson and Foyn Bruun[61]: Female, 1.6 Mb 3p26.1-p25.3 deletion of 24 genes including *SETD5* (chr3:8305426-9885334), *de novo*. The girl was born at 37 weeks, with birth measures below –2 SD. She was noted to have dysmorphic facial features (hypertelorism, ptosis, strabism, flat and broad nasal root, long philtrum, downturned corner of the mouth, low set ears), bilateral overlap of the 2$^{nd}$ and 4$^{th}$ toes over the 3$^{rd}$ and 5$^{th}$ toes, atrial and ventricular septal defects, hypotonia, and severe developmental delay, with absent language at the age of 4 y. She showed autistic behavior; she smiled and laughed, but had poor eye contact and did not interact with other children, focusing on objects in her close vicinity. Vision, hearing, and growth were normal. At the age of 4, she had transient seizures, documented with EEG; a MRI showed asymmetry of thalamus. (Corresponds to subject 253231 in DECIPHER.)
- Riess et al.[62]: Female, 1.24 Mb 3p26.1-p25.3 deletion of 12 genes including *SETD5* (chr3:8250541-9491586), *de novo*. Normal birth measures, hypotonia during the first year of life, delayed psychomotor development. She started to walk and talk at the age of 2 y. On examination at the age of 3, she had large head (75$^{th}$ centile) compared to her height (10$^{th}$ centile), bilateral strabismus, large fontanelle, prominent forehead, depressed nasal bridge, thin upper lip and prominent philtrum. Ultrasound examinations of the heart and abdominal organs were normal.
- Peltekova et al.[63]: Female, 643 kb 3p25.3 deletion of 23 genes including *SETD5* (chr3:9367274-10010209), inheritance unknown (not maternal, father's DNA not available). The proband was born via cesarean section for breech presentation at 37 weeks gestation; birth weight 3$^{rd}$–15$^{th}$ centile, length 5$^{th}$ centile, and head circumference 25$^{th}$ centile. She presented polydactyly in all four extremities, cleft palate, atrial septal defect, and bowel malrotation. At 1 year of age she developed tonic–clonic seizures and had episodes of aspiration pneumonia requiring placement of a gastrostomy tube. She was non-verbal as an adult, but smiled, laughed and was interactive. At 22 y her height and weight were <3$^{rd}$ centile, and head circumference was <5$^{th}$ centile. She had low anterior hairline, short forehead, thick eyebrows with synophrys, microphthalmia, downslanting palpebral fissures, blepharophimosis with mild ptosis, hypotelorism, esotropia, low set and cupped ears, and scoliosis. Hands and feet were small, with bilateral syndactyly of the 2$^{nd}$ and 3$^{rd}$ toes. Brain MRI showed parenchymal volume loss and atrophy of all structures except the brain stem.
- Kellogg et al.[64]: Female, 684 kb 3p25.3 deletion of 7 genes including *SETD5* (chr3:8980098-9664733), *de novo*. The proband was born at full term, with normal birth weight (50$^{th}$ centile) and length (90$^{th}$ centile). She had strabismus (corrected surgically) and developmental delay. Obsessive-compulsive disorder, with repetitively smelling of various objects, was noted at 10 y. On examination at age 11, she had dysmorphic features, including prominent ear lobes, right ear pit, depressed nasal bridge, anteverted nares, long philtrum, and proximally placed thumbs. Height, weight and head circumference were 50$^{th}$, 10$^{th}$ and 50$^{th}$ centile, respectively; the brain MRI was normal. Assessment at 5 y 8 mo placed her in the borderline to intellectually disabled range (Stanford-Binet IV), and she met criteria for autism spectrum on ADOS.

  The authors compared this individual with three previously reported cases with interstitial 3p25 deletions with ID and characteristic facial features[61-63] and identified a region of overlap including only three genes: *THUMPD3, SETD5* and *SETD5-AS1*, which could play a critical role in the neurocognitive phenotype of the 3p deletion syndrome.
- Ellery et al.[65]: Male, 486 kb 3p25.3 deletion of 6 genes including *SETD5* (chr3:8965201-9450984), *de novo*. The proband was born at term with a birth weight of 4600 g, after a pregnancy marked by gestational diabetes. Bilateral postaxial polydactyly, a single palmar crease, right preauricular pit, mild hypertelorism, anteverted nostrils and micrognathia were noted at birth. Hypotonia, feeding difficulties and developmental delay became evident afterwards. A diagnosis of Simpson–Golabi–Behmel syndrome was considered. He walked independently at 4 y but tired easily; at 6 y his vocabulary was limited to two-dozen words and occasional

short sentences. He developed grand mal seizures at 7 y, treated with carbamazepine. He had a disturbed sleep pattern with frequent waking. On examination at 8 y of age, he had epicanthic folds, mild hypertelorism, high palate, myopathic facies with coarse features, and pectus excavatum. He had poor muscle bulk and showed a partial Gower's sign. These features persisted into adulthood.

- DECIPHER: Subject 248715, female, 1.1 Mb 3p26.1-p25.3 deletion of 12 genes including *SETD5* (chr3:8330935-9450984), *de novo*. Phenotype: ID/DD, macrocephaly, coarse facial features, synophrys, gum hypertrophy, low posterior/trident hairline, general abnormalities of hair texture. Only CNV reported in the subject.
- DECIPHER: Subject 250021, male, 946 kb 3p25.3 deletion of 10 genes including *SETD5* (chr3:8724500-9671040), *de novo*. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 253820, male, 700 kb 3p25.3 deletion of 9 genes including *SETD5* (chr3:9061621-9763580), *de novo*. Phenotype: ID, hypertelorism, micrognathia, low-set ears, hydrocephalus, horseshoe kidney. Only CNV reported in the subject.
- DECIPHER: Subject 280545, male, 1.1 Mb 3p25.3 deletion of 33 genes including *SETD5* (chr3:9186364-10290795), *de novo*. Phenotype: global developmental delay, stereotypic behavior, absent speech, abnormality of the corpus callosum, abnormality of the hypothenar eminence, small thenar eminence, facial asymmetry and blepharophimosis. Only CNV reported in the subject.
- Note that several *SETD5* exonic deletions are reported in controls from DGV among non-BAC based studies, all from Shaik et al.;[66] these deletions are likely to be a study-specific artifact.

**SNV:**

- Neale et al.[5] (exome sequencing in ASD): Subject 09C98906, male, *SETD5* missense variant, *de novo*. The variant is predicted to be either benign (PolyPhen2, SNAP, Mutation Taster) or damaging (SIFT); the substituted residue is variable according to ConSurf (1/9); GERP score 1.91 (constrained >2). Phenotype: ASD, verbal IQ 93, performance IQ 112, full scale IQ 97.
- Iossifov et al.[4] (exome sequencing in ASD): Subject 13576, female, *SETD5* missense variant (gene listed as *KIAA1757*), *de novo*. The variant is predicted to be damaging (PolyPhen2, SNAP and Mutation Taster; not scored by SIFT). The substituted residue is variable according to ConSurf (1/9); GERP score 5.69. Phenotype: ASD (no other information provided).
- Rauch et al.[41] (exome sequencing in nonsyndromic sporadic ID): Subject ER14209, female, *SETD5* nonsense mutation, *de novo*. Phenotype: Born full term, weight –0.95 SD, head circumference –1.46 SD, sitting at 8 mo, walking at 20 mo, first words at 48 mo, IQ 70, mild attention deficit disorder, no ASD, no seizures, strabism, recurrent infections, constipation, prominent finger joints, facial dysmorphisms. When last evaluated at 9 y, height –0.78 SD, head circumference –1.16 SD, spoke in fluent sentences.

### 3) Functional evidence

- The gene is predicted to be a methyltransferase; several other genes encoding methyltransferases have been shown to be altered in ASD/ID (see examples below).
- Expressed in the brain.

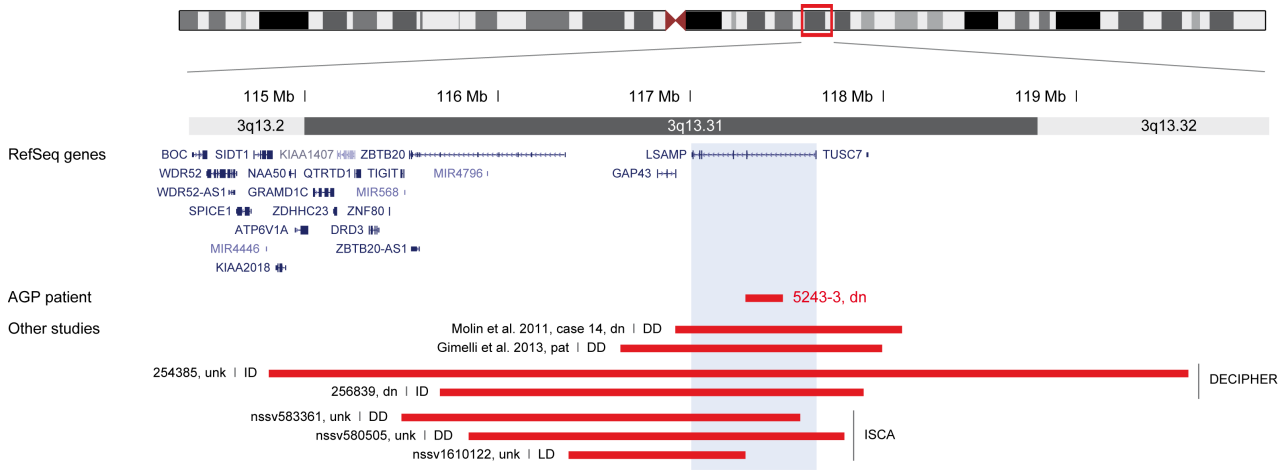
### 4) Other evidence

- % HI = 21.3 (0%-10%: likely to be haploinsufficient; 90%-100%: not likely to be haploinsufficient)
- Haploinsufficiency of *NSD1* (nuclear receptor binding SET domain protein 1), encoding a histone methyltransferase, causes Sotos syndrome, a neurodevelopmental disorder characterized by overgrowth, distinctive craniofacial appearance, and variable ID, sometimes associated with ASD.[67,68]
- Another gene encoding a histone methyltransferase, *EHMT1* (euchromatic histone-lysine N-methyltransferase), is involved in Kleefstra syndrome through deletions (9q34.3 deletion syndrome) or mutations. *EHMT1* haploinsufficiency is associated with ID and sometimes with ASD.[69] An AGP proband with a 9q34.3 deletion encompassing *EHMT1* was identified in the present study (proband 6259-3, **Table S7B**).

### 5) Comment

The minimal region of overlap between our proband and those of 9 other cases with deletions involving 3p25.3 reviewed here contains only *SETD5*. Remarkably, all deletions are *de novo*. This finding, together with the recent report of a *de novo* loss-of-function mutation in a subject with language delay and borderline IQ (70, like in our proband)[41] suggest that *SETD5* is involved in cognitive, social, and language development. Therefore, *SETD5* could be associated with these features in 3p deletion syndrome. Two other *de novo* variants, identified in whole exome studies in ASD,[4,5] are missense changes and their functional effect is difficult to predict.

Interestingly, the AGP proband with the intragenic *SETD5* deletion, as well as the subject with the *SETD5* loss-of-function mutation, have a borderline IQ (70), whereas other individuals with larger deletions have ID ranging from severe to moderate, suggesting that other genes in the region contribute to the cognitive deficit associated with 3p deletions. Similarly, our proband lacks some of the characteristic but variable features associated with the 3p deletion syndrome, including growth retardation and dysmorphic features, such as ptosis and depressed nasal bridge.




## Figure S10. CNV identified in *LSAMP* (limbic system-associated membrane protein) in chromosome 3q13.31

Chr3:114,400,000-120,000,000 (hg18). Abbreviations: dn, *de novo*; DD, developmental delay; ID, intellectual disability; LD, learning disability; pat, paternal; unk, unknown inheritance

### 1) AGP

- Proband 5245_3, male, 192 kb deletion including only *LSAMP* (chr3:117285007-117477191), *de novo*. The deletion was shown to be mosaic; the percentage of mosaicism was estimated at 50% of deleted cells using a formula based on the deviation of the B allele frequency distribution.[70] Phenotype: born at 29 wks, intraventricular hemorrhage, mild cerebral palsy; autism (based on ADI-R and ADOS), low non-verbal IQ (<1$^{st}$ centile), language delay (1$^{st}$ centile), apraxia, abnormal sleep EEG without seizures; alopecia areata, no dysmorphic features. Multiplex family, a sister with ASD doesn't carry the CNV.
- No exonic deletions of *LSAMP* among 4,768 controls and 4,875 parents (1 control and 1 mother have intronic deletions).

### 2) Other human genetic evidence

**CNV:**

- Molin et al.[71]: Subject 14, male, 1.18 Mb 3q13.31 deletion (chr3:116922662-118098190) containing three genes, *GAP43*, *LSAMP* and *TUSC7*. Uncertain inheritance (absent in the mother and father not tested). Five year old boy with developmental delay; no other phenotype information provided. *GAP43* is found in growth cones of extending axons in the central nervous system. (Corresponds to subject 252520 in DECIPHER.)
- Gimelli et al.[72]: Girl with 1.36 Mb 3q13.31 deletion (chr3:116640577-118002810) containing three genes, *GAP43, LSAMP,* and *TUSC7*. The deletion was inherited from the father, who had slightly delayed psychomotor development but his cognitive level was not tested. The proband had developmental delay, clumsiness and attention deficit, associated with renal, vascular and skeletal anomalies.
- DECIPHER: Subject 254385, female, 4.8 Mb 3q13.2-q13.32 deletion involving *LSAMP* (chr3:114813585-119579912), inheritance unknown. Phenotype: ID, muscular hypotonia, downslanted palpebral fissures, joint laxity, open mouth. Only CNV reported in the subject.
- DECIPHER: Subject 256839, female, 2.2 Mb 3q13.31 deletion involving *LSAMP* (chr3:115701890-117895614), *de novo*. Phenotype: ID, broad nasal tip, short nose, anteverted nares. Only CNV reported in the subject.
- ISCA: Subject nssv583361, unknown gender, 2 Mb 3q13.31 deletion involving *LSAMP* (chr3:115499855-117567970), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv580505, unknown gender, 1.9 Mb 3q13.31 deletion involving *LSAMP* (chr3:115850559-117797499), inheritance unknown. Phenotype: global developmental delay.
- ISCA: Subject nssv1610122, unknown gender, 918 kb 3q13.31 deletion involving *LSAMP* and *GAP43* (chr3:116367264-117285252), inheritance unknown. Phenotype: specific learning disability.

### 3) Functional evidence

- *LSAMP* encodes a neuronal surface glycoprotein belonging to the cell adhesion molecule (CAM) family, found in cortical and subcortical regions of the limbic system. During development of the limbic system, it is found on the surface of axonal membranes and growth cones, where it acts as a selective homophilic adhesion molecule, and guides the development of specific

patterns of neuronal connections. LSAMP mediates selective neuronal growth and axon targeting and also contributes to the guidance of developing axons and remodeling of mature circuits in the limbic system. This protein is essential for normal growth of the hippocampal mossy fiber projection.[73] *Lsamp*$^{-/-}$ mice have normal gross neuroanatomical organization but display heightened reactivity to novelty, reduced anxiety-like behaviors, impaired synaptic plasticity, and spatial memory deficit.[74,75] *LSAMP* has also been shown to function as a tumor suppressor gene.

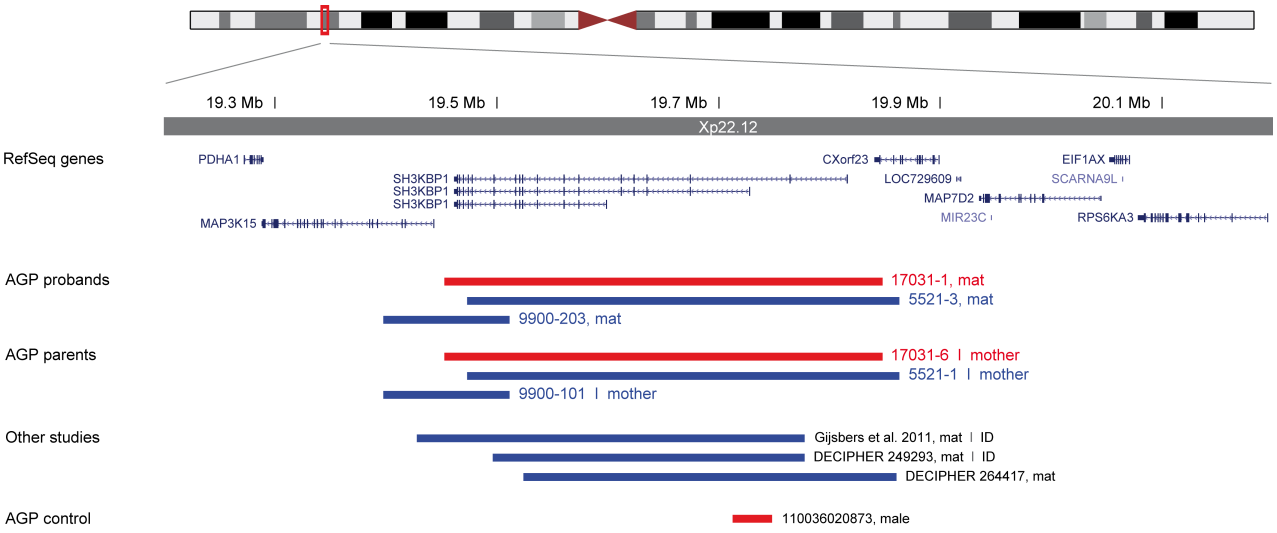
- Brain expression: *LSAMP* is expressed on limbic regions but also, less intensely, in midbrain and hindbrain regions.

### 4) Other evidence

- % HI = 10.5 (likely to be haploinsufficient)

### 5) Comment

The minimal region of overlap between our proband and seven other cases with 3q13.31 deletions reviewed here contains only the *LSAMP* gene. All cases with known inheritance are *de novo* and would thus support involvement of deletions in this region in neurodevelopmental disorders. Other cases with deletions involving only *LSAMP* or deleterious mutations are needed to implicate this gene in ASD and ID.




## Figure S11. CNV identified in *SH3KBP1* (SH3-domain kinase binding protein 1) in chromosome Xp22.12

ChrX:19,200,000-20,200,000 (hg18). Abbreviations: ID, intellectual disability; mat, maternal

### 1) AGP

- Proband 17031_1, male, deletion involving *SH3KBP1* and *CXorf23* (chrX:19450969-19845766), maternal. *CXorf23* encodes a protein of unknown function. Phenotype: autism (based on ADI-R and ADOS), language delay, functional language, verbal IQ 79; physical examination at 5 y 8 mo revealed no dysmorphic features, neurological examination was normal except for a deficit in coordination and gross and fine motor development; no seizures, normal sleep EEG. Sporadic case, mother unaffected, negative family history of neuropsychiatric disorders.
- Proband 5521_3, male, partial duplication of *SH3KBP1* and *CXorf23* (chrX:19471138-19861338), maternal. Phenotype: autism (based on ADI-R and ADOS), below average IQ (<1$^{st}$ centile), nonverbal, seizure disorder, coarse facial features. Sporadic case, mother unaffected.
- Proband 9900_203, male, partial duplication of *SH3KBP1* and *MAP3K15* (chrX:19396116-19509785), maternal. Expression study in cell lines showed no increased expression of *SH3KBP1*. The CNV also involves *MAP3K15*, encoding a mitogen-activated protein (MAP) kinase expressed in the brain. Phenotype: autism (meets criteria on ADI-R and ADOS), language delay (first words 18 mo, first phrases 42 mo), limited language, moderate ID (Bayley Scales of Infant Development II, mental developmental index 45), strabismus (like his mother), normal physical exam, no dysmorphic features, no epilepsy. Sporadic case, mother unaffected.
- 1 of 2,022 male controls carries a *SH3KBP1* deletion overlapping only exon 1 of transcript variant 2, not present in other isoforms (chrX:19710729-19746114); no deletions among 2,441 fathers.

### 2) Other human genetic evidence

**CNV:**

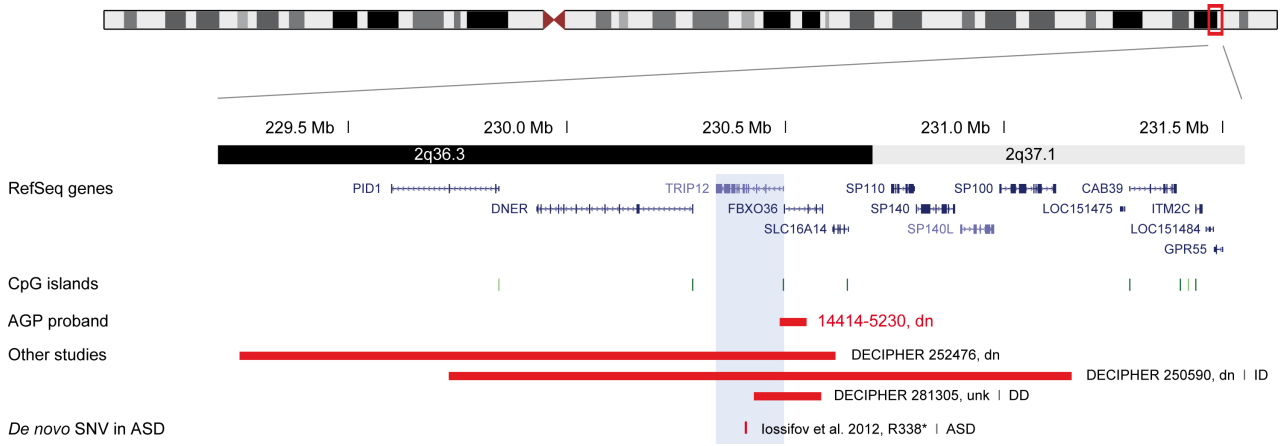
- Gijsbers et al.[76]: Male, 349 kb duplication overlapping partially 2 genes, *SH3KBP1* and *MAP3K15* (chrX:19425768-19775308), maternal. Phenotype: growth retardation, severe ID, absent or delayed speech, stereotypic movement of head and hands, bitemporal narrowing, narrow palpebral fissures, deep-set eyes, large mouth, widely spaced teeth. The healthy mother and grandmother carried the same Xp22.12 duplication and showed skewed X inactivation. The authors interpreted the CNV as potentially pathogenic.
- DECIPHER: Subject 249293, male, intragenic *SH3KBP1* duplication (chrX:19494754-19775308), maternal. Phenotype: ID, narrow forehead, abnormality of the eyebrow, deep set eyes, abnormality of the mouth, widely spaced teeth, proportionate short stature, abnormal CNS myelination. Only CNV reported in the subject.
- DECIPHER: Subject 264417, male, duplication overlapping partially 2 genes, *SH3KBP1* and *CXorf23* (chrX:19521667-19858019), maternal. No phenotype information; only CNV reported in the subject.

### 3) Functional evidence

- The *SH3KBP1* gene encodes CIN85, an endocytic scaffold protein that facilitates protein-protein interactions and has been implicated in numerous cellular processes including apoptosis, cytoskeletal rearrangement, cell adhesion and clathrin-dependent endocytosis. CIN85 plays a role in receptor internalization, including dopamine receptors and epidermal growth factor receptor.[77,78] CIN85 localizes at synapses and interacts with the scaffold protein S-SCAM via dendrin.[79]
- Mice lacking CIN85 exon 2 (present in both isoforms expressed in the central nervous system) show hyperactivity and increased exploratory behavior, but no alterations in synaptic plasticity or learning and memory. These mice show increased dopamine and dopamine D2 receptors in the striatum, due to impaired endocytic internalization of D2 receptors.[77]
- Expressed in the brain.

### 4) Comment

The deletion in proband 17031_1 results in a *SH3KBP1* null allele. The intragenic duplication in DECIPHER subject 249293 is also likely to disrupt the gene, acting as a deletion. In contrast, the functional consequence of the partial duplications observed in AGP probands 5521-3 and 9900-203, and in the two other subjects ([76] and DECIPHER 264417) are difficult to predict and thus can not be counted as evidence in favor of the implication of *SH3KBP1* alterations in neurodevelopmental disorders, in the absence of expression studies. In one proband (9900_203) in whom we evaluated *SH3KBP1* mRNA expression in cell lines, no alteration was observed. Other affected males with deletions, intragenic duplications or deleterious mutations of *SH3KBP1* are required to implicate loss of function of this gene in ASD and ID.




## Figure S12. CNV and SNV identified in *TRIP12* (thyroid hormone receptor interactor 12) in chromosome 2q36.3

Chr2:229,200,000-231,550,000 (hg18). Abbreviations: ASD, autism spectrum disorder; DD, developmental delay; dn, de *novo*; ID, intellectual disability; unk, unknown inheritance

### 1) AGP

- Proband 14414_5230, male, 60 kb deletion involving 2 genes, *TRIP12* and *FBXO36* (chr2:230486629-230547253), *de novo*. The deletion only involves the first exon of *TRIP12*, which is non-coding, but seems to be part of the promoter of the gene since a CpG island is located in this region. The other deleted gene, *FBXO36*, encodes a F-box protein that plays a role in ubiquitination and is

not likely to be haploinsufficient (% HI = 52.9). Phenotype: autism (ADI-R and ADOS positive), no language delay (first words 12 mo, first phrases 24 mo), verbal, Griffiths at 5 y 9 mo: language DQ 87, performance DQ 78, global DQ 79.

- No *TRIP12* deletions among 4,768 controls and 4,875 parents.

### 2) Other human genetic evidence

**CNV:**

- DECIPHER: Subject 252476, female, 1.4 Mb deletion involving *TRIP12* (chr2:229250832-230614988), *de novo*. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 250590, female, 1.4 Mb deletion involving *TRIP12* (chr2:229728861-231153046), *de novo*. Phenotype: ID, delayed speech and language development, epicanthus, hypermetropia, low hanging columella, palpebral edema, broad philtrum, thin upper lip vermilion, wide mouth. Only CNV reported in the subject.
- DECIPHER: Subject 281305, male, 144 kb deletion involving *TRIP12* (chr2:230432282-230576372), inheritance unknown. Phenotype: global developmental delay and cystic renal dysplasia. Only CNV reported in the subject.

**SNV:**

- Iossifov et al.[4] (exome sequencing in ASD): Subject 12867, female, *TRIP12 de novo* nonsense mutation. Phenotype: ASD (no other information provided).

### 3) Functional evidence

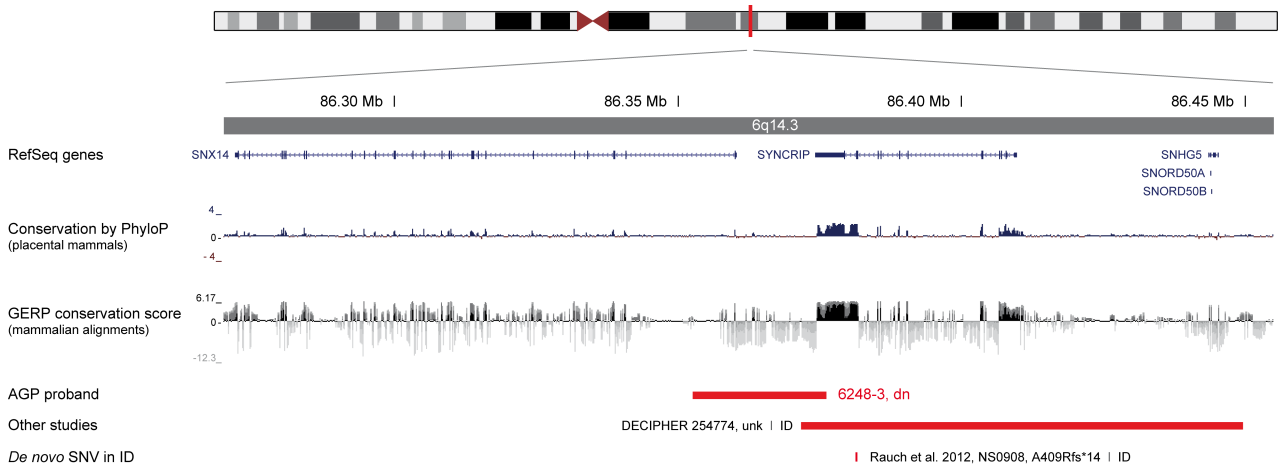
- *TRIP12* encodes a HECT-type E3 ubiquitin-ligase, which plays a role in degradation of ubiquitin fusion substrates and can regulate chromatin function to maintain genome integrity.[80]
- Expressed in the brain.

### 4) Other evidence

- % HI = 5.3 (highly likely to be haploinsufficient)
- *HUWE1*, mutated in X-linked ID, also encodes a HECT-type E3 ubiquitin-ligase involved in the ubiquitin-fusion degradation (UFD) pathway. Double knock-down of *HUWE1* and *TRIP12* results in additive stabilization of an UFD substrate, suggesting functional redundancy between both proteins.[81]

### 5) Comment

The evidence from the AGP subject is not very strong; although the *TRIP12* deletion is *de novo*, only the first non-coding exon is involved and the effect on the protein is difficult to predict. Expression studies are necessary to interpret the clinical significance of this CNV. The support from other human genetic studies is limited, but the *de novo* loss-of function *TRIP12* mutation in a female with ASD identified in an exome sequencing study, together with the function of the protein, make this gene an interesting candidate for ASD.




## Figure S13. CNV and SNV identified in *SYNCRIP* (synaptotagmin binding, cytoplasmic RNA interacting protein) in chromosome 6q14.3

Chr6:86,270,000-86,455,000 (hg18). Abbreviations: dn, *de novo*; ID, intellectual disability; unk, unknown inheritance

### 1) AGP

- Proband 6248_3, male, 23 kb deletion involving 2 genes, *SYNCRIP* and *SNX14* (chr6:86352577-86376159), *de novo*. The deletion only involves the 3' untranslated region (UTR) of *SYNCRIP*; this region might contain regulatory elements that are crucial for gene expression. *SNX14* encodes a member of the sorting nexin family that are involved in intracellular trafficking; this gene is not predicted to be haploinsufficient (% HI = 58.6), but its contribution to the phenotype can not be ruled out. Phenotype: autism, severe ID, no language; born by C-section with increased height and head circumference. One absence seizure at 11 y, with abnormal EEG, probably related to medications taken at that time, no other seizures afterwards. At 15 y: height >+3 SD, macrocephaly (head circumference >+3 DS) (his unaffected father and 2 brothers are also very tall and have macrocephaly); long and narrow hands and feet, 2 café-au-lait spots, normal neurological exam.
- No *SYNCRIP* deletions among 4,768 controls; 1 carries a partial *SYNCRIP* duplication (chr6:86382351-87814038); no CNV overlapping *SYNCRIP* in 4,875 parents.

### 2) Other human genetic evidence

**CNV:**

- DECIPHER: Subject 254774, male, 78 kb deletion including *SYNCRIP* (chr6:86371713-86449627), inheritance unknown. The deletion also involves *SNHG5* (snoRNA host gene) as well as *SNORD50A* and *SNORD50B* (snoRNAs). Only CNV reported in the subject. Phenotype: ID.

  Note that there are several large deletions (>5 Mb) overlapping this gene in DECIPHER and ISCA, not reviewed here.

**SNV:**

- Rauch et al.[41] (exome sequencing in nonsyndromic sporadic ID): Subject NS0908, female, *SYNCRIP* frameshift mutation, *de novo*. Phenotype: severe nonsyndromic sporadic ID (IQ <50), myoclonic astatic seizures (onset at 13 mo), no ASD, MRI at 24 mo: prominent lateral ventricles.

### 3) Functional evidence

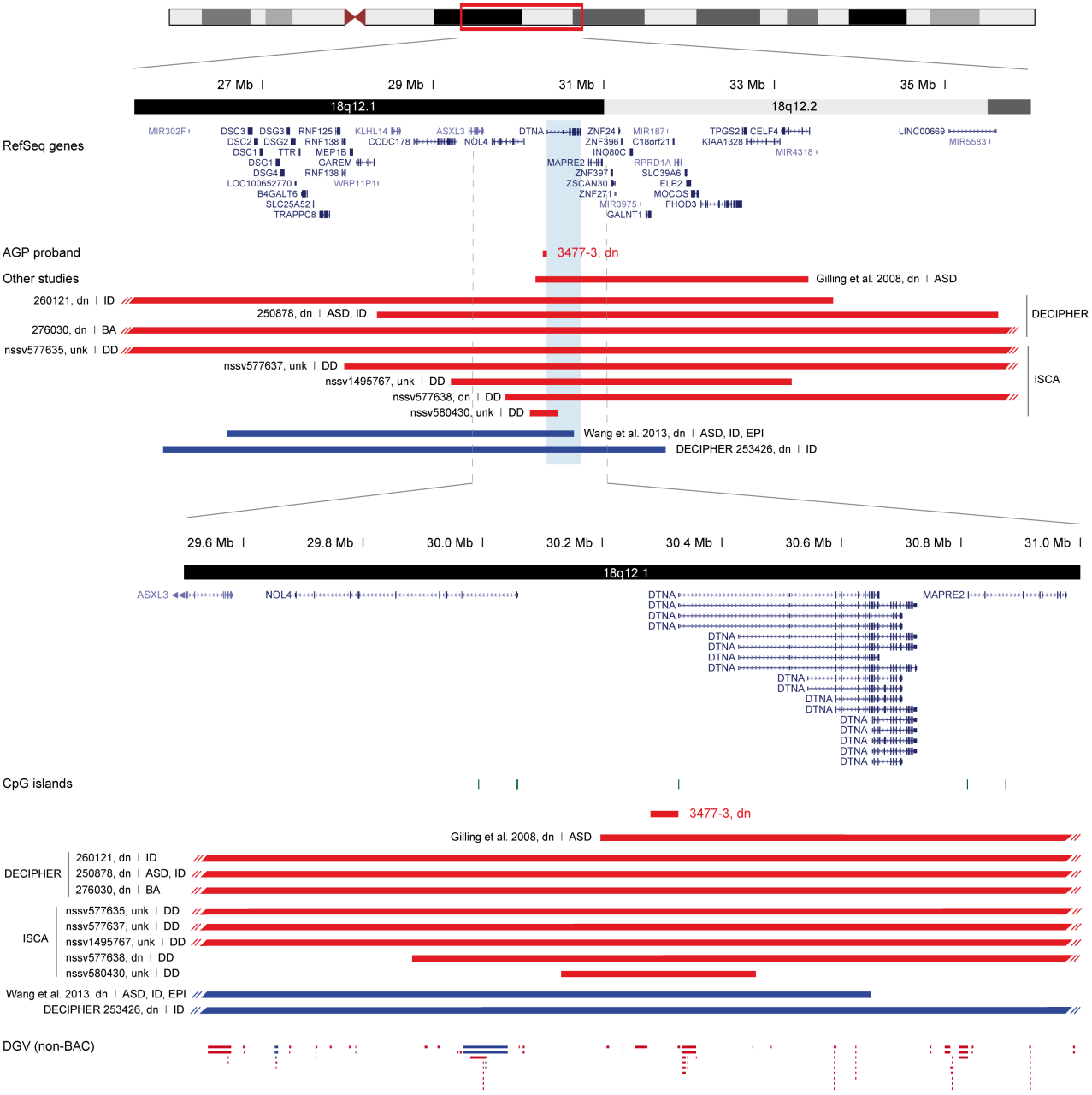
- *SYNCRIP* encodes a nuclear ribonucleoprotein implicated in mRNA processing mechanisms including mRNA stability and transport, RNA editing and splicing and localized mRNA translation. SYNCRIP is a component of mRNA transport granules in dendrites.[82] Selective mRNA transport, local translation and subsequent protein synthesis in neuronal dendrites are part of the fundamental mechanisms involved in synaptic plasticity, learning and memory.
- Other RNA binding proteins have been implicated in ID and ASD, including FMRP, involved in fragile X syndrome, the most frequent monogenic cause of ID and ASD, and ZC3H14, involved in recessive non-syndromic ID.[83]
- Expressed in the brain.

### 4) Other evidence

- % HI = 1.5 (highly likely to be haploinsufficient)

### 5) Comment

The AGP proband carries a *de novo SYNCRIP* deletion that only affects the highly conserved 3'UTR of the gene. Expression studies are necessary to determine the effects of the deletion. Although this gene does not have support from many other CNV studies, the *de novo* loss-of function *SYNCRIP* mutation reported in an individual with ID and the function of the encoded protein make this gene a strong candidate for involvement in neurodevelopmental disorders. Additional cases with small deletions affecting *SYNCRIP* and/or with deleterious mutations are required to implicate this gene definitely in ASD, ID and other neurodevelopmental phenotypes.




## Figure S14. CNV identified in *DTNA* (dystrobrevin alpha) in chromosome 18q12.1

Chr18:25,500,000-36,000,000 (hg18). Abbreviations: ASD, autism spectrum disorder; BA, behavioral/psychiatric abnormality; dn, *de novo*; DD, developmental delay; EPI, epilepsy; ID, intellectual disability; unk, unknown inheritance

### 1) AGP

- Proband 3477_3, male, 47 kb deletion involving *DTNA* (chr18:30280260-30327512), *de novo*. Only the first non-coding exon of the longer isoforms of *DTNA* (2, 5, 7 and 17) is disrupted by the deletion. It may be part of the promoter of the gene since a CpG island is located in the region. Phenotype: ASD (ASD on ADI-R, autism on ADOS), verbal; WISC-R at 5 y 7 mo: verbal IQ 88, performance IQ 88, full scale IQ 86; no cardiovascular or neuromuscular abnormalities.
- No *DTNA* deletions among 4,768 controls and 4,875 parents.

### 2) Other human genetic evidence

**CNV:**

- Gilling et al.[84]: Female, *de novo* translocation t(5;18)(q34;q12.2), with a 3.2 Mb deletion at the 18q breakpoint encompassing 20 genes including *DTNA* (chr18:30197000-33392000); the breakpoint on 5q did not contain any known genes. She also carried a

1.27 Mb deletion on chromosome 4q35, containing two genes (*MTNR1A* and *FAT*); inherited from her father. Phenotype: born at term, prolonged delivery with asphyxia noted at birth. She presented mild cerebral palsy, language delay, autism (met criteria on ADI-R and ADOS), no ID (WAIS-R at 34 y: verbal IQ 78, performance IQ 105, full scale IQ 88), high-grade myopia, no dysmorphism, hyperflexible joints, no seizures.

- DECIPHER: Subject 260121, female, 13.4 Mb 18q12.1 deletion (chr18:20286120-33684898), *de novo*. Phenotype: ID, abnormality of the face, malformation of the heart and great vessels. Only CNV reported in the subject.
- DECIPHER: Subject 250878, male, 7.3 Mb deletion including *DTNA* (chr18:28338083-35619727), *de novo*. Phenotype: ID (full scale IQ 40-50), autism, delayed speech and language development, flat occiput, hypotelorism, narrow nasal bridge, narrow nares, narrow mouth, high palate, wide intermamillary distance, absent nipples, proximal placement of thumb, hypotonia, tall stature, abnormality of the male genitalia. This individual also carries a duplication involving 3 genes, *HAO1*, *TMX4* and *PLCB1* (chr20:7022125-8482355), inherited from a normal parent.
- DECIPHER: Subject 276030, male, 14.5 Mb deletion including *DTNA* (chr18:25269813-39765123), *de novo*. The deletion is mosaic (percentage of mosaicism not indicated). Phenotype: behavioral/psychiatric abnormality, autoagression and motor delay. Only CNV reported in the subject.
- ISCA: Subject nssv577635, unknown gender, 10.9 Mb 18q12.1 deletion including *DTNA* (chr18:25278473-36237614), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv577637, unknown gender, 11.9 Mb 18q12.1 deletion including *DTNA* (chr18:27945491-39904182), inheritance unknown. Phenotype: global developmental delay, muscular hypotonia, short stature.
- ISCA: Subject nssv1495767, unknown gender, 4 Mb deletion including *DTNA* (chr18:29207760-33198709), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv577638, unknown gender, 7.3 Mb 18q12.1 deletion including *DTNA* (chr18:29881080-37228316), *de novo*. Phenotype: global developmental delay, strabismus, stridor.
- ISCA: Subject nssv580430, unknown gender, deletion involving only *DTNA* (chr18:30131834-30456329), inheritance unknown. Only the first non-coding exon of the longer isoforms (2, 5, 7 and 17) is disrupted by the deletion, which is very similar to the one observed in the AGP proband 3477-3. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- Wang et al.[85]: Female, 4.1 Mb duplication including *DTNA* (chr18:26587700-30649280), *de novo*. Phenotype: autistic disorder (met criteria on ADI-R and ADOS), language delay, non-verbal, mild ID, focal epilepsy, short stature (5$^{th}$ centile), no dysmorphic features, mild myopia.
- DECIPHER: Subject 253426, male, 5.9 Mb duplication including *DTNA* (chr18:25834124-31721175), *de novo*. Phenotype: moderate ID (WAIS-IV full scale IQ 51), no ASD, recurrent seizures, facial dysmorphism, short stature, Chiari malformation, cryptorchidism, strabismus. Only CNV reported in the subject.

The smallest region of overlap among these 11 cases contains only the *DTNA* gene.

**SNV:**

- Ichida et al.[86]: Heterozygous missense mutation in *DTNA* in a large pedigree with left ventricular non compaction (family NVLNC-09), no neurological/behavioral phenotype described. No other *DTNA* mutations reported in the literature.

### 3) Functional evidence

- *DTNA* belongs to the dystrobrevin subfamily of the dystrophin family, like *DMD*, involved in Duchenne's/Becker muscular dystrophy, sometimes associated with ID/ASD.[87,88] Dystrobrevin alpha is part of the transmembrane dystrobrevin-associated protein complex (DPC), which participates in synaptic transmission at the neuromuscular junction, long-term memory and synaptic plasticity.[89]
- α-dystrobrevin knockout mice exhibit mild muscular dystrophy but show no obvious CNS defects, likely reflecting coexpression of the homolog β-dystrobrevin, which is predominantly expressed in the brain.[90] Double mutants lacking both α-dystrobrevin and β-dystrobrevin exhibit synaptic and behavioral defects similar to those seen in dystrophin-deficient mice.[91] Both dystrobrevin isoforms are required for the maturation and function of a subset of inhibitory synapses in the cerebellum and for correct execution of motor behaviors that depend on cerebellar integrity.[91]
- Expressed in the brain.

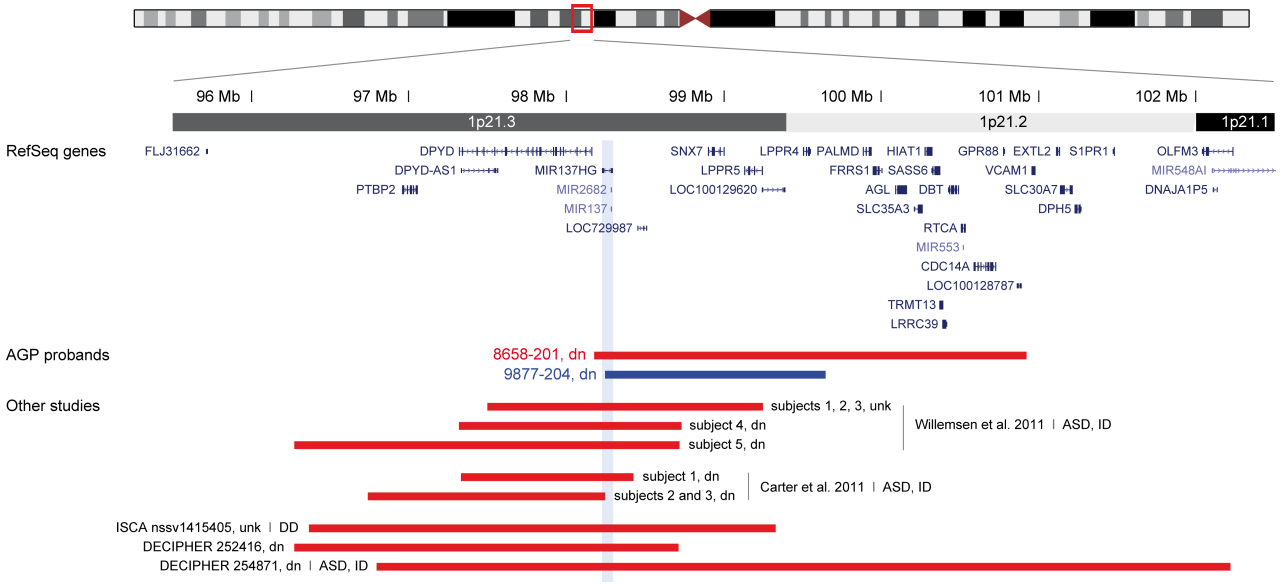
### 4) Other evidence

- % HI = 3.2 (highly likely to be haploinsufficient)
- Upregulation of *DTNA* has been reported in the temporal cortex of subjects with autism[92] and in the prefrontal cortex of individuals with bipolar disorder.[93]

- Dystrobrevin immunostaining is severely reduced at the sarcolemma of individuals with Duchenne muscular dystrophy and to a lesser extent in individuals with Becker muscular dystrophy.[94]

### 5) Comment

The gene function and the description of overlapping CNV would make this an excellent candidate gene, potentially pathogenic. However, the deletion in the AGP proband only affects the first non-coding exon of the longer isoforms, so it is difficult to know if it is deleterious. Expression studies are necessary to assess the effect on mRNA in this individual. Furthermore, description of other cases with small CNV overlapping *DTNA* or SNV are required to implicate this gene specifically. Indeed, given that the majority of the overlapping CNV are very large and contain many genes, it is difficult to ascribe pathogenicity to alterations of a single gene in the interval.




## Figure S15. CNV overlapping *MIR137* (microRNA 137) in chromosome 1p21.3

Chr1:95,500,000-102,500,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; DD, developmental delay; ID, intellectual disability; unk, unknown inheritance

### 1) AGP

- Proband 8658_201, female, 2.7 Mb 1p21.3-p21.2 deletion containing 22 genes including *MIR137* (chr1:98175622-100923952), *de novo*. Phenotype: autism on ADI and ADOS, comorbid ADHD, no language delay; low average IQ (WASI at 21 y: verbal IQ 78, performance IQ 88, full scale IQ 81); overweight, height and head circumference 50$^{th}$ centile, high pain tolerance, no epilepsy.
- Proband 9877_204, male, 1.4 Mb 1p21.3-p21.2 duplication containing 8 genes including *MIR137* (chr1:98247355-99645560), *de novo*. Phenotype: autism on ADI-R and ADOS, language delay (first words 36 mo, first phrases 60 mo), functional language, mild ID (WISC-R at 11 y: verbal IQ 51, performance IQ 80, full scale IQ 64), normal height and head circumference, weight -1.6 SD, no dysmorphic features, normal physical exam, no epilepsy.

  Both CNV in AGP probands also involve the *LPPR4* and *LPPR5* genes (lipid phosphate phosphatase-related proteins, types 4 and 5). *LPPR4* is specifically expressed in neurons and involved in axonal outgrowth during development and regenerative sprouting;[95] no haploinsufficiency score available (no information for *LPPR5*).
- No *MIR137* deletions among 4,768 controls; 1 carries a *MIR137* duplication (chr1:97673140-98319409); no CNV in 4,875 parents.

### 2) Other human genetic evidence

**CNV:**

- Willemsen et al.[96]: Chromosome 1p21.3 microdeletions comprising *DPYD* and *MIR137* associated with ID in 3 sibs and 2 unrelated subjects; the minimal region of overlap includes only *DPYD* and *MIR137*. The individuals displayed decreased expression of both precursor and mature miR-137 levels, as well as increased expression of the downstream targets *MITF*, *EZH2*, and *KLF4*. *DPYD* is involved in autosomal recessive dihydropyrimidine dehydrogenase deficiency; the significance of a defect in only one allele is uncertain.

Subjects 1, 2 and 3: siblings carrying a 1.75 Mb 1p21.3 deletion (chr1:97500000-99250000), inheritance unknown (parents deceased). Subject 1: male, borderline-mild ID (verbal IQ 65, performance IQ 90, full scale IQ 73), features of ASD, tendency to overeat, remarkably shy and friendly behavior, weight 90$^{th}$ centile, deep set eyes, broad nasal tip, long ears, thick lower lip, myopia. Subject 2: male, mild-moderate ID (verbal IQ < performance IQ 70, full scale IQ 52), features of ASD, tendency to overeat, self mutilation, aggressive outbursts, remarkably shy and friendly behavior, speech deficits, weight >98$^{th}$ centile, broad nasal tip, long ears, macrostomia, thick lower lip, astigmatism. Subject 3: female, mild-moderate ID (no other information available).
Subject 4: male, 1.41 Mb 1p21.3 deletion (chr1:97320000-98730000), *de novo*. Phenotype: mild ID (verbal IQ 59, performance IQ 71, global IQ 62), features of ASD, tendency to overeat, remarkably shy and friendly behavior, speech deficits, weight 98$^{th}$ centile, deep set eyes, astigmatism, myopia, broad nasal tip, full cheeks, thick lower lip, micrognathia and long ears.
Subject 5: female, 2.45 Mb 1p21.3 deletion (chr1:96270000-98720000), *de novo*. Phenotype: mild ID (global IQ 66), aggressive outbursts, remarkably shy and friendly behavior, weight >98$^{th}$ centile, full cheeks, long ears, thick lower lip.

- Carter et al.[97]: Subject 1: male, 1.1 Mb 1p21.3 deletion (chr1:97332167-98424667), *de novo*. Phenotype: severely delayed language, ID, autism. The individual also carries a balanced translocation t(9;21)(p13.3;q22.1) and has a *PTCHD1* missense mutation. Both abnormalities are inherited from the mother. The translocation was also transmitted to a healthy sister. Subjects 2 and 3: siblings with a 1.5 Mb 1p21.3 deletion (chr1:96742150-98243813), *de novo*. In their mother, the deleted region from chromosome 1p21.3 was inserted into chromosome 10. Subject 2: female, severe language delay, adaptive skills severely delayed, autism. Subject 3: male, language delay, no ID (full scale IQ 99), ASD. Both sibs had mild dysmorphic features, including upslanting palpebral fissures and small joint hypermobility.
- ISCA: Subject nssv1415405, unknown gender, 3 Mb 1p21.3 deletion (chr1:96362589-99332669), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- DECIPHER: Subject 252416, female, 2.4 Mb 1p21.3 deletion (chr1:96274145-98715464), *de novo*. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 254871, male, 5.4 Mb 1p21.3-p21.1 deletion (chr1:96792350-102220124), *de novo*. Phenotype: autism, ID, spotty hyperpigmentation, precocious puberty. Only CNV reported in the subject.

#### 3) Functional evidence

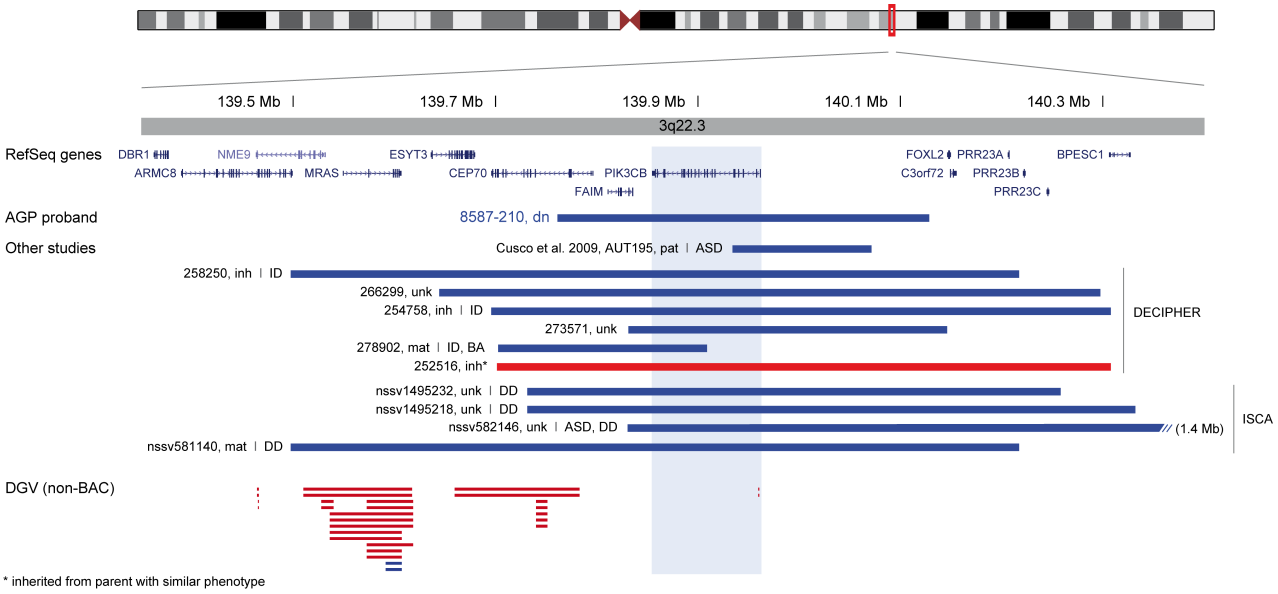
- The mature microRNA transcript miR-137 regulates neuronal maturation: overexpression of miR-137 inhibits dendritic and spine morphogenesis in newborn cells in the adult hippocampus and in cultured hippocampal neurons, whereas a reduction in miR-137 had opposite effects,[98] miR-137 has also been shown to modulate neurogenesis in adult neural stem cells.[99] Significant enrichment of miR-137 at the synapses of cortical and hippocampal neurons suggests a role in regulating local synaptic protein synthesis machinery.[96]
- Expressed in the brain, enriched in neurons, at the synaptic compartment.

#### 4) Other evidence

- An intronic SNP in *MIR137* was strongly associated with schizophrenia in a mega-analysis combining genome-wide association study data from over 40,000 individuals.[100]

#### 5) Comment

1p21.3 deletions of variable sizes overlapping *MIR137* reported in 11 individuals with ASD and/or ID. All cases in which inheritance is known originated *de novo*. Interestingly, both overexpression and inhibition of miR-137 had significant but opposite effects on dendritic development of hippocampal neurons, suggesting that the *MIR137* gene may be dosage sensitive, and that both the deletion and duplication observed in AGP probands could interfere with neuronal maturation.




## Figure S16. CNV identified in *PIK3CB* (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit beta) in chromosome 3q22.3

Chr3:139,350,000-140,400,000 (hg18). Abbreviations: ASD, autism spectrum disorder; BA, behavioral/psychiatric abnormality; dn, *de novo*; DD, developmental delay; inh, inherited; ID, intellectual disability; mat, maternal; pat, paternal; unk, unknown inheritance

### 1) AGP

- Proband 8587_210, female, 366 kb duplication involving *PIK3CB, CEP70* and *FAIM* (chr3:139760896-140127703), *de novo*. The entire *PIK3CB* gene is duplicated, which could result in increased expression, leading to excessive phosphatidylinositol 3-kinase (PI3K) activity. Phenotype: autism (ADI-R and ADOS positive), no language delay (first words 10 mo, first phrases 12 mo), verbal, WISC-III at 13 y 9 mo: verbal IQ 92, performance IQ 78, full scale IQ 84.
- No deletions or duplications of *PIK3CB* among 4,768 controls and 4,875 parents.

### 2) Other human genetic evidence

**CNV:**

- Cusco et al.[101]: Subject AUT195, male, 3q22.3 *PIK3CB* partial duplication (chr3:139934042-140070771), paternal. Phenotype: autism, mild ID, unilateral sensorineural deafness, no dysmorphism, no seizures. The functional consequences of a partial duplication are unknown.
- DECIPHER: Subject 258250, female, 720 kb 3q22.3 duplication involving *PIK3CB* (chr3:139497567-140217021), inherited from normal parent. Phenotype: cognitive impairment, Dandy-Walker malformation. Only CNV reported in the subject.
- DECIPHER: Subject 266299, male, 650 kb 3q22.3 duplication involving *PIK3CB* (chr3:139644292-140297350), inheritance unknown. Phenotype: microtia. This individual carries another duplication (chr3:160280710-160878458), inheritance unknown.
- DECIPHER: Subject 254758, male, 610 kb 3q22.3 duplication involving *PIK3CB* (chr3:139695765-140307747), inherited from normal parent. Phenotype: ID, short attention span, microcephaly, long face, high anterior hairline, downslanted palpebral fissures, depressed nasal tip, prominent nasal bridge, macrotia, micrognathia, abnormality of the pinna, prominent ears, slender build, scoliosis, mitral regurgitation, atrial septum defect, cryptorchidism, large hands. Only CNV reported in the subject.
- DECIPHER: Subject 273571, male, 315 kb 3q22.3 duplication involving *PIK3CB* (chr3:139830639-140145866), inheritance unknown. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 252516, female, 606 kb 3q22.3 deletion involving *PIK3CB* (chr3:139701603-140307606), inherited from parent with similar phenotype. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 278902, female, 207 kb 3q22.3 duplication involving *PIK3CB* (chr3:139701603-139908842), maternal. Phenotype: moderate ID and behavioral/psychiatric abnormality. The phenotype of the mother is unknown. Only CNV reported in the subject.

- ISCA: Subject nssv1495232, unknown gender, 527 kb 3q22.3 duplication involving *PIK3CB* (chr3:139730880-140257978), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv1495218, unknown gender, 600 kb 3q22.3 duplication involving *PIK3CB* (chr3:139730880-140331637), inheritance unknown. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv582146, unknown gender, 1.4 Mb 3q22.3-q23 duplication involving *PIK3CB* (chr3:139830647-141256588), inheritance unknown. Phenotype: autism, developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.
- ISCA: Subject nssv581140, unknown gender, 719 kb 3q22.3 duplication involving *PIK3CB* (chr3:139497567-140217080), maternal. Phenotype: developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing.

#### 3) Functional evidence

- *PIK3CB* encodes an isoform of the catalytic subunit PI3Kbeta of phosphoinositide 3-kinase (PI3K). These signaling molecules activate a wide range of downstream targets that regulate multiple cellular processes, including intracellular trafficking of proteins, cell proliferation, migration and survival.
- Expressed in the brain.

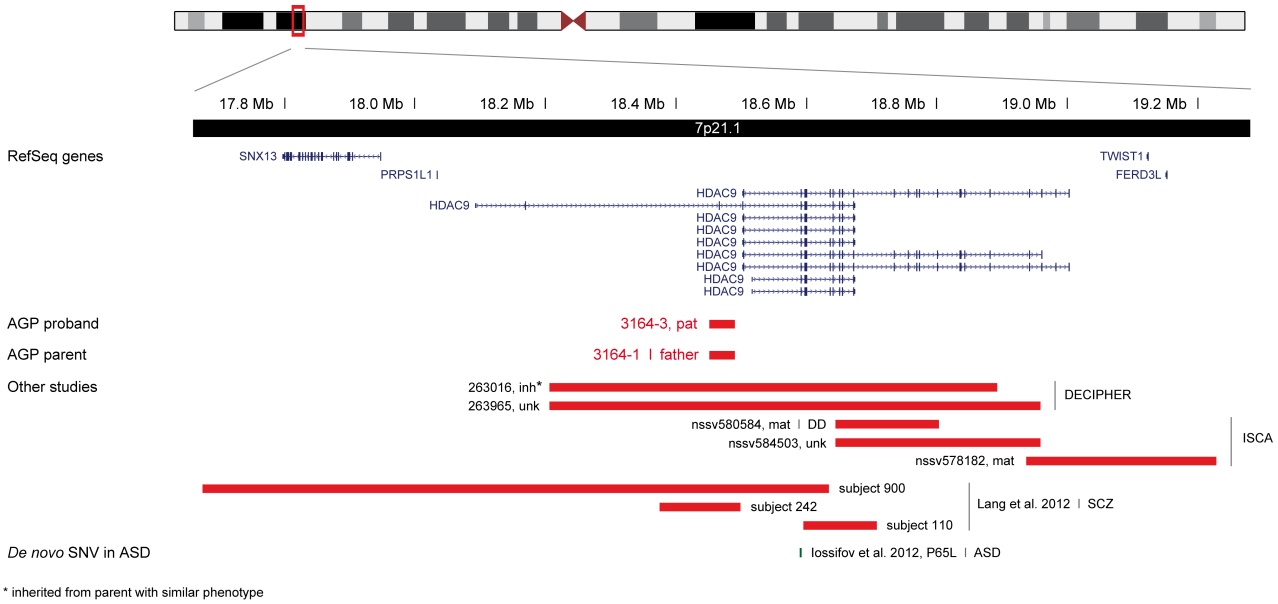
#### 4) Other evidence

- % HI = 0.5 (highly likely to be haploinsufficient)
- PI3K is regulated by the fragile X mental retardation protein (FMRP),[10] and is elevated in fragile X syndrome Fmr1-knockout mice,[102,103] suggesting that dysregulated PI3K signaling may underlie the synaptic impairments in fragile X syndrome. Accordingly, PI3K antagonists rescue fragile X syndrome phenotypes, including dysregulated synaptic protein synthesis, excess AMPA receptor internalization, and increased spine density.[102]

#### 5) Comment

The only report identified in the literature is a partial *PIK3CB* duplication in a subject with ASD, paternal;[101] the functional consequences of a partial duplication are difficult to predict. Eight cases of whole *PIK3CB* duplication are described in DECIPHER and ISCA, 4 are inherited and for the other 4 the inheritance is unknown. Thus, the duplication of *PIK3CB* identified in the AGP proband 8587-210 is the only one *de novo*. No similar duplications overlapping this gene were found among AGP controls, parents or population controls in DGV. Taken together these findings suggest that *PIK3CB* duplications could represent a risk factor for ASD/ID, associated with incomplete penetrance and/or variable expressivity. Further studies in larger samples of cases and controls are necessary to confirm this hypothesis.

## Examples of candidate genes affected by inherited CNV in AGP probands




## Figure S17. CNV and SNV identified in *HDAC9* (histone deacetylase 9) in chromosome 7p21.1

Chr7:17,660,000-19,280,000 (hg18). Abbreviations: ASD, autism spectrum disorder; dn, *de novo*; DD, developmental delay; inh, inherited; mat, maternal; pat, paternal; SCZ, schizophrenia; unk, unknown inheritance

### 1) AGP

- Proband 3164_3, male, rare 40 kb deletion involving only *HDAC9* (chr7:18450792-18490822), paternal. Phenotype: autism (ADI-R and ADOS positive), language delay (first words 54 mo, first sentences 60 mo), verbal, mild ID (PPVT-III at 6 y 10 m: verbal IQ 67); multiplex family, affected sibling not yet tested. Phenotype information about father not available.
- No *HDAC9* exonic deletions among 4,768 controls and 4,874 parents (1 deletion in the father of proband 3164-003).

### 2) Other human genetic evidence

**CNV:**

- DECIPHER: Subject 263016, female, *HDAC9* deletion (chr7:18206711-18892382), inherited from parent with similar phenotype. No phenotype information; only CNV reported in the subject.
- DECIPHER: Subject 263965, female, *HDAC9* deletion (chr7:18206712-18958442), inheritance unknown. No phenotype information; only CNV reported in the subject.
- ISCA: Subject nssv580584, unknown gender, *HDAC9* deletion (chr7:18644447-18803445), maternal. Phenotype: global developmental delay, seizures.
- ISCA: Subject nssv584503, unknown gender, *HDAC9* deletion (chr7:18644447-18958471), inheritance unknown. Phenotype: abnormal facial shape, facial asymmetry.
- ISCA: Subject nssv578182, unknown gender, 7p21.1 deletion of 3 genes: *HDAC9, TWIST1* and *FERD3L* (chr7:18937678-19227544), maternal. Phenotype: craniosynostosis (defects in *TWIST1* cause autosomal dominant craniosynostosis type 1).
- Lang et al.[104]: Three schizophrenia individuals with exonic *HDAC9* deletions among 3391 cases (inheritance unknown); no *HDAC9* deletion in 3181 controls.

**SNV:**

- Iossifov et al.[4] (exome sequencing in ASD): Subject 13076, male, *HDAC9* missense variant, *de novo*. The variant appears to be damaging (PolyPhen2, SIFT, PANTHER), affecting a highly conserved residue (GERP 5.93, ConSurf 9/9). Phenotype: ASD (no other information provided).

### 3) Functional evidence

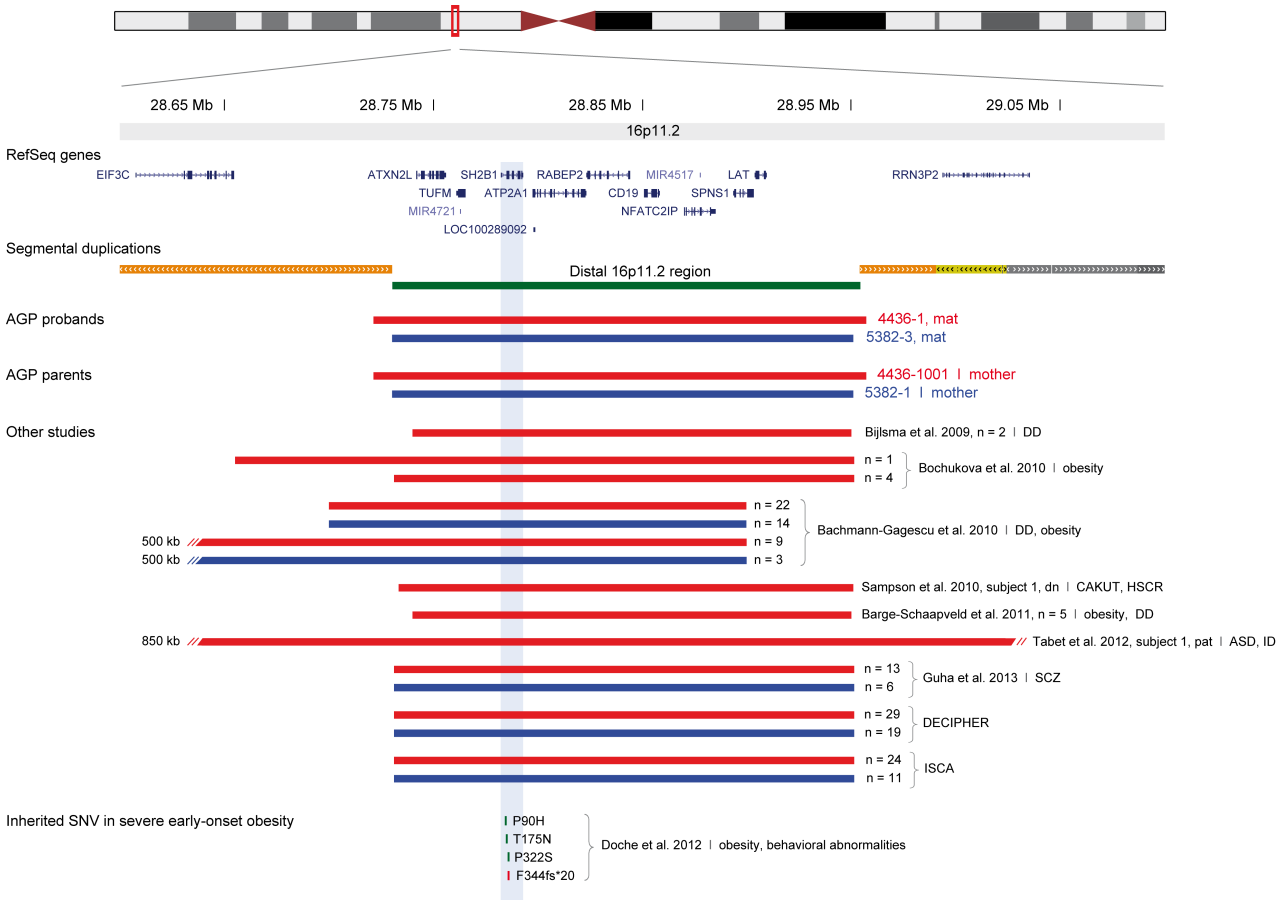
- *HDAC9* encodes a histone deacetylase, expressed in the brain. Histones play a critical role in transcriptional regulation, cell cycle progression, and development.
- Expressed in the brain.

### 4) Other evidence

- % HI = 2.9 (highly likely to be haploinsufficient)
- The *HDAC* family of genes has already been involved in ASD/ID through *HDAC4* (involved in brachydactyly-mental retardation syndrome) and *HDAC8* (X-linked ID, mutations are responsible for Cornelia de Lange syndrome).

### 5) Comment

In addition to the *HDAC9* paternal deletion identified in AGP proband 3164-003, we identified 5 other overlapping deletions in subjects in DECIPHER and ISCA, as well as three deletions in schizophrenia. The deletion was inherited in 3 subjects; no information was available for the others. No *HDAC9* exonic deletions were observed among the AGP 4,768 controls and 4,874 parents; in addition, no deletions overlapping this gene were found in DGV. Taken together these findings suggest that *HDAC9* deletions could represent a risk factor for ASD, ID and schizophrenia, associated with incomplete penetrance/variable expressivity. Further studies in larger samples of cases and controls are necessary to confirm this hypothesis.




## Figure S18. CNV identified in the distal 16p11.2 region containing *SH2B1* (SH2B adaptor protein 1)

Chr16:28,600,000-29,100,000 (hg18). Abbreviations: ASD, autism spectrum disorder; CAKUT, congenital anomalies of the kidney and urinary tract; DD, developmental delay; HSCR, Hirschsprung disease; ID, intellectual disability; mat, maternal; pat, paternal; SCZ, schizophrenia.

Longer CNV encompassing the distal 16p11.2 region and the proximal 16p11.2 region (29.5-30.1 Mb) involved in ASD, ID, and regulation of body mass index are not shown here.

### 1) AGP

- Proband 4436_1, male, distal 16p11.2 deletion including *SH2B1* (chr16:28721599-28957155), maternal. Phenotype: ASD diagnosis, no language delay (first words and phrases 15 mo), verbal, WISC-IV at 8 y 8 mo: verbal IQ 95, performance IQ 67, full scale IQ 76. Sporadic case.

- Proband 5382_3, male, distal 16p11.2 duplication including *SH2B1* (chr16:28730274-28950951), maternal. Phenotype: autism (ADI-R and ADOS positive), language delay (first words 42 mo, first phrases 48 mo), verbal; PPVT-IV verbal IQ 82, Leiter-R Brief performance IQ 97 (both at 11 y 10 m). Sporadic case.
- 1 distal 16p11.2 deletion among 4,768 controls and 1 deletion among 4,875 parents; no reciprocal duplications in controls or parents (the mother of proband 5382_3 was excluded from the microarray analyses after quality checks, her duplication was identified during qPCR validation of the CNV in her son).

### 2) Other human genetic evidence

**CNV:**

- Deletions at distal 16p11.2, with a minimal common region of 220 kb (28.73–28.95 Mb), have been implicated in early-onset obesity and developmental delay,[105-108] and in other variable phenotypes, including behavioral problems such as ASD and ADHD, anomalies of the kidney and urinary tract and Hirschsprung disease.[109,110] Whereas deletions appeared to be significantly enriched in populations with early-onset obesity or with developmental delay,[105,108] reciprocal duplications were not enriched in cases compared to controls.[105]
- A recent meta-analysis in large clinical cohorts with developmental delay, ID, ASD and congenital malformations referred for genetic testing found deletions at distal 16p11.2 in 23 of 31516 cases and in 2 of 13696 controls (OR 5, *P* = 0.01).[111] Reciprocal duplications were found in 25 of 31516 cases and in 3 of 13696 controls (OR 3.62, *P* = 0.02). Analysis of three ASD cohorts (AGP, SSC, and AGRE; n = 3955) found 1 deletion (OR 1.73, *P* = 0.53) and 1 duplication (OR 1.15, *P* = 1). The lack of a significant effect of these CNV in ASD was suggested to be due to the relatively small sample size.[111]
- Tabet et al.[110]: Male, 847 kb 16p11.2 distal deletion containing *SH2B1* (chr16:28401454-29249055), paternal. Autism, severely delayed speech, childhood-onset obesity, IQ 47. At age 19, he was tall (+2.5 SD), with troncular obesity (+4 SD). The father was described as being non talkative, introverted and having few social relationships.
- Guha et al.[112]: Deletions at distal 16p11.2 were reported in schizophrenia, in 13 of 13850 cases (0.094%) and 3 of 19954 controls (0.015%) (OR 6.25 [95% CI, 1.78-21.93]; *P* = 0.001). The rate of duplications in the region was not significantly different between cases and controls: 6 of 13850 cases (0.043%) vs 13 of 19954 controls (0.065%).

**SNV:**

- Doche et al.[113]: The minimal deleted interval contains nine genes, including *SH2B1*, which plays a role in the regulation of body weight and glucose metabolism in mice (see below). Mutation screening of *SH2B1* in 300 individuals with severe early-onset obesity revealed five mutations, one frameshift and three missense (including one found in two subjects). Mutation carriers exhibited childhood-onset obesity, hyperphagia, insulin resistance and short stature as adults. Neurobehavioral phenotypes included social isolation, speech and language delay and aggression. All mutations were inherited from overweight/obese parents reported to also have variable behavioral abnormalities. The mutations were absent from 500 controls.

### 3) Functional evidence

- *SH2B1* encodes an adaptor protein that binds to a large range of receptor tyrosine kinases and is thus involved in multiple biological pathways, including leptin and insulin signaling. The widely expressed scaffold protein SH2B1 binds to the receptors for nerve growth factor, insulin and insulin-growth factor 1, and has been implicated in neuronal differentiation and neurite outgrowth.[114,115]
- Expressed in the brain.
- *Sh2b1* deficient mice develop obesity and diabetes, a phenotype rescued by neuron-specific expression of *SH2B1*.[116]

### 4) Other evidence

- % HI = 18.7 (likely to be haploinsufficient)

### 5) Comment

*SH2B1* haploinsufficiency is clearly implicated in early-onset obesity. Recent evidence suggests that distal 16p11.2 deletions could also be involved in neurodevelopmental phenotypes, associated with incomplete penetrance and variable expressivity. Although a significant enrichment has been reported in samples with developmental delay/ID, the risk effect appears to be weak compared to other recurrent CNV.[111] The implication of deletions at distal 16p11.2 in schizophrenia[112] and the description of maladaptive behaviors in individuals carrying *SH2B1* mutations,[113] lend further support to their role as risk factors. The involvement of distal 16p11.2 duplications in ID and ASD is difficult to assess at present, since they have not been found to be consistently enriched in cases. Further studies, comparing the frequency of distal 16p11.2 deletions and duplications in larger samples of cases and controls, are needed to clarify the impact of these CNV in neurodevelopmental disorders.

# SUPPLEMENTAL TABLES

## Table S1A. Autism strict and spectrum classifications

| ASD diagnostic category | Phenotype classifications | |
|---|---|---|
| | ADI-R | ADOS |
| Strict | Autism | Autism |
| Spectrum | Autism | NA |
| | Autism | ASD |
| | ASD | Autism |
| | NA | Autism |

NA, not available or not administered

## Table S1B. Quality control – Family and control sample breakdown

| Quality control filters | Initial | Filter 1 Low call rate | Filter 2 Mendelian errors | Filter 3 Gender-mismatch | Filter 4 Duplicates | Filter 5 High LRR/ BAF SD | Filter 6 Excess calls | Filter 7 Excess *de novos* | Filter 8 Peri-centrom. | Filter 9 Large chrom. abnorm. | Filter 10 Incomplete phenotype data | Filter 11 All ancestries | Filter 12 European-only |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # Single probands | 51 | 56 | 55 | 55 | 55 | 90 | 106 | 106 | 106 | 106 | 106 | 106 | 102 |
| # Proband + mother duo | 10 | 30 | 30 | 30 | 30 | 100 | 125 | 125 | 125 | 131 | 126 | 126 | 106 |
| # Proband + father duo | 12 | 31 | 31 | 31 | 31 | 113 | 127 | 127 | 127 | 142 | 137 | 137 | 119 |
| # Complete trios* | 2772 | 2677 | 2620 | 2613 | 2606 | 2268 | 2161 | 2158 | 2155 | 2126 | 2077 | 2077* | 1820 |
| **Total # families** | **2845** | **2794** | **2736** | **2729** | **2722** | **2571** | **2519** | **2516** | **2513** | **2505** | **2446** | **2446** | **2147** |
| # Father + mother only (with or without relatives) | 258 | 277 | 273 | 273 | 273 | 309 | 324 | 324 | 326 | 332 | 380 | 199 | 196 |
| # Fathers only | 12 | 21 | 22 | 22 | 22 | 60 | 69 | 69 | 69 | 68 | 73 | 28 | 28 |
| # Mothers only | 16 | 27 | 27 | 27 | 27 | 52 | 60 | 60 | 61 | 62 | 67 | 32 | 32 |
| # Relatives only | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 4 |
| # Technical controls | 9 | 9 | 9 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| **Control datasets** | | | | | | | | | | | | | |
| **SAGE** (dbGaP)[a] | 1847 | 1847 | – | 1847 | 1829 | 1815 | 1793 | – | 1792 | 1769 | – | 1769 | 1166[#] |
| **OC** (European-only)[b#] | 511 | 509 | – | 509 | 501 | 475 | 440 | – | 437 | 433 | – | 433 | 234[#] |
| **HABC** (dbGaP)[c] | 2860 | 2860 | – | 2857 | 2809 | 2658 | 2571 | – | 2570 | 2566 | – | 2566 | 1240 |

**CNV detection and quality control evaluation**: For samples that passed the SNP and intensity QC, genome-wide CNVs were detected using a multiple-algorithm approach to maximize sensitivity and specificity of CNV calling. For a detailed description see the Supplemental information of Pinto et al.[117] Briefly, CNVs were identified by using QuantiSNP,[118] iPattern,[119] and PennCNV;[120] the family-based CNV detection option of PennCNV was used to confirm inheritance.

We excluded CNVs when they failed stringent QC criteria: <5 probes and low confidence score (QuantiSNP log Bayes factor <15); if CNVs resided in regions of extreme GC content (>70%); or if they were within centromere proximal cytobands. CNVs detected by QuantiSNP and iPattern in one individual with a minimum of 5 consecutive probes covering at least 5 kb of sequence were merged using outer probe boundaries (i.e., union of the CNVs). All CNVs by any algorithm with size larger than 1 Mb were inspected manually, and all samples that passed all above QC filters were inspected for the presence of large abnormalities in chromosomes X and Y (that is, in addition to the algorithm calling). As a final step, we joined CNVs that appeared to be artificially split by either of the calling algorithms and also removed CNVs that spanned known large assembly gaps in hg18 (greater than 200 kb).

**Filter descriptions**: A total of 9,050 individuals from 2,845 ASD families were genotyped as part of Stages 1([117]) and 2 and those passing QC filters were used in the rare CNV analysis. Incomplete families, where proband-father/mother duos passed QC filters were also analysed for CNVs. The number and the composition of families remaining after each filtering step is indicated. Counts may increase or decrease after each step, as removal of individuals in some instances will break complete trios into proband+parent duos or single probands. **Filter 1**: Low call rate or high missingness. **Filter 2**: High Mendelian error rate (with or without proband gender mismatch); families with unresolved gender mismatches were excluded; families where both parents had a gender mismatch without high Mendelian error rate were kept and the parents' gender swapped. Families with high Mendelian rate for one parent only were excluded. **Filter 3**: Proband with gender mismatch but no Mendelian error (i.e., another sib of the same family was genotyped instead) resulted in exclusion of the whole family. Technical controls with gender mismatches were excluded. **Filter 4**: Any duplicate samples/families had one sample/family excluded. **Filter 5**: Samples with high standard deviation (SD) of log R ratio of intensities (LRR) and/or B allele frequency (BAF), or showing extreme/wide intensities. **Filter 6**: Samples with excess of CNV calls by at least one of the algorithms, except those with fragmented calls due to large chromosome abnormalities. **Filter 7**: Samples with excess of *de novos* (which were confirmed to be false positives with experimental validation). **Filter 8**: Lack of CNV calls after three filters (removing pericentromeric calls, <30 kb size, 50% overlap with segmental duplication

blocks). **Filter 9**: Large chromosomal abnormalities >7.5 Mb. A list of chromosome abnormalities detected in probands can be found in **Table S1C**, and for parents and controls in **Table S17B**. Subjects were removed if one or more CNVs were found to be cell-line artifacts after experimental validation. If a parent failed QC at this step, his/her family was not excluded. **Filter 10**: No phenotype data in database or proband did not meet full criteria for ASD. **Filter 11**: Parents that passed QC but are parents of probands that failed QC. **Filter 12**: European ancestry only.

**Ancestry:** Ancestry for each of the four samples (AGP cases, and SAGE, HABC, and OC controls) was inferred by eigenvector decomposition and clustering. To identify European subjects from the Ontario controls, we used the multidimensional scaling (MDS) function of PLINK to cluster the OC subjects with HapMap-CEU. The remaining three samples had been described and analyzed for ancestry in previous studies.[117,121,122] We used the results from those published studies to identify subjects of inferred European ancestry for this study. In each of the published studies, ancestry was inferred by using SpectralGem[123] to analyze thousands of high quality SNPs genotyped for all subjects. After the clustering step from SpectralGem, it was simple to identify groups of European ancestry because a substantial fraction of the contributing AGP sites were European.

Data from 2,446 families passed all QC steps (13% of subjects excluded), adding 1,359 new families to the combined analysis. Of the new families, 1,168 families were European and 191 were of other ancestries. As described in detail in Pinto et al.[117], to avoid confounding by ancestry, all downstream CNV analyses used European-only cases (n=2,147) and controls (n=2,640). For the analyses presented in **Figure S1**, we extended the number of controls to include 1,843 subjects from other ancestries —517 SAGE and 1,326 HABC non-Europeans controls that passed QC— giving a total of 4,768 control subjects (2,022 males and 2,746 females) from all ancestries to be compared to ASD cases of all ancestries.

***** *de novo* CNVs were detected and confirmed in a total of 2077 complete trios passing array QC plus 19 families in which at least one of the parents failed initial array QC but additional experimental validation in both parents confirmed the presence of *de novo* CNVs, giving a total of 2096 complete trios of all ancestries (1838 European trios = 1820 + 18) studied.

[a] **Study of Addiction: Genetics and Environment (SAGE) cohort**: Both raw intensities and genotypes were obtained from 1,847 SAGE control subjects from NHGRI-dbGaP (accession: phs000092.v1.p1) as part of the larger SAGE case-control study[124]. The consented sample included 31% males and 69% females, with a mean age of 39.2 y (SD 9.1); 73% of subjects self-identified as European-American, 26% as African-American and 1% as other (http://zork.wustl.edu/gei /study_description.html). Subjects may have had exposure to alcohol (and possibly to other drugs), but did not meet criteria for alcohol or other drug dependence. The subset of control dataset used in the specific CNV analyses in this paper is composed of 1,166 unrelated European control samples that passed all quality control filters (75% had DNA extracted from whole blood and 25% from cell lines), composed of 370 males and 796 females.

[b] **Ontario Colorectal Cancer case-control study cohort (OC)**: 433 unrelated European control subjects from the population-based Colorectal Cancer case-control study, recruited randomly from the province of Ontario in Canada (Ontario Familial Colorectal Cancer Registry, OFCCR) as described elsewhere[125] and genotyped with the Illumina 1M single array.[126] The OC control sample consisted of 199 females and 234 males with mean age of 61.8 y (range: 27-78); all subjects were self-identified as non-Hispanic whites and estimated to be of European ancestry from their genotypes. All DNA samples were extracted from whole blood. The GenomeStudio v. 2010.3, with the clustering algorithm GenTrain2 and a GenCall cutoff of 0.15 was used to generate genotypes. The same quality control procedures applied to the ASD family samples and SAGE controls were used here. For the main CNV analysis, we used only the 234 OC males.

[c] **HealthABC (HABC)**: 1,240 unrelated European control subjects from the whole-genome-association study of visceral adiposity in the Health, Aging, and Body Composition (HealthABC or HABC)[127] were used in the main analyses. The HABC cohort studied the factors that contribute to disability and the decline in function of healthier older persons, with a particular emphasis on changes in body composition. The HealthABC study recruited 3,075 70-79 year-old community-dwelling adults (41% African-American, remainder were white-European), who were initially free of mobility and activities of daily living disability. Genotyping was performed by the Center for Inherited Disease Research (CIDR) using the Illumina Human1Mv3 (duo) BeadChip system like most of the AGP data, providing excellent comparability with the case dataset. Both raw intensities and genotypes were obtained from JAAMH-dbGaP (accession: phs000169.v1.p1) for 2,860 samples, resulting in 1,240 European (637 males and 603 females) and 1,326 non-European samples after QC. Samples were excluded from the dataset in case of sample failure, genotypic sex mismatch, or first-degree relative of an included individual based on genotype data.

[#] To try to balance the number of male and female controls used in the various analyses (by stage and by platform), we only included male samples from the OC dataset (n=234), and excluded 86 female SAGE samples of European ancestry (i.e., all males [n=370] and 796 females were included).

Abbreviations: chrom. abnorm., chromosomal abnormalitities; pericentrom., pericentromeric; M, males.

## Table S1C. Quality control – Chromosome abnormalities detected in probands

(Chromosome abnormalities in parents and controls are listed in **Table S17B**)

| Stage | Sample ID | Sex | Family type | DNA source | Cytoband | Anomaly type | Karyotype | Comment |
|---|---|---|---|---|---|---|---|---|
| **Confirmed chromosomal abnormalities** | | | | | | | | |
| 1 | 5467_3 | M | SPX | Blood | 1q42.3-q44 | High intensity, BAF split | N/A | 13.7 Mb 1q42-q44 duplication (1:233476547-247165725), *de novo*; confirmed by qPCR |
| 1 | 14270_3930 | F | SPX | Blood | 6q25.3-q27 | High intensity, BAF split | 46,XX.ish der(22)t(6;22)(6q25.3;p11.2)pat(6qtel+) | 10 Mb 6q25.3-q27 duplication (6:160773919-170761395); confirmed by FISH, resulting from a balanced translocation in the father |
| 1 | 13137_1543 | F | UNK | Blood | 8p12-8q12.1 | 8p-q duplication | 47,XX,+r[10]/46,XX[70] | 27 Mb 8p12-8q12.1 duplication (8:31928590-58996070); karyotype: mosaic supernumerary ring chromosome |
| 1 | 5420_3 | M | UNK | CL | Whole chr 21 | High chr 21 intensity, BAF split | 47,XY+21 | Down syndrome, confirmed by karyotype |
| 1 | 5266_4 | F | UNK | Blood | Whole chr 2 | Run of homozygosity entire chr 2 | N/A | *De novo* uniparental disomy chr 2, confirmed in blood DNA |
| 2 | 21020_1 | M | SPX | CL (array), blood (qPCR & FISH) | 4p16 | 4p16.3p16.1 DUP | 46,XY, der(8)t(4;8)(4p16.1→ter; 8p23.1→ter) | *De novo* unbalanced translocation t(4p16;8p23) leading to 4p16.3p16.1 duplication (4:53403-9016339) and 8p23.3p23.1 deletion; maternal origin by microsatellite markers. Normal karyotype, translocation validated by FISH and qPCR; the 4p terminal duplication (8.9 Mb) includes the Wolf-Hirschhorn syndrome region |
| 2 | 21020_1 | M | SPX | (see above) | 8p23 | 8p23.3p23.1 DEL | (see above) | *De novo* unbalanced translocation t(4p16;8p23) (see above). The 6.8 Mb 8p23.3p23.1 terminal deletion (8:154984-6994825) overlaps numerous deletions described in individuals with ID |
| 2 | 2175_1 | M | SPX | Blood | 22q13.2-qter | 22q13.2-qter BAF split (no visible LRR deflection) | N/A | *De novo* 10 Mb 22q13.2-qter uniparental disomy (22:39729380-49582267), mosaicism in approximately 24% of cells; normal MLPA |
| 2 | 4316_1 | F | UNK | Blood | Whole chr X | Whole chr BAF split (no visible LRR deflection) | N/A | Triple X syndrome, *de novo*, mosaic |
| 1 | 5257_3 | M | SPX | Blood | Whole chr Y | Male, high chr Y intensity | 47,XYY | XYY syndrome, *de novo*, confirmed by karyotype |
| 1 | 5515_3 | M | SPX | Blood | Whole chr Y | Male, high chr Y intensity | 47,XYY | XYY syndrome, *de novo*, confirmed by karyotype |
| **Other abnormalities resulting from cell line artifacts or not validated in blood** | | | | | | | | |
| 1 | 5010_3 | M | MPX | CL | 1q43-qter DEL, 9q13-qter DUP | 1q43-q44 low intensity & BAF split; 9q arm high intensity & little BAF split | 47,XY | 1q43-qter 9.4 Mb DEL, mosaic (1:237816283-247249719), 9q13-qter 69.8 Mb DUP, mosaic (9:70400178-140273252); found in the same DNA batch by Affy 500K-EA; normal karyotype; parents normal; AffyCytoScanHD on blood DNA showed normal results: cell line artifact |
| 1 | 5321_3 | M | SPX | CL | Whole chr 4 | Run of homozygosity entire chr 4 | N/A | Uniparental disomy chr 4, confirmed in cell line; no blood DNA available so we can't exclude a cell line artifact |
| 1 | 6379_4 | M | SPX | CL | Whole chr 14 | High intensity, BAF split | 46,XY | Whole chr 14 duplication by one algorithm only; karyotype excluded a chr14 trisomy: cell line artifact |

Samples that passed QC filters but showed CNVs by any algorithm larger than 7.5 Mb, long runs of homozygosity, or CNVs encompassing/or close to the centromere on any of the autosomes or chromosome X were further inspected manually by plotting their log2 ratio intensities as well as allelic genotype ratios. A cutoff of >7.5 Mb was selected to be consistent with large cytogenetically visible chromosome abnormalities. Samples containing such alterations were excluded from the main burden analyses, except for two AGP probands with XYY syndrome that were retained since chromosome Y markers were not used in the CNV analysis, and two probands with uniparental disomy (proband 2175_1 with a segmental uniparental disomy of chr 22q was excluded during QC because of excessive number of calls).

Abbreviations: BAF, B allele frequency; CL, cell line; DEL, deletion; DUP, duplication; EA: early access (Affy500K-EA vs. Affy500K); F, female; LRR; log R ratio; M, male; MPX, multiplex; N/A, not available; SPX, simplex; UNK, unknown family type (extended family not evaluated for ASD).

## Table S2A. Sample characteristics

| Classification | | Sex | Stage 1 (n=1087) | | Stage 2 (new cases, n=1359) | | Combined (n=2446) | |
|---|---|---|---|---|---|---|---|---|
| | | | European n=979[a] | Other n=108 | European n=1168 | Other n=191 | European n=2147 | Other n=299 |
| ASD[b] | | M | 826 | 90 | 1027 | 171 | 1853 | 261 |
| | | F | 153 | 18 | 141 | 20 | 294 | 38 |
| Developmental impairment[c] | Without DI | M | 433 | 51 | 569 | 65 | 1002 | 116 |
| | | F | 67 | 11 | 63 | 9 | 130 | 20 |
| | With DI | M | 361 | 35 | 413 | 103 | 774 | 138 |
| | | F | 84 | 7 | 72 | 11 | 156 | 18 |
| Family type[d] | Simplex | M | 366 | 33 | 660 | 132 | 1026 | 165 |
| | | F | 62 | 3 | 78 | 14 | 140 | 17 |
| | Multiplex | M | 330 | 53 | 208 | 30 | 538 | 83 |
| | | F | 59 | 14 | 43 | 5 | 102 | 19 |
| | Unknown | M | 131 | 4 | 159 | 9 | 290 | 13 |
| | | F | 31 | 1 | 20 | 1 | 51 | 2 |

[a] 17 of the 996 (1.7%) European cases used in Pinto et al.[117] were excluded from the combined sample after applying additional QC steps.

[b] Subjects met criteria for strict autism or autism spectrum according to the ADI-R and/or ADOS (see Methods for detailed description).

[c] Developmental impairment is a hierarchical classification based on scores on full-scale IQ, performance IQ, verbal IQ and the Vineland Adaptive Behavior Scales composite score (see Methods for detailed description). A cut-off score of 70 was applied on all measures. Some data are missing on this item, ranging from 2%-4%.

[d] Family-history reports were taken to inform on the family type. Multiplex families had at least two individuals receiving validated ASD diagnoses who were first to third degree relatives (for third degree, only cousins were considered). Simplex families had only one known individual with ASD in first to third (cousin) degree relatives. Families that did not fall into the multiplex or simplex criteria above were classified as unknown.

The ratio of simplex to multiplex families was 1.82 (1348:742), with 356 (14.5%) families of unknown status. The ratio of males to females was 6.4:1; 46% of cases (n=1,086 out of 2,354) showed developmental impairment (DI). There were 1.43 times more females than males with DI compared to no-DI (M:F for DI is 5.2 vs. 7.45 for no-DI, 95% CI 1.12-1.82, chi square p=0.003). When considering family type, there were 1.47 times more males than females in simplex compared to multiplex families (95% CI 1.13-1.91, chi square p=0.0039), and simplex cases had 1.57 fold more DI compared to multiplex cases (95% CI 1.30-1.90, chi-square p=0.004).

Simplex males: 1026+165=1191
Simplex females: 140+17=157
Simplex M:F ratio: 1191/157= 7.59:1

Multiplex males: 538+83=621
Multiplex females: 102+19=121
Multiplex M:F ratio: 621/121= 5.13:1

# Table S2B. Sample characteristics (*continued*)

## Cases

| | European | European Males | European Females | Other | Total |
|---|---|---|---|---|---|
| Stage 1 | 979 | 826 | 153 | 108 | 1,087 |
| Stage 2 | 1,168 | 1,027 | 141 | 191 | 1,359 |
| Stage 1 + 2 (all cases) | 2,147 | 1,853 | 294 | 299 | 2,446 |

## Complete Trios

| | European | European Males | European Females | Other | Total (1) | Total (2) |
|---|---|---|---|---|---|---|
| Stage 1 | 862 | 733 | 129 | 91 | 953 | n/a |
| Stage 2 | 958 | 845 | 113 | 166 | 1,124 | n/a |
| Stage 1 + 2 (all trios) | 1,820 | 1,578 | 242 | 257 | 2,077 | 2,096 |

(1) trios with array data after QC available from both parents

(2) trios in (1) plus additional trios with laboratory validation data from both parents;
n/a: not applicable

## Cases, breakdowns

| | Stage 1+2 Males | | Stage 1+2 Females | | Stage 1+2 All | |
|---|---|---|---|---|---|---|
| | All ancestries | European | All ancestries | European | All ancestries | European |
| Single probands | 92 | 88 | 14 | 14 | 106 | 102 |
| Proband+mother duo | 101 | 83 | 25 | 23 | 126 | 106 |
| Proband+father duo | 120 | 104 | 17 | 15 | 137 | 119 |
| Complete trios | 1,801 | 1,578 | 276 | 242 | 2,077 | 1,820 |
| Total probands | 2,114 | 1,853 | 332 | 294 | 2,446 | 2,147 |

## Cases with rare CNVs >30 kb

| | European (all) | European (genic)* | European (exonic) |
|---|---|---|---|
| Stage 1 | 868 | 747 | 691 |
| Stage 2 | 1,046 | 871 | 795 |
| Stage 1 + 2 (all cases) | 1,914 | 1,618 | 1,486 |
| Stage 1 + 2 (all controls) | 2,359 | 1,971 | 1,820 |
| Total sample (cases + controls) | 4,273 | 3,589 | 3,306 |

* with 10 kb-flanking

## Controls

| PRIMARY controls | European | European Males | European Females |
|---|---|---|---|
| Stage 1: 1166 SAGE | 1,166 | 370 | 796 |
| Stage 2: 1240 HABC + 234 OC | 1,474 | 871 | 603 |
| Totals | 2,640 | 1,241 | 1,399 |

## Additional controls

| SECONDARY controls | Other ancestries | Other ancestries Males | Other ancestries Females |
|---|---|---|---|
| Stage 1: SAGE | 517 | 185 | 332 |
| Stage 2: HABC | 1,326 | 596 | 730 |
| Totals | 1,843 | 781 | 1,062 |
| Controls All ancestries - totals | 4,483 | 2,022 | 2,461 |

# Tables S3A-S3C. CNV burden

## Table S3A. Platform comparison

### CNV rate

| Type | Group | 1M single: 1382 cases / 1400 controls # rare CNVs | Baseline CNV rate (Ctrl) | Case/ Ctrl ratio | P value | 1M duo: 765 cases / 1240 controls # rare CNVs | Baseline CNV rate (Ctrl) | Case/ Ctrl ratio | P value |
|---|---|---|---|---|---|---|---|---|---|
| All | All | 6,506 | 2.32 | 1.02 | 0.27621 | 4,407 | 2.16 | 1.05 | 0.06357 |
| DEL | All | 3,353 | 1.17 | 1.07 | 0.03401 | 2,248 | 1.10 | 1.05 | 0.13466 |
| DUP | All | 3,153 | 1.15 | 0.96 | 0.83521 | 2,159 | 1.06 | 1.05 | 0.14956 |
| All | 30 – 500 kb | 6,088 | 2.18 | 1.01 | 0.39300 | 4,201 | 2.08 | 1.02 | 0.25742 |
| | ≥ 500 kb | 418 | 0.14 | 1.15 | 0.09388 | 206 | 0.08 | 1.79 | 0.00005 |
| | ≥ 1 Mb | 127 | 0.04 | 1.24 | 0.13127 | 76 | 0.03 | 2.00 | 0.00214 |
| DEL | 30 – 500 kb | 3,235 | 1.12 | 1.07 | 0.03287 | 2,187 | 1.08 | 1.02 | 0.34081 |
| | ≥ 500 kb | 118 | 0.04 | 1.01 | 0.50979 | 61 | 0.02 | 2.87 | 0.00005 |
| | ≥ 1 Mb | 42 | 0.01 | 1.65 | 0.08491 | 24 | 0.00 | 4.86 | 0.00026 |
| DUP | 30 – 500 kb | 2,853 | 1.06 | 0.94 | 0.93625 | 2,014 | 1.00 | 1.02 | 0.31434 |
| | ≥ 500 kb | 300 | 0.10 | 1.20 | 0.06567 | 145 | 0.06 | 1.47 | 0.01390 |
| | ≥ 1 Mb | 85 | 0.03 | 1.09 | 0.39456 | 52 | 0.02 | 1.39 | 0.15035 |

### Total CNV size (kb)

| Type | Group | 1M single: 1382 cases / 1400 controls Baseline size (Ctrl) | Case/ Ctrl ratio | P value | 1M duo: 765 cases / 1240 controls Baseline size (Ctrl) | Case/ Ctrl ratio | P value |
|---|---|---|---|---|---|---|---|
| All | All | 442.7 | 1.10 | 0.03171 | 365.1 | 1.39 | 0.00001 |
| DEL | All | 222.2 | 1.12 | 0.06035 | 190.1 | 1.35 | 0.00011 |
| DUP | All | 362.8 | 1.08 | 0.10991 | 305.1 | 1.41 | 0.00005 |
| All | 30 – 500 kb | 298.9 | 0.99 | 0.56721 | 285.8 | 1.06 | 0.07551 |
| | ≥ 500 kb | 995.1 | 1.24 | 0.00201 | 998.7 | 1.44 | 0.00534 |
| | ≥ 1 Mb | 1,621.0 | 1.27 | 0.00651 | 1,519.0 | 1.59 | 0.00514 |
| DEL | 30 – 500 kb | 167.9 | 1.02 | 0.28813 | 168.7 | 0.97 | 0.70923 |
| | ≥ 500 kb | 948.0 | 1.45 | 0.00841 | 964.5 | 1.40 | 0.08070 |
| | ≥ 1 Mb | 1,551.0 | 1.51 | 0.02177 | 1,411.0 | 1.42 | 0.12728 |
| DUP | 30 – 500 kb | 233.9 | 0.98 | 0.72926 | 223.3 | 1.11 | 0.01821 |
| | ≥ 500 kb | 985.7 | 1.12 | 0.08485 | 980.7 | 1.45 | 0.01585 |
| | ≥ 1 Mb | 1,649.0 | 1.11 | 0.15156 | 1,487.0 | 1.77 | 0.00600 |

### Gene count

| Type | Group | 1M single: 1382 cases / 1400 controls # Rare genic CNVs | # Genes inters. by rare CNVs | Baseline gene rate (Ctrl) | Case/ Ctrl gene ratio | P value | Pcorr | 1M duo: 765 cases / 1240 controls # Rare genic CNVs | # Genes inters. by rare CNVs | Baseline gene rate (Ctrl) | Case/ Ctrl gene ratio | P value | Pcorr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| All | All | 4,180 | 5,210 | 4.11 | 1.22 | 0.00140 | 0.00120 | 2,681 | 3,587 | 3.12 | 1.64 | 0.00001 | 0.00001 |
| DEL | All | 1,806 | 2,040 | 1.37 | 1.33 | 0.00122 | 0.01068 | 1,164 | 1,469 | 1.16 | 1.40 | 0.00146 | 0.05123 |
| DUP | All | 2,374 | 4,002 | 2.74 | 1.17 | 0.03831 | 0.00824 | 1,517 | 2,597 | 1.95 | 1.78 | 0.00001 | 0.00004 |
| All | 30 – 500 kb | 3,800 | 3,795 | 3.12 | 1.03 | 0.30338 | 0.33000 | 2,496 | 2,780 | 2.71 | 1.12 | 0.02734 | 0.02645 |
| | ≥ 500 kb | 380 | 2,011 | 0.99 | 1.84 | 0.00034 | 0.00234 | 185 | 1,039 | 0.41 | 5.07 | 0.00001 | 0.00019 |
| | ≥ 1 Mb | 122 | 1,075 | 0.37 | 2.65 | 0.00096 | 0.00231 | 71 | 641 | 0.22 | 7.10 | 0.00001 | 0.00329 |
| DEL | 30 – 500 kb | 1,700 | 1,407 | 1.09 | 1.15 | 0.02952 | 0.10020 | 1,113 | 1,028 | 1.08 | 0.92 | 0.84269 | 0.86370 |
| | ≥ 500 kb | 106 | 732 | 0.28 | 2.05 | 0.00992 | 0.01546 | 51 | 485 | 0.08 | 7.80 | 0.00001 | 0.00589 |
| | ≥ 1 Mb | 41 | 504 | 0.13 | 2.80 | 0.01122 | 0.11560 | 23 | 339 | 0.04 | 10.00 | 0.00005 | 0.05489 |
| DUP | 30 – 500 kb | 2,100 | 2,849 | 2.03 | 0.96 | 0.70648 | 0.43300 | 1,383 | 2,045 | 1.63 | 1.26 | 0.00239 | 0.00397 |
| | ≥ 500 kb | 274 | 1,561 | 0.72 | 1.76 | 0.00545 | 0.01727 | 134 | 697 | 0.33 | 4.38 | 0.00008 | 0.01109 |
| | ≥ 1 Mb | 81 | 721 | 0.24 | 2.57 | 0.01518 | 0.00677 | 48 | 391 | 0.17 | 6.38 | 0.00086 | 0.08588 |

Baseline CNV rate (Ctrl): average number of CNVs per control subject
Baseline size (Ctrl): average total size per control subject in kb
Baseline gene rate (Ctrl): average number of genes intersected by CNVs per control subject

Pcorr: corrected for global differences in CNV size and rate
100,000 permutations

Global burden analyses for rare CNVs were performed using PLINK v1.0730, R stats and custom scripts, as previously described in detail.[117] We tested for global increased burden in a combined set of 2,147 European ASD cases compared to 2,640 European controls for three measures: CNV rate, CNV size (**Tables S3A-S3B**) and the average number of genes affected by CNVs (gene-count) (**Table 1**, main text). We observed a significantly increased burden in the number of genes affected by rare CNVs in cases over controls (**Table 1**). This enrichment for rare genic CNVs was apparent for both deletions and duplications, and remained after we further controlled for potential case-control differences that could be present due to biological differences or technical biases (**Table 1**). Similar findings were obtained when data were broken down by array type (**Table S3A**), or when each stage was considered separately (**Table S3B**).

# Table S3B. Stage 1 (Pinto et al.) versus Stage 2 (new cases) versus Combined (all 2147 European cases)

## CNV rate

| | | Stage 1: 979 cases / 1166 controls | | | | Stage 2: 1168 cases / 1474 controls | | | | Combined stage 1 + stage 2: 2147 cases / 2640 controls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | Group | # rare CNVs | Baseline CNV rate (Ctrl) | Case/ Ctrl ratio | P value | # rare CNVs | Baseline CNV rate (Ctrl) | Case/ Ctrl ratio | P value | # rare CNVs | Baseline CNV rate (Ctrl) | Case/ Ctrl ratio | P value |
| All | All | 5,153 | 2.44 | 0.97 | 0.85304 | 5,923 | 2.16 | 1.09 | 0.00066 | 11,076 | 2.28 | 1.03 | 0.05804 |
| DEL | All | 2,664 | 1.23 | 1.02 | 0.29184 | 3,063 | 1.11 | 1.09 | 0.00554 | 5,727 | 1.16 | 1.06 | 0.01238 |
| DUP | All | 2,489 | 1.21 | 0.91 | 0.98261 | 2,860 | 1.05 | 1.08 | 0.02640 | 5,349 | 1.12 | 1.00 | 0.49532 |
| All | 30 – 500 kb | 4,830 | 2.29 | 0.96 | 0.90681 | 5,634 | 2.08 | 1.06 | 0.01827 | 10,464 | 2.17 | 1.01 | 0.27705 |
| | ≥ 500 kb | 323 | 0.14 | 1.11 | 0.19004 | 289 | 0.08 | 1.88 | 0.00002 | 612 | 0.11 | 1.43 | 0.00003 |
| | ≥ 1 Mb | 91 | 0.04 | 1.02 | 0.50026 | 106 | 0.02 | 2.45 | 0.00003 | 197 | 0.03 | 1.62 | 0.00066 |
| DEL | 30 – 500 kb | 2,575 | 1.19 | 1.03 | 0.27934 | 2,978 | 1.09 | 1.07 | 0.03504 | 5,553 | 1.14 | 1.05 | 0.04012 |
| | ≥ 500 kb | 89 | 0.04 | 0.97 | 0.59513 | 85 | 0.02 | 2.71 | 0.00002 | 174 | 0.03 | 1.59 | 0.00173 |
| | ≥ 1 Mb | 27 | 0.01 | 1.28 | 0.32376 | 38 | 0.01 | 4.73 | 0.00005 | 65 | 0.01 | 2.58 | 0.00024 |
| DUP | 30 – 500 kb | 2,255 | 1.11 | 0.89 | 0.99523 | 2,656 | 0.99 | 1.05 | 0.13460 | 4,911 | 1.04 | 0.97 | 0.82515 |
| | ≥ 500 kb | 234 | 0.10 | 1.17 | 0.13809 | 204 | 0.06 | 1.63 | 0.00037 | 438 | 0.08 | 1.37 | 0.00084 |
| | ≥ 1 Mb | 64 | 0.03 | 0.93 | 0.66292 | 68 | 0.02 | 1.80 | 0.01192 | 132 | 0.02 | 1.31 | 0.07407 |

## Total CNV size (kb)

| | | Stage 1: 979 cases / 1166 controls | | | Stage 2: 1168 cases / 1474 controls | | | Combined stage 1 + stage 2: 2147 cases / 2640 controls | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Type | Group | Baseline size (Ctrl) | Case/ Ctrl ratio | P value | Baseline size (Ctrl) | Case/ Ctrl ratio | P value | Baseline size (Ctrl) | Case/ Ctrl ratio | P value |
| All | All | 453.1 | 1.03 | 0.26520 | 368.9 | 1.41 | 0.00001 | 406 | 1.22 | 0.00001 |
| DEL | All | 222.5 | 1.02 | 0.40872 | 196.4 | 1.38 | 0.00006 | 207.9 | 1.21 | 0.00028 |
| DUP | All | 365.2 | 1.06 | 0.20712 | 306.1 | 1.36 | 0.00001 | 333.3 | 1.21 | 0.00007 |
| All | 30 – 500 kb | 304.4 | 0.97 | 0.78403 | 290.9 | 1.06 | 0.04013 | 296.8 | 1.02 | 0.22737 |
| | ≥ 500 kb | 1009.0 | 1.14 | 0.04687 | 963.5 | 1.49 | 0.00027 | 990.1 | 1.32 | 0.00005 |
| | ≥ 1 Mb | 1632.0 | 1.18 | 0.04903 | 1,533.0 | 1.54 | 0.00144 | 1,591.0 | 1.38 | 0.00013 |
| DEL | 30 – 500 kb | 168.3 | 0.99 | 0.56454 | 173.4 | 0.99 | 0.63485 | 171.2 | 0.99 | 0.64394 |
| | ≥ 500 kb | 952.3 | 1.12 | 0.19717 | 967.3 | 1.64 | 0.01571 | 957.4 | 1.42 | 0.00371 |
| | ≥ 1 Mb | 1560.0 | 1.17 | 0.16605 | 1,483.0 | 1.61 | 0.03369 | 1,531.0 | 1.44 | 0.01435 |
| DUP | 30 – 500 kb | 233.9 | 0.97 | 0.75751 | 226.2 | 1.08 | 0.03083 | 229.7 | 1.03 | 0.18388 |
| | ≥ 500 kb | 997.0 | 1.11 | 0.13966 | 940.2 | 1.37 | 0.00872 | 972.4 | 1.24 | 0.00429 |
| | ≥ 1 Mb | 1658.0 | 1.15 | 0.14467 | 1,493.0 | 1.52 | 0.00919 | 1,586.0 | 1.34 | 0.00380 |

## Gene count

| | | Stage 1: 979 cases / 1166 controls | | | | | Stage 2: 1168 cases / 1474 controls | | | | | Combined stage 1 + stage 2: 2147 cases / 2640 controls | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | Group | # Rare genic CNVs | # Genes inters. by rare CNVs | Baseline gene rate (Ctrl) | Case/ Ctrl gene ratio | P value | Pcorr | # Rare genic CNVs | # Genes inters. by rare CNVs | Baseline gene rate (Ctrl) | Case/ Ctrl gene ratio | P value | Pcorr | # Rare genic CNVs | # Genes inters. by rare CNVs | Baseline gene rate (Ctrl) | Case/ Ctrl gene ratio | P value | Pcorr |
| All | All | 3,252 | 4,319 | 4.16 | 1.17 | 0.02076 | 0.00462 | 3,607 | 4,514 | 3.08 | 1.66 | 0.00001 | 0.00001 | 6,859 | 6,745 | 3.55 | 1.41 | 0.00001 | 0.00001 |
| DEL | All | 1,378 | 1,505 | 1.32 | 1.25 | 0.01603 | 0.01603 | 1,568 | 1,924 | 1.16 | 1.55 | 0.00001 | 0.01498 | 2,946 | 2,804 | 1.23 | 1.40 | 0.00001 | 0.00049 |
| DUP | All | 1,874 | 3,417 | 2.84 | 1.14 | 0.10573 | 0.01342 | 2,039 | 3,244 | 1.92 | 1.73 | 0.00001 | 0.00001 | 3,913 | 5,217 | 2.32 | 1.41 | 0.00001 | 0.00001 |
| All | 30 – 500 kb | 2,962 | 3,211 | 3.14 | 0.98 | 0.60840 | 0.29800 | 3,345 | 3,414 | 2.70 | 1.15 | 0.00384 | 0.02582 | 6,307 | 5,163 | 2.89 | 1.07 | 0.03752 | 0.03628 |
| | ≥ 500 kb | 290 | 1,549 | 1.02 | 1.75 | 0.00488 | 0.00700 | 262 | 1,485 | 0.38 | 5.29 | 0.00001 | 0.00001 | 552 | 2,491 | 0.66 | 2.88 | 0.00001 | 0.00001 |
| | ≥ 1 Mb | 86 | 805 | 0.38 | 2.39 | 0.01245 | 0.00648 | 101 | 865 | 0.17 | 8.12 | 0.00001 | 0.00001 | 187 | 1,337 | 0.26 | 4.48 | 0.00001 | 0.00001 |
| DEL | 30 – 500 kb | 1,300 | 1,137 | 1.07 | 1.16 | 0.03994 | 0.03134 | 1,495 | 1,296 | 1.06 | 0.99 | 0.55163 | 0.79270 | 2,795 | 2,014 | 1.07 | 1.07 | 0.11389 | 0.20110 |
| | ≥ 500 kb | 78 | 454 | 0.25 | 1.63 | 0.09459 | 0.08104 | 73 | 696 | 0.09 | 7.91 | 0.00001 | 0.00051 | 151 | 947 | 0.16 | 3.60 | 0.00001 | 0.00051 |
| | ≥ 1 Mb | 26 | 278 | 0.11 | 2.22 | 0.08343 | 0.13570 | 37 | 492 | 0.07 | 7.84 | 0.00002 | 0.10500 | 63 | 647 | 0.08 | 4.58 | 0.00004 | 0.02289 |
| DUP | 30 – 500 kb | 1,662 | 2,411 | 2.07 | 0.9 | 0.92581 | 0.62180 | 1,850 | 2,491 | 1.63 | 1.26 | 0.00094 | 0.00364 | 3,512 | 3,934 | 1.83 | 1.08 | 0.08690 | 0.03750 |
| | ≥ 500 kb | 212 | 1,325 | 0.77 | 1.79 | 0.01117 | 0.01837 | 189 | 979 | 0.28 | 4.43 | 0.00001 | 0.00026 | 401 | 1,896 | 0.50 | 2.64 | 0.00001 | 0.00026 |
| | ≥ 1 Mb | 60 | 632 | 0.27 | 2.46 | 0.03874 | 0.01247 | 64 | 471 | 0.11 | 8.30 | 0.00006 | 0.00176 | 124 | 890 | 0.18 | 4.43 | 0.00004 | 0.00036 |

Baseline CNV rate (Ctrl): average number of CNVs per control subject
Baseline size (Ctrl): average total size per control subject in kb
Baseline gene rate (Ctrl): average number of genes intersected by CNVs per control subject

Pcorr: corrected for global differences in CNV size and rate
100,000 permutations

## Table S3C. Characteristics of rare CNVs in 2,147 European ASD probands and 2,640 European controls

| | ASD probands, European (n = 2,147 probands; 1,838 trios) | | | | Controls, European (n = 2,640) | | |
|---|---|---|---|---|---|---|---|
| | **Stringent CNVs**[a] (all sizes) | **Stringent CNVs**[b] (≥30 kb size, no peri-centromeric + no segdup) | **Rare CNVs**[c] | **Rare *de novo* CNVs**[d] | **Stringent CNVs**[a] (all sizes) | **Stringent CNV**[b] (≥30 kb size, no peri-centromeric + no segdup) | **Rare CNVs**[c] |
| Samples | 2,147 | 2,147 | 1,941 | 86 | 2,640 | 2,640 | 2,359 |
| # CNVs | 36,034 | 15,423 | 5,054 | 89 | 46,101 | 20,341 | 6,022 |
| Mean / median CNVs per genome[e] | 16.8 / 16 | 7.18 / 7 | 2.64 / 2 | 1.03 / 1 | 17.5 / 17 | 7.70 / 8 | 2.55 / 2 |
| Mean / median CNV size (kb) | 107.0 / 44.8 | 162.5 / 94.5 | 187.9 / 88.4 | 1,244.5 / 477.4 | 91.0 / 39.3 | 150.6 / 89.9 | 159.1 / 87.0 |
| % Gain / loss | 26.4 / 73.6 | 39.6 / 60.4 | 47.9 / 52.1 | 33.7 / 66.3 | 24.8 / 75.2 | 36.9 / 63.1 | 49.3 / 50.7 |
| # Recurrent/overlapping CNVs (%) / # loci[f] | 33,803 (93.8%) / 1,536 | 13,812 (89.5%) / 912 | 3,431 (67.9%) / 822 | 34 (38.2%) / 11 | 43,905 (95.2%) / 1,613 | 18,641 (91.6%) / 1,019 | 4,321 (71.8%) / 936 |
| # CNVs >1 Mb (%) | 365 (1.0%) | 277 (1.8%) | 112 (2.2%) | 28 (31.5%) | 264 (0.6%) | 226 (1.1%) | 85 (1.4%) |
| # CNVs >100-999 kb (%) | 9,782 (27.1%) | 7,019 (45.5%) | 2,445 (48.4%) | 42 (47.2%) | 11,945 (25.9%) | 9,341 (45.9%) | 2,537 (42.1%) |

[a] Stringent CNVs are CNV called by two or more algorithms. CNVs detected in the same individual by at least two algorithms were merged with the outside probes used as boundaries. All sizes, no filter applied.

[b] Stringent CNVs ≥30 kb and filtered for pericentromeric calls as well as calls overlapped by segmental duplications (segdup) for >50% of their length.

[c] Rare stringent CNVs ≥30 kb (filtered for pericentromeric and segdup calls) present at a frequency <1% in the total sample of 2,147 European cases and 2,640 European controls**.**

[d] Inheritance state was estimated for CNVs detected in 1,838 European probands from complete trios, including 1,820 complete trios with array data after QC available from both parents plus 18 families in which at least one of the parents failed initial array QC but additional laboratory validation was obtained for both parents. Laboratory validation confirmed that at least 4.7% (86/1,838) of European families carried at least one *de novo* CNV (average of 1 verified *de novo* CNV/sample). Similar rates were obtained for families of all ancestries, with 4.7% (98/2,096) carrying at least one *de novo* event.

[e] Probands with CNVs larger than 7.5 Mb are listed in **Table S1C**; they were excluded from the main burden analyses.

[f] Number and percentage of recurrent and/or overlapping CNVs in the dataset (%), and corresponding number of CNV loci.

## Table S4. Rare *de novo* CNVs in probands confirmed experimentally

| Sample ID | Sex | Ancestry | CNV type | Chr#:start-end (hg18) | Size | Cytoband | RefSeq gene(s) | Location |
|---|---|---|---|---|---|---|---|---|
| 14310_4270 | M | Eur | Gain | 1:105623065-105795181 | 172,117 | 1p21.1 | — | intergenic |
| 14310_4270 | M | Eur | Gain | 1:105896243-106480692 | 584,450 | 1p21.1 | — | intergenic |
| 16035_1571013001 | M | Eur | Gain | 1:144482933-146325557 | 1,842,625 | 1q21.1 | 15 genes (includes *CHD1L, GJA5, GJA8*) | gene/exonic |
| 1952_301 | M | Other | Gain | 1:144500467-146336720 | 1,836,254 | 1q21.1 | 15 genes (includes *CHD1L, GJA5, GJA8*) | gene/exonic |
| 8635_201 | M | Eur | Loss | 1:144500467-146377870 | 1,877,404 | 1q21.1 | 17 genes (includes *CHD1L, GJA5, GJA8*) | gene/exonic |
| 13135_1523 | F | Eur | Gain | 1:144838594-146308287 | 1,469,694 | 1q21.1 | 14 genes (includes *CHD1L, GJA5, GJA8*) | gene/exonic |
| 6356_5 | M | Eur | Loss | 1:2118508-2325536 | 207,029 | 1p36.33-p36.32 | *MORN1, SKI, LOC100129534, RER1, C1orf86* | gene/exonic |
| 8658_201 | F | Eur | Loss | 1:98175622-100923952 | 2,748,331 | 1p21.3-p21.2 | 18 genes (includes *CDC14A, MIR137, RTCD1, SNX7*) | gene/exonic |
| 9877_204 | M | Other | Gain | 1:98247355-99645560 | 1,398,206 | 1p21.3-p21.2 | *LOC100129620, LPPR4, LPPR5, MIR137, SNX7* | gene/exonic |
| 5089_5 | M | Other | Loss | 2:102292943-102345460 | 52,518 | 2q12.1 | *IL1RL1* | gene/exonic |
| 13082_963 | M | Eur | Gain | 2:11712589-11741036 | 28,448 | 2p25.1 | *NTSR2* | gene/exonic |
| 4260_1 | M | Eur | Gain | 2:119603876-119882278 | 278,403 | 2q14.2 | *C1QL2, C2orf76, DBI, STEAP3* | gene/exonic |
| 13027_353 | F | Eur | Gain | 2:144837809-145315383 | 477,575 | 2q22.3 | *DKFZp686O1327, ZEB2* | gene/exonic |
| 14414_5230 | M | Eur | Loss | 2:230486629-230547253 | 60,625 | 2q36.3 | *FBXO36, TRIP12* | gene/exonic |
| 14068_1180 | M | Eur | Gain | 2:50493827-50677835 | 184,009 | 2p16.3 | *NRXN1* | gene/exonic |
| 13017_223 | F | Eur | Loss | 2:50539877-50730546 | 190,670 | 2p16.3 | *NRXN1* | gene/exonic |
| 13216_2383 | M | Eur | Loss | 2:50968208-51214171 | 245,964 | 2p16.3 | *NRXN1* | gene/exonic |
| 13153_1703 | M | Eur | Loss | 2:50990306-51222043 | 231,738 | 2p16.3 | *NRXN1* | gene/exonic |
| 13037_463 | M | Eur | Loss | 2:51002576-51157742 | 155,167 | 2p16.3 | *NRXN1* | gene/exonic |
| 3616_3 | M | Eur | Loss | 3:9474320-9498362 | 24,043 | 3p25.3 | *SETD5* | gene/exonic |
| 5220_3 | F | Eur | Gain | 3:19127998-19640299 | 512,302 | 3p24.3 | *KCNH8* | gene/exonic |
| 3174_3 | M | Eur | Loss | 3:19921061-20096832 | 175,772 | 3p24.3 | *C3orf48, EFHB, KAT2B, RAB5A* | gene/exonic |
| 5245_3 | M | Eur | Loss | 3:117285007-117477191 | 192,185 | 3q13.31 | *LSAMP* (50% mosaic) | gene/exonic |
| 8587_210 | F | Eur | Gain | 3:139760896-140127703 | 366,808 | 3q22.3 | *CEP70, FAIM, PIK3CB* | gene/exonic |
| 8588_201 | M | Eur | Loss | 3:195971510-197675831 | 1,704,322 | 3q29 | 20 genes (includes *PCYT1A, TFRC, TNK2*) | gene/exonic |
| 8677_201 | M | Other | Gain | 4:94659049-95289667 | 630,619 | 4q22.2 | *ATOH1, GRID2* | gene/exonic |
| 6402_3 | M | Eur | Gain | 4:98781694-99627129 | 845,436 | 4q22.3-q23 | *C4orf37, RAP1GDS1, TSPAN5* | gene/exonic |
| 8709_201 | F | Eur | Loss | 5:88588170-89505509 | 917,340 | 5q14.3 | *MIR3660* | gene/exonic |
| 5386_3 | M | Other | Loss | 6:156785155-158489874 | 1,704,720 | 6q25.3 | *ARID1B, MIR3692, SERAC1, SNX9, SYNJ2, ZDHHC14* | gene/exonic |
| 6164_3 | F | Eur | Gain | 6:160023074-160081618 | 58,545 | 6q25.3 | *SOD2, WTAP* | gene/exonic |
| 8612_201 | M | Eur | Loss | 6:162588879-162629938 | 41,060 | 6q26 | *PARK2* | gene/exonic |
| 8404_201 | M | Eur | Loss | 6:169136788-170761395 | 1,624,608 | 6q27 | 14 genes (includes *PSMB1, TBP, TCTE3, THBS2, WDR27*) | gene/exonic |
| 5353_3 | F | Eur | Loss | 6:33399849-33512042 | 112,194 | 6p21.32 | *CUTA, KIFC1, PHF1, SYNGAP1* | gene/exonic |
| 18133_302 | M | Eur | Gain | 6:69540429-73375020 | 3,834,592 | 6q12-q13 | *B3GAT2, BAI3, C6orf155, C6orf57, COL19A1, COL9A1, FAM135A, LMBRD1, MIR30A, MIR30C2, OGFRL1, RIMS1, SMAP1* | gene/exonic |
| 6248_3 | M | Eur | Loss | 6:86352577-86376159 | 23,583 | 6q14.3 | *SNX14, SYNCRIP* | gene/exonic |
| 1960_301 | F | Eur | Loss | 7:102699832-102798745 | 98,914 | 7q22.1 | *DNAJC2, DPY19L2P2, PMPCB, PSMC2, SLC26A5* | gene/exonic |
| 5370_3 | M | Eur | Loss | 7:153775586-153844747 | 69,162 | 7q36.2 | *DPP6* | gene/exonic |
| 8446_201 | M | Eur | Loss | 7:72344426-73782113 | 1,437,688 | 7q11.23 | 28 genes (includes *GTF2I, GTF2IRD1, LAT2, LIMK1, STX1A*) | gene/exonic |
| 1142_4 | F | Eur | Gain | 8:48631388-48802529 | 171,142 | 8q11.21 | *KIAA0146* | gene/exonic |
| 6321_3 | M | Eur | Loss | 8:65354366-66254869 | 900,504 | 8q12.3-q13.1 | *BHLHE22, CYP7B1, LOC100130155, LOC401463, MIR124-2* | gene/exonic |
| 5290_3 | M | Other | Gain | 8:704383-1521910 | 817,528 | 8p23.3 | *DLGAP2, LOC286083* | gene/exonic |
| 8500_201 | M | Eur | Loss | 8:74692898-74967280 | 274,383 | 8q21.11 | *STAU2, UBE2W* | gene/exonic |
| 14237_2650 | M | Eur | Loss | 9:103456776-107140273 | 3,683,498 | 9q31.1 | *ABCA1, CYLC2, GRIN3A, LOC286367, NIPSNAP3A, NIPSNAP3B, OR13C2, OR13C3, OR13C4, OR13C5, OR13C8, OR13C9, OR13D1, OR13F1, SLC44A1, SMC2* | gene/exonic |
| 14417_5260 | M | Eur | Loss | 9:117311405-117727764 | 416,360 | 9q33.1 | *C9orf27* (or *LINC00474*) | gene/exonic |
| 9756_201 | M | Other | Loss | 9:132572435-132597937 | 25,503 | 9q34.12 | *ABL1* | gene/exonic |
| 6259_3 | M | Eur | Loss | 9:139516033-140208462 | 692,430 | 9q34.3 | *ARRDC1, C9orf37, CACNA1B, EHMT1, FLJ40292* (or *EHMT1-IT1*) , *MIR602, MRPL41, PNPLA7, TUBBP5, WDR85, ZMYND19* | gene/exonic |
| 6246_4 | M | Eur | Loss | 9:9399606-9631169 | 231,564 | 9p23 | *PTPRD* | gene/exonic |
| 5032_4 | M | Eur | Loss | 9:98998-334508 | 235,511 | 9p24.3 | *C9orf66, CBWD1, DOCK8, FOXD4* | gene/exonic |
| 13123_1403 | F | Eur | Loss | 9:98998-3682923 | 3,583,926 | 9p24.3-p24.2 | *C9orf66, CBWD1, DMRT1, DMRT2, DMRT3, DOCK8, FLJ35024, FOXD4, KANK1, KCNV2, KIAA0020, RFX3, SMARCA2, VLDLR* | gene/exonic |
| 18172_302 | M | Other | Gain | 10:34963905-35536952 | 573,048 | 10p11.21 | *CREM, CUL2, PARD3* | gene/exonic |
| 8534_201 | M | Eur | Loss | 10:45633089-51564756 | 5,931,668 | 10q11.21-q11.23 | 56 genes (includes *CHAT, ERCC6, MAPK8, SLC18A3*) | gene/exonic |

| Sample ID | Sex | Ancestry | CNV type | Chr#:start-end (hg18) | Size | Cytoband | RefSeq gene(s) | Location |
|---|---|---|---|---|---|---|---|---|
| 14393_5020 | M | Eur | Loss | 10:74801551-75278951 | 477,401 | 10q22.2 | *AGAP5, ANXA7, BMS1P4, CAMK2G, CHCHD1, FUT11, KIAA0913, MYOZ1, NDST2, PPP3CB, SEC24C, SYNPO2L, USP54, ZMYND17* | gene/exonic |
| 6240_4 | M | Eur | Loss | 11:126633939-132060374 | 5,426,436 | 11q24.2-q25 | 20 genes (includes *ARHGAP32, SNX19*) | gene/exonic |
| 6325_3 | M | Eur | Loss | 11:70077507-70506315 | 428,809 | 11q13.3 | *MIR3664, SHANK2* | gene/exonic |
| 6319_3 | M | Eur | Loss | 11:70119917-70187872 | 67,956 | 11q13.3 | *SHANK2* | gene/exonic |
| 5237_3 | M | Eur | Loss | 11:70154458-70220632 | 66,175 | 11q13.3 | *SHANK2* | gene/exonic |
| 6053_3 | M | Eur | Gain | 12:54218922-58779615 | 4,560,694 | 12q13.2-q14.1 | 101 genes (includes *CDK2, CDK4, SMARCC2*) | gene/exonic |
| 5272_3 | M | Other | Loss | 12:98445422-98540678 | 95,257 | 12q23.1 | *ANKS1B* | gene/intronic |
| 1050_3 | F | Eur | Gain | 14:20279711-20345174 | 65,464 | 14q11.2 | *EDDM3A, EDDM3B, RNASE1, RNASE6* | gene/exonic |
| 8638_201 | M | Eur | Loss | 14:35203374-35369853 | 166,480 | 14q13.2 | *BRMS1L, RALGAPA1* | gene/exonic |
| 4272_1 | M | Eur | Loss | 14:63824114-65347410 | 1,523,297 | 14q23.2-q23.3 | 18 genes (includes *ESR2, MAX, MTHFD1, SPTB*) | gene/exonic |
| 17009_1 | M | Eur | Loss | 14:78575296-78596793 | 21,498 | 14q31.1 | *NRXN3* | gene/intronic |
| 20187_1464001 | M | Eur | Gain | 15:18811937-26209270 | 7,397,334 | 15q11.2-q13.1 | 125 genes (includes *CYFIP1, UBE3A, HERC2*) | gene/exonic |
| 8630_201 | M | Eur | Gain | 15:19800798-26209270 | 6,408,473 | 15q11.2-q13.1 | 117 genes (includes *CYFIP1, UBE3A, HERC2*) | gene/exonic |
| 20069_1328001 | M | Eur | Gain | 15:20203578-26209270 | 6,005,693 | 15q11.2-q13.1 | 112 genes (includes *CYFIP1, UBE3A, HERC2*) | gene/exonic |
| 17035_1 | F | Eur | Gain | 15:20274130-26120360 | 5,846,231 | 15q11.2-q13.1 | 111 genes (includes *CYFIP1, UBE3A, HERC2*) | gene/exonic |
| 8430_204 | M | Other | Loss | 15:20301669-20777695 | 476,027 | 15q11.2 | *CYFIP1, NIPA1, NIPA2, TUBGCP5, WHAMML1* | gene/exonic |
| 13050_593 | M | Eur | Gain | 15:21190624-26203954 | 5,013,331 | 15q11.2-q13.1 | 101 genes (includes *UBE3A, HERC2*) | gene/exonic |
| 16040_1571029001 | M | Eur | Loss | 15:28450423-30303265 | 1,852,843 | 15q13.2-q13.3 | *ARHGAP11B, CHRFAM7A, CHRNA7, FAM7A1, FAM7A2, FAN1, KLF13, LOC100288637, MIR211, MTMR10, OTUD7A, TRPM1* | gene/exonic |
| 6101_4 | M | Other | Loss | 15:74735339-74929817 | 194,479 | 15q24.3 | *SCAPER* | gene/exonic |
| 8695_201 | M | Eur | Loss | 15:81728085-82623936 | 895,852 | 15q25.2 | *ADAMTSL3, BNC1, LOC648809, SH3GL3* | gene/exonic |
| 14181_2940 | M | Eur | Gain | 15:82906265-82985247 | 78,983 | 15q25.2 | *LOC100506874, SCAND2, UBE2Q2P1, ZSCAN2* | gene/exonic |
| 14070_1230 | M | Eur | Loss | 15:91200007-91283004 | 82,998 | 15q26.1 | *CHD2, LOC100507217, MIR3175* | gene/exonic |
| 2204_1 | M | Eur | Loss | 16:29466569-30147029 | 680,461 | 16p11.2 | 42 genes (includes *KCTD13, MAPK3, SEZ6L2*) | gene/exonic |
| 20089_1391001 | M | Eur | Loss | 16:29502984-30107306 | 604,323 | 16p11.2 | 29 genes (includes *KCTD13, MAPK3, SEZ6L2*) | gene/exonic |
| 5068_3 | F | Eur | Loss | 16:29502984-30127026 | 624,043 | 16p11.2 | 40 genes (includes *KCTD13, MAPK3, SEZ6L2*) | gene/exonic |
| 5262_4 | M | Eur | Gain | 16:29502984-30210849 | 707,866 | 16p11.2 | 43 genes (includes *KCTD13, MAPK3, SEZ6L2*) | gene/exonic |
| 4030_1 | M | Eur | Gain | 16:29554843-30107306 | 552,464 | 16p11.2 | 28 genes (includes *KCTD13, MAPK3, SEZ6L2*) | gene/exonic |
| 5359_4 | M | Eur | Loss | 16:29554843-30195224 | 640,382 | 16p11.2 | 42 genes (includes *KCTD13, MAPK3, SEZ6L2*) | gene/exonic |
| 3439_3 | M | Eur | Loss | 17:17156307-18262979 | 1,106,673 | 17p11.2 | 22 genes (includes *RAI1, SREBF1, LLGL1, TOP3A*) | gene/exonic |
| 2211_1 | F | Eur | Loss | 17:17169258-20101517 | 2,932,260 | 17p11.2 | 53 genes (includes *MAPK7, RAI1, SREBF1, FAM83G, LLGL1, TOP3A*) | gene/exonic |
| 5056_4 | M | Eur | Gain | 17:34612208-34732327 | 120,120 | 17q12 | *FBXL20, RPL19, STAC2* | gene/exonic |
| 8463_202 | M | Eur | Loss | 17:52774693-52895975 | 121,283 | 17q22 | *MSI2* | gene/exonic |
| 5444_3 | M | Eur | Gain | 17:76953064-77782267 | 829,204 | 17q25.3 | 38 genes (includes *ACTG1, ARHGDIA, FASN, RAC3, HGS, DUS1L*) | gene/exonic |
| 5444_3 | M | Eur | Loss | 17:77785939-77849717 | 63,779 | 17q25.3 | *CSNK1D, SLC16A3* | gene/exonic |
| 3477_3 | M | Eur | Loss | 18:30280260-30327512 | 47,253 | 18q12.1 | *DTNA* | gene/exonic |
| 14331_4450 | M | Eur | Loss | 18:72085223-73670156 | 1,584,934 | 18q23 | *GALR1, LOC284276, MBP, ZNF236, ZNF516* | gene/exonic |
| 6358_6 | M | Eur | Loss | 19:4548413-5287389 | 738,977 | 19p13.3 | *ARRDC5, C19orf10, DPP9, FEM1A, KDM4B, MIR7-3, NCRNA00306, PLIN3, PTPRS, TICAM1, TNFAIP8L1, UHRF1* | gene/exonic |
| 5335_3 | M | Eur | Loss | 20:14545734-14948785 | 403,052 | 20p12.1 | *MACROD2, MACROD2-AS1* | gene/exonic |
| 5046_3 | M | Other | Loss | 20:8607242-8637441 | 30,200 | 20p12.3 | *PLCB1* | gene/exonic |
| 20180_1704001 | M | Eur | Gain | 21:43018846-43444308 | 425,463 | 21q22.3 | *CBS, NDUFV3, PDE9A, PKNOX1, U2AF1, WDR4* | gene/exonic |
| 3183_7 | M | Eur | Loss | 22:17241748-19819918 | 2,578,171 | 22q11.21 | 63 genes (includes *TBX1, CRKL, TSSK2*) | gene/exonic |
| 8627_201 | M | Eur | Gain | 22:17257787-19793730 | 2,535,944 | 22q11.21 | 63 genes (includes *TBX1, CRKL, TSSK2*) | gene/exonic |
| 17015_1 | M | Eur | Loss | 22:17257787-19795780 | 2,537,994 | 22q11.21 | 63 genes (includes *TBX1, CRKL, TSSK2*) | gene/exonic |
| 4271_1 | M | Eur | Gain | 22:17257787-19795780 | 2,537,994 | 22q11.21 | 63 genes (includes *TBX1, CRKL, TSSK2*) | gene/exonic |
| 8630_201 | M | Eur | Loss | 22:32346124-32413987 | 67,864 | 22q12.3 | *LARGE* | gene/exonic |
| 2072_1 | M | Eur | Loss | 22:45159185-49582267 | 4,423,083 | 22q13.31-q13.33 | 47 genes (includes *SHANK3*) | gene/exonic |
| 6130_4 | F | Eur | Loss | 22:47996161-49512530 | 1,516,370 | 22q13.32-q13.33 | 37 genes (includes *SHANK3*) | gene/exonic |
| 14291_4120 | F | Eur | Loss | 22:49468716-49485255 | 16,540 | 22q13.33 | *SHANK3* | gene/exonic |
| 16079_1571066001 | M | Eur | Loss | 22:49470371-49567383 | 97,013 | 22q13.33 | *ACR, RABL2B, RPL23AP82, SHANK3* | gene/exonic |
| 20013_1075001 | M | Eur | Gain | X:153239048-153521797 | 282,750 | Xq28 | 20 genes (includes *EMD, FLNA, GDI1, IKBKG, PLXNA3, RPL10*) | gene/exonic |
| 16076_1571045001 | M | Eur | Loss | X:22768481-23133948 | 365,468 | Xp22.11 | *DDX53* (includes *PTCHD1-AS*) | gene/exonic |

We identified 102 rare *de novo* CNVs in 99 cases; three individuals (5444_3, 8630_201 and 14310_4270) have 2 *de novo* CNV each. Of the 99 subjects with *de novo* events, 60 (60%) are simplex, 26 (27%) multiplex, and 13 (13%) of unknown status. Eleven *de novo* chromosomal abnormalities >7.5 Mb identified in probands are listed in **Table S1C**. Abbreviations: Eur, European; F, female; M, male.

## Table S5A. Parent of origin for rare *de novo* validated CNVs in probands

| Sample ID | Chr#:start-end (hg18) | CNV type | Family type | Parent of origin | SNPs: \|#paternal\|#maternal\|total |
|---|---|---|---|---|---|
| 14310_4270 | 1:105623065-105795181 | Gain | SPX | Paternal | SNPs: \|20\|0\|41 |
| 14310_4270 | 1:105896243-106480692 | Gain | SPX | Paternal | Custom-designed microsatellite (14310b) |
| 16035_1571013001 | 1:144482933-146325557 | Gain | MPX | Unknown | No informative SNPs or microsatellites |
| 1952_301 | 1:144500467-146336720 | Gain | MPX | Paternal | SNPs: \|24\|0\|534 |
| 8635_201 | 1:144500467-146377870 | Loss | MPX | Paternal | SNPs: \|73\|0\|617 |
| 13135_1523 | 1:144838594-146308287 | Gain | UNK | Paternal | SNPs: \|121\|56\|667 |
| 6356_5 | 1:2118508-2325536 | Loss | SPX | Maternal | SNPs: \|0\|11\|92 |
| 8658_201 | 1:98175622-100923952 | Loss | SPX | Maternal | SNPs: \|0\|199\|1018 |
| 9877_204 | 1:98247355-99645560 | Gain | SPX | Unknown | Ambiguous, SNPs: \|74\|27\|488; microsat ?D1S2739 |
| 5089_5 | 2:102292943-102345460 | Loss | MPX | Unknown | No informative SNPs or microsatellites |
| 13082_963 | 2:11712589-11741036 | Gain | UNK | Maternal | Custom-designed microsatellite (13082b) |
| 4260_1 | 2:119603876-119882278 | Gain | SPX | Maternal | SNPs: \|1\|26\|118 (Griswold et al. 2012)[128] |
| 13027_353 | 2:144837809-145315383 | Gain | UNK | Maternal | SNPs: \|0\|40\|124 |
| 14414_5230 | 2:230486629-230547253 | Loss | SPX | Unknown | No informative SNPs |
| 14068_1180 | 2:50493827-50677835 | Gain | SPX | Maternal | SNPs: \|0\|15\|75 |
| 13017_223 | 2:50539877-50730546 | Loss | UNK | Paternal | SNPs: \|19\|0\|89 |
| 13216_2383 | 2:50968208-51214171 | Loss | UNK | Maternal | SNPs: \|0\|10\|88 |
| 13153_1703 | 2:50990306-51222043 | Loss | UNK | Paternal | Custom-designed microsatellite (13153a) |
| 13037_463 | 2:51002576-51157742 | Loss | SPX | Paternal | SNPs: \|10\|0\|47 |
| 5245_3 | 3:117285007-117477191 | Loss | MPX | Unknown | No informative SNPs |
| 8587_210 | 3:139760896-140127703 | Gain | UNK | Paternal | SNPs: \|5\|0\|105 |
| 5220_3 | 3:19127998-19640299 | Gain | SPX | Paternal | SNPs: \|38\|0\|205 |
| 8588_201 | 3:195971510-197675831 | Loss | UNK | Maternal | SNPs: \|0\|35\|633 |
| 3174_3 | 3:19921061-20096832 | Loss | MPX | Maternal | SNPs: \|3\|0\|72 |
| 3616_3 | 3:9474320-9498362 | Loss | UNK | Unknown | No informative SNPs |
| 8677_201 | 4:94659049-95289667 | Gain | SPX | Paternal | SNPs: \|28\|1\|195 |
| 6402_3 | 4:98781694-99627129 | Gain | SPX | Maternal | SNPs: \|0\|28\|230 |
| 8709_201 | 5:88588170-89505509 | Loss | SPX | Paternal | SNPs: \|48\|0\|249 |
| 5386_3 | 6:156785155-158489874 | Loss | MPX | Paternal | SNPs: \|115\|0\|605 |
| 6164_3 | 6:160023074-160081618 | Gain | SPX | Unknown | Ambiguous, SNPs: \|2\|5\|23 * |
| 8612_201 | 6:162588879-162629938 | Loss | UNK | Paternal | Custom-designed microsatellite (8612a) |
| 8404_201 | 6:169136788-170761395 | Loss | MPX | Maternal | SNPs: \|0\|104\|648 |
| 5353_3 | 6:33399849-33512042 | Loss | SPX | Paternal | SNPs: \|1\|0\|63 ** |
| 18133_302 | 6:69540429-73375020 | Gain | SPX | Maternal | SNPs: \|0\|182\|1516 |
| 6248_3 | 6:86352577-86376159 | Loss | SPX | Unknown | No informative SNPs |
| 1960_301 | 7:102699832-102798745 | Loss | MPX | Paternal | D7S2509 |
| 5370_3 | 7:153775586-153844747 | Loss | SPX | Maternal | SNPs: \|0\|6\|43 |
| 8446_201 | 7:72344426-73782113 | Loss | SPX | Maternal | SNPs: \|0\|66\|615 |
| 1142_4 | 8:48631388-48802529 | Gain | MPX | Paternal | SNPs: \|3\|0\|27 |
| 6321_3 | 8:65354366-66254869 | Loss | SPX | Unknown | No informative SNPs or microsatellites |
| 5290_3 | 8:704383-1521910 | Gain | SPX | Paternal | SNPs: \|85\|37\|361 |
| 8500_201 | 8:74692898-74967280 | Loss | MPX | Maternal | SNPs: \|0\|22\|91 |
| 14417_5260 | 9:117311405-117727764 | Loss | SPX | Paternal | SNPs: \|30\|0\|141 |
| 9756_201 | 9:132572435-132597937 | Loss | SPX | Unknown | No informative SNPs |
| 6259_3 | 9:139516033-140208462 | Loss | SPX | Paternal | SNPs: \|23\|0\|316 |
| 6246_4 | 9:9399606-9631169 | Loss | SPX | Unknown | No informative SNPs; no informative microsatellites*** |
| 5032_4 | 9:98998-334508 | Loss | MPX | Paternal | SNPs: \|15\|0\|128 |
| 13123_1403 | 9:98998-3682923 | Loss | UNK | Paternal | SNPs: \|204\|0\|1992 |
| 18172_302 | 10:34963905-35536952 | Gain | SPX | Paternal | SNPs: \|61\|0\|179 |
| 8534_201 | 10:45633089-51564756 | Loss | SPX | Maternal | SNPs: \|0\|161\|1197 |
| 14393_5020 | 10:74801551-75278951 | Loss | SPX | Paternal | SNPs: \|1\|0\|225 and patD10S188 |
| 6240_4 | 11:126633939-132060374 | Loss | SPX | Paternal | SNPs: \|351\|0\|2144 |
| 6325_3 | 11:70077507-70506315 | Loss | SPX | Maternal | SNPs: \|0\|15\|138 (Leblond et al 2012)[129] |
| 6319_3 | 11:70119917-70187872 | Loss | SPX | Paternal | SNPs: \|3\|0\|24 |
| 5237_3 | 11:70154458-70220632 | Loss | SPX | Maternal | SNPs: \|0\|3\|15 |

| Sample ID | Chr#:start-end (hg18) | CNV type | Family type | Parent of origin | SNPs: \|#paternal\|#maternal\|total |
|---|---|---|---|---|---|
| 6053_3 | 12:54218922-58779615 | Gain | MPX | Maternal | SNPs: \|0\|241\|1766 |
| 5272_3 | 12:98445422-98540678 | Loss | MPX | Maternal | SNPs: \|0\|5\|32 |
| 1050_3 | 14:20279711-20345174 | Gain | MPX | Paternal | SNPs: \|14\|0\|49 |
| 8638_201 | 14:35203374-35369853 | Loss | UNK | Paternal | Custom designed microsatellite (8638b) |
| 4272_1 | 14:63824114-65347410 | Loss | SPX | Maternal | SNPs: \|0\|47\|563 (Griswold et al. 2012)[128] |
| 17009_1 | 14:78575296-78596793 | Loss | SPX | Unknown | No informative SNPs or microsatellites |
| 20187_1464001 | 15:18811937-26209270 | Gain | SPX | Maternal | D15S1002, D15S128 |
| 8630_201 | 15:19800798-26209270 | Gain | MPX | Paternal | SNPs: \|328\|9\|2265 |
| 20069_1328001 | 15:20203578-26209270 | Gain | SPX | Maternal | SNPs: \|0\|467\|2230 |
| 17035_1 | 15:20274130-26120360 | Gain | SPX | Maternal | D15S1002, D15S128 |
| 8430_204 | 15:20301669-20777695 | Loss | SPX | Paternal | SNPs: \|16\|0\|187 |
| 13050_593 | 15:21190624-26203954 | Gain | UNK | Maternal | D15S128 |
| 16040_1571029001 | 15:28450423-30303265 | Loss | MPX | Paternal | SNPs: \|56\|0\|481 |
| 6101_4 | 15:74735339-74929817 | Loss | MPX | Paternal | SNPs: \|5\|0\|36 |
| 8695_201 | 15:81728085-82623936 | Loss | SPX | Maternal | SNPs: \|0\|27\|299 |
| 14181_2940 | 15:82906265-82985247 | Gain | SPX | Maternal | SNPs: \|0\|4\|18 |
| 14070_1230 | 15:91200007-91283004 | Loss | MPX | Paternal | SNPs: \|1\|0\|22 |
| 2204_1 | 16:29466569-30147029 | Loss | SPX | Maternal | SNPs: \|0\|28\|248 |
| 20089_1391001 | 16:29502984-30107306 | Loss | SPX | Maternal | SNPs: \|0\|21\|306 |
| 5068_3 | 16:29502984-30127026 | Loss | MPX | Paternal | SNPs: \|5\|0\|249 |
| 5262_4 | 16:29502984-30210849 | Gain | SPX | Paternal | SNPs: \|69\|4\|249 |
| 4030_1 | 16:29554843-30107306 | Gain | MPX | Unknown | Ambiguous, SNPs: \|22\|22\|253 |
| 5359_4 | 16:29554843-30195224 | Loss | SPX | Maternal | SNPs: \|0\|1\|248; no informative microsatellites |
| 3439_3 | 17:17156307-18262979 | Loss | SPX | Paternal | SNPs: 184\|0\|439 |
| 2211_1 | 17:17169258-20101517 | Loss | SPX | Paternal | SNPs: \|79\|0\|927 |
| 5056_4 | 17:34612208-34732327 | Gain | MPX | Maternal | SNPs: \|0\|1\|57; no informative microsatellites |
| 8463_202 | 17:52774693-52895975 | Loss | MPX | Paternal | SNPs: \|5\|0\|66 |
| 5444_3 | 17:76953064-77782267 | Gain | SPX | Paternal | SNPs: \|20\|5\|276 |
| 5444_3 | 17:77785939-77849717 | Loss | SPX | Paternal | Custom-designed microsatellite |
| 3477_3 | 18:30280260-30327512 | Loss | SPX | Maternal | Custom-designed microsatellite (20xAC, 22xAC) |
| 14331_4450 | 18:72085223-73670156 | Loss | SPX | Paternal | SNPs: \|168\|0\|900 |
| 6358_6 | 19:4548413-5287389 | Loss | SPX | Maternal | SNPs: \|0\|57\|310 |
| 5335_3 | 20:14545734-14948785 | Loss | SPX | Maternal | SNPs: \|0\|7\|151 |
| 5046_3 | 20:8607242-8637441 | Loss | MPX | Maternal | Custom-designed microsatellite (5046) |
| 20180_1704001 | 21:43018846-43444308 | Gain | SPX | Maternal | SNPs: \|7\|33\|225 |
| 3183_7 | 22:17241748-19819918 | Loss | MPX | Maternal | SNPs: \|0\|184\|1251 |
| 8627_201 | 22:17257787-19793730 | Gain | SPX | Paternal | SNPs:\|292\|94\|1350 |
| 17015_1 | 22:17257787-19795780 | Loss | SPX | Maternal | SNPs: \|0\|190\|1444 |
| 4271_1 | 22:17257787-19795780 | Gain | SPX | Maternal | SNPs: \|105\|249\|1311 |
| 8630_201 | 22:32346124-32413987 | Loss | MPX | Unknown | Ambiguous, SNPs: \|21\|25\|291; no informative microsatellites |
| 2072_1 | 22:45159185-49582267 | Loss | MPX | Maternal | Gonadal mosaicism, two affected sibs (Moessner et al. 2008)[130] |
| 6130_4 | 22:47996161-49512530 | Loss | SPX | Paternal | SNPs: \|170\|0\|875 |
| 14291_4120 | 22:49468716-49485255 | Loss | SPX | Paternal | SNPs: \|1\|0\|5 |
| 16079_1571066001 | 22:49470371-49567383 | Loss | SPX | Paternal | SNPs: \|7\|0\|39 |
| 20013_1075001 | X:153239048-153521797 | Gain | SPX | Unknown | — |
| 16076_1571045001 | X:22768481-23133948 | Loss | SPX | Unknown | — |

For 85 of 102 *de novo* events it was possible to determine the parent-of-origin from SNPs or microsatellite genotypes. Of the 85 *de novo* CNVs where parental origin could be assigned, 45 (53%) were paternally-derived and 40 (47%) were maternally-derived. These findings do not confirm the results from Hehir-Kwa et al.[131] of an increased rate of paternally-derived *de novo* events. Furthermore, paternal age was not found to be increased in fathers of ASD probands (with or without *de novo* CNVs) when compared to fathers of controls from that study (**Table S5H**). Parental bias was also not observed when we separately considered *de novo* events by type of CNV (pathogenic, uncertain significance, intergenic/intronic), mechanism of causality (mediated/flanked by segmental duplications or not) and type of family (simplex or multiplex).

* Similar results when using a larger 500 kb-flanking SNP window: ambiguous (SNPs: |8|17|203)
** Similar results when using a larger 500 kb-flanking SNP window: pat (SNPs: |1|0|369)
*** When using a 500 kb-flanking window: ambiguous (SNPs: |0|1|208)
Abbreviations: microsat, microsatellite marker; MPX, multiplex; SPX, simplex; UNK, unknown family type.

## Tables S5B-S5H. Parent of origin of *de novo* CNVs – breakdown by family type and CNV characteristics

**Table S5B. Breakdown by type of CNV**

| | Paternal | Maternal | Total |
|---|---|---|---|
| Pathogenic | 23 | 20 | 43 |
| Uncertain | 20 | 19 | 39 |
| Intergenic/ intronic | 2 | 1 | 3 |
| Total | 45 | 40 | 85 |

Classification according to **Tables S5A** and **S7A**

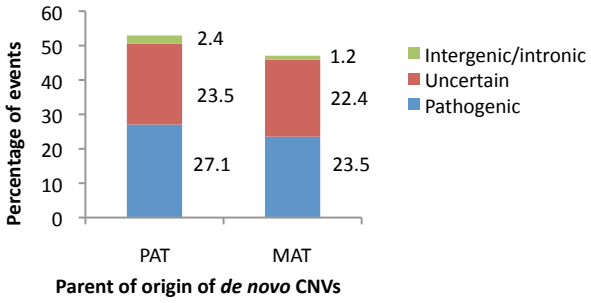



**Table S5C. Breakdown by type of CNV, including large chromosomal abnormalities**

| | Paternal | Maternal | Total |
|---|---|---|---|
| Pathogenic | 25 | 22 | 47 |
| Uncertain | 20 | 19 | 39 |
| Intergenic/ intronic | 2 | 1 | 3 |
| Total | 47 | 42 | 88 |

Classification according to **Tables S5A** and **S7A**

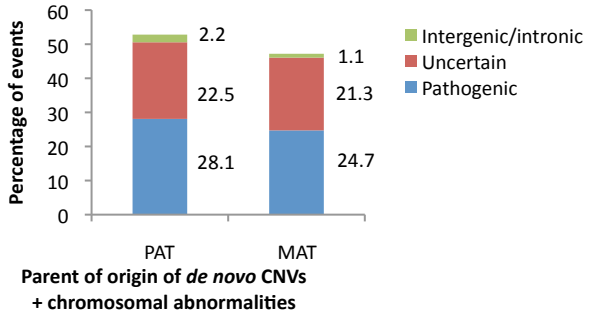



**Table S5D. Breakdown by mechanism (CNV flanked or not by segmental duplications)**

| | Paternal | | Maternal | | Total |
|---|---|---|---|---|---|
| | SD | no SD | SD | no SD | |
| Pathogenic | 10 | 13 | 13 | 7 | 43 |
| Uncertain | 1 | 19 | 1 | 18 | 39 |
| Intergenic/ intronic | 0 | 2 | 0 | 1 | 3 |
| Total | 11 | 34 | 14 | 26 | 85 |

Classification according to **Tables S5A** and **S7A**
SD, segmental duplication

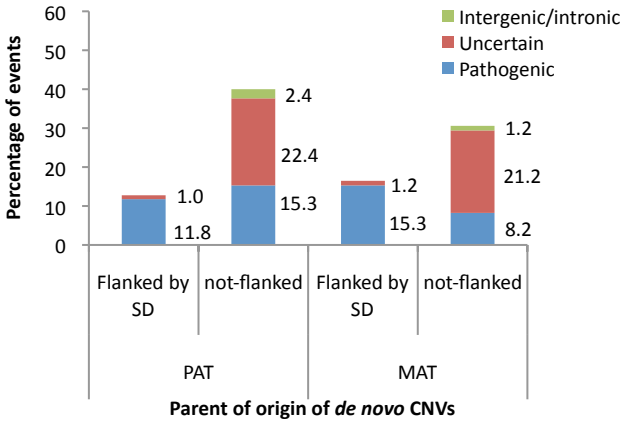


**Table S5E. Breakdown of CNVs by type of family**

| | # Probands without de novo events | # Probands with *de novo* events | Total | % |
|---|---|---|---|---|
| AGP SPX | 1142 | 59 | 1201 | 4.9 |
| AGP MPX | 600 | 26 | 626 | 4.2 |
| AGP Unknown | 256 | 13 | 269 | 4.8 |
| Total | 1998 | 98 | 2096 | 4.7 |
| SSC | 1059 | 65 | 1124 | 5.8 |

Number of simplex (SPX) and multiplex (MPX) families after QC is 1201 and 626, respectively. SSC, Simons Simplex Collection[132]

**Table S5F. Breakdown of *de novo* events by type of family**

| | Paternal | Maternal | Total |
|---|---|---|---|
| AGP SPX | 23 | 26 | 49 |
| AGP MPX | 13 | 9 | 22 |
| SSC | 27 | 34 | 61 |
| Total | 63 | 69 | 132 |

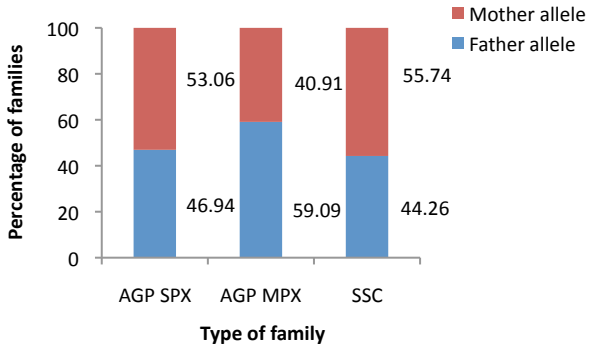



**Table S5G. Breakdown of *de novo* CNVs by size in different family types**

| | <30 kb | 30 kb-500 kb | 500 kb-1 Mb | >1 Mb |
|---|---|---|---|---|
| AGP SPX | 4 | 23 | 15 | 17 |
| AGP MPX | 0 | 14 | 2 | 10 |
| Total | 4 | 37 | 17 | 27 |

Average CNV size is 1,199,751 bp for SPX, 1,180,129 bp for MPX, and 1,041,851 bp for families classified as unknown.

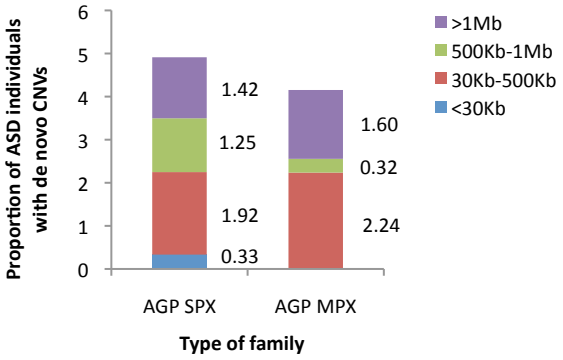



**Table S5H. Breakdown of *de novo* CNVs by paternal age**

| | # Fathers in group* | Median paternal age ± SD |
|---|---|---|
| ASD without *de novo* CNVs | 1659 | 32.33 ± 5.74 |
| ASD with *de novo* CNVs | 76 | 32.17 ± 6.57 |
| ASD with *de novo* CNVs, paternal in origin, with flanking SD | 6 | 33.75 ± 6.03 |
| ASD with *de novo* CNVs, paternal in origin, without flanking SD | 24 | 33.71 ± 6.08 |

*45 *de novo* CNVs were paternally derived from 43 fathers (2 affected individuals had 2 paternally-derived *de novo* CNVs each), and parental age was available for 30 of them. The average maternal and paternal age at childbirth in our group of rare *de novo* CNVs was 29.6 and 32.2 years, respectively (76/99 had information on parental age). These age ranges were similar to parental ages at childbirth in the control cohort used by Hehir-Kwa et al.[131] (31.4 and 32.1 years, respectively).

# Tables S6A-S6D. List of genes and loci implicated in ASD and ID

## Table S6A. Genes implicated in ASD

| | Gene | Chr. band | Disorder | Inheritance pattern |
|---|---|---|---|---|
| 1 | *POMGNT1* | 1p34.1 | Muscle-eye-brain disease | AR |
| 2 | *RPE65* | 1p31.3 | Leber congenital amaurosis | AR |
| 3 | *DPYD* | 1p21.3 | Dihydropyrimidine dehydrogenase deficiency | AR |
| 4 | *NRXN1* | 2p16.3 | Disrupted in ASD, ID, schizophrenia (dominant); Pitt-Hopkins-like syndrome 2 (recessive) | AD/AR |
| 5 | *NPHP1* | 2q13 | Joubert syndrome type 4, nephronophthisis | AR |
| 6 | *MBD5* | 2q23.1 | Autosomal dominant ID, 2q23.1 microdeletion syndrome | AD |
| 7 | *SCN1A* | 2q24.3 | Severe myoclonic epilepsy of infancy (Dravet syndrome) | AD |
| 8 | *SCN2A* | 2q24.3 | Benign familial neonatal-infantile seizures, intractable childhood epilepsy | AD |
| 9 | *SATB2* | 2q33.1 | Cleft palate and ID; implicated in the 2q33.1 microdeletion syndrome | AD |
| 10 | *BTD* | 3p24.3 | Biotinidase deficiency | AR |
| 11 | *FOXP1* | 3p14.1 | Non-syndromic ID with language impairment and ASD | AD |
| 12 | *PRSS12* | 4q26 | Autosomal recessive non-syndromic ID | AR |
| 13 | *NIPBL* | 5p13.2 | Cornelia de Lange syndrome | AD |
| 14 | *MEF2C* | 5q14.3 | ID, stereotypic movements, epilepsy, and/or cerebral malformations; 5q14.3 microdeletion syndrome | AD |
| 15 | *ALDH7A1* | 5q23.2 | Pyridoxine-dependent epilepsy | AR |
| 16 | *NSD1* | 5q35.2-q35.3 | Sotos syndrome | AD |
| 17 | *ALDH5A1* | 6p22.2 | Succinic semialdehyde dehydrogenase deficiency | AR |
| 18 | *SYNGAP1* | 6p21.32 | Autosomal dominant ID | AD |
| 19 | *AHI1* | 6q23.3 | Joubert syndrome 3 | AR |
| 20 | *PEX7* | 6q23.3 | Refsum disease, Rhizomelic chondrodysplasia punctata, type 1 | AR |
| 21 | *ARID1B* | 6q25.3 | ID, speech impairment, minor anomalies and variable corpus callosum abnormalities; Coffin-Siris syndrome | AD |
| 22 | *HOXA1* | 7p15.2 | HOXA1 syndrome, Bosley-Salih-Alorainy variant | AR |
| 23 | *BRAF* | 7q34 | Cardio-facio-cutaneous syndrome | AD |
| 24 | *CNTNAP2* | 7q35-q36.1 | Cortical dysplasia-focal epilepsy syndrome, Pitt-Hopkins-like syndrome 1 (recessive); the clinical significance of the disruption of 1 allele is unknown | AR |
| 25 | *HGSNAT* | 8p11.21 | Mucopolysaccharidosis type IIIC (Sanfilippo syndrome C) | AR |
| 26 | *TUSC3* | 8p22 | Autosomal recessive non-syndromic ID | AR |
| 27 | *CHD7* | 8q12.2 | CHARGE syndrome | AD |
| 28 | *VPS13B* | 8q22.2 | Cohen syndrome | AR |
| 29 | *SMARCA2* | 9p24.3 | Nicolaides-Baraitser syndrome; Coffin-Siris syndrome | AD |
| 30 | *STXBP1* | 9q34.11 | Non-syndromic epilepsy, ID and autism; early infantile epileptic encephalopathy | AD |
| 31 | *POMT1* | 9q34.13 | Limb-girdle muscular dystrophy with ID; Walker-Warburg syndrome | AR |
| 32 | *TSC1* | 9q34.13 | Tuberous sclerosis | AD |
| 33 | *EHMT1* | 9q34.3 | Kleefstra syndrome (9q subtelomeric deletion syndrome) | AD |
| 34 | *PTEN* | 10q23.31 | PTEN hamartoma-tumor syndrome, ID and ASD with macrocephaly | AD |
| 35 | *FGFR2* | 10q26.13 | Apert syndrome | AD |
| 36 | *HRAS* | 11p15.5 | Costello syndrome | AD |
| 37 | *IGF2* | 11p15.5 | Aberrant imprinting of *IGF2* is associated with Beckwith–Wiedermann syndrome and Silver–Russell syndrome | AD |
| 38 | *KCNJ11* | 11p15.1 | DEND syndrome (developmental delay, epilepsy, and neonatal diabetes) | AD |
| 39 | *SHANK2* | 11q13.3 | Non-syndromic ID and ASD | AD |
| 40 | *DHCR7* | 11q13.4 | Smith-Lemli-Opitz syndrome | AR |
| 41 | *FOLR1* | 11q13.4 | Cerebral folate transport deficiency | AR |
| 42 | *HEPACAM* | 11q24.2 | Megalencephalic leukoencephalopathy with subcortical cysts (recessive); leukodystrophy and macrocephaly (dominant) | AR/AD |
| 43 | *CACNA1C* | 12p13.33 | Timothy syndrome | AD |
| 44 | *GRIN2B* | 12p13.1 | Autosomal dominant ID | AD |
| 45 | *KRAS* | 12p12.1 | Cardio-facio-cutaneous syndrome | AD |
| 46 | *SCN8A* | 12q13.13 | Early infantile epileptic encephalopathy | AD |
| 47 | *GNS* | 12q14.3 | Mucopolysaccharidosis type IIID (Sanfilippo disease D) | AR |
| 48 | *BBS10* | 12q21.2 | Bardet-Biedl syndrome | AR |
| 49 | *CEP290* | 12q21.32 | Joubert syndrome 5, Leber congenital amaurosis, Bardet-Biedl syndrome, Meckel syndrome | AR |
| 50 | *PAH* | 12q23.2 | Phenylketonuria | AR |
| 51 | *PTPN11* | 12q24.13 | Noonan syndrome | AD |
| 52 | *CHD8* | 14q11.2 | Autosomal dominant ASD and ID | AD |
| 53 | *FOXG1* | 14q12 | Congenital variant of Rett syndrome | AD |
| 54 | *L2HGDH* | 14q22.1 | L-2-hydroxyglutaric aciduria | AR |
| 55 | *UBE3A* | 15q11.2 | Angelman syndrome | AD |
| 56 | *SPRED1* | 15q14 | Neurofibromatosis type 1-like syndrome (Legius syndrome) | AD |
| 57 | *GATM* | 15q21.1 | Arginine:glycine amidinotransferase (AGAT) deficiency | AR |
| 58 | *MAP2K1* | 15q22.31 | Cardio-facio-cutaneous syndrome | AD |
| 59 | *TSC2* | 16p13.3 | Tuberous sclerosis | AD |
| 60 | *CREBBP* | 16p13.3 | Rubinstein-Taybi syndrome | AD |
| 61 | *SRCAP* | 16p11.2 | Floating-Harbor syndrome | AD |
| 62 | *BCKDK* | 16p11.2 | Autosomal recessive autism, ID & epilepsy/abnormal EEG | AR |
| 63 | *RPGRIP1L* | 16q12.2 | Joubert syndrome 7, Meckel syndrome, COACH syndrome | AR |

| | Gene | Chr. band | Disorder | Inheritance pattern |
|---|---|---|---|---|
| 64 | *ANKRD11* | 16q24.3 | KBG syndrome; 16q24.3 microdeletion syndrome | AD |
| 65 | *YWHAE* | 17p13.3 | Miller-Dieker syndrome | AD |
| 66 | *PAFAH1B1* | 17p13.3 | Isolated lissencephaly, Miller-Dieker syndrome | AD |
| 67 | *GUCY2D* | 17p13.1 | Leber congenital amaurosis | AR |
| 68 | *RAI1* | 17p11.2 | Smith-Magenis syndrome (deletion, mutation), Potocki-Lupski syndrome (duplication) | AD |
| 69 | *RNF135* | 17q11.2 | Overgrowth syndrome; haploinsufficiency of *RNF135* contributes to the phenotype of the NF1 microdeletion syndrome | AD |
| 70 | *NF1* | 17q11.2 | Neurofibromatosis type 1 | AD |
| 71 | *NAGLU* | 17q21.31 | Mucopolysaccharidosis type IIIB (Sanfilippo syndrome B) | AR |
| 72 | *SGSH* | 17q25.3 | Sanfilippo syndrome A (mucopolysaccharidosis III A) | AR |
| 73 | *SETBP1* | 18q12.3 | Haploinsufficiency of *SETBP1* causes the core clinical features of the del(18)(q12.2q21.1) syndrome | AD |
| 74 | *SMAD4* | 18q21.2 | Myhre syndrome | AD |
| 75 | *TCF4* | 18q21.2 | Pitt-Hopkins syndrome | AD |
| 76 | *NFIX* | 19p13.13 | Sotos-like overgrowth syndrome, Marshall-Smith syndrome | AD |
| 77 | *GAMT* | 19p13.3 | Guanidine acetate methyltransferase (GAMT) deficiency | AR |
| 78 | *DMPK* | 19q13.32 | Myotonic dystrophy type 1 (Steinert disease) | AD |
| 79 | *MKKS* | 20p12.2 | Bardet-Biedl syndrome | AR |
| 80 | *DYRK1A* | 21q22.13 | Autosomal dominant ID | AD |
| 81 | *ADSL* | 22q13.1 | Adenylosuccinate lyase deficiency | AR |
| 82 | *SHANK3* | 22q13.33 | Phelan-mcdermid syndrome (22q13 deletion syndrome) | AD |
| 83 | *NLGN4X* | Xp22.31-p22.32 | Non-syndromic X-linked ID and ASD | XLR |
| 84 | *MID1* | Xp22.2 | Opitz syndrome (Opitz/BBB syndrome) | XLR |
| 85 | *AP1S2* | Xp22.2 | Syndromic X-linked ID, Fried type; non-syndromic X-linked ID | XLR |
| 86 | *NHS* | Xp22.13 | Nance-Horan syndrome | XLD |
| 87 | *CDKL5* | Xp22.13 | Early infantile epileptic encephalopathy | XLD |
| 88 | *PTCHD1* | Xp22.11 | Non-syndromic X-linked ID and ASD | XLR |
| 89 | *ARX* | Xp21.3 | X-linked lissencephaly and abnormal genitalia, West syndrome, Partington syndrome, non-syndromic X-linked ID | XLR |
| 90 | *IL1RAPL1* | Xp21.2-p21.3 | Non-syndromic X-linked ID and ASD | XLR |
| 91 | *DMD* | Xp21.1-21.2 | Muscular dystrophy, Duchenne and Becker types | XLR |
| 92 | *OTC* | Xp11.4 | Ornithine transcarbamylase deficiency | XLD/XLR |
| 93 | *CASK* | Xp11.4 | Syndromic and non-syndromic X-linked ID | XLD |
| 94 | *NDP* | Xp11.3 | Norrie disease | XLR |
| 95 | *KDM6A* | Xp11.3 | Kabuki syndrome | XLD |
| 96 | *SYN1* | Xp11.23 | X-linked epilepsy and ID | XLR |
| 97 | *FTSJ1* | Xp11.23 | Non-syndromic X-linked ID | XLR |
| 98 | *PQBP1* | Xq11.23 | Renpenning syndrome, non-syndromic ID | XLR |
| 99 | *CACNA1F* | Xp11.23 | X-linked incomplete congenital stationary night blindness | XLR |
| 100 | *KDM5C* | Xp11.22 | Syndromic X-linked ID, Claes-Jensen type; non-syndromic X-linked ID | XLR |
| 101 | *IQSEC2* | Xp11.22 | Non-syndromic X-linked ID | XLD |
| 102 | *SMC1A* | Xp11.22 | Cornelia de Lange syndrome | XLD |
| 103 | *HSD17B10* | Xp11.22 | 17-beta-hydroxysteroid dehydrogenase X deficiency, X-linked syndromic ID with choreoathetosis and abnormal behavior | XLD |
| 104 | *PHF8* | Xp11.22 | Siderius-Hamel syndrome | XLR |
| 105 | *FGD1* | Xp11.22 | Aarskog-Scott syndrome, non-syndromic X-linked ID | XLR |
| 106 | *OPHN1* | Xq12 | X-linked ID with cerebellar hypoplasia and distinctive facial appearance | XLR |
| 107 | *MED12* | Xq13.1 | Lujan-Fryns syndrome, Opitz-Kaveggia syndrome | XLR |
| 108 | *NLGN3* | Xq13.1 | Non-syndromic X-linked ID and ASD | XLR |
| 109 | *ATRX* | Xq21.1 | Alpha-thalassemia/mental retardation syndrome, non-syndromic X-linked ID | XLD |
| 110 | *PCDH19* | Xq22.1 | Female-limited epilepsy with ID, early infantile epileptic encephalopathy | XL - affected females |
| 111 | *ACSL4* | Xq22.3 | Non-syndromic X-linked ID | XLD |
| 112 | *DCX* | Xq22.3 | Type 1 lissencephaly | XLD |
| 113 | *UPF3B* | Xq24 | Non-syndromic X-linked ID, Opitz-Kaveggia/Lujan-Fryns phenotype | XLR |
| 114 | *LAMP2* | Xq24 | Danon disease | XLD |
| 115 | *GRIA3* | Xq25 | Syndromic X-linked ID, Wu type; non-syndromic X-linked ID | XLR |
| 116 | *OCRL* | Xq25 | Lowe syndrome | XLR |
| 117 | *PHF6* | Xq26.2 | Borjeson-Forssman-Lehmann syndrome | XLR |
| 118 | *SLC9A6* | Xq26.3 | Syndromic X-linked ID, Christianson type | XLD? |
| 119 | *FMR1* | Xq27.3 | Fragile X syndrome | XLD |
| 120 | *AFF2* | Xq28 | Fragile X mental retardation 2 | XLR |
| 121 | *SLC6A8* | Xq28 | Creatine deficiency syndrome, non-syndromic X-linked ID | XLR |
| 122 | *L1CAM* | Xq28 | MASA (mental retardation, aphasia, shuffling gait, and adducted thumbs) syndrome | XLR |
| 123 | *MECP2* | Xq28 | Rett syndrome, non-syndromic X-linked ID (mutation, deletion; XL dominant); *MECP2* duplication syndrome (XL recessive) | XLD/XLR |
| 124 | *RAB39B* | Xq28 | Non-syndromic X-linked ID | XLR |

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ASD, autism spectrum disorder; ID, intellectual disability; XL, X-linked; XLD, X-linked dominant; XLR, X-linked recessive

## Table S6B. Loci implicated in ASD

| | Disorder | Chr. band | Start (hg18) | End (hg18) | Genes involved | Inheritance pattern |
|---|---|---|---|---|---|---|
| 1 | 1p36 deletion syndrome | 1p36.32-p36.33 | 823,946 | 5,308,621 | contiguous gene syndrome | AD |
| 2 | 1q21.1 deletion/duplication syndrome | 1q21.1 | 145,044,110 | 145,861,130 | ? | AD |
| 3 | 1q41q42 microdeletion syndrome | 1q41q42 | 219,500,000 | 223,000,000 | ? | AD |
| 4 | 1q43q44 microdeletion syndrome | 1q43q44 | 238,000,000 | 247,249,718 | ? | AD |
| 5 | 2p16.1p15 microdeletion syndrome | 2p16.1p15 | 57,595,300 | 61,591,838 | ? | AD |
| 6 | 2q13 deletion/duplication | 2q13 | 111,158,601 | 112,782,250 | ? | AD |
| 7 | 2q23.1 microdeletion syndrome | 2q23.1 | 148,932,242 | 148,987,514 | *MBD5* | AD |
| 8 | 2q33.1 deletion syndrome | 2q32.3-q33.2 | 196,633,334 | 204,915,185 | *SATB2* (maybe other gene(s) contribute too) | AD |
| 9 | 2q37 deletion syndrome (brachydactyly-mental retardation syndrome) | 2q37.3 | 239,619,630 | 242,579,273 | *HDAC4* (other distal gene(s) contribute too) | AD |
| 10 | 3q13.31 microdeletion syndrome | 3q12.3q21.3 | 115,335,356 | 115,916,848 | ? | AD |
| 11 | 3q29 microdeletion/microduplication syndrome | 3q29 | 197,240,451 | 198,829,062 | ? (candidates: *PAK2* and *DLG1*) | AD |
| 12 | Wolf-Hirschhorn syndrome | 4p16.3 | 62,448 | 2,297,002 | contiguous gene syndrome | AD |
| 13 | 4q21 microdeletion syndrome | 4q21.21-q21.22 | 82,228,875 | 83,182,488 | ? | AD |
| 14 | Cri du Chat syndrome (5p deletion) | 5p15.2-p15.33 | 90,693 | 11,400,262 | ? | AD |
| 15 | 5q14.3 microdeletion syndrome | 5q14.3 | 88,049,814 | 88,235,678 | *MEF2C* | AD |
| 16 | Sotos syndrome (deletion), 5q35.2q35.3 duplication | 5q35.2-q35.3 | 175,661,584 | 176,946,567 | *NSD1* | AD |
| 17 | 6p subtelomere deletion syndrome | 6p25 | 100,000 | 3,000,000 | ? (*FOXC1* involved in ophthalmologic findings) | AD |
| 18 | Williams syndrome (deletion), 7q11.23 duplication syndrome | 7q11.23 | 72,382,390 | 73,780,449 | contiguous gene syndrome | AD |
| 19 | 8p23.1 deletion/duplication syndrome | 8p23.1 | 8,156,705 | 11,803,128 | ? (*GATA4* involved in heart defects) | AD |
| 20 | 8q21.11 microdeletion syndrome | 8q21.11 | 77,389,019 | 77,928,794 | ? | AD |
| 21 | Kleefstra syndrome (9q subtelomeric deletion syndrome) | 9q34.3 | 139,523,184 | 140,273,252 | *EHMT1* | AD |
| 22 | 10p14p15 deletion syndrome | 10p14-p15.1 | 4,700,001 | 10,600,000 | ? *(GATA3* involved in hypoparathyroidism, deafness, renal disease) | AD |
| 23 | 10p12p11 microdeletion | 10p12.31p11.21 | 28,833,195 | 29,138,742 | ? | AD |
| 24 | 10q22-q23 deletion | 10q22.3-q23.2 | 81,682,644 | 88,931,994 | ? | AD |
| 25 | Distal 10q deletion syndrome | 10q26.2-q26.3 | 128,000,000 | 135,374,737 | ? | AD |
| 26 | 11p15.5 duplication: Beckwith-Wiedemann/Silver-Russell syndromes | 11p15.4-p15.5 | 1,970,000 | 2,870,000 | *H19, IGF2* | AD |
| 27 | WAGR syndrome (11p13 deletion syndrome) | 11p13 | 31,760,085 | 32,467,564 | ? | AD |
| 28 | Potocki-Shaffer syndrome (11p11.2 deletion syndrome) | 11p11.2 | 43,941,853 | 46,021,136 | *PHF21A (EXT2* and *ALX4* involved in bone defects) | AD |
| 29 | Jacobsen syndrome (11q deletion syndrome) | 11q23.3-qter | 115,400,001 | 134,452,384 | ? | AD |
| 30 | 12q14 microdeletion syndrome | 12q14 | 63,358,186 | 66,931,792 | ? (*HMGA2* involved in short stature) | AD |
| 31 | Terminal deletion 14q syndrome | 14q32.31-q32.33 | 101,500,000 | 105,000,000 | ? | AD |
| 32 | Angelman syndrome (maternal deletion), Prader-Willi syndrome (paternal deletion), 15q11-q13 duplication syndrome | 15q11.2-q13.1 | 21,309,483 | 26,230,781 | Angelman: maternal *UBE3A*; Prader-Willi: paternally expressed genes (HBII-85 snoRNA cluster; *SNURF-SNRPN, NDN* and *MAGEL2*?) | AD |
| 33 | 15q13.3 deletion syndrome (duplication = uncertain significance) | 15q13.2-q13.3 | 28,924,396 | 30,232,700 | *CHRNA7* | AD |
| 34 | 15q24 microdeletion syndrome | 15q24.1-q24.2 | 72,164,227 | 73,949,332 | ? | AD |
| 35 | Distal 15q25.2q25.3 microdeletion | 15q25.2q25.3 | 82,944,098 | 83,484,862 | ? | AD |
| 36 | 15q26 overgrowth syndrome | 15q26.3 | 97,175,493 | 100,218,756 | *IGF1R* | AD |
| 37 | Rubinstein-Taybi syndrome (deletion), 16p13.3 duplication syndrome | 16p13.3 | 3,715,057 | 3,870,122 | *CREBBP* | AD |
| 38 | 16p13.11 microdeletion syndrome (duplication = uncertain significance) | 16p13.11 | 15,411,955 | 16,199,769 | ? | AD |
| 39 | 16p11.2-p12.2 microdeletion/microduplication syndrome | 16p11.2-p12.2 | 21,521,457 | 28,949,693 | ? | AD |
| 40 | 16p11.2 microdeletion/microduplication syndrome | 16p11.2 | 29,557,497 | 30,107,356 | ? | AD |
| 41 | Miller-Dieker syndrome/isolated lissencephaly (deletion), 17p13.3 microduplication | 17p13.3 | 1,129,706 | 2,535,659 | *PAFAH1B1, YWHAE* | AD |
| 42 | Smith-Magenis syndrome (deletion), Potocki-Lupski syndrome (duplication) | 17p11.2 | 16,697,836 | 20,160,243 | *RAI1* | AD |
| 43 | NF1 microdeletion/microduplication syndrome | 17q11.2 | 26,186,948 | 27,242,780 | *NF1* (*RNF135* contributes to the overgrowth, facial dysmorphism and ID) | AD |
| 44 | 17q12 deletion syndrome (renal cysts and diabetes syndrome), 17q12 duplication syndrome | 17q12 | 31,930,169 | 33,323,031 | ? (*HNF1B* involved in renal cysts and diabetes syndrome) | AD |
| 45 | Koolen-De Vries syndrome (17q21.31 microdeletion syndrome), 17q21.31 microduplication syndrome | 17q21.31 | 41,060,949 | 41,544,225 | *KANSL1* | AD |
| 46 | del(18)(q12.2q21.1) syndrome | 18q12.2q21.1 | 39,890,000 | 41,700,000 | *SETBP1* | AD |
| 47 | Down syndrome (trisomy 21) | whole chr 21 | 1 | 46,944,323 | contiguous gene syndrome | AD |

| | Disorder | Chr. band | Start (hg18) | End (hg18) | Genes involved | Inheritance pattern |
|---|---|---|---|---|---|---|
| 48 | 22q11.2 deletion syndrome (velocardiofacial/DiGeorge syndrome), 22q11.2 duplication syndrome | 22q11.21-q11.22 | 17,041,725 | 18,691,904 | contiguous gene syndrome, *TBX1* is responsible for most of the physical malformations | AD |
| 49 | Phelan-McDermid syndrome (22q13 deletion syndrome), 22q13 duplication | 22q13.33 | 49,392,382 | 49,525,811 | *SHANK3* | AD |
| 50 | Xq28 duplication syndrome (*MECP2* duplication syndrome) | Xq28 | 152,403,094 | 153,044,193 | *MECP2* | XLR |
| 51 | Turner syndrome (X0) | whole chr X | 1 | 154,913,754 | contiguous gene syndrome | |
| 52 | Klinefelter syndrome (XXY) | whole chr X | 1 | 154,913,754 | contiguous gene syndrome | |
| 53 | XYY syndrome | whole chr Y | 1 | 57,772,954 | contiguous gene syndrome | |
| 54 | XXYY syndrome | whole chr X-Y | | | contiguous gene syndrome | |
| 55 | 45,X/46,XY mosaicism | whole chr X | 1 | 154,913,754 | contiguous gene syndrome | |

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ASD, autism spectrum disorder; ID, intellectual disability; XLR, X-linked recessive

## Table S6C. Genes implicated in ID

| | Gene | Chr. band | Disorder | Inheritance pattern |
|---|---|---|---|---|
| 1 | *SKI* | 1p36.33 | Shprintzen-Goldberg syndrome | AD |
| 2 | *GALE* | 1p36.11 | Galactose epimerase deficiency (galactosemia III) | AR |
| 3 | *FUCA1* | 1p36.11 | Fucosidosis | AR |
| 4 | *ARID1A* | 1p36.11 | Coffin-Siris syndrome | AD |
| 5 | *PIGV* | 1p36.11 | Hyperphosphatasia mental retardation syndrome | AR |
| 6 | *SLC2A1* | 1p34.2 | Glucose transport defect | AD |
| 7 | *ST3GAL3* | 1p34.3 | Autosomal recessive non-syndromic ID | AR |
| 8 | *STIL* | 1p33 | Primary microcephaly | AR |
| 9 | *ALG6* | 1p31.3 | Congenital disorder of glycosylation, type Ic | AR |
| 10 | *DBT* | 1p21.2 | Maple syrup urine disease, type II | AR |
| 11 | *AP4B1* | 1p13.2 | Autosomal recessive ID with spastic paraplegia | AR |
| 12 | *NRAS* | 1p13.2 | Noonan syndrome | AD |
| 13 | *GATAD2B* | 1q21.3 | Autosomal dominant ID | AD |
| 14 | *KCNJ10* | 1q23.2 | SESAME syndrome (seizures, sensorineural deafness, ataxia, mental retardation, and electrolyte imbalance) | AR |
| 15 | *ASPM* | 1q31 | Microcephaly and ID | AR |
| 16 | *SYT14* | 1q32.2 | ID with adult-onset spinocerebellar ataxia | AR |
| 17 | *RAB3GAP2* | 1q41 | Martsolf syndrome (congenital cataracts, hypogonadism, and ID) | AR |
| 18 | *TBCE* | 1q42.3 | Hypoparathyroidism-retardation-dysmorphism syndrome | AR |
| 19 | *FH* | 1q43 | Fumarase deficiency | AR |
| 20 | *MYCN* | 2p24.3 | Feingold syndrome (microcephaly-oculo-digito-esophageal-duodenal syndrome), Microcephaly and digital abnormalities with normal intelligence | AD |
| 21 | *SOS1* | 2p22.1 | Noonan syndrome | AD |
| 22 | *ERCC3* | 2q14.3 | Trichothiodystrophy; Xeroderma pigmentosum, group B | AR |
| 23 | *RAB3GAP1* | 2q21.3 | Warburg Micro syndrome 1 | AR |
| 24 | *ZEB2* | 2q22.3 | Mowat-Wilson syndrome (Hirschsprung disease-mental retardation syndrome) | AD |
| 25 | *BBS5* | 2q31.1 | Bardet-Biedl syndrome | AR |
| 26 | *GAD1* | 2q31.1 | Spastic quadriplegic cerebral palsy | AR |
| 27 | *TMEM237* | 2q33.1 | Joubert syndrome 14 | AR |
| 28 | *HDAC4* | 2q37.3 | Brachydactyly mental retardation syndrome (2q37 deletion syndrome) | AD |
| 29 | *D2HGDH* | 2q37.3 | D-2-hydroxyglutaric aciduria | AR |
| 30 | *CRBN* | 3p26.2 | Autosomal recessive non-syndromic ID | AR |
| 31 | *SUMF1* | 3p26.2 | Multiple sulfatase deficiency | AR |
| 32 | *TSEN2* | 3p25.1 | Pontocerebellar hypoplasia type 2B | AR |
| 33 | *RAF1* | 3p25.1 | Noonan syndrome, LEOPARD syndrome | AD |
| 34 | *TGFBR2* | 3p24.1 | Loeys–Dietz syndrome | AD |
| 35 | *GLB1* | 3p22.3 | GM1-gangliosidosis, Mucopolysaccharidosis IVB | AR |
| 36 | *CTNNB1* | 3p22.1 | Autosomal dominant ID | AD |
| 37 | *GTDC2* | 3p22.1 | Walker-Warburg syndrome | AR |
| 38 | *LZTFL1* | 3p21.31 | Bardet-Biedl syndrome with situs inversus and polydactyly | AR |
| 39 | *DAG1* | 3p21.31 | Limb-girdle muscular dystrophy | AR |
| 40 | *ARL13B* | 3q11.2 | Joubert syndrome 8 | AR |
| 41 | *ARL6* | 3q11.2 | Bardet-Biedl syndrome | AR |
| 42 | *CEP63* | 3q22.2 | Primary microcephaly | AR |
| 43 | *ATR* | 3q23 | Seckel syndrome | AR |
| 44 | *ALG3* | 3q27.1 | Congenital disorder of glycosylation, type Id | AR |
| 45 | *KIAA0226* | 3q29 | Syndromic ID with ataxia, dysarthria and epilepsy | AR |
| 46 | *IDUA* | 4p16.3 | Mucopolysaccharidosis Ih (Hurler syndrome); mucopolysaccharidosis Is (Scheie syndrome) | AR |
| 47 | *CC2D2A* | 4p15.3 | Joubert syndrome 9, Meckel syndrome, COACH syndrome | AR |
| 48 | *QDPR* | 4p15.32 | Hyperphenylalaninemia due to dihydropteridine reductase deficiency | AR |
| 49 | *SRD5A3* | 4q12 | Kahrizi syndrome; congenital disorder of glycosylation, type Iq | AR |
| 50 | *TMEM165* | 4q12 | Congenital disorder of glycosylation, type iik | AR |
| 51 | *CEP135* | 4q12 | Autosomal-recessive primary microcephaly | AR |

| | Gene | Chr. band | Disorder | Inheritance pattern |
|---|---|---|---|---|
| 52 | *SLC4A4* | 4q13.3 | Renal tubular acidosis, proximal, with ocular abnormalities | AR |
| 53 | *BBS7* | 4q27 | Bardet-Biedl syndrome | AR |
| 54 | *BBS12* | 4q27 | Bardet-Biedl syndrome | AR |
| 55 | *AGA* | 4q34.3 | Aspartylglucosaminuria | AR |
| 56 | *NSUN2* | 5p15.31 | Autosomal-recessive syndromic ID, Dubowitz syndrome | AR |
| 57 | *ANKH* | 5p15.2 | Chondrocalcinosis 2, Craniometaphyseal dysplasia | AR |
| 58 | *C5orf42* | 5p13.2 | Joubert syndrome 17 | AR |
| 59 | *MOCS2* | 5q11.2 | Molybdenum cofactor deficiency, type B | AR |
| 60 | *ERCC8* | 5q12.1 | Cockayne syndrome type A | AR |
| 61 | *SIL1* | 5q31.2 | Marinesco-Sjogren syndrome | AR |
| 62 | *TUBB2B* | 6p25.2 | Asymmetric polymicrogyria | AD |
| 63 | *NEU1* | 6p21.3 | Sialidosis type I, Sialidosis type II | AR |
| 64 | *MOCS1* | 6p21.2 | Molybdenum cofactor deficiency, type A | AR |
| 65 | *SLC17A5* | 6q13 | Salla disease; Sialic acid storage disorder, infantile | AR |
| 66 | *ELOVL4* | 6q14.1 | Ichthyosis, spastic quadriplegia, and mental retardation (recessive); Macular dystrophy (dominant) | AR/AD |
| 67 | *BCKDHB* | 6q14.1 | Maple syrup urine disease, type Ib | AR |
| 68 | *RARS2* | 6q15 | Pontocerebellar hypoplasia | AR |
| 69 | *GRIK2* | 6q16.3 | Autosomal recessive non-syndromic ID | AR |
| 70 | *SOBP* | 6q21 | Autosomal recessive syndromic and non-syndromic ID | AR |
| 71 | *LAMA2* | 6q22.33 | Merosin-deficient congenital muscular dystrophy type 1A | AR |
| 72 | *ARG1* | 6q23.2 | Argininemia | AR |
| 73 | *MED23* | 6q23.2 | Autosomal recessive non-syndromic ID | AR |
| 74 | *GTF2H5* | 6q25.3 | Trichothiodystrophy | AR |
| 75 | *ACTB* | 7p22.1 | Baraitser-Winter syndrome | AD |
| 76 | *FAM126A* | 7p15.3 | Hypomyelinating leukodystrophy | AR |
| 77 | *BBS9* | 7p14.3 | Bardet-Biedl syndrome | AR |
| 78 | *C7orf11* | 7p14.1 | Trichothiodystrophy | AR |
| 79 | *GUSB* | 7q11.21 | Mucopolysaccharidosis VII | AR |
| 80 | *AP4M1* | 7q22.1 | Autosomal recessive tetraplegic cerebral palsy with ID | AR |
| 81 | *RELN* | 7q22 | Lissencephaly | AR |
| 82 | *DLD* | 7q31.1 | Maple syrup urine disease, type III | AR |
| 83 | *CEP41* | 7q32.2 | Joubert syndrome 15 | AR |
| 84 | *TPK1* | 7q35 | Thiamine metabolism dysfunction syndrome (episodic encephalopathy type) | AD |
| 85 | *EZH2* | 7q36.1 | Weaver syndrome | AD |
| 86 | *MCPH1* | 8p23 | Microcephaly and ID | AR |
| 87 | *ERLIN2* | 8p12 | Autosomal recessive ID, motor dysfunction and multiple joint contractures | AR |
| 88 | *CA8* | 8q12.1 | Cerebellar ataxia, quadrupedal locomotion and ID | AR |
| 89 | *TMEM67* | 8q21 | Joubert syndrome 6, Meckel-Gruber syndrome | AR |
| 90 | *RAD21* | 8q24.11 | Cornelia de Lange syndrome | AD |
| 91 | *KCNK9* | 8q24.3 | Birk-Barel mental retardation dysmorphism syndrome (imprinting defect) | AD |
| 92 | *TRAPPC9* | 8q24.3 | Autosomal recessive non-syndromic ID | AR |
| 93 | *RECQL4* | 8q24.3 | Baller-Gerold syndrome, Rothmund-Thomson syndrome and RAPADILINO syndrome | AR |
| 94 | *VLDLR* | 9p24.2 | Cerebellar ataxia and ID | AR |
| 95 | *SLC1A1* | 9p24.2 | Dicarboxylic aminoaciduria | AR |
| 96 | *PIGO* | 9p13.3 | Hyperphosphatasia with mental retardation syndrome | AR |
| 97 | *EXOSC3* | 9p13.2 | Pontocerebellar hypoplasia and spinal motor neuron degeneration | AR |
| 98 | *TGFBR1* | 9q22.33 | Loeys–Dietz syndrome | AD |
| 99 | *FKTN* | 9q31.2 | Fukuyama congenital muscular dystrophy with type 2 lissencephaly, Walker-Warburg syndrome | AR |
| 100 | *TRIM32* | 9q33.1 | Bardet-Biedl syndrome | AR |
| 101 | *CDK5RAP2* | 9q33.2 | Microcephaly vera | AR |
| 102 | *SPTAN1* | 9q34.11 | West syndrome with severe cerebral hypomyelination, spastic quadriplegia and ID | AD |
| 103 | *KCNT1* | 9q34.3 | Malignant migrating partial seizures of infancy; nocturnal frontal lobe epilepsy, ID | AD |
| 104 | *INPP5E* | 9q34.3 | Joubert syndrome 1 | AR |
| 105 | *MAN1B1* | 9q34.3 | Autosomal recessive non-syndromic ID | AR |
| 106 | *RAB18* | 10p12.1 | Warburg micro syndrome | AR |
| 107 | *ERCC6* | 10q11.23 | Cockayne syndrome type B, Cerebro-oculo-facio-skeletal syndrome | AR |
| 108 | *KIAA1279* | 10q21.3 | Goldberg-Shprintzen megacolon syndrome | AR |
| 109 | *ADK* | 10q22.2 | Hypermethioninemia due to adenosine kinase deficiency | AR |
| 110 | *KAT6B* | 10q22.2 | Say-Barber-Biesecker-Young-Simpson syndrome (SBBYSS syndrome) | AD |
| 111 | *POLR3A* | 10q22.3 | Hypomyelinating leukodystrophy with or without oligodontia and/or hypogonadotropic hypogonadism | AR |
| 112 | *KIF11* | 10q23.33 | Microcephaly variably associated with congenital lymphedema, chorioretinopathy and learning difficulties | AD |
| 113 | *TCTN3* | 10q23.33 | Joubert syndrome 18, orofaciodigital syndrome IV | AR |
| 114 | *SMC3* | 10q25.2 | Cornelia de Lange syndrome | AD |
| 115 | *SHOC2* | 10q25.2 | Noonan syndrome | AD |
| 116 | *SLC25A22* | 11p15.5 | Autosomal recessive neonatal epileptic encephalopathy | AR |
| 117 | *PAX6* | 11p13 | Isolated and syndromic aniridia, including Gillespie syndrome (aniridia, cerebellar ataxia and ID) | AD |
| 118 | *SLC35C1* | 11p11.2 | Congenital disorder of glycosylation, type iic | AR |
| 119 | *PHF21A* | 11p11.2 | Potocki-Shaffer syndrome (11p11.2 deletion) | AD |
| 120 | *TMEM138* | 11q12.2 | Joubert syndrome 16 | AR |
| 121 | *TMEM216* | 11q12.2 | Joubert syndrome 2 | AR |
| 122 | *BBS1* | 11q13.1 | Bardet-Biedl syndrome | AR |

| | Gene | Chr. band | Disorder | Inheritance pattern |
|---|---|---|---|---|
| 123 | *ALG8* | 11q14.1 | Congenital disorder of glycosylation type Ih | AR |
| 124 | *MED17* | 11q21 | Primary microcephaly of postnatal onset, spasticity, epilepsy, and ID | AR |
| 125 | *ALG9* | 11q23.1 | Congenital disorder of glycosylation, type Il | AR |
| 126 | *CBL* | 11q23.3 | Noonan syndrome-like disorder | AD |
| 127 | *PVRL1* | 11q23.3 | Cleft lip/palate ectodermal dysplasia syndrome | AR |
| 128 | *MLL2* | 12q13.12 | Kabuki syndrome | AD |
| 129 | *TUBA1A* | 12q13.12 | Lissencephaly | AD |
| 130 | *DIP2B* | 12q13.13 | Mental retardation, FRA12A type | AD |
| 131 | *SUOX* | 12q13.2 | Sulfite oxidase deficiency | AR |
| 132 | *GNPTAB* | 12q23.2 | Mucolipidosis III alpha/beta | AR |
| 133 | *IGF1* | 12q23.2 | Growth retardation with deafness and mental retardation due to IGF1 deficiency | AR |
| 134 | *POLR3B* | 12q23.3 | Leukodystrophy, hypomyelinating, 8, with or without oligodontia and/or hypogonadotropic hypogonadism | AR |
| 135 | *ATP6V0A2* | 12q24.31 | Cutis laxa with epilepsy and mental retardation | AR |
| 136 | *CENPJ* | 13q12.12 | Microcephaly vera, Seckel syndrome | AR |
| 137 | *SLC25A15* | 13q14.11 | Orinithine translocase deficiency | AR |
| 138 | *MIR17HG* | 13q31.3 | Feingold syndrome | AD |
| 139 | *ERCC5* | 13q33.1 | Cockayne syndrome, Cerebro-oculo-facio-skeletal syndrome | AR |
| 140 | *COL4A1* | 13q34 | Porencephaly | AD |
| 141 | *AP4S1* | 14q12 | Autosomal recessive ID with spastic paraplegia | AR |
| 142 | *MGAT2* | 14q22.1 | Congenital disorder of glycosylation, type iia | AR |
| 143 | *GCH1* | 14q22.2-q22.3 | Tetrahydrobiopterin-deficient hyperphenylalaninemia B | AR |
| 144 | *POMT2* | 14q24.3 | Walker-Warburg syndrome | AR |
| 145 | *GALC* | 14q31.3 | Krabbe disease | AR |
| 146 | *ZC3H14* | 14q31.3 | Autosomal recessive non-syndromic ID | AR |
| 147 | *TTC8* | 14q31.3 | Bardet-Biedl syndrome | AR |
| 148 | *VRK1* | 14q32.2 | Pontocerebellar hypoplasia type 1 | AR |
| 149 | *DYNC1H1* | 14q32.31 | Severe ID with neuronal migration defects; Charcot-Marie-Tooth disease, axonal, type 20 | AD |
| 150 | *SLC12A6* | 15q14 | Andermann syndrome | AR |
| 151 | *CEP152* | 15q21.1 | Primary microcephaly | AR |
| 152 | *AP4E1* | 15q21.2 | Spastic paraplegia 51, autosomal recessive | AR |
| 153 | *BBS4* | 15q24.1 | Bardet-Biedl syndrome | AR |
| 154 | *KIF7* | 15q26.1 | Joubert syndrome 12 | AR |
| 155 | *IDH2* | 15q26.1 | D-2-hydroxyglutaric aciduria 2 | AD |
| 156 | *GNPTG* | 16p13.3 | Mucolipidosis III gamma | AR |
| 157 | *TBC1D24* | 16p13.3 | Autosomal recessive syndrome of focal epilepsy, dysarthria, and ID | AR |
| 158 | *ZNF423* | 16q12.1 | Joubert syndrome 19 (dominant); nephronophthisis (recessive) | AD/AR |
| 159 | *TUBB3* | 16q34.3 | Cortical dysplasia, complex, with other brain malformations | AD |
| 160 | *PMM2* | 16p13.2 | Congenital disorder of glycosylation, type Ia | AR |
| 161 | *GRIN2A* | 16p13.2 | Variable neurodevelopmental phenotypes, including ID and/or epilepsy | AD |
| 162 | *NDE1* | 16p13.11 | Lissencephaly | AR |
| 163 | *BBS2* | 16q13 | Bardet-Biedl syndrome | AR |
| 164 | *GPR56* | 16q13 | Autosomal recessive bilateral frontoparietal polymicrogyria | AR |
| 165 | *COG8* | 16q22.1 | Congenital disorder of glycosylation, type iih | AR |
| 166 | *TMEM231* | 16q23.1 | Joubert syndrome 20 | AR |
| 167 | *CDH15* | 16q24.3 | Autosomal dominant non-syndromic ID | AD |
| 168 | *CHMP1A* | 16q24.3 | Pontocerebellar hypoplasia, microcephaly, ID | AR |
| 169 | *WDR81* | 17p13.3 | Cerebellar hypoplasia, quadrupedal locomotion and ID | AR |
| 170 | *MPDU1* | 17p13.1 | Congenital disorder of glycosylation, type If | AR |
| 171 | *PIGL* | 17p11.2 | CHIME syndrome (colobomas, heart defects, ichthyosiform dermatosis, mental retardation, and ear anomalies, including conductive hearing loss) | AR |
| 172 | *ALDH3A2* | 17p11.2 | Sjogren-Larsson syndrome | AR |
| 173 | *SLC46A1* | 17q11.2 | Folate malabsorption | AR |
| 174 | *SMARCE1* | 17q21.2 | Coffin-Siris syndrome | AD |
| 175 | *EFTUD2* | 17q21.31 | Mandibulofacial dysostosis with microcephaly | AD |
| 176 | *GFAP* | 17q21.31 | Alexander disease | AD |
| 177 | *MKS1* | 17q22 | Bardet-Biedl syndrome, Meckel syndrome | AR |
| 178 | *COG1* | 17q25.1 | Congenital disorder of glycosylation, type iig | AR |
| 179 | *TSEN54* | 17q25.1 | Pontocerebellar hypoplasia type 2A | AR |
| 180 | *ACTG1* | 17q25.3 | Baraitser-Winter syndrome | AD |
| 181 | *RBBP8* | 18q11.2 | Seckel syndrome, Jawad syndrome | AR |
| 182 | *IER3IP1* | 18q21.1 | Microcephaly with simplified gyration, epilepsy, and infantile diabetes | AR |
| 183 | *RTTN* | 18q22.2 | Polymicrogyria, ID | AR |
| 184 | *MAP2K2* | 19p13.3 | Cardio-facio-cutaneous syndrome | AD |
| 185 | *MCOLN1* | 19p13.2 | Mucolipidosis IV | AR |
| 186 | *SMARCA4* | 19p13.2 | Coffin-Siris syndrome | AD |
| 187 | *CC2D1A* | 19p13.12 | Autosomal recessive non-syndromic ID | AR |
| 188 | *GPSN2* | 19p13.12 | Autosomal recessive non-syndromic ID | AR |
| 189 | *WDR62* | 19q13.12 | Severe brain malformations, including microcephaly, pachygyria and hypoplasia of the corpus callosum | AR |
| 190 | *BCKDHA* | 19q13.2 | Maple syrup urine disease, type Ia | AR |
| 191 | *ERCC2* | 19q13.32 | Cockayne syndrome, Trichothiodystrophy, Cerebro-oculo-facio-skeletal syndrome | AR |
| 192 | *ERCC1* | 19q13.32 | Cerebro-oculo-facio-skeletal syndrome | AR |

| | Gene | Chr. band | Disorder | Inheritance pattern |
|---|---|---|---|---|
| 193 | *FKRP* | 19q13.32 | Congenital muscular dystrophy 1C, limb-girdle muscular dystrophy type 2I, muscle-eye-brain disease, Walker-Warburg syndrome | AR |
| 194 | *PNKP* | 19q13.33 | Microcephaly, seizures and defects in DNA repair | AR |
| 195 | *ASXL1* | 20q11.21 | Bohring-Opitz syndrome | AD |
| 196 | *DNMT3B* | 20q11.2 | Immunodeficiency-centromeric instability-facial anomalies syndrome 1 | AR |
| 197 | *CTSA* | 20q13.12 | Galactosialidosis | AR |
| 198 | *ARFGEF2* | 20q13.13 | Autosomal recessive periventricular heterotopia with microcephaly | AR |
| 199 | *DPM1* | 20q13.13 | Congenital disorder of glycosylation, type Ie | AR |
| 200 | *CBS* | 21q22.3 | Homocystinuria | AR |
| 201 | *PCNT* | 21q22.3 | Seckel syndrome, Majewski osteodysplastic primordial dwarfism type II | AR |
| 202 | *SNAP29* | 22q11.21 | Cerebral dysgenesis, neuropathy, ichthyosis, and palmoplantar keratoderma syndrome | AR |
| 203 | *SMARCB1* | 22q11.23 | Coffin-Siris syndrome | AD |
| 204 | *LARGE* | 22q12.3 | Congenital muscular dystrophy | AR |
| 205 | *EP300* | 22q13.2 | Rubinstein-Taybi syndrome | AD |
| 206 | *CYB5R3* | 22q13.2 | Methemoglobinemia type II | AR |
| 207 | *ALG12* | 22q13.33 | Congenital disorder of glycosylation, type Ig | AR |
| 208 | *CHKB* | 22q13.33 | Congenital muscular dystrophy, mitochondrial structural abnormalities and ID | AR |
| 209 | *HCCS* | Xp22.2 | Microphthalmia with linear skin defects syndrome | XLD |
| 210 | *OFD1* | Xp22.2 | Oral-facial-digital syndrome type I (XL dominant), Simpson-Golabi-Behmel syndrome, type 2 (XL recessive), Joubert syndrome 10 | XLD/XLR |
| 211 | *FANCB* | Xp22.2 | VACTERL with hydrocephalus, Fanconi anemia of complementation group B | XLR |
| 212 | *PDHA1* | Xp22.12 | Pyruvate decarboxylase deficiency | XLD |
| 213 | *RPS6KA3* | Xp22.12 | Coffin-Lowry syndrome, non-syndromic ID | XLD |
| 214 | *MBTPS2* | Xp22.12 | Ichthyosis follicularis, atrichia and photophobia syndrome | XLR |
| 215 | *SMS* | Xp22.11 | X-linked ID, Snyder-Robinson type | XLR |
| 216 | *GK* | Xp21.2 | Glycerol kinase deficiency | XLD |
| 217 | *TSPAN7* | Xp11.4 | Non-syndromic X-linked ID | XLR |
| 218 | *BCOR* | Xp11.4 | Syndromic Lenz microphthalmia-2, oculofaciocardiodental syndrome | XLD |
| 219 | *ATP6AP2* | Xp11.4 | X-linked ID with epilepsy | XLR |
| 220 | *MAOA* | Xp11.3 | Brunner syndrome (Monoamine oxidase A deficiency) | XLR |
| 221 | *PORCN* | Xp11.23 | Focal dermal hypoplasia | XLD |
| 222 | *SYP* | Xp11.23 | Non-syndromic X-linked ID | XLD |
| 223 | *SHROOM4* | Xp11.22 | Stocco dos Santos X-linked ID syndrome, non-syndromic X-linked ID | XLR |
| 224 | *HUWE1* | Xp11.22 | Non-syndromic X-linked ID (duplications, XL recessive); syndromic X-linked ID, Turner type (point mutations, XL dominant) | XLD/XLR |
| 225 | *ARHGEF9* | Xq11.1 | Syndromic X-linked ID, hyperekplexia and epilepsy | XLR |
| 226 | *DLG3* | Xq13.1 | Non-syndromic X-linked ID | XLR |
| 227 | *HDAC8* | Xq13.1 | Cornelia de Lange syndrome; Wilson-Turner syndrome | XLD |
| 228 | *SLC16A2* | Xq13.2 | T3 transporter deficiency; syndromic and non-syndromic ID | XLD |
| 229 | *ATP7A* | Xq21.1 | Menkes disease, occipital horn syndrome | XLR |
| 230 | *PGK1* | Xq21.1 | Phosphoglycerate kinase-1 deficiency | XLR |
| 231 | *BRWD3* | Xq21.1 | Non-syndromic X-linked ID | XLR |
| 232 | *ZNF711* | Xq21.1 | Non-syndromic X-linked ID | XLR |
| 233 | *TIMM8A* | Xq22.1 | Mohr-Tranebjaerg syndrome, Jensen syndrome | XLR |
| 234 | *RAB40AL* | Xq22.1 | Syndromic X-linked ID, Martin-Probst type | XLR |
| 235 | *PLP1* | Xq22.2 | Pelizaeus-Merzbacher disease | XLR |
| 236 | *PRPS1* | Xq22.3 | Phosphoribosylpyrophosphate synthetase superactivity, Arts syndrome | XLR |
| 237 | *PAK3* | Xq22.3 | Non-syndromic X-linked ID | XLR |
| 238 | *UBE2A* | Xq24 | Syndromic X-linked ID, Nascimento-type | XLR |
| 239 | *NDUFA1* | Xq24 | Mitochondrial complex I deficiency (syndromic X-linked ID) | XLD |
| 240 | *CUL4B* | Xq24 | Syndromic X-linked ID, Cabezas type | XLR |
| 241 | *ZDHHC9* | Xq25 | Syndromic X-linked ID, Raymond type; non-syndromic X-linked ID | XLD |
| 242 | *GPC3* | Xq26.2 | Simpson-Golabi-Behmel syndrome type 1 | XLR |
| 243 | *HPRT1* | Xq26.2 | Lesch-Nyhan syndrome | XLR |
| 244 | *SOX3* | Xq27.1 | Isolated GH deficiency, short stature and ID | XLR |
| 245 | *IDS* | Xq28 | Mucopolysaccharidosis II (Hunter syndrome) | XLR |
| 246 | *NSDHL* | Xq28 | CK syndrome (XL recessive); CHILD syndrome (Congenital Hemidysplasia with Ichthyosiform Nevus and Limb Defects) (XL dominant) | XLR/XLD |
| 247 | *ABCD1* | Xq28 | Adrenoleukodystrophy | XLD |
| 248 | *HCFC1* | Xq28 | Non-syndromic X-linked ID | XLR |
| 249 | *FLNA* | Xq28 | Bilateral periventricular nodular heterotopia, otopalatodigital syndrome, frontometaphyseal dysplasia | XLD |
| 250 | *GDI1* | Xq28 | Non-syndromic X-linked ID | XLD |
| 251 | *IKBKG* | Xq28 | Incontinentia pigmenti | XLD |
| 252 | *DKC1* | Xq28 | Dyskeratosis congenita | XLR |

Note that all the ASD genes listed in **Table S6A** are also involved in ID.
Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ID, intellectual disability; XL, X-linked; XLD, X-linked dominant; XLR, X-linked recessive

## Table S6D. Loci implicated in ID

| | Disorder | Chr. band | Start (hg18) | End (hg18) | Genes involved | Inheritance pattern |
|---|---|---|---|---|---|---|
| 1 | Thrombocytopenia-absent radius (TAR) syndrome | 1q21.1 | 144,110,432 | 144,305,571 | *RBM8A* | AR |
| 2 | 3p deletion syndrome | 3p26.3p25.3 | 1 | 9,000,000 | contiguous gene syndrome? | AD |
| 3 | Proximal 15q25.2 microdeletion | 15q25 | 80,451,495 | 82,719,635 | ? | AD |
| 4 | ATR-16 syndrome (alpha-thalassemia/mental retardation syndrome) | 16p13.3 | 30,843 | 774,373 | ? | AD |
| 5 | 17p13.1 microdeletion syndrome | 17p13.1 | 7,429,371 | 7,937,620 | ? | AD |
| 6 | 19p13.13 microdeletion/microduplication syndrome | 19p13.13 | 12,793,474 | 13,104,643 | ? | AD |
| 7 | Cat-Eye syndrome | 22p13-22q11.21 | 15,772,953 | 16,971,860 | ? | AD |
| 8 | 22q11.2 distal deletion syndrome | 22q11.2 | 20,445,848 | 22,026,229 | ? | AD |
| 9 | Pelizaeus-Merzbacher disease | Xq22.2 | 102,918,095 | 102,934,203 | *PLP1* | XLR |

Note that all the ASD loci listed in **Table S6B** are also involved in ID.
Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ID, intellectual disability; XLR, X-linked recessive

## Table S7A. CNVs overlapping ASD or ID genes and loci in affected and control subjects (all ancestries)

### Cases

| AGP ID | Sex | Chr#:start-end (hg18) | Size | CNV type | Inheri-tance | Classification | Validation | CNV description |
|---|---|---|---|---|---|---|---|---|
| 16035_1571013001 | M | 1:144482933-146325557 | 1,842,625 | Gain | *dn* | Pathogenic | qPCR | 1q21.1 duplication syndrome |
| 8635_201 | M | 1:144500467-146377870 | 1,877,404 | Loss | *dn* | Pathogenic | qPCR | 1q21.1 deletion syndrome |
| 1952_301 | M | 1:144500467-146336720 | 1,836,254 | Gain | *dn* | Pathogenic | qPCR | 1q21.1 duplication syndrome |
| 13135_1523 | F | 1:144838594-146308287 | 1,469,694 | Gain | *dn* | Pathogenic | qPCR | 1q21.1 duplication syndrome |
| 4291_1 | M | 1:144967972-146317915 | 1,349,944 | Gain | mat | Pathogenic | qPCR | 1q21.1 duplication syndrome |
| 16074_1571042001 | M | 2:50037898-50407550 | 369,653 | Loss | mat | Pathogenic | qPCR | *NRXN1* exonic deletion |
| 14068_1180 | M | 2:50493827-50677835 | 184,009 | Gain | *dn* | Pathogenic | qPCR | *NRXN1* intragenic duplication, predicted to result in a premature truncated protein |
| 13017_223 | F | 2:50539877-50730546 | 190,670 | Loss | *dn* | Pathogenic | qPCR | *NRXN1* exonic deletion |
| 13216_2383 | M | 2:50968208-51214171 | 245,964 | Loss | *dn* | Pathogenic | qPCR | *NRXN1* exonic deletion |
| 13153_1703 | M | 2:50990306-51222043 | 231,738 | Loss | *dn* | Pathogenic | qPCR | *NRXN1* exonic deletion |
| 13037_463 | M | 2:51002576-51157742 | 155,167 | Loss | *dn* | Pathogenic | qPCR | *NRXN1* exonic deletion |
| 17027_1 | M | 2:51076611-51147600 | 70,990 | Loss | pat | Pathogenic | qPCR | *NRXN1* exonic deletion |
| 5353_3 | F | 6:33399849-33512042 | 112,194 | Loss | *dn* | Pathogenic | qPCR | *SYNGAP1* exonic deletion |
| 5386_3 | M | 6:156785155-158489874 | 1,704,720 | Loss | *dn* | Pathogenic | qPCR, Illumina 1M | *ARID1B* exonic deletion |
| 8446_201 | M | 7:72344426-73782113 | 1,437,688 | Loss | *dn* | Pathogenic | qPCR | Williams syndrome (7q11.23 deletion) |
| 13123_1403 | F | 9:98998-3682923 | 3,583,926 | Loss | *dn* | Pathogenic | qPCR | Terminal 9p deletion, 3.58 Mb (14 genes) |
| 6259_3 | M | 9:139516033-140208462 | 692,430 | Loss | *dn* | Pathogenic | qPCR | Kleefstra syndrome (9q34.3 deletion including *EHMT1*) |
| 6325_3 | M | 11:70077507-70506315 | 428,809 | Loss | *dn* | Pathogenic | qPCR | *SHANK2* exonic deletion |
| 6319_3 | M | 11:70119917-70187872 | 67,956 | Loss | *dn* | Pathogenic | qPCR | *SHANK2* exonic deletion |
| 5237_3 | M | 11:70154458-70220632 | 66,175 | Loss | *dn* | Pathogenic | qPCR, Agilent 1M | *SHANK2* exonic deletion |
| 6240_4 | M | 11:126633939-132060374 | 5,426,436 | Loss | *dn* | Pathogenic | qPCR | Chromosome 11q deletion syndrome (Jacobsen syndrome) |
| 20187_1464001 | M | 15:18811937-26209270 | 7,397,334 | Gain | *dn* | Pathogenic | qPCR | 15q11-q13 duplication syndrome, maternally derived |
| 8630_201 | M | 15:19800798-26209270 | 6,408,473 | Gain | *dn* | Pathogenic | qPCR | 15q11-q13 duplication syndrome, paternally derived |
| 20069_1328001 | M | 15:20203578-26209270 | 6,005,693 | Gain | *dn* | Pathogenic | qPCR | 15q11-q13 duplication syndrome, maternally derived |
| 17035_1 | F | 15:20274130-26120360 | 5,846,231 | Gain | *dn* | Pathogenic | qPCR | 15q11-q13 duplication syndrome, maternally derived |
| 8117_202 | M | 15:21168391-26217954 | 5,049,564 | Gain | mat | Pathogenic | qPCR | 15q11-q13 duplication syndrome, maternally derived |
| 8741_201 | F | 15:21168391-26315093 | 5,146,703 | Gain | mat | Pathogenic | qPCR | 15q11-q13 duplication syndrome, maternally derived |
| 13050_593 | M | 15:21190624-26203954 | 5,013,331 | Gain | *dn* | Pathogenic | qPCR | 15q11-q13 duplication syndrome, maternally derived |
| 1950_301 | M | 15:26762141-30436163 | 3,674,023 | Loss | mat | Pathogenic | Illumina 550K | 15q13.3 deletion syndrome |
| 16040_1571029001 | M | 15:28450423-30303265 | 1,852,843 | Loss | *dn* | Pathogenic | qPCR | 15q13.3 deletion syndrome |
| 14167_2720 | M | 15:28705540-30436163 | 1,730,624 | Loss | pat | Pathogenic | qPCR | 15q13.3 deletion syndrome |
| 18100_302 | M | 15:28714502-30303265 | 1,588,764 | Loss | pat | Pathogenic | qPCR | 15q13.3 deletion syndrome |
| 5537_3 | M | 15:82722026-83529838 | 807,813 | Loss | mat | Pathogenic | Affy 500K, Illumina 2.5M, qPCR | Distal 15q25.2 deletion syndrome |
| 14283_4060 | M | 16:14771033-16307313 | 1,536,281 | Loss | mat | Pathogenic | qPCR | 16p13.11 deletion syndrome |
| 3441_3 | M | 16:14808871-16215852 | 1,406,982 | Loss | pat | Pathogenic | qPCR | 16p13.11 deletion syndrome |
| 14412_5210 | M | 16:14960247-16307313 | 1,347,067 | Loss | mat | Pathogenic | qPCR, LR-PCR | 16p13.11 deletion syndrome |
| 2204_1 | M | 16:29466569-30147029 | 680,461 | Loss | *dn* | Pathogenic | qPCR | 16p11.2 deletion syndrome |
| 3544_3 | M | 16:29499858-30107306 | 607,449 | Gain | mat | Pathogenic | — | 16p11.2 duplication syndrome |
| 5068_3 | F | 16:29502984-30127026 | 624,043 | Loss | *dn* | Pathogenic | qPCR, Affy 500K, Agilent 1M | 16p11.2 deletion syndrome, 50% mosaicism |
| 20089_1391001 | M | 16:29502984-30107306 | 604,323 | Loss | *dn* | Pathogenic | qPCR | 16p11.2 deletion syndrome |
| 3211_3 | M | 16:29502984-30127026 | 624,043 | Gain | mat | Pathogenic | qPCR | 16p11.2 duplication syndrome |
| 5262_4 | M | 16:29502984-30210849 | 707,866 | Gain | *dn* | Pathogenic | qPCR, Affy 500K | 16p11.2 duplication syndrome |
| 5359_4 | M | 16:29554843-30195224 | 640,382 | Loss | *dn* | Pathogenic | qPCR, Affy 500K, Agilent 1M | 16p11.2 deletion syndrome |
| 20127_4014001 | M | 16:29554843-30130862 | 576,020 | Loss | pat | Pathogenic | qPCR | 16p11.2 deletion syndrome |
| 4030_1 | M | 16:29554843-30107306 | 552,464 | Gain | *dn* | Pathogenic | qPCR | 16p11.2 duplication syndrome |
| 3439_3 | M | 17:17156307-18262979 | 1,106,673 | Loss | *dn* | Pathogenic | qPCR | Smith-Magenis syndrome (17p11.2 deletion including *RAI1*) |
| 2211_1 | F | 17:17169258-20101517 | 2,932,260 | Loss | *dn* | Pathogenic | qPCR | Smith-Magenis syndrome (17p11.2 deletion including *RAI1*) |
| 14315_4320 | M | 17:31621634-33323919 | 1,702,286 | Gain | mat | Pathogenic | qPCR | 17q12 duplication syndrome |
| 3183_7 | M | 22:17241748-19819918 | 2,578,171 | Loss | *dn* | Pathogenic | qPCR | 22q11.2 deletion syndrome |
| 17015_1 | M | 22:17257787-19795780 | 2,537,994 | Loss | *dn* | Pathogenic | qPCR | 22q11.2 deletion syndrome |
| 3127_4 | M | 22:17257787-19793730 | 2,535,944 | Gain | pat | Pathogenic | qPCR | 22q11.2 duplication syndrome |
| 4271_1 | M | 22:17257787-19795780 | 2,537,994 | Gain | *dn* | Pathogenic | qPCR | 22q11.2 duplication syndrome |
| 5261_4 | F | 22:17257787-19795780 | 2,537,994 | Gain | pat | Pathogenic | qPCR, Illumina 1M | 22q11.2 duplication syndrome |
| 16074_1571042001 | M | 22:17257787-19793730 | 2,535,944 | Gain | pat | Pathogenic | qPCR | 22q11.2 duplication syndrome |
| 8627_201 | M | 22:17257787-19793730 | 2,535,944 | Gain | *dn* | Pathogenic | qPCR | 22q11.2 duplication syndrome |
| 2072_1 | M | 22:45159185-49582267 | 4,423,083 | Loss | *dn* | Pathogenic | qPCR | Phelan-McDermid syndrome (22q13 deletion including *SHANK3*) |
| 6130_4 | F | 22:47996161-49512530 | 1,516,370 | Loss | *dn* | Pathogenic | qPCR, MLPA | Phelan-McDermid syndrome (22q13 deletion including *SHANK3*) |
| 16079_1571066001 | M | 22:49470371-49567383 | 97,013 | Loss | *dn* | Pathogenic | qPCR | Phelan-McDermid syndrome (22q13 deletion including *SHANK3*) |
| 5240_4 | M | X:23116188-23280628 | 164,441 | Loss | mat | Pathogenic | qPCR, Illumina 1M | *PTCHD1* exonic deletion |
| 5126_4 | M | X:28931559-29478966 | 547,408 | Gain | mat | Pathogenic | qPCR, Agilent 1M | *IL1RAPL1* intragenic duplication of exons 3-5 |
| 8597_201 | M | X:31303978-32025062 | 721,085 | Loss | mat | Pathogenic | qPCR | *DMD* deletion of exons 45-60 |
| 3019_3 | M | X:32100618-32315937 | 215,320 | Gain | mat | Pathogenic | qPCR | *DMD* duplication of exons 31-44 |
| 20013_1075001 | M | X:153239048-153521797 | 282,750 | Gain | *dn* | Pathogenic | qPCR, LR-PCR | Xq28 duplication encompassing 20 genes, including 3 involved in ID: *FLNA, GDI1, IKBKG*. Corresponds to the recurrent Xq28 duplication reported in XLID families;[55] *GDI1* is the most likely candidate gene (sequencing coordinates: chrX:153222048-153514311, size 292 kb) |
| 14216_3470 | M | X:153263157-153474401 | 211,245 | Gain | mat | Pathogenic | qPCR | Xq28 duplication encompassing 18 genes, including 2 involved in ID: *GDI1* and *IKBKG.* Recurrent Xq28 duplication involved in XLID (see above) |
| 6356_5 | M | 1:2118508-2325536 | 207,029 | Loss | *dn* | Uncertain | qPCR | Deletion encompassing 5 genes, including *SKI,* involved in Shprintzen-Goldberg syndrome through dominant negative mutations |
| 6317_5 | M | 1:26767083-26941756 | 174,674 | Gain | pat | Uncertain | qPCR | *ARID1A* partial duplication; *ARID1A* mutations reported recently in Coffin-Siris syndrome |
| 13027_353 | F | 2:144837809-145315383 | 477,575 | Gain | *dn* | Uncertain | qPCR | *ZEB2* duplication; mutations and deletions in *ZEB2* cause Mowat-Wilson syndrome |

| AGP ID | Sex | Chr#:start-end (hg18) | Size | CNV type | Inheri-tance | Classification | Validation | CNV description |
|---|---|---|---|---|---|---|---|---|
| 3424_3 | M | 2:148881443-149078468 | 197,026 | Gain | mat | Uncertain | qPCR | *MBD5* partial duplication; *MBD5* is implicated in autosomal dominant ID through deletions (2q23.1 deletion syndrome) and mutations |
| 3599_3 | F | 6:155906594-157336808 | 1,430,215 | Gain | pat | Uncertain-likely benign | qPCR | *ARID1B* partial duplication; haploinsufficiency of *ARID1B* causes ID and Coffin-Siris syndrome |
| 13037_463 | M | 9:137682721-137840339 | 157,619 | Loss | mat | Uncertain | — | Deletion encompassing 4 genes, including *KCNT1*; gain-of-function *KCNT1* mutations cause epilepsy |
| 16072_1571036001 | M | 11:381049-1019320 | 638,272 | Gain | pat | Uncertain | qPCR | Duplication encompassing 35 genes, including *HRAS* (involved in Costello syndrome through activating mutations); although duplication of *HRAS* is not expected to be pathogenic, the contribution of other genes in the interval is unknown |
| 20070_1331001 | M | 16:14808871-15935225 | 1,126,355 | Gain | mat | Uncertain | qPCR | 16p13.11 microduplication, significance unknown. Recurrent 16p13.11 deletions are associated with variable phenotype and incomplete penetrance, and have been reported in subjects with diverse neuropsychiatric disorders, including ID, ASD, schizophrenia and epilepsy, sometimes inherited from unaffected parents. Duplications have been described in neurodevelopmental disorders and in controls, with studies reporting either no enrichment or a small but significant enrichment in cases. Further studies in larger case populations and controls are needed to clarify their role as risk factors |
| 9766_202 | M | 16:15032942-16199484 | 1,166,543 | Gain | mat | Uncertain | qPCR | 16p13.11 microduplication (see above) |
| 14142_2400 | M | 16:15387380-16256106 | 868,727 | Gain | pat | Uncertain | qPCR | 16p13.11 microduplication (see above) |
| 5258_3 | M | 16:15387380-16270740 | 883,361 | Gain | pat | Uncertain | qPCR, Illumina 1M | 16p13.11 microduplication (see above) |
| 4182_1 | M | 16:15387380-16199484 | 812,105 | Gain | mat | Uncertain | — | 16p13.11 microduplication (see above) |
| 2265_1 | M | 16:15387380-18176669 | 2,789,290 | Gain | mat | Uncertain | qPCR | 16p13.11 microduplication (see above) |
| 8703_201 | M | 17:1092256-1249222 | 156,967 | Gain | pat | Uncertain-likely benign | qPCR | Duplication of 3 genes, partially overlapping *YWHAE*; whole duplications are pathogenic (17p13.3 duplication syndrome); there is an AGP control with a *YWHAE* partial duplication; several partial duplications reported in DGV |
| 5444_3 | M | 17:76953064-77782267 | 829,204 | Gain | *dn* | Uncertain | qPCR | 829 kb duplication encompassing 38 genes, including *ACTG1* (involved in Baraitser-Winter syndrome through dominant-negative or gain-of-function missense mutations); although duplication of *ACTG1* is not expected to be pathogenic, the contribution of the other genes in the interval is unknown |
| 14242_3660 | F | 22:21995356-22598120 | 602,765 | Gain | mat | Uncertain | — | 602 kb duplication encompassing 15 genes, including *SMARCB1*; *SMARCB1* mutations reported recently in Coffin-Siris syndrome |
| 6381_3 | M | X:30521738-30789831 | 268,094 | Gain | mat | Uncertain | qPCR | Duplication of 2 genes, partially overlapping *GK*; *GK* mutations and deletions cause XLID |
| 5126_4 | M | X:32948977-33330592 | 381,616 | Gain | mat | Uncertain | Agilent 1M | *DMD* duplication of exon 1 of the Dp427 transcripts; the effect of this duplication is difficult to predict; experimental evidence at the RNA level is required to interpret the significance |
| 4356_1 | M | X:38001148-38346471 | 345,324 | Gain | mat | Uncertain-likely benign | — | Duplication of *RPGR, OTC,* and *TSPAN7*; similar duplication found in an AGP male control. *OTC* mutations and deletions cause ornithine transcarbamylase deficiency; *TSPAN7* duplications are found in healthy controls |
| 4152_1 | M | X:40260354-40372806 | 112,453 | Gain | mat | Uncertain | — | Duplication encompassing 3 genes, including *ATP6AP2*; *ATP6AP2* silent mutation affecting splicing described in XLID with epilepsy |
| 3240_3 | M | X:44706702-44919064 | 212,363 | Gain | mat | Uncertain | — | Duplication of 2 genes, partially overlapping *KDM6A* partial duplication; *KDM6A* mutations and deletions cause X-linked Kabuki syndrome |
| 14314_4310 | M | X:70865247-71509736 | 644,490 | Gain | mat | Uncertain | qPCR | 644 kb duplication encompassing 13 genes, partially overlapping*HDAC8*; *HDAC8* mutations described recently in Cornelia de Lange syndrome and XLID resembling Wilson-Turner syndrome; similar partial duplication of *HDAC8* observed in a male control |
| 1348_301 | M | X:147163528-147758700 | 595,173 | Gain | mat | Uncertain | qPCR | *AFF2* partial duplication; involved in XLID through trinucleotide expansion or deletion |
| 5036_4 | M | X:148075334-148617551 | 542,218 | Gain | mat | Uncertain-likely benign | — | 542 kb duplication encompassing 9 genes, including *IDS*; mutations and deletions cause mucopolysaccharidosis type II; duplication reported in unaffected males. Although duplication of *IDS* is not expected to be pathogenic, the contribution of the other genes in the interval is unknown. The duplication is also present in the proband's unaffected brother, suggesting this CNV is unlikely to be pathogenic |
| 4354_1 | M | X:148344051-148707925 | 363,875 | Gain | mat | Uncertain-likely benign | — | 363 kb duplication encompassing 9 genes, including *IDS*; mutations and deletions cause mucopolysaccharidosis type II; duplication reported in unaffected males (see above) |
| 4166_1 | M | X:151620401-151805387 | 184,987 | Gain | mat | Uncertain | — | Xq28 duplication of 10 genes, including *NSDHL*; mutations and deletions cause syndromic XLID (CK syndrome and CHILD syndrome); similar duplication including *NSDHL* in a male AGP control |
| 9901_201 | M | 3:12608293-12786824 | 178,532 | Gain | pat | Benign | — | Duplication of 2 genes, partially overlapping *RAF1*; *RAF1* is involved in Noonan syndrome through activating mutations; similar duplication present in 2 AGP controls |
| 20087_1386001 | M | 3:12610706-12786824 | 176,119 | Gain | mat | Benign | — | Duplication of 2 genes, partially overlapping *RAF1*; *RAF1* is involved in Noonan syndrome through activating mutations; similar duplication present in 2 AGP controls |
| 4457_1 | M | X:38375788-38515190 | 139,403 | Gain | mat | Benign | — | *TSPAN7* partial duplication; mutations and deletions cause XLID, duplications are found in healthy controls |
| 5524_3 | M | X:38375788-38515190 | 139,403 | Gain | mat | Benign | — | *TSPAN7* partial duplication (see above) |
| 6034_3 | M | X:38375788-38515190 | 139,403 | Gain | mat | Benign | — | *TSPAN7* partial duplication (see above) |
| 20033_1227001 | M | X:38375788-38515190 | 139,403 | Gain | mat | Benign | — | *TSPAN7* partial duplication (see above) |
| 20141_1396001 | M | X:38375788-38515190 | 139,403 | Gain | mat | Benign | qPCR | *TSPAN7* partial duplication (see above) |
| 5089_5 | M | X:53555568-53640902 | 85,335 | Gain | mat | Benign | — | *HUWE1* partial duplication. Whole gene duplications cause non-syndromic XLID; similar recurrent partial duplications of *HUWE1* reported recently, considered polymorphic[133] |
| 17018_1 | M | X:53568262-53640902 | 72,641 | Gain | mat | Benign | qPCR | *HUWE1* partial duplication (see above) |

## Controls

| ID | Sex | Chr#:start-end (hg18) | Size | CNV type | Classification | CNV description |
|---|---|---|---|---|---|---|
| HABC_902399_902399 | F | 1:144614719-146470277 | 1,855,559 | Loss | "Pathogenic" | 1q21.1 deletion syndrome |
| B436528_1007852654 | M | 1:144627859-146546371 | 1,918,513 | Loss | "Pathogenic" | 1q21.1 deletion syndrome |
| B984152_1007842480 | F | 1:144800611-145863421 | 1,062,811 | Gain | "Pathogenic" | 1q21.1 duplication syndrome |
| B618929_1007875266 | M | 1:144933825-145518117 | 584,293 | Gain | "Pathogenic" | 1q21.1 duplication syndrome |
| B666224_1007871687 | M | 1:144967972-145863421 | 895,450 | Gain | "Pathogenic" | 1q21.1 duplication syndrome |
| HABC_902895_902895 | F | 2:50829989-51064129 | 234,141 | Loss | "Pathogenic" | *NRXN1* exonic deletion |
| B964957_1007872180 | F | 7:72344426-73782113 | 1,437,688 | Gain | "Pathogenic" | 7q11.23 duplication syndrome (Williams syndrome region) |
| B914224_1007874975 | M | 10:89620404-89723400 | 102,997 | Loss | "Pathogenic" | *PTEN* exonic deletion |
| HABC_902475_902475 | M | 15:36401817-36434987 | 33,171 | Loss | "Pathogenic" | *SPRED1* deletion; gene involved in Legius syndrome (phenotypic overlap with neurofibromatosis 1); most individuals don't have ID and only present dermatologic findings |
| B675955_1007841005 | F | 16:29415871-30239704 | 823,834 | Gain | "Pathogenic" | 16p11.2 duplication syndrome |
| B416484_1007875540 | M | 16:29502984-30127026 | 624,043 | Loss | "Pathogenic" | 16p11.2 deletion syndrome |
| B879700_1007854073 | F | 16:29502984-30127026 | 624,043 | Gain | "Pathogenic" | 16p11.2 duplication syndrome |
| HABC_900681_900681 | M | 16:29554843-30180288 | 625,446 | Gain | "Pathogenic" | 16p11.2 duplication syndrome |

| ID | Sex | Chr#:start-end (hg18) | Size | CNV type | Classification | CNV description |
|---|---|---|---|---|---|---|
| B121881_1007874637 | M | 17:31621634-33323919 | 1,702,286 | Gain | "Pathogenic" | 17q12 duplication syndrome |
| HABC_901636_901636 | M | 22:17248170-19795780 | 2,547,611 | Gain | "Pathogenic" | 22q11.2 duplication syndrome |
| B928258_1007854097 | F | 22:17257787-18693299 | 1,435,513 | Gain | "Pathogenic" | 22q11.2 duplication syndrome |
| 110036016178_ | M | 2:148753482-148819104 | 65,623 | Loss | Uncertain | *MBD5* exonic deletion; deletions of *MBD5* cause autosomal dominant ID, but this one overlaps only the long isoform, which has not been fully characterized and contains 5 additional non-coding exons (this deletion overlaps 1 of the non-coding exons) |
| B183736_1007853714 | F | 8:117910382-117948637 | 38,256 | Gain | Uncertain-likely benign | *RAD21* partial duplication; mutations and deletions cause Cornelia de Lange syndrome |
| HABC_900405_900405 | M | 8:117910403-117948935 | 38,533 | Gain | Uncertain-likely benign | *RAD21* partial duplication; mutations and deletions cause Cornelia de Lange syndrome |
| HABC_900854_900854 | M | 9:2018757-2080718 | 61,962 | Loss | Uncertain-likely benign | *SMARCA2* deletion of exons 2-19, removes the translation start site in exon 2; mutations resulting in Nicolaides-Baraitser syndrome are thought to act through a dominant–negative or gain-of-function manner and cluster in exons 15–25; deletions encompassing *SMARCA2* do not cause this syndrome, except for one reported in-frame deletion overlapping the mutation-clustering region. The deletion in the AGP control involves exons 2-19 and removes the translation start site in exon 2. |
| B116679_1007853952 | M | 9:129475725-129875601 | 399,877 | Gain | Uncertain | Duplication encompassing 19 genes, partially overlapping *STXBP1*; mutations and deletions of *STXBP1* cause nonsyndromic ID with epilepsy and infantile epileptic encephalopathy |
| B246752_1007872634 | F | 9:139616009-139725155 | 109,147 | Gain | Uncertain | Duplication encompassing 3 genes, partially overlapping *EHMT1*; *EHMT1* mutations and deletions cause Kleefstra syndrome |
| B252606_1007874475 | M | 12:1889583-2538831 | 649,249 | Gain | Uncertain | Duplication encompassing 4 genes, partially overlapping *CACNA1C*, involved in Timothy syndrome through activating mutations |
| B936611_1007853579 | F | 12:25087733-25286046 | 198,314 | Gain | Uncertain | Duplication encompassing 4 genes, partially overlapping *KRAS*; *KRAS* is involved in cardio-facio-cutaneous syndrome through activating mutations |
| B978305_1007874920 | M | 13:109744730-110078003 | 333,274 | Gain | Uncertain | Duplication encompassing 4 genes, partially overlapping *COL4A1*; only *COL4A1* missense mutations reported thus far, no deletions or duplications |
| B777599_1007853701 | F | 14:101564897-101608061 | 43165 | Gain | Uncertain | *DYNC1H1* partial duplication; all mutations identified so far (ID with neuronal migration defects and motor neuropathies) are heterozygous missense mutations, suggesting a dominant-negative effect |
| HABC_900744_900744 | F | 15:28723577-30232287 | 1,508,711 | Gain | Uncertain | 15q13.3 microduplication |
| HABC_901557_901557 | F | 15:28730804-30389965 | 1,659,162 | Gain | Uncertain | 15q13.3 microduplication |
| B833125_0057060983 | F | 16:1995854-2052977 | 57,124 | Gain | Uncertain | Duplication encompassing 5 genes, partially overlapping *TSC2*; *TSC2* mutations and deletions cause tuberous sclerosis |
| HABC_902940_902940 | M | 16:14831165-16199484 | 1,368,320 | Gain | Uncertain | 16p13.11 microduplication |
| HABC_901197_901197 | F | 16:14882793-16199484 | 1,316,692 | Gain | Uncertain | 16p13.11 microduplication |
| HABC_902864_902864 | M | 16:14977368-16190572 | 1,213,205 | Gain | Uncertain | 16p13.11 microduplication |
| HABC_902897_902897 | M | 16:15387380-18075924 | 2,688,545 | Gain | Uncertain | 16p13.11 microduplication |
| HABC_901863_901863 | F | 16:87658525-87896679 | 238,155 | Gain | Uncertain | *CDH15* duplication and *ANKRD11* partial duplication; *CDH15* is involved in non syndromic ID through mutations and *ANKRD11* in KBG syndrome through mutations and deletions |
| HABC_901829_901829 | F | 16:87729306-87962759 | 233,454 | Gain | Uncertain | *CDH15* duplication and *ANKRD11* partial duplication (see above) |
| HABC_902145_902145 | F | 17:1136077-1211814 | 75,738 | Gain | Uncertain-likely benign | Duplication of 2 genes, partially overlapping *YWHAE*; *YWHAE* whole gene duplications are pathogenic; several partial duplications in DGV |
| HABC_902046_902046 | M | 17:2437871-2476338 | 38,468 | Gain | Uncertain | *PAFAH1B1* partial duplication; whole gene duplications are pathogenic |
| HABC_901640_901640 | M | 22:21320275-23360745 | 2,040,471 | Gain | Uncertain | 2 Mb duplication encompassing 44 genes, including *SMARCB1*; *SMARCB1* mutations reported in Coffin-Siris syndrome |
| B278753_1007874641 | M | 22:21995356-22676385 | 681,030 | Gain | Uncertain | 681 kb duplication encompassing 20 genes, including *SMARCB1*; *SMARCB1* mutations reported in Coffin-Siris syndrome |
| HABC_900402_900402 | F | X:11038333-11069582 | 31,250 | Gain | Uncertain-likely benign | Duplication of 2 genes, including *HCCS*; *HCCS* mutations and deletions cause a syndromic form of XLID |
| HABC_902971_902971 | F | X:13633067-13700254 | 67,188 | Gain | Uncertain-likely benign | Duplication encompassing 4 genes, including *OFD1*; mutations and deletions cause syndromic XLID |
| HABC_901634_901634 | M | X:17239813-17435795 | 195,983 | Gain | Uncertain-likely benign | *NHS* partial duplication; mutations and deletions cause Nance-Horan syndrome |
| HABC_902647_902647 | M | X:18574793-18780863 | 206,071 | Gain | Uncertain-likely benign | Xp22.13 duplication encompassing 3 genes, partially overlapping *CDKL5*; *CDKL5* mutations and deletions cause epileptic encephalopathy |
| HABC_900416_900416 | M | X:38013482-38643203 | 629,722 | Gain | Uncertain-likely benign | Duplication encompassing 4 genes, including *OTC* and *TSPAN7*. *OTC* mutations and deletions cause ornithine transcarbamylase deficiency; no cases have been reported with a duplication. *TSPAN7* mutations and deletions cause XLID, duplications are found in healthy controls |
| B345605_1007844543 | M | X:71239825-71490721 | 250,897 | Gain | Uncertain-likely benign | Duplication encompassing 8 genes, partially overlapping *HDAC8*; *HDAC8* mutations described recently in Cornelia de Lange syndrome and XLID |
| HABC_900366_900366 | F | X:76924341-77030430 | 106,090 | Gain | Uncertain-likely benign | Partial duplication of *ATRX* and *MAGT1*; *ATRX* is involved in XLID through mutations or deletions |
| HABC_900333_900333 | F | X:134903813-135264655 | 360,843 | Gain | Uncertain | Xq26.3 duplication encompassing 4 genes, partially overlapping *SLC9A6*; mutations and deletions cause syndromic XLID (Christianson syndrome) |
| B418695_1007840289 | M | X:151644548-151839695 | 195,148 | Gain | Uncertain | Xq28 duplication encompassing 10 genes, including *NSDHL*; *NSDHL* mutations and deletions cause XLID (CK syndrome and CHILD syndrome); several overlapping duplications reported in DGV |
| HABC_902725_902725 | F | X:153340432-153435070 | 94,639 | Gain | Uncertain-likely benign | 95 kb duplication encompassing 7 genes, partially overlapping *IKBKG*; only mutations, deletions and intragenic duplications described |
| HABC_902313_902313 | M | X:153440007-153622054 | 182,048 | Gain | Uncertain-likely benign | 182 kb duplication encompassing 7 genes, partially overlapping *IKBKG*; only mutations, deletions and intragenic duplications described |
| HABC_901807_901807 | M | 3:12608293-12781123 | 172,831 | Gain | Benign | Duplication encompassing 2 genes, partially overlapping *RAF1*; *RAF1* is involved in Noonan syndrome through activating mutations |
| B260038_1007841400 | F | 3:12610706-12781123 | 170,418 | Gain | Benign | Duplication encompassing 2 genes, partially overlapping *RAF1*; *RAF1* is involved in Noonan syndrome through activating mutations |
| B630497_1007872229 | F | 12:2085709-2127756 | 42,048 | Loss | Benign | *CACNA1C* deletion; *CACNA1C* causes Timothy syndrome through activating mutations |
| B131548_1007842659 | F | 12:2663790-2714138 | 50,349 | Gain | Benign | *CACNA1C* partial duplication; *CACNA1C* causes Timothy syndrome through activating mutations |
| HABC_902939_902939 | F | 12:2675893-2717980 | 42,088 | Gain | Benign | *CACNA1C* partial duplication; *CACNA1C* causes Timothy syndrome through activating mutations |
| 110036016517_ | M | X:38375788-38515190 | 139,403 | Gain | Benign | *TSPAN7* partial duplication; mutations and deletions cause XLID, duplications are found in healthy controls |
| B818627_1007854359 | M | X:38375788-38515190 | 139,403 | Gain | Benign | *TSPAN7* partial duplication; mutations and deletions cause XLID, duplications are found in healthy controls |

This table shows the CNVs overlapping ASD/ID genes and loci interpreted as pathogenic, uncertain or benign in cases and controls of all ancestries. Phenotype information and CNV segregation in siblings can be found in **Table S8**. In order to compare the burden of CNVs overlapping ASD/ID genes and loci in cases and controls (**Figures 1** and **S1**), CNVs were interpreted irrespective of affected status. A small number of CNVs that would have been considered pathogenic in an affected individual were identified among controls; all were CNVs known to be associated with incomplete penetrance/variable expressivity.

Abbreviations: DGV, Database of Genomic Variants; dn, *de novo*; F, female; ID, intellectual disability; LR-PCR, long range PCR; M, male; mat, maternal; MLPA, multiplex ligation-dependent probe amplification; pat, paternal; qPCR, quantitative PCR; XLID, X-linked intellectual disability. —, no validation attempted, CNV confirmed by visual inspection.

## Table S7B. Pathogenic CNVs in affected subjects (all ancestries)

### Pathogenic CNVs overlapping ASD/ID genes or loci (stringent CNV, >30 kb) n = 64

| AGP ID | Sex | Family type | Cytoband | Chr#:start-end (hg18) | Size | CNV type | Inheritance | CNV description | Classification penetrance/ expressivity[§] |
|---|---|---|---|---|---|---|---|---|---|
| 16035_1571013001 | M | Familial | 1q21.1 | 1:144482933-146325557 | 1,842,625 | Gain | *De novo* | 1q21.1 duplication syndrome | VE/IP |
| 1952_301 | M | Familial | 1q21.1 | 1:144500467-146336720 | 1,836,254 | Gain | *De novo* | 1q21.1 duplication syndrome | VE/IP |
| 8635_201 | M | Familial | 1q21.1 | 1:144500467-146377870 | 1,877,404 | Loss | *De novo* | 1q21.1 deletion syndrome | VE/IP |
| 13135_1523 | F | Unknown | 1q21.1 | 1:144838594-146308287 | 1,469,694 | Gain | *De novo* | 1q21.1 duplication syndrome | VE/IP |
| 4291_1 | M | Familial | 1q21.1 | 1:144967972-146317915 | 1,349,944 | Gain | Maternal | 1q21.1 duplication syndrome | VE/IP |
| 16074_1571042001† | M | Sporadic | 2p16.3 | 2:50037898-50407550 | 369,653 | Loss | Maternal | *NRXN1* exonic deletion | VE/IP |
| 14068_1180 | M | Sporadic | 2p16.3 | 2:50493827-50677835 | 184,009 | Gain | *De novo* | *NRXN1* intragenic duplication, predicted to result in a premature truncated protein | VE/IP |
| 13017_223 | F | Sporadic | 2p16.3 | 2:50539877-50730546 | 190,670 | Loss | *De novo* | *NRXN1* exonic deletion | VE/IP |
| 13216_2383 | M | Unknown | 2p16.3 | 2:50968208-51214171 | 245,964 | Loss | *De novo* | *NRXN1* exonic deletion | VE/IP |
| 13153_1703 | M | Sporadic | 2p16.3 | 2:50990306-51222043 | 231,738 | Loss | *De novo* | *NRXN1* exonic deletion | VE/IP |
| 13037_463 | M | Sporadic | 2p16.3 | 2:51002576-51157742 | 155,167 | Loss | *De novo* | *NRXN1* exonic deletion | VE/IP |
| 17027_1 | M | Familial | 2p16.3 | 2:51076611-51147600 | 70,990 | Loss | Paternal | *NRXN1* exonic deletion | VE/IP |
| 5353_3 | F | Sporadic | 6p21.32 | 6:33399849-33512042 | 112,194 | Loss | *De novo* | *SYNGAP1* exonic deletion | HP |
| 5386_3 | M | Familial | 6q25.3 | 6:156785155-158489874 | 1,704,720 | Loss | *De novo* | *ARID1B* exonic deletion | HP |
| 8446_201 | M | Sporadic | 7q11.23 | 7:72344426-73782113 | 1,437,688 | Loss | *De novo* | Williams syndrome (7q11.23 deletion) | HP |
| 13123_1403 | F | Sporadic | 9p24.3-p24.2 | 9:98998-3682923 | 3,583,926 | Loss | *De novo* | Terminal 9p deletion, 3.58 Mb (14 genes) | HP |
| 6259_3 | M | Sporadic | 9q34.3 | 9:139516033-140208462 | 692,430 | Loss | *De novo* | Kleefstra syndrome (9q34.3 deletion including *EHMT1*) | HP |
| 6325_3 | M | Sporadic | 11q13.3 | 11:70077507-70506315 | 428,809 | Loss | *De novo* | *SHANK2* exonic deletion[129] | HP |
| 6319_3 | M | Sporadic | 11q13.3 | 11:70119917-70187872 | 67,956 | Loss | *De novo* | *SHANK2* exonic deletion[117,129] | HP |
| 5237_3 | M | Sporadic | 11q13.3-q13.4 | 11:70154458-70220632 | 66,175 | Loss | *De novo* | *SHANK2* exonic deletion[117,134] | HP |
| 6240_4 | M | Sporadic | 11q24.2-q25 | 11:126633939-132060374 | 5,426,436 | Loss | *De novo* | Chromosome 11q deletion syndrome (Jacobsen syndrome) | HP |
| 20187_1464001 | M | Sporadic | 15q11.2-q13.1 | 15:18811937-26209270 | 7,397,334 | Gain | *De novo* | 15q11-q13 duplication syndrome, maternally derived | VE/IP |
| 8630_201 | M | Familial | 15q11.2-q13.1 | 15:19800798-26209270 | 6,408,473 | Gain | *De novo* | 15q11-q13 duplication syndrome, paternally derived | VE/IP |
| 20069_1328001 | M | Sporadic | 15q11.2-q13.1 | 15:20203578-26209270 | 6,005,693 | Gain | *De novo* | 15q11-q13 duplication syndrome, maternally derived | VE/IP |
| 17035_1 | F | Sporadic | 15q11.2-q13.1 | 15:20274130-26120360 | 5,846,231 | Gain | *De novo* | 15q11-q13 duplication syndrome, maternally derived | VE/IP |
| 8117_202 | M | Familial | 15q11.2-q13.1 | 15:21168391-26217954 | 5,049,564 | Gain | Maternal | 15q11-q13 duplication syndrome, maternally derived | VE/IP |
| 8741_201 | F | Familial | 15q11.2-q13.1 | 15:21168391-26315093 | 5,146,703 | Gain | Maternal | 15q11-q13 duplication syndrome, maternally derived | VE/IP |
| 13050_593 | M | Sporadic | 15q11.2-q13.1 | 15:21190624-26203954 | 5,013,331 | Gain | *De novo* | 15q11-q13 duplication syndrome, maternally derived | VE/IP |
| 1950_301 | M | Familial | 15q13.1-q13.3 | 15:26762141-30436163 | 3,674,023 | Loss | Maternal | 15q13.3 deletion syndrome | VE/IP |
| 16040_1571029001 | M | Familial | 15q13.2-q13.3 | 15:28450423-30303265 | 1,852,843 | Loss | *De novo* | 15q13.3 deletion syndrome | VE/IP |
| 14167_2720 | M | Sporadic | 15q13.2-q13.3 | 15:28705540-30436163 | 1,730,624 | Loss | Paternal | 15q13.3 deletion syndrome | VE/IP |
| 18100_302 | M | Familial | 15q13.2-q13.3 | 15:28714502-30303265 | 1,588,764 | Loss | Paternal | 15q13.3 deletion syndrome | VE/IP |
| 5537_3 | M | Familial | 15q13.2-q13.3 | 15:82722026-83529838 | 807,813 | Loss | Maternal | Distal 15q25.2 deletion syndrome | VE/IP |
| 14283_4060 | M | Sporadic | 16p13.11 | 16:14771033-16307313 | 1,536,281 | Loss | Maternal | 16p13.11 deletion syndrome | VE/IP |
| 3441_3 | M | Sporadic | 16p13.11 | 16:14808871-16215852 | 1,406,982 | Loss | Paternal | 16p13.11 deletion syndrome | VE/IP |
| 14412_5210 | M | Sporadic | 16p13.11 | 16:14960247-16307313 | 1,347,067 | Loss | Maternal | 16p13.11 deletion syndrome | VE/IP |
| 2204_1 | M | Sporadic | 16p11.2 | 16:29466569-30147029 | 680,461 | Loss | *De novo* | 16p11.2 deletion syndrome | VE/IP |
| 3544_3 | M | Familial | 16p11.2 | 16:29499858-30107306 | 607,449 | Gain | Maternal | 16p11.2 duplication syndrome | VE/IP |
| 20089_1391001 | M | Sporadic | 16p11.2 | 16:29502984-30107306 | 604,323 | Loss | *De novo* | 16p11.2 deletion syndrome | VE/IP |
| 5068_3 | F | Familial | 16p11.2 | 16:29502984-30127026 | 624,043 | Loss | *De novo* | 16p11.2 deletion syndrome, 50% mosaicism | VE/IP |
| 5262_4 | M | Sporadic | 16p11.2 | 16:29502984-30210849 | 707,866 | Gain | *De novo* | 16p11.2 duplication syndrome | VE/IP |
| 3211_3 | M | Familial | 16p11.2 | 16:29502984-30127026 | 624,043 | Gain | Maternal | 16p11.2 duplication syndrome | VE/IP |
| 5359_4 | M | Sporadic | 16p11.2 | 16:29554843-30195224 | 640,382 | Loss | *De novo* | 16p11.2 deletion syndrome | VE/IP |
| 20127_4014001 | M | Sporadic | 16p11.2 | 16:29554843-30130862 | 576,020 | Loss | Paternal | 16p11.2 deletion syndrome | VE/IP |
| 4030_1 | M | Familial | 16p11.2 | 16:29554843-30107306 | 552,464 | Gain | *De novo* | 16p11.2 duplication syndrome | VE/IP |
| 3439_3 | M | Sporadic | 17p11.2 | 17:17156307-18262979 | 1,106,673 | Loss | *De novo* | Smith-Magenis syndrome (17p11.2 deletion including *RAI1*) | HP |
| 2211_1 | F | Sporadic | 17p11.2 | 17:17169258-20101517 | 2,932,260 | Loss | *De novo* | Smith-Magenis syndrome (17p11.2 deletion including *RAI1*) | HP |
| 14315_4320 | M | Sporadic | 17q12 | 17:31621634-33323919 | 1,702,286 | Gain | Maternal | 17q12 duplication syndrome | VE/IP |
| 3183_7 | M | Familial | 22q11.21 | 22:17241748-19819918 | 2,578,171 | Loss | *De novo* | 22q11.2 deletion syndrome | VE/IP |
| 17015_1 | M | Sporadic | 22q11.21 | 22:17257787-19795780 | 2,537,994 | Loss | *De novo* | 22q11.2 deletion syndrome | VE/IP |
| 8627_201 | M | Sporadic | 22q11.21 | 22:17257787-19793730 | 2,535,944 | Gain | *De novo* | 22q11.2 duplication syndrome | VE/IP |
| 4271_1 | M | Sporadic | 22q11.21 | 22:17257787-19795780 | 2,537,994 | Gain | *De novo* | 22q11.2 duplication syndrome | VE/IP |
| 16074_1571042001† | M | Sporadic | 22q11.21 | 22:17257787-19793730 | 2,535,944 | Gain | Paternal | 22q11.2 duplication syndrome | VE/IP |
| 3127_4 | M | Familial | 22q11.21 | 22:17257787-19793730 | 2,535,944 | Gain | Paternal | 22q11.2 duplication syndrome | VE/IP |
| 5261_4 | F | Familial | 22q11.21 | 22:17257787-19795780 | 2,537,994 | Gain | Paternal | 22q11.2 duplication syndrome | VE/IP |
| 2072_1 | M | Familial | 22q13.31-q13.33 | 22:45159185-49582267 | 4,423,083 | Loss | *De novo* | Phelan-McDermid syndrome (22q13 deletion including *SHANK3*)[130,135] | HP |
| 6130_4 | F | Sporadic | 22q13.32-q13.33 | 22:47996161-49512530 | 1,516,370 | Loss | *De novo* | Phelan-McDermid syndrome (22q13 deletion including *SHANK3*) | HP |
| 16079_1571066001 | M | Sporadic | 22q13.33 | 22:49470371-49567383 | 97,013 | Loss | *De novo* | Phelan-McDermid syndrome (22q13 deletion including *SHANK3*) | HP |
| 5240_4 | M | Familial | Xp22.11 | X:23116188-23280628 | 164,441 | Loss | Maternal | *PTCHD1* exonic deletion[136,137] | — (sex chr) |
| 5126_4 | M | Familial | Xp21.3 | X:28931559-29478966 | 547,408 | Gain | Maternal | *IL1RAPL1* intragenic duplication of exons 3-5 | — (sex chr) |
| 8597_201 | M | Sporadic | Xp21.2-p21.1 | X:31303978-32025062 | 721,085 | Loss | Maternal | *DMD* deletion of exons 45-60 | — (sex chr) |
| 3019_3 | M | Familial | Xp21.1 | X:32100618-32315937 | 215,320 | Gain | Maternal | *DMD* duplication of exons 31-44 (sequencing coordinates: chrX:32097213-32321945, size: 224733 bp)[138] | — (sex chr) |
| 20013_1075001 | M | Sporadic | Xq28 | X:153239048-153521797 | 282,750 | Gain | *De novo* | Xq28 duplication encompassing 20 genes, including 3 involved in ID: *FLNA, GDI1*, *IKBKG*. Recurrent Xq28 duplication implicated in XLID; GDI1 is the most likely candidate gene[55] (sequencing coordinates: chrX:153,222,048-153,514,311, size 292 kb) | — (sex chr) |
| 14216_3470 | M | Sporadic | Xq28 | X:153263157-153474401 | 211,245 | Gain | Maternal | Xq28 duplication encompassing 18 genes, including 2 involved in ID: *GDI1* & *IKBKG.* Recurrent Xq28 duplication implicated in XLID; *GDI1* is the most likely candidate gene | — (sex chr) |

**Other pathogenic CNVs, including large *de novo* CNVs as well as CNVs not included in the main analyses (chromosomal abnormalities, CNVs <30 kb, non stringent CNVs or CNVs identified with 1 algorithm only and validated), n = 20**

| AGP ID | Sex | Family type | Cytoband | Chr#:start-end (hg18) | Size | CNV type | Inheritance | CNV description | Classification penetrance/ expressivity[§] |
|---|---|---|---|---|---|---|---|---|---|
| 8658_201 | F | Sporadic | 1p21.3-p21.2 | 1:98175622-100923952 | 2,748,331 | Loss | *De novo* | 2.7 Mb 1p21.3-p21.2 deletion, *de novo* (18 genes, including *DPYD* and *MIR137*). 1p21.3 microdeletions comprising *DPYD* and *MIR137* have been reported in subjects with ID | — (novel region) |
| 5467_3 | M | Sporadic | 1q42.3-q44 | 1:233476547-247165725 | 13,689,179 | Gain | *De novo* | 13.7 Mb 1q42.3-q44 duplication, *de novo* (114 genes). This CNV overlaps the critical region of the 1q43-q44 deletion syndrome. Distal duplications of the long arm of chromosome 1 are rare, reported in few individuals in its pure form (chromosome abnormality) | HP |
| 5236_3 | F | Familial | 2p16.3 | 2:50705521-50719594 | 14,074 | Loss | Maternal | *NRXN1* exonic deletion (CNV is exonic according to RefSeq, but not to UCSC) (<30 kb, non stringent: QSNP\|PCNV) | VE/IP |
| 5328_3 | M | Sporadic | 2p16.3 | 2:51044181-51120644 | 76,464 | Loss | Paternal | *NRXN1* exonic deletion (non stringent: QSNP\|PCNV) | VE/IP |
| 16037_1571015001 | M | Sporadic | 2q37.3 | 2:239765200-239777909 | 12,710 | Loss | Maternal | *HDAC4* small intragenic deletion. *HDAC4* is responsible for some of the features of the 2q37 deletion syndrome (brachydactyly mental retardation syndrome). There are many 2q37 deletions reported in individuals with ASD, but no mutations or single gene deletions of *HDAC4* had been reported in ASD. (sequencing coordinates: chr2:239766528-239778481, size 11954 bp) (<30 kb) | VE/IP |
| 21020_1‡ | M | Sporadic | 4p16.3-p16.1 | 4:53403-9016339 | 8,962,937 | Gain | *De novo* | *De novo* unbalanced translocation leading to 4p16.3-p16.1 duplication and 8p23.3-p23.1 deletion. The 4p duplication spans 8.9 Mb and involves the region implicated in Wolf-Hirschhorn syndrome. Normal 46,XY karyotype, unbalanced translocation shown by FISH (chromosome abnormality) | HP |
| 21020_1‡ | M | Sporadic | 8p23.3-p23.1 | 8:154984-6994825 | 6,839,842 | Loss | *De novo* | *De novo* unbalanced translocation leading to 4p16.3-p16.1 duplication and 8p23.3-p23.1 deletion (see above). The 8p terminal deletion spans 6.8 Mb; numerous terminal deletions of 8p have been described in the literature in subjects with ID, behavioral issues and mild dysmorphic features (chromosome abnormality) | HP |
| 14270_3930 | F | Sporadic | 6q25.3-q27 | 6:160773919-170761395 | 9,987,477 | Gain | *De novo* | 10 Mb duplication 6q25.3-q27, *de novo*, 46,XX.ish der(22)t(6;22)(6q25.3;p11.2)pat(6qtel+). "6q duplication syndrome" is a rare cytogenetic abnormality reported in few individuals in its pure form; the duplications can affect any part of the 6q arm and few overlap. Most of the reported cases are the result of abnormal segregation of a balanced translocation carried by a parent, like in this case (chromosome abnormality) | HP |
| 8404_201 | M | Familial | 6q27 | 6:169136788-170761395 | 1,624,608 | Loss | *De novo* | 1.6 Mb *de novo* deletion in 6q27 region (14 genes). Terminal 6q deletion syndrome. The CNV contains *DLL1*, an ASD candidate gene: *de novo* frameshift variant identified in ASD exome study[4] | HP |
| 13137_1543 | F | Sporadic | 8p12-8q12.1 | 8:31928590-58996070 | 27,067,481 | Gain | *De novo* | 27 Mb pericentromeric duplication 8p12-8q12.1, *de novo*. The karyotype revealed a mosaic supernumerary ring chromosome (47, XX, +r[10]/46, XX[70]) of unknown origin, shown to involve chr 8 by SNP array. Trisomy 8 syndrome is characterized by mild to severe mental and growth deficiency, facial dysmorphisms, and limb abnormalities. In cases of supernumerary marker chromosomes or supernumerary ring chromosomes derived from chr 8, mosaic or non-mosaic, the clinical presentation varies from normal phenotype to features overlapping the trisomy 8 syndrome. (chromosome abnormality) | HP |
| 8534_201 | M | Sporadic | 10q11.21-q11.23 | 10:45633089-51564756 | 5,931,668 | Loss | *De novo* | 5.9 Mb 10q11.22-q11.23 deletion (56 genes), *de novo*. Recurrent deletions in this region are associated with variable clinical features, with ID as the only feature present in the majority of individuals; most deletions are inherited from apparently normal parents, indicating variable expressivity/incomplete penetrance[139] | VE/IP |
| 4312_1 | M | Sporadic | 10q11.21-q11.23 | 10:45550419-51496386 | 5,945,968 | Loss | Paternal | 5.9 Mb 10q11.22-q11.23 deletion (56 genes), paternal. Recurrent deletion (see above) | VE/IP |
| 6053_3 | M | Familial | 12q13.3-q14.1 | 12:54218922-58779615 | 4,560,694 | Gain | *De novo* | 4.5 Mb *de novo* duplication in 12q13 (101 genes); no similar CNV reported | — (novel region) |
| 14070_1230 | M | Familial | 15q26.1 | 15:91200007-91283004 | 82,998 | Loss | *De novo* | *CHD2* exonic deletion. *CHD2 de novo* mutations reported recently in epileptic encephalopathy and ID; several deletions reported in individuals with ID, ASD and epilepsy | HP |
| 5420_3* | M | Sporadic | Whole chr 21 | 21:1-247249719 | 247,249,719 | Gain | *De novo* | Down syndrome (47,XY+21) (chromosome abnormality) | HP |
| 14291_4120 | F | Sporadic | 22q13.33 | 22:49470371-49480446 | 10,076 | Loss | *De novo* | *SHANK3* deletion of exons 9-13 (<30 kb, called with 1 algorithm only: PCNV) (sequencing coordinates: 49468716-49485255, size 16540 bp) | HP |
| 5241_3 | M | Familial | Xp21.1 | X:31793278-31822704 | 29,427 | Loss | Maternal | *DMD* deletion of exon 48 (predicted to lead to an in-frame deletion) (<30 kb) | — (sex chr) |
| 9861_202 | M | Sporadic | Xp11.4 | X: 41248675-41259467 | 10,793 | Gain | Maternal | *CASK* partial duplication of the 5' UTR of exon 1, decreased mRNA expression in cell line (<30 kb, called with 1 algorithm only: PCNV) | — (sex chr) |
| 5257_3* | M | Sporadic | Whole chr Y | Y:1-57772954 | 57,772,954 | Gain | *De novo* | XYY syndrome (chromosome abnormality) | — (sex chr) |
| 5515_3* | M | Sporadic | Whole chr Y | Y:1-57772954 | 57,772,954 | Gain | *De novo* | XYY syndrome (chromosome abnormality) | — (sex chr) |

This table includes pathogenic CNVs overlapping ASD/ID genes or loci as well as other pathogenic CNVs in cases of all ancestries. The latter included chromosome abnormalities (>7.5 Mb) that had been excluded from the main analyses, selected large rare *de novo* events, as well as experimentally validated smaller CNVs (<30 kb) or CNVs not considered stringent (called by only one algorithm).

With the exception of the 16p11.2 duplication found in proband 3544_3 (no validation attempted), all the CNVs reported in this table were experimentally validated (see details in **Table S8**).

§ For the analysis shown in **Figure 2E**, autosomal pathogenic CNVs were classified as highly penetrant (HP) or associated with variable expressivity and/or incomplete penetrance (VE/IP). CNVs on sex chromosomes or affecting novel regions were not classified for this analysis.

† Proband 16074_1571042001 carries two pathogenic CNV: a *NRXN1* exonic deletion inherited from his mother and a 22q11.2 duplication inherited from his father.

‡ Proband 21020_1 has a *de novo* unbalanced translocation leading to 4p duplication and 8p deletion, both considered pathogenic.

* Three cases showed whole chromosome aneuploidies: one case with Down syndrome and two with XYY syndrome.

Abbreviations: F, female; ID, intellectual disability; M, male; UTR, untranslated region; XLID, X-linked intellectual disability

## Table S8. Phenotypes in ASD subjects with pathogenic CNVs or with selected CNVs of uncertain significance

mmc2.xlsx (Excel workbook)

This file also contains information on CNV validation and segregation in siblings, when available.

## Table S9. Meta-analysis of loci and genes affected by rare CNVs in large ASD cohorts

| Locus/gene | AGP, Stages 1+2 [a] Cases (1853 M, 294 F) Controls (1241 M, 1399 F) | | | | SSC [b] Cases (968 M, 156 F) Controls (403 M, 469 F) | | | | AGRE [c] Cases (n=1835)* Cases (n=837; 755 M)** | | Combined AGP + SSC + AGRE Cases (2821 M, 450 F + 1835* or 837; 755 M**) Controls (1644 M, 1868 F) | | | | Combined AGP+SSC+AGRE Del | Combined AGP+SSC+AGRE Dup | Del P-value† | Dup P-value† | Del Freq % | Dup Freq % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cases | | Controls | | Cases | | Controls | | Cases | | Cases | | Controls | | Cases/Controls | Cases/Controls | | | | |
| | Del | Dup | Del | Dup | Del | Dup | Del | Dup | Del | Dup | Del | Dup | Del | Dup | | | | | | |
| 16p11.2 | 5 (4) | 4 (2) | – | 2 | 8 (7) | 6 (4) | – | – | 3 (3)* | 2 (0) * | 16 (14) | 12 (6) | – | 2 | 16/5106; 0/3512 | 12/5106; 2/3512 | 2.28E-04 | 3.52E-02 | 0.313 | 0.235 |
| 15q13.3 (BP3-BP5, BP4-BP5)# | 4 (1) | – | – | 1 | 2 (1) | 1 (1) | – | – | 2 (0)* | 1 (0)* | 8 (2) | 2 (1) | – | 1 | 8/5106; 0/3512 | 2/5106; 1/3512 | 1.52E-02 | 0.637 | 0.157 | 0.039 |
| 16p13.11# | 3 (0) | 5 (0) | – | 2 | 1 (0) | 2 (1) | 1 | 3 | 3 (0)* | 4 (0; 1 unk)** | 7 (0) | 11 (1; 1 unk) | 1 | 5 | 7/5106; 1/3512 | 11/4108; 5/3512 | 9.86E-02 | 0.174 | 0.137 | 0.268 |
| 15q11q13 (BP1-BP3, BP2-BP3)$ | – | 6 (5) | – | – | – | 1 (1) | – | – | – | 6 (2)* | – | 13 (8) | – | – | 0 | 13/5106; 0/3512 | – | 1.10E-03 | – | 0.255 |
| 22q11.2 (DiGeorge syndrome) | 2 (2) | 5 (2) | – | 1 | 1 (1) | – | – | – | – | 3 (1)* | 3 (3) | 8 (3) | – | 1 | 3/5106; 0/3512 | 8/5106; 1/3512 | 0.208 | 6.46E-02 | 0.059 | 0.157 |
| 17p11.2 (Smith-Magenis syndrome) | 2 (2) | – | – | – | – | – | – | – | – | – | 2 (2) | – | – | – | 2/4108**; 0/3512 | 0 | 0.291 | – | 0.049 | – |
| 9q34.3 (Kleefstra syndrome) | 1 (1) | – | – | – | 1 (1) | – | – | – | – | – | 2 (2) | – | – | – | 2/4108**; 0/3512 | 0 | 0.291 | – | 0.049 | – |
| 1q21.1 | 1 (1) | 3 (2) | – | 1 | – | 3 (2) | – | – | 1 (1)* | 2 (0)* | 2 (2) | 8 (4) | – | 1 | 2/5106; 0/3512 | 8/5106; 1/3512 | 0.351 | 6.46E-02 | 0.039 | 0.157 |
| 17q12 | – | 1 (0) | – | – | 2 (1) | – | – | – | – | – | 2 (1) | 1 (0) | – | – | 2/5106; 0/3512 | 1/5106; 0/3512 | 0.351 | 0.593 | 0.039 | 0.020 |
| 7q11.23 (Williams syndrome) | 1 (1) | – | – | – | – | 4 (4) | – | – | – | – | 1(1) | 4 (4) | – | – | 1/4108; 0/3512 | 4/4108; 0/3512 | 0.539 | 8.44E-02 | 0.024 | 0.097 |
| | | | | | | | | | | | | | | | | | – | – | – | – |
| *PTCHD1/PTCHD1AS* (Xp22.11)‡+ | 8 (1) | – | 1 | – | 3 (1) | – | 1 | – | 1 (0)** | – | 12 (2) | – | 2 | – | 12/3576**; 2/1644 M | 0 | 0.133 | – | 0.336 | – |
| *NRXN1* (2p16.3) | 6 (4) | 1 (1)§ | 1 | – | 3 (1) | – | 1 | – | 4 (0)** | – | 13 (5) | 1 (1) | 2 | – | 13/4108**; 2/3512 | 1/4108; 0/3512 | 8.50E-03 | 0.539 | 0.316 | 0.024 |
| *SHANK3* (22q13.33)& | 4 (4) | – | – | – | – | – | – | – | – | – | 4 (4) | – | – | – | 4/4108**; 0/3512 | 0 | 8.44E-02 | – | 0.097 | – |
| *SHANK2* (11q13.3) | 3 (3) | – | – | – | – | – | – | – | – | 1 (0)** | 3 (3) | 1 (0) | – | – | 3/4108**; 0/3512 | 1/4108; 0/3512 | 0.157 | 0.539 | 0.073 | 0.024 |
| *NLGN3* (Xq13.1) ‡ | – | – | – | – | 1 (0) | – | – | – | – | – | 1 (0) | – | – | – | 1/3576**; 0/1644 M | 0 | 0.685 | – | 0.028 | – |
| *NLGN4X* (Xp22.3) ‡ | – | – | – | – | – | 1 (0) | – | – | – | 1 (1 unk)** | – | 2 (1 unk) | – | – | 0 | 2/3576; 0/1644 M | – | 0.469 | – | 0.056 |

CNVs are ordered according to frequency. Numbers in parentheses indicate *de novo* events. For *NRXN1*, *SHANK2*, *SHANK3*, *NLGN3* and *NLGN4X* CNVs are only counted in cases and controls if they affect one or more exons; chromosome X events are only counted for males.

*Number of AGRE cases in Moreno-de-Luca et al.[111] (n=1835 cases). A few of the regions in this table were not listed in Moreno-de-Luca et al.'s paper; we obtained the counts of subjects with deletions/duplications in these additional regions directly from the authors - the total number of inspected cases after quality control for the additional regions was 837 unrelated probands from 1105 families (or a total of 1472 all affected): 161 families with multiple males and females affected children, 594 families with only one male affected child and 82 families with only one female affected child. The total number of AGRE families with at least one male affected child was 161+594=755, and the total number of AGRE families with at least one female affected child was 161+82=243. The combined number of cases used to get estimates for this table was: i) 5106 = (AGP+SSC) + AGRE= 3271+1835 or ii) 4108 (**) = 3271+837.

† Fisher exact test, one-sided P-value

[#] Duplications of the 15q13.3 and 16p13.11 regions are of uncertain clinical significance, because they have not been found to be consistently enriched in cases compared to controls.

[$] Of the 13 duplications of the 15q11q13 region, 5 were maternally inherited (1 AGP, 4 AGRE), and 8 were *de novo*. Of the 8 *de novo* duplications, 5 were of maternal origin (4 AGP, 1 SSC), two were of paternal origin (AGP: 8630_201 and AGRE: AU1135202), and one without information (1 AGRE: AU1042303).

[§] *NRXN1* intragenic duplication, predicted to result in a premature truncated protein.

‡ Only males counted. The combined number of males used to get estimates for this table was 3576 = (AGP+SSC) + AGRE= 2821+755

+ The two *de novo* deletions in this locus were found in male probands from an SSC family (12561) and AGP family (subject 5240_4)

[&] Includes *SHANK3* gene exonic deletions as well as 22q13 deletion syndrome (Phelan-McDermid syndrome).

Abbreviations: AGP, Autism Genome Project; AGRE, Autism Genetic Resource Exchange; BP, breakpoint; Del, deletion; Dup, duplication; F, female; Freq, frequency; M, male; SSC, Simons Simplex Collection; unk, unknown.

[a] Ref [117] + Stage 2 (this study) (only subjects of European ancestry included)

[b] Ref [132]

[c] Ref [111]

## Table S10. FMRP targets affected by deletions in probands and not yet implicated in ASD or ID

| FMRP targets not yet implicated in ASD or ID | |
|---|---|
| *ABR* | *MAPK8IP3* |
| *AGTPBP1* | *MBP* |
| *ARHGAP32* | *NBEA* |
| *BAI1* | *NRXN3* |
| *BRSK1* | *PCDH9* |
| *CAMSAP1* | *PLCB1* |
| *DTNA* | *PLEC* |
| *EML2* | *PRKACB* |
| *EP400* | *PTPRD* |
| *FAM115A* | *PTPRT* |
| *FAM21A* | *R3HDM1* |
| *FAM91A1* | *RALGAPA1* |
| *FAT4* | *RALGDS* |
| *GRM5* | *RPRD2* |
| *KCNH1* | *SHANK1* |
| *KIAA0430* | *SPTB* |
| *KIAA0913* | *TCF25* |
| *LLGL1* | *TRIP12* |
| *LPHN3* | *TRPM3* |
| *LPPR4* | *ULK2* |
| *LYNX1* | *UNC13C* |
| *MADD* | |

Of the FMRP targets[9] (n=842, one without correspondence in our annotation file) affected by exonic deletions in ASD subjects, 58 genes correspond to known ASD/ID genes (**Tables S6A-S6D**). This table lists the remaining 43 FMRP targets not yet implicated in ASD or ID, considered ASD candidate genes.

## Tables S11A-S11E. Multigene analyses – various models

### Table S11A. Primary analysis

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| **Model treating CNVs within the same individual as independent** (pseudo $R^2$ 0.007) | | | | |
| (Intercept) | -0.195271 | 0.027722 | -7.044 | 1.87e-12*** |
| factor(CNV)Dup | -0.120601 | 0.039779 | -3.032 | 0.00243** |
| ngene | 0.030912 | 0.005768 | 5.359 | 8.37e-08*** |
| RPKM | 0.057893 | 0.024207 | 2.392 | 0.01677* |
| **Model treating CNVs within the same individual as independent, and analyzing deletions only** (pseudo $R^2$ 0.007) | | | | |
| (Intercept) | -0.21920 | 0.02979 | -7.359 | 1.85e-13*** |
| ngene | 0.02857 | 0.01030 | 2.773 | 0.005549** |
| RPKM | 0.14595 | 0.04153 | 3.515 | 0.000441*** |
| **Model treating CNVs within the same individual as independent, and analyzing duplications only** (pseudo $R^2$ 0.008) | | | | |
| (Intercept) | -0.282964 | 0.036126 | -7.833 | 4.78e-15*** |
| ngene | 0.031361 | 0.006981 | 4.492 | 7.04e-06*** |
| RPKM | 0.011110 | 0.029966 | 0.371 | 0.711 |
| **Model with the number of genes and the average RPKM nested within deletion or duplication status** (pseudo $R^2$ 0.008) | | | | |
| (Intercept) | -0.219196 | 0.029785 | -7.359 | 1.85e-13*** |
| factor(CNV)Dup | -0.063768 | 0.046821 | -1.362 | 0.173217 |
| factor(CNV)Del:ngene | 0.028566 | 0.010301 | 2.773 | 0.005549** |
| factor(CNV)Dup:ngene | 0.031361 | 0.006981 | 4.492 | 7.04e-06*** |
| factor(CNV)Del:RPKM | 0.145950 | 0.041528 | 3.515 | 0.000441*** |
| factor(CNV)Dup:RPKM | 0.011110 | 0.029966 | 0.371 | 0.710813 |

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

In our primary analysis we treated CNVs falling within the same individual as independent (main **Figures 3C-D** and **Table S11A**). **Figures 3C-D** show the pattern of increased burden with the increased number of brain-expressed genes affected (1-hit, 2-hit, 3-hit, 4- to 10-hit) by deletions or duplications. The percentage of cases and controls with CNVs overlapping genes is shown for deletions and duplications separately. To account for the fact that some individuals carry a multiplicity of rare CNVs, we also performed analyses counting all genes in all rare CNVs per individual, as well as the brain expression of those genes, and entered those in the model (**Table S11B)**; and we evaluated the impact of the largest CNV and its gene expression. Results from these alternative models were similar, showing that the dependence due to CNVs occurring in the same individual has only a minor impact on inference. For each model, Nagelkerke's pseudo-$R^2$ was calculated to provide a description of how well the model fitted the data.

In **Figure 3C-D** (main text), the expected odds ratios (depicted by stars) were estimated by fitting a logit model of the case status (case/control) with CNV type and three covariates, the deletion/duplication status, the number of genes covered by each CNV and their average brain expression value for the genes covered by the CNV. Gene level expression values for the neocortex were obtained from the BrainSpan RNA sequencing resource, transformed to log(1+RPKM), and the average of the transformed RPKM value for the genes covered by the CNV was used in the model (i.e. total RPKM of all genes covered by the CNV divided by the total number of genes within the CNV). Differences between the effect of the number of genes and expression values between deletion and duplication CNVs were further evaluated by analyzing each subset of the data separately.

In **Table S11A**, we fit a logit model with the following covariates: deletion/duplication status of the CNV, the number of genes covered by each CNV and the average of the transformed RPKM value (i.e. RPKM value transformed to log(1+RPKM)) for the genes covered by the CNV (total RPKM of all genes covered by the CNV divided by the total number of genes). We ignored the fact that individuals could have multiple CNVs. All effects in the model are significant. There is higher risk for autism when the number of genes covered by the CNV is larger and when the average (transformed) RPKM value is higher. Also notice that the risk for autism is slightly lower for duplications as opposed to deletions.

We also checked whether there were differences between the effect of the number of genes and RPKM between deletions and duplications by analyzing each subset of the data separately. Notice that the increased risk due to the number of covered genes for autism is fairly similar for duplications and deletions, while the increased risk due to an increase of the average RPKM is significant for deletions but not for duplications.

Another way to assess the differences between duplications and deletions would be to fit the effect of the number of genes and the average RPKM nested within deletion and duplication CNV status. The results are virtually the same for this model as the previous split data analysis.

## Table S11B. Models accounting for the fact that some individuals carry a multiplicity of rare CNVs

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| **Model considering the total number of genes covered by all deletions or duplications in an individual (ngeneDel, ngeneDup) and the average RPKM value for the genes in deleted or duplicated regions (mRPKMDel, mRPKMDup)** (pseudo $R^2$ 0.016) | | | | |
| (Intercept) | -0.37032 | 0.04326 | -8.560 | <2e-16*** |
| ngeneDel | 0.03292 | 0.01006 | 3.273 | 0.001066** |
| ngeneDup | 0.03506 | 0.01010 | 3.471 | 0.000519*** |
| mRPKMDel | 0.17159 | 0.05014 | 3.422 | 0.000621*** |
| mRPKMDup | 0.01307 | 0.03796 | 0.344 | 0.730597 |
| **Model when selecting CNVs within individuals based on the maximum number of genes covered by the CNVs** (pseudo $R^2$ 0.018) | | | | |
| (Intercept) | -0.365583 | 0.042510 | -8.600 | <2e-16*** |
| maxGeneDel | 0.035244 | 0.010939 | 3.222 | 0.001273** |
| maxGeneDup | 0.027772 | 0.006957 | 3.992 | 6.55e-05*** |
| RPKMDel | 0.171068 | 0.048600 | 3.520 | 0.000432*** |
| RPKMDup | -0.000702 | 0.036458 | -0.019 | 0.984637 |
| **Model when selecting the CNVs based on the maximum average standardized RPKM** (pseudo $R^2$ 0.015) | | | | |
| (Intercept) | -0.36725 | 0.04326 | -8.490 | <2e-16*** |
| nGeneDel | 0.03289 | 0.01173 | 2.805 | 0.00503** |
| nGeneDup | 0.02543 | 0.00795 | 3.199 | 0.00138** |
| maxRPKMDel | 0.17640 | 0.04392 | 4.016 | 5.91e-05*** |
| maxRPKMDup | 0.01741 | 0.03205 | 0.543 | 0.58705 |

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

To take into account the effect of multiple CNVs per individual we decided to re-analyze the data using a model in which we fit the total number of genes covered by all deletions in an individual (ngeneDel), the total number of genes covered by duplications in an individual (ngeneDup) as well as the average RPKM value for these genes in a deletion region (mRPKMDel) and the average RPKM for those in a duplication region (mRPKMDup) **(upper panel)**. All effects, except the one for the average RPKM of genes covered by duplications (mRPKMDup) were significant. Notice the similarity of the results of this model with the one in which we analyzed the duplications and deletions separately. In both analyses the risk due to the number of genes is similar in duplications and deletions. However, only higher mean RPKM values for deletions increase risk for ASD.

Instead of looking at the average number of genes and average RPKM covered by deletions and duplications, one can also identify the CNVs for which the maximum value is obtained. This table shows the results of the generalized linear model when CNVs are selected within individuals based on the maximum number of genes covered by the CNVs **(middle panel)**, as well as the results obtained when selecting the CNVs based on the maximum average standardized RPKM **(lower panel)**. Notice the great similarity between these two analyses as well as the analysis in which we took the average of multiple CNVs per individual. In all three cases the effect of the number of genes covered by the CNVs is significant and the magnitude of the risk is very similar. Also, for all three analyses, the value of the average RPKM is only significant for deletions.

## Table S11C. Removing cases with validated *de novo* CNVs

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| **Model considering the average number of genes and the average RPKM of genes covered by the CNVs in the remaining individuals** (pseudo $R^2$ 0.005) | | | | |
| (Intercept) | -0.343950 | 0.043740 | -7.863 | 3.74e-15*** |
| ngeneDel | 0.010159 | 0.011975 | 0.848 | 0.39625 |
| ngeneDup | 0.024238 | 0.010586 | 2.290 | 0.02204* |
| mRPKMDel | 0.137438 | 0.051433 | 2.672 | 0.00754** |
| mRPKMDup | 0.006056 | 0.038418 | 0.158 | 0.87474 |
| **Model considering the CNVs with the maximum number of genes covered, in the remaining individuals** (pseudo $R^2$ 0.006) | | | | |
| (Intercept) | -0.343072 | 0.043122 | -7.956 | 1.78e-15*** |
| maxGeneDel | 0.010173 | 0.013176 | 0.772 | 0.44005 |
| maxGeneDup | 0.020914 | 0.007687 | 2.721 | 0.00651** |
| RPKMDel | 0.139397 | 0.049723 | 2.803 | 0.00506** |
| RPKMDup | -0.004947 | 0.036953 | -0.134 | 0.89351 |
| **Model considering the CNVs with the maximum RPKM, in the remaining individuals** (pseudo $R^2$ 0.006) | | | | |
| (Intercept) | -0.342025 | 0.043733 | -7.821 | 5.25e-15*** |
| nGeneDel | 0.009579 | 0.014296 | 0.670 | 0.50285 |
| nGeneDup | 0.020248 | 0.008438 | 2.399 | 0.01642* |
| maxRPKMDel | 0.131945 | 0.045620 | 2.892 | 0.00382** |
| maxRPKMDup | 0.001850 | 0.032626 | 0.057 | 0.95477 |

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

To assess the impact of validated *de novo* CNVs we removed 77 cases that had at least one validated *de novo* CNV. (Note that the total dataset contained 90 validated *de novo* CNVs in 87 cases of European ancestry; of these, 85 *de novo* CNVs in 82 unique European cases had sizes ≥30 kb; individuals with chromosomal abnormalities were excluded from the main analyses). Because of the nature of the control data, *de novo* CNV status was not available in the controls.

Results for the analysis of the average number of genes and the average RPKM of genes covered by the CNVs in the remaining individuals showed that the average number of genes covered by deletions is no longer significant **(upper panel)**. In this analysis, only the average number of genes covered by duplications and the average RPKM of genes covered by deletions are significant. A similar pattern of results is found when analyzing the CNVs with the maximum number of genes covered **(middle panel)**, and the CNVs selected based on the maximum RPKM **(lower panel)**. The results in this table show that removing only 77 out of 1,914 cases, or 4% of the case sample, results in a large decrease in signal, suggesting that most of the risk traces to *de novo* CNVs.

## Table S11D. Removing subjects with CNVs considered pathogenic

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| **Model considering the average number of genes and the average RPKM of genes covered by the CNVs in the remaining individuals** (pseudo $R^2$ 0.004) | | | | |
| (Intercept) | -0.329996 | 0.043973 | -7.505 | 6.17e-14*** |
| ngeneDel | 0.006021 | 0.012452 | 0.484 | 0.62871 |
| ngeneDup | 0.015398 | 0.011366 | 1.355 | 0.17551 |
| mRPKMDel | 0.159530 | 0.051078 | 3.123 | 0.00179** |
| mRPKMDup | 0.013352 | 0.038353 | 0.348 | 0.72774 |
| **Model considering the CNVs with the maximum number of genes covered, in the remaining individuals** (pseudo $R^2$ 0.005) | | | | |
| (Intercept) | -0.326948 | 0.043399 | -7.533 | 4.94e-14*** |
| maxGeneDel | 0.003602 | 0.013877 | 0.260 | 0.795200 |
| maxGeneDup | 0.015129 | 0.008447 | 1.791 | 0.073290 . |
| RPKMDel | 0.164659 | 0.049298 | 3.340 | 0.000838*** |
| RPKMDup | -0.001081 | 0.037003 | -0.029 | 0.976694 |
| **Model considering the CNVs with the maximum RPKM, in the remaining individuals** (pseudo $R^2$ 0.005) | | | | |
| (Intercept) | -0.330966 | 0.044004 | -7.521 | 5.43e-14*** |
| nGeneDel | 0.004337 | 0.014825 | 0.293 | 0.769840 |
| nGeneDup | 0.011546 | 0.009555 | 1.208 | 0.226890 |
| maxRPKMDel | 0.156664 | 0.045120 | 3.472 | 0.000516*** |
| maxRPKMDup | 0.011904 | 0.032612 | 0.365 | 0.715084 |

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

The AGP list of CNVs was curated to identify a subset that could be considered pathogenic (**Table S7B**). These CNVs were carried by 82 unique individuals, and of these 69 were of European ancestry and thus included in these analyses (missing individuals include other ancestries as well as individuals with chromosomal abnormalities [**Table S1C**], excluded from the main analyses).

After removing the individuals with pathogenic CNVs we reran the same three analyses as before. In general, these analyses show good agreement with the results from the comparable models obtained by removing the validated *de novo* CNVs only. This is not surprising because the two sets of events overlap fairly substantially. Here again it is shown that the value of RPKM for deletion CNVs is the most important irrespective of CNV status (i.e., even after removing both *de novo* and inherited pathogenic CNVs), significant in all three analyses. The magnitude of the risk associated with the deletion RPKMs is again very consistent to what was found in the previous analysis. Altogether these results are consistent with a genetic interaction model where imbalance of multiple genes intersected by rare *de novo* and inherited pathogenic CNVs contributes to risk. Furthermore, our findings are also in line with a previous report suggesting that deletions have larger effects on transcriptional level and contained more genes with altered expression compared to duplications.[140]

## Table S11E. Model including sex

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| **Males** (pseudo $R^2$ 0.011) | | | | |
| (Intercept) | 0.23934 | 0.05445 | 4.396 | 1.1e-05*** |
| nGeneDel | 0.03734 | 0.01417 | 2.635 | 0.00842** |
| nGeneDup | 0.03118 | 0.01300 | 2.398 | 0.01649* |
| maxRPKMDel | 0.12490 | 0.06376 | 1.959 | 0.05010 . |
| maxRPKMDup | 0.03293 | 0.04898 | 0.672 | 0.50133 |
| **Females** (pseudo $R^2$ 0.010) | | | | |
| (Intercept) | -1.69525 | 0.09428 | -17.980 | <2e-16*** |
| nGeneDel | 0.04850 | 0.01853 | 2.617 | 0.00886** |
| nGeneDup | 0.02393 | 0.01777 | 1.347 | 0.17803 |
| maxRPKMDel | 0.18177 | 0.10508 | 1.730 | 0.08366 . |
| maxRPKMDup | 0.01009 | 0.07995 | 0.126 | 0.89953 |
| **Model selecting the CNV covering the largest number of genes in males** (pseudo $R^2$ 0.013) | | | | |
| (Intercept) | 0.237196 | 0.053568 | 4.428 | 9.51e-06*** |
| maxGeneDel | 0.033687 | 0.014671 | 2.296 | 0.02167* |
| maxGeneDup | 0.027932 | 0.009389 | 2.975 | 0.00293** |
| RPKMDel | 0.132569 | 0.061994 | 2.138 | 0.03248* |
| RPKMDup | 0.027329 | 0.047174 | 0.579 | 0.56238 |
| **Model selecting the CNV covering the largest number of genes in females** (pseudo $R^2$ 0.012) | | | | |
| (Intercept) | -1.68903 | 0.09260 | -18.240 | <2e-16*** |
| maxGeneDel | 0.05820 | 0.02053 | 2.835 | 0.00458** |
| maxGeneDup | 0.01797 | 0.01216 | 1.477 | 0.13955 |
| RPKMDel | 0.19642 | 0.10003 | 1.964 | 0.04957* |
| RPKMDup | -0.01833 | 0.07737 | -0.237 | 0.81267 |
| **Model selecting the CNV with the largest average standardized RPKM in males** (pseudo $R^2$ 0.011) | | | | |
| (Intercept) | 0.23799 | 0.05448 | 4.369 | 1.25e-05*** |
| maxGeneDel | 0.02668 | 0.01536 | 1.737 | 0.08239 . |
| maxGeneDup | 0.02680 | 0.01086 | 2.467 | 0.01361* |
| maxRPKMDel | 0.16081 | 0.05596 | 2.874 | 0.00406** |
| maxRPKMDup | 0.03207 | 0.04131 | 0.776 | 0.43749 |
| **Model selecting the CNV with the largest average standardized RPKM in females** (pseudo $R^2$ 0.011) | | | | |
| (Intercept) | -1.684587 | 0.094701 | -17.788 | <2e-16*** |
| maxGeneDel | 0.065273 | 0.021786 | 2.996 | 0.00273** |
| maxGeneDup | 0.008598 | 0.017921 | 0.480 | 0.63137 |
| maxRPKMDel | 0.147739 | 0.094098 | 1.570 | 0.11640 |
| maxRPKMDup | 0.019179 | 0.068047 | 0.282 | 0.77806 |

Abbreviations: ngene, number of genes; RPKM, reads per kb per million reads

We ran the same three models, except this time the data were separated by sex: 1,641 cases and 1,102 controls were male, while 273 cases and 1,257 controls were female. There is general agreement in the parameter estimates between males and females for all three sets of analyses. We also ran tests for significant differences in predictors when they interacted with sex, but none of these interactions were significant (data not shown).

## Tables S12A-S12D. GO terms, pathways, and MPO enrichment in affecteds versus control subjects

mmc3.xlsx (Excel workbook)

## Table S13A. Gene Ontology terms and pathways used to generate a list of neurodevelopmental functions

| Gs ID | Gs Name |
|---|---|
| GO:0007399 | Nervous system development |
| GO:0050877 | Neurological system process |
| GO:0043025 | Neuronal cell body |
| GO:0043005 | Neuron projection |
| KEGG:04725 | Cholinergic synapse |
| KEGG:04724 | Glutamatergic synapse |
| KEGG:04728 | Dopaminergic synapse |
| KEGG:04727 | GABAergic synapse |
| KEGG:04726 | Serotonergic synapse |
| KEGG:04721 | Synaptic vesicle cycle |
| KEGG:04723 | Retrograde endocannabinoid signaling |
| KEGG:05030 | Cocaine addiction |
| KEGG:05031 | Amphetamine addiction |
| KEGG:05032 | Morphine addiction |
| KEGG:05033 | Nicotine addiction |
| KEGG:04722 | Neurotrophin signaling pathway |
| REACT:708 | REACT: Neuronal system |
| REACT:675 | REACT: NCAM signaling for neurite out-growth |
| REACT:138 | REACT: Axon guidance |
| NCI:142 | NCI: netrin pathway |
| NCI:180 | NCI: reelin pathway |
| KEGG:04360 | Axon guidance |
| KEGG:04720 | Long-term potentiation |
| KEGG:04730 | Long-term depression |

## Table S13B. Effect size for neurobiological-related clusters

**B1. FDR 15%**

| Logistic regression-FDR 15%[a] | Combined (% all subjects) | Stage 1 | Stage 2 |
|---|---|---|---|
| Cases | 89 (4.70%) | 57 (4.65%) | 32 (4.79%) |
| Controls | 35 (1.49%) | 17 (1.37%) | 18 (1.63%) |
| OR | 3.15 | 3.39 | 2.94 |
| Cases (max 10 genes/CNV) | 56 (2.96%) | 37 (3.02%) | 19 (2.84%) |
| Controls (max 10 genes/CNV) | 31 (1.32%) | 14 (1.13%) | 17 (1.54%) |
| OR (max 10 genes/CNV) | 2.24 | 2.67 | 1.85 |

[a]REACT: Neuronal system; REACT: Transmission across chemical synapses; KEGG: Glutamatergic synapse; KEGG: Cholinergic synapse; GO: Generation of neurons; GO: Neuron projection morphogenesis; GO: Neuron differentiation; GO: Neuron development; GO: Cell morphogenesis involved in neuron differentiation; GO: Axonogenesis; GO: Neuron projection development; GO: Axon guidance.
FDR, false-discovery rate; OR, odds ratio; max-10 genes-CNVs, estimates considering CNVs with maximum 10 genes.

**B2. FDR 20%**

| Logistic regression-FDR 20%[a] | Combined (% all subjects) | Stage 1 | Stage 2 |
|---|---|---|---|
| Cases | 99 (5.23%) | 64 (5.23%) | 35 (5.24%) |
| Controls | 49 (2.09%) | 26 (2.10%) | 23 (2.08%) |
| OR | 2.50 | 3.39 | 2.51 |
| Cases (max 10 genes/CNV) | 66 (3.49%) | 44 (3.59%) | 22 (3.29%) |
| Controls (max 10 genes/CNV) | 45 (1.92%) | 23 (1.86%) | 22 (1.99%) |
| OR (max 10 genes/CNV) | 1.82 | 1.93 | 1.65 |

[a]REACT: Neuronal system; REACT: Transmission across chemical synapses; KEGG: Glutamatergic synapse; KEGG: Cholinergic synapse; GO: Generation of neurons; GO: Neuron projection morphogenesis; GO: Neuron differentiation; GO: Neuron development; GO: Cell morphogenesis involved in neuron differentiation; GO: Axonogenesis; GO: Neuron projection development; GO: Axon guidance; REACT: Neurotransmitter receptor binding and downstream transmission in the postsynaptic cell; KEGG: Retrograde endocannabinoid signaling; REACT: Axon guidance; KEGG: Dopaminergic synapse; KEGG: Neurotrophin signaling pathway; GO: Learning or memory.
FDR, false-discovery rate; OR, odds ratio; max-10 genes-CNVs, estimates considering CNVs with maximum 10 genes.

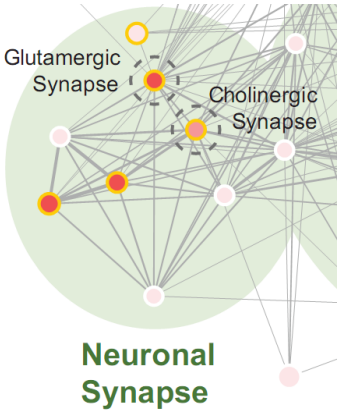
## Table S13C. Neuronal synapse main cluster

**C1. Statistics for all subjects with exonic CNVs (deletions and duplications) at FDR 15%**

| | Subject N (% all subjects) | Stage 1 | Stage 2 |
|---|---|---|---|
| Cases | 36 (1.96%) | 20 (1.63%) | 16 (2.40%) |
| Controls | 9 (0.38%) | 5 (0.40%) | 4 (0.36%) |
| OR [95% CI] | **4.93 [2.31-11.72]** | 4.04 | 6.62 |

36 cases (25 supporting genes) *vs*. 9 controls (9 supporting genes). Stage 2 is more enriched, but Stage 1 also displays a respectable signal.

FDR, false-discovery rate; CI, confidence interval; OR: odds ratio




**C2. Detailed statistics for pathways included within the Neuronal synapse cluster at FDR 15%**

| | All Cases | All Controls | Stage 1 Cases | Stage 1 Controls | Stage 2 Cases | Stage 2 Controls |
|---|---|---|---|---|---|---|
| KEGG Glutamatergic | 19 (1.00%) | 2 (0.09%) | 10 (0.82%) | 1 (0.08%) | 9 (1.35%) | 1 (0.09%) |
| KEGG Cholinergic | 17 (0.90%) | 2 (0.09%) | 9 (0.74%) | 2 (0.16%) | 8 (1.20%) | 0 (0.00%) |
| REACT Neuronal System | 26 (1.37%) | 6 (0.26%) | 14 (1.14%) | 3 (0.24%) | 12 (1.80%) | 3 (0.27%) |
| REACT Synaptic Transmission | 21 (1.11%) | 3 (0.13%) | 11 (0.90%) | 2 (0.16%) | 10 (1.50%) | 1 (0.09%) |

Statistics for all subjects with exonic CNVs (deletions and duplications included). Calculation of % of subjects is based on subjects with at least one genic exonic loss; for max-10 this is limited to subjects with a maximum of 10 deleted genes.

## Table S13D. Genes and subjects represented in the enriched cholinergic and glutamergic synapse subclusters

### D1. Genes

| | Cholinergic | | Both | | Glutamergic | |
|---|---|---|---|---|---|---|
| | **Gene name** | **# cases; # controls** | **Gene name** | **# cases; # controls** | **Gene name** | **# cases; # controls** |
| Neurotransmitters (receptors and metabolism) | *CHAT** | 2; 0 (2*) | – | – | *GRIK2* | 1; 0 |
| | *CHRNA7* | 4; 0 | – | – | *GRM5* | 1; 0 |
| | *KCNJ12* | 1; 0 | – | – | – | – |
| | *SLC18A3 (VAChT)** | 2; 0 (2*) | – | – | – | – |
| Scaffolds (synaptic organizers) | – | – | – | – | *SHANK1* | 1; 0 |
| | – | – | – | – | *SHANK2* | 3; 0 |
| | – | – | – | – | *SHANK3* | 3; 0 |
| Signalling (downstream neurotransmitter receptors and ion channels) | *CAMK2G*** | 1; 0 (1**) | *GNG2**** | 2; 0 (1***) | *PPP3CB*** | 1; 0 (1**) |
| | – | – | *GNG13* | 1; 0 | – | – |
| | – | – | *MAPK3**** | 5; 0 (1***) | – | – |
| | – | – | *PRKACB* | 1; 0 | – | – |
| | – | – | *PLCB1* | 1; 0 | – | – |

26 cases (16 supporting genes) *vs.* 3 controls (3 genes). Numbers between parentheses represent the number of *de novo* events.
** Two genes within the same event/same sample (multigene); *** two genes in two different events in same sample (double-CNV-hit).

### D2. ASD subjects

| Sample ID | Gene(s) in CNV | CNV inheritance | Biological function | Functional sub-cluster |
|---|---|---|---|---|
| 8534_201 | *CHAT*; SLC18A3** | dn (5.9 Mb del) | Neurotransmitters (receptors and metabolism) | Cholinergic |
| 4312_1 | *CHAT*; SLC18A3** | pat (5.9 Mb del) | | Cholinergic |
| 8465_202 | *GRIK2* | pat (39 kb del) | | Glutamergic |
| 8549_201 | *GRM5* | pat (1.98 Mb del) | | Glutamergic |
| 3567_4 | *KCNJ12* | pat (324 kb del) | | Cholinergic |
| 18100_302 | *CHRNA7* | pat (1.6 Mb del) | | Cholinergic |
| 1950_301 | *CHRNA7* | mat (1.7 Mb del) | | Cholinergic |
| 14167_2720 | *CHRNA7* | pat (1.7 Mb del) | | Cholinergic |
| 16040_157102900 | *CHRNA7* | dn (1.8 Mb del) | | Cholinergic |
| 14393_5020 | *CAMK2G**; PPP3CB** | dn (477 kb del) | Signalling (downstream neurotransmitter receptors and ion channels) | Cholinergic; Glutamergic |
| 13204_883 | *GNG13* | unk[$] (81 kb del) | | Both |
| 2204_1 | *GNG2**** | mat (101 kb del) | | Both |
| | *MAPK3**** | dn (680 kb del) | | Both |
| 20057_1290002 | *GNG2* | pat (102 kb del) | | Both |
| 20089_1391001 | *MAPK3* | dn (680 kb del) | | Both |
| 20127_4014001 | *MAPK3* | pat (680 kb del) | | Both |
| 5359_4 | *MAPK3* | dn (680 kb del) | | Both |
| 5068_3 | *MAPK3* | dn (680 kb del) | | Both |
| 5451_3 | *PRKACB* | dn (80 kb del) | | Both |
| 5046_3 | *PLCB1* | dn (30 kb del) | | Both |
| 5237_3 | *SHANK2** | dn del | Scaffolds (synaptic organizers) | Glutamergic |
| 6319_3 | *SHANK2** | dn del | | Glutamergic |
| 6325_3 | *SHANK2** | dn del | | Glutamergic |
| 16079_1571066001 | *SHANK3** | dn del | | Glutamergic |
| 2072_1 | *SHANK3** | dn del | | Glutamergic |
| 6130_4 | *SHANK3** | dn del | | Glutamergic |
| 5340_3 | *SHANK1* | mat del | | Glutamergic |

* Same gene in two or more samples (recurrent gene); ** two genes within the same deletion/same sample (multigene); *** two genes in two different events in same sample (double-CNV-hit).
[$] Both parental samples failed array QC.
del, deletion; dn, *de novo*; mat; maternal; pat, paternal; unk, inheritance unknown.

## Tables S14A-S14E. Characterization of genes selected by NETBAG

mmc4.xlsx (Excel workbook)

## Table S15. Functional-group enrichment for DAPPLE results

mmc5.*xlsx* (Excel workbook)

## Table S16. List of 97 high-confidence CNV/SNV genes

| Gene symbol | Type | Gene symbol | Type | Gene symbol | Type | Gene symbol | Type |
|---|---|---|---|---|---|---|---|
| *ABCA1* | CNV | *DTNA** | CNV | *MDM2* | LoF-SNV | *SKI* | CNV |
| *ABL1* | CNV | *DRP2* | LoF XL-SNV in males | *MED13L* | LoF-SNV | *SMARCC2* | LoF-SNV |
| *ANK2* | LoF-SNV | *DST* | LoF-SNV | *NCKAP1* | LoF-SNV | *SMC2* | CNV |
| *ARHGAP32* | CNV | *EPHB2* | LoF-SNV | *NFIA* | LoF-SNV | *SNRPN* | CNV |
| *ARHGDIA* | CNV | *ERCC6* | CNV | *NRXN3* | CNV | *SNX9* | CNV |
| *ATOH1* | CNV | *ESR2* | CNV | *PARD3* | CNV | *SOD2* | CNV |
| *ATP1B1* | LoF-SNV | *ETS1* | CNV | *PARK2* | CNV | *SPAST* | LoF-SNV |
| *BAZ1B* | CNV | *FLNA* | CNV | *PAX5* | LoF-SNV | *SPATA13* | LoF-SNV |
| *BCL11A* | LoF-SNV | *GARNL1* | CNV | *PIK3CB** | CNV | *SREBF1* | CNV |
| *BRWD1* | LoF-SNV | *HAUS7* | LoF XL-SNV in males | *PIR* | LoF XL-SNV in males | *STAU2* | CNV |
| *CACNA1B* | CNV | *IKBKG* | CNV | *PLCB1* | CNV | *SVIL* | LoF-SNV |
| *CAMK2G* | CNV | *IQGAP2* | LoF-SNV | *PPM1D* | LoF-SNV | *SYNCRIP** | CNV |
| *CBX4* | LoF-SNV | *ITGA5* | LoF-SNV | *PPP3CB* | CNV | *SYNJ2* | CNV |
| *CDK2* | CNV | *KAT2B* | CNV | *PRKAB2* | CNV | *TAOK2* | CNV |
| *CDK4* | CNV | *KIAA0232* | LoF-SNV | *PSMB1* | CNV | *TBC1D23* | LoF-SNV |
| *CHD2** | CNV | *KLF13* | CNV | *PSMC2* | CNV | *TBP* | CNV |
| *CNOT3* | LoF-SNV | *LIMK1* | CNV | *PTPRD* | CNV | *TBR1* | LoF-SNV |
| *CRKL* | CNV | *LRP2* | LoF-SNV | *RAB5A* | CNV | *TCF3* | LoF-SNV |
| *CSDE1* | LoF-SNV | *MAPK11* | CNV | *RAC3* | CNV | *TNK2* | CNV |
| *CUBN* | LoF-SNV | *MAPK12* | CNV | *RAP1GDS1* | CNV | *VCP* | LoF-SNV |
| *CUL2* | CNV | *MAPK3* | CNV | *RER1* | CNV | *ZMYND11* | LoF-SNV |
| *CUL3* | LoF-SNV | *MAPK7* | CNV | *RIMS1* | CNV, LoF-SNV | *ZNF292* | LoF-SNV |
| *CYFIP1* | CNV | *MAPK8* | CNV | *RPS6KA3* | LoF-SNV | | |
| *DLGAP2* | CNV | *MAX* | CNV | *SETD2* | LoF-SNV | | |
| *DLL1* | LoF-SNV | *MCF2* | LoF XL-SNV in males | *SH3GL3* | CNV | | |

This table lists 97 CNV or SNV genes present in the DAPPLE network, but not yet implicated in ASD or ID (i.e. not yet in Tables S6A-S6D), considered high-confidence ASD candidate genes. *Genes in section 'Highlighted genes'.

## Table S17A. Listing of CNV calls in affected subjects

mmc6.xlsx (Excel workbook)

## Table S17B. Chromosome abnormalities in parents and control subjects

mmc7.xlsx (Excel workbook)

Chromosome abnormalities in probands are listed in **Table S1C**.

## Table S17C. Experimentally validated CNVs

mmc8.xlsx (Excel workbook)

The 456 validated CNVs in **Table S17C** comprise all CNVs validated experimentally across the 2446 ASD cases, including 315 CNVs confirmed in stage 1 samples and 141 CNVs confirmed in stage 2 samples.

## SUPPLEMENTAL DATAFILES

Listing of separate Excel workbooks.

### Table S8. Phenotypes in ASD subjects with pathogenic CNVs or with selected CNVs of uncertain significance

mmc2.xlsx

### Tables S12A-S12D. GO terms, pathways, and MPO enrichment in affecteds versus control subjects

mmc3.xlsx

### Tables S14A-S14E. Characterization of genes selected by NETBAG

mmc4.xlsx

### Table S15. Functional-group enrichment for DAPPLE results

mmc5.xlsx

### Table S17A. Listing of CNV calls in affected subjects

mmc6.xlsx

### Table S17B. Chromosome abnormalities in parents and control subjects

mmc7.xlsx

### Table S17C. Experimentally validated CNVs

mmc8.xlsx

## SUPPLEMENTAL ACKNOWLEDGEMENTS

Additional support for individual groups of the Autism Genome Project was provided by the US National Institutes of Health (NIH grants: HD34565, HD055751, HD055782, MH06359, MH52708, MH057881, MH061009, MH082121, MH089924, MH094293, MH094382, MH094400, NS043540, NS049261), Assistance Publique-Hôpitaux de Paris (France), Autism Speaks Canada, Autistica, Canada Foundation for Innovation, Canadian Institutes for Advanced Research, The Centre for Applied Genomics (Canada), The Center for Applied Genomics–Children's Hospital of Philadelphia (USA); Deutsche Forschungsgemeinschaft (Germany), Fondation de France, Fondation FondaMental (France), Fondation pour la Recherche Médicale (France), Foundation for Autism Research (USA), Fundação Calouste Gulbenkian (Portugal), Fundação para Ciência e Tecnologia (SFRH/BPD/74739/2010, SFRH/BPD/64281/2009, SFRH/BD/79081/2011, Portugal), the Gulbenkian Programme for Advanced Medical Education and the Portuguese Foundation for Science and Technology (SFRH/BDINT/51549/2011, Portugal), the Hospital for Sick Children Foundation (Canada), INSERM (France), The Margaret Q. Landenberger Foundation (USA), The Medical Charities Research Group (Ireland), The Mindich Child Health and Development Institute (USA), Nancy Lurie Marks Family Foundation (USA); The National Children's Research Centre (Ireland), National Institute of Health Research (UK), NeuroDevNet (Canada), the Ontario government and Ontario Brain Institute (Canada), Our Lady's Children's Hospital, Dublin (Ireland), the Pennsylvania Department of Health (4100047863), the Seaver Foundation (USA), the University of Toronto McLaughlin Centre (Canada), and the Wellcome Trust (core award 090532/Z/09/Z, UK).

Acknowledgement NIH grant No. U24 CA074783 to S. Gallinger: This work was made possible through collaboration and cooperative agreements with the Colon Cancer Family Registry and P.I.s.

This study makes use of data generated by the DECIPHER Consortium. A full list of centres who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

Lead AGP investigators: D. Pinto, E. M. Wijsman, V. J. Vieland, A. M. Vicente, G. D. Schellenberg, M. Pericak-Vance, A. D. Paterson, J. R. Parr, G. Oliveira, J. I. Nurnberger, A. P. Monaco, E. Maestrini, S. M. Klauck, H. Hakonarson, J. L. Haines, D. H. Geschwind, C. M. Freitag, S. E. Folstein, S. Ennis, H. Coon, A. Battaglia, P. Szatmari, J. S. Sutcliffe, J. Hallmayer, M. Gill, E. H. Cook, J. D. Buxbaum, B. Devlin, L. Gallagher, C. Betancur, and S. W. Scherer.

## SUPPLEMENTAL REFERENCES

1. Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., and Vitkup, D. (2011). Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron *70*, 898-907.
2. Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288-289.
3. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res *40*, D109-114.
4. Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. Neuron *74*, 285-299.
5. Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242-245.
6. O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246-250.
7. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237-241.
8. Kielinen, M., Rantala, H., Timonen, E., Linna, S.L., and Moilanen, I. (2004). Associated medical disorders and disabilities in children with autistic disorder: a population-based study. Autism *8*, 49-60.
9. Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247-261.
10. Ascano, M., Jr., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., et al. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature *492*, 382-386.
11. Bayes, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D., Whittle, I.R., Choudhary, J.S., and Grant, S.G. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat Neurosci *14*, 19-21.
12. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet *6*, e1001154.
13. Robinson, P.N., Kohler, S., Bauer, S., Seelow, D., Horn, D., and Mundlos, S. (2008). The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet *83*, 610-615.
14. Eppig, J.T., Blake, J.A., Bult, C.J., Kadin, J.A., and Richardson, J.E. (2012). The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. Nucleic Acids Res *40*, D881-886.
15. BrainSpan. (2012). Atlas of the Developing Human Brain, Technical white paper; Section: Measuring Expression. In. (
16. Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483-489.
17. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.
18. Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J.A., and Karayiorgou, M. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat Genet *44*, 1365-1369.
19. Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Cotsapas, C., and Daly, M.J. (2011). Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet *7*, e1001273.
20. Lage, K., Karlberg, E.O., Storling, Z.M., Olason, P.I., Pedersen, A.G., Rigina, O., Hinsby, A.M., Tumer, Z., Pociot, F., Tommerup, N., et al. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol *25*, 309-316.
21. Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res *1380*, 42-77.

22. Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., et al. (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron *77*, 235-242.
23. Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B., et al. (2013). The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res *41*, D64-69.
24. Khurana, E., Fu, Y., Chen, J., and Gerstein, M. (2013). Interpretation of genomic variants using a unified biological network approach. PLoS Comput Biol *9*, e1002886.
25. Li, M.M., Nimmakayalu, M.A., Mercer, D., Andersson, H.C., and Emanuel, B.S. (2008). Characterization of a cryptic 3.3 Mb deletion in a patient with a "balanced t(15;22) translocation" using high density oligo array CGH and gene expression arrays. Am J Med Genet A *146*, 368-375.
26. Veredice, C., Bianco, F., Contaldo, I., Orteschi, D., Stefanini, M.C., Battaglia, D., Lettori, D., Guzzetta, F., and Zollino, M. (2009). Early onset myoclonic epilepsy and 15q26 microdeletion: observation of the first case. Epilepsia *50*, 1810-1815.
27. Dhamija, R., Breningstall, G., Wong-Kisiel, L., Dolan, M., Hirsch, B., and Wirrell, E. (2011). Microdeletion of chromosome 15q26.1 in a child with intractable generalized epilepsy. Pediatr Neurol *45*, 60-62.
28. Capelli, L.P., Krepischi, A.C., Gurgel-Giannetti, J., Mendes, M.F., Rodrigues, T., Varela, M.C., Koiffmann, C.P., and Rosenberg, C. (2012). Deletion of the RMGA and CHD2 genes in a child with epilepsy and mental deficiency. Eur J Med Genet *55*, 132-134.
29. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet *84*, 524-533.
30. Kaminsky, E.B., Kaul, V., Paschall, J., Church, D.M., Bunke, B., Kunig, D., Moreno-De-Luca, D., Moreno-De-Luca, A., Mulle, J.G., Warren, S.T., et al. (2011). An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med *13*, 777-784.
31. Feenstra, I., Fang, J., Koolen, D.A., Siezen, A., Evans, C., Winter, R.M., Lees, M.M., Riegel, M., de Vries, B.B., Van Ravenswaaij, C.M., et al. (2006). European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA); an online database for rare chromosome abnormalities. Eur J Med Genet *49*, 279-291.
32. Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., et al. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet *40*, 1253-1260.
33. Carvill, G.L., Heavin, S.B., Yendle, S.C., McMahon, J.M., O'Roak, B.J., Cook, J., Khan, A., Dorschner, M.O., Weaver, M., Calvert, S., et al. (2013). Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet *45*, 825-830.
34. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods *7*, 248-249.
35. Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res *13*, 2129-2141.
36. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res *31*, 3812-3814.
37. Bromberg, Y., and Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res *35*, 3823-3835.
38. Schwarz, J.M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods *7*, 575-576.
39. Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res *15*, 901-913.
40. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res *38*, W529-533.
41. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet *380*, 1674-1682.
42. Kulkarni, S., Nagarajan, P., Wall, J., Donovan, D.J., Donell, R.L., Ligon, A.H., Venkatachalam, S., and Quade, B.J. (2008). Disruption of chromodomain helicase DNA binding protein 2 (CHD2) causes scoliosis. Am J Med Genet A *146A*, 1117-1127.
43. Marfella, C.G., Henninger, N., LeBlanc, S.E., Krishnan, N., Garlick, D.S., Holzman, L.B., and Imbalzano, A.N. (2008). A mutation in the mouse Chd2 chromatin remodeling enzyme results in a complex renal phenotype. Kidney Blood Press Res *31*, 421-432.
44. Vissers, L.E., van Ravenswaaij, C.M., Admiraal, R., Hurst, J.A., de Vries, B.B., Janssen, I.M., van der Vliet, W.A., Huys, E.H., de Jong, P.J., Hamel, B.C., et al. (2004). Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. Nat Genet *36*, 955-957.
45. Johansson, M., Rastam, M., Billstedt, E., Danielsson, S., Stromland, K., Miller, M., and Gillberg, C. (2006). Autism spectrum disorders and underlying brain pathology in CHARGE association. Dev Med Child Neurol *48*, 40-50.
46. O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., et al. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science *338*, 1619-1622.

47. Doherty, E.S., and Lacbawan, F. (2013). 2q37 microdeletion syndrome. GeneReviews, http://www.ncbi.nlm.nih.gov/books/NBK1158/.
48. Williams, S.R., Aldred, M.A., Der Kaloustian, V.M., Halal, F., Gowans, G., McLeod, D.R., Zondag, S., Toriello, H.V., Magenis, R.E., and Elsea, S.H. (2010). Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. Am J Hum Genet *87*, 219-228.
49. Villavicencio-Lorini, P., Klopocki, E., Trimborn, M., Koll, R., Mundlos, S., and Horn, D. (2012). Phenotypic variant of Brachydactyly-mental retardation syndrome in a family with an inherited interstitial 2q37.3 microdeletion including HDAC4. Eur J Hum Genet *21*, 743-748.
50. Morris, B., Etoubleau, C., Bourthoumieu, S., Reynaud-Perrine, S., Laroche, C., Lebbar, A., Yardin, C., and Elsea, S.H. (2012). Dose dependent expression of HDAC4 causes variable expressivity in a novel inherited case of brachydactyly mental retardation syndrome. Am J Med Genet A *158A*, 2015-2020.
51. Sando, R., 3rd, Gounko, N., Pieraut, S., Liao, L., Yates, J., 3rd, and Maximov, A. (2012). HDAC4 governs a transcriptional program essential for synaptic plasticity and memory. Cell *151*, 821-834.
52. Kim, M.S., Akhtar, M.W., Adachi, M., Mahgoub, M., Bassel-Duby, R., Kavalali, E.T., Olson, E.N., and Monteggia, L.M. (2012). An essential role for histone deacetylase 4 in synaptic plasticity and memory formation. J Neurosci *32*, 10879-10886.
53. Bienvenu, T., des Portes, V., Saint Martin, A., McDonell, N., Billuart, P., Carrie, A., Vinet, M.C., Couvert, P., Toniolo, D., Ropers, H.H., et al. (1998). Non-specific X-linked semidominant mental retardation by mutations in a Rab GDP-dissociation inhibitor. Hum Mol Genet *7*, 1311-1315.
54. D'Adamo, P., Menegon, A., Lo Nigro, C., Grasso, M., Gulisano, M., Tamanini, F., Bienvenu, T., Gedeon, A.K., Oostra, B., Wu, S.K., et al. (1998). Mutations in GDI1 are responsible for X-linked non-specific mental retardation. Nat Genet *19*, 134-139.
55. Vandewalle, J., Van Esch, H., Govaerts, K., Verbeeck, J., Zweier, C., Madrigal, I., Mila, M., Pijkels, E., Fernandez, I., Kohlhase, J., et al. (2009). Dosage-dependent severity of the phenotype in patients with mental retardation due to a recurrent copy-number gain at Xq28 mediated by an unusual recombination. Am J Hum Genet *85*, 809-822.
56. Clayton-Smith, J., Walters, S., Hobson, E., Burkitt-Wright, E., Smith, R., Toutain, A., Amiel, J., Lyonnet, S., Mansour, S., Fitzpatrick, D., et al. (2009). Xq28 duplication presenting with intestinal and bladder dysfunction and a distinctive facial appearance. Eur J Hum Genet *17*, 434-443.
57. Madrigal, I., Rodriguez-Revenga, L., Badenas, C., Sanchez, A., Martinez, F., Fernandez, I., Fernandez-Burriel, M., and Mila, M. (2007). MLPA as first screening method for the detection of microduplications and microdeletions in patients with X-linked mental retardation. Genet Med *9*, 117-122.
58. Sakai, Y., Shaw, C.A., Dawson, B.C., Dugas, D.V., Al-Mohtaseb, Z., Hill, D.E., and Zoghbi, H.Y. (2011). Protein interactome reveals converging molecular pathways among autism disorders. Sci Transl Med *3*, 86ra49.
59. Bianchi, V., Farisello, P., Baldelli, P., Meskenaite, V., Milanese, M., Vecellio, M., Muhlemann, S., Lipp, H.P., Bonanno, G., Benfenati, F., et al. (2009). Cognitive impairment in Gdi1-deficient mice is associated with altered synaptic vesicle pools and short-term synaptic plasticity, and can be corrected by appropriate learning training. Hum Mol Genet *18*, 105-117.
60. Shuib, S., McMullan, D., Rattenberry, E., Barber, R.M., Rahman, F., Zatyka, M., Chapman, C., Macdonald, F., Latif, F., Davison, V., et al. (2009). Microarray based analysis of 3p25-p26 deletions (3p- syndrome). Am J Med Genet A *149A*, 2099-2105.
61. Gunnarsson, C., and Foyn Bruun, C. (2010). Molecular characterization and clinical features of a patient with an interstitial deletion of 3p25.3-p26.1. Am J Med Genet A *152A*, 3110-3114.
62. Riess, A., Grasshoff, U., Schaferhoff, K., Bonin, M., Riess, O., Horber, V., and Tzschach, A. (2012). Interstitial 3p25.3-p26.1 deletion in a patient with intellectual disability. Am J Med Genet A *158A*, 2587-2590.
63. Peltekova, I.T., Macdonald, A., and Armour, C.M. (2012). Microdeletion on 3p25 in a patient with features of 3p deletion syndrome. Am J Med Genet A *158A*, 2583-2586.
64. Kellogg, G., Sum, J., and Wallerstein, R. (2013). Deletion of 3p25.3 in a patient with intellectual disability and dysmorphic features with further definition of a critical region. Am J Med Genet A *161*, 1405-1408.
65. Ellery, P.M., Ellis, R.J., and Holder, S.E. (2014). Interstitial 3p25 deletion in a patient with features of 3p deletion syndrome: further evidence for the role of SRGAP3 in mental retardation. Clin Dysmorphol *23*, 29-31.
66. Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'Arcy, M., et al. (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res *19*, 1682-1690.
67. Schaefer, G.B., and Lutz, R.E. (2006). Diagnostic yield in the clinical genetic evaluation of autism spectrum disorders. Genet Med *8*, 549-556.
68. Ververi, A., Vargiami, E., Papadopoulou, V., Tryfonas, D., and Zafeiriou, D.I. (2012). Clinical and laboratory data in a sample of Greek children with autism spectrum disorders. J Autism Dev Disord *42*, 1470-1476.
69. Kleefstra, T., van Zelst-Stams, W.A., Nillesen, W.M., Cormier-Daire, V., Houge, G., Foulds, N., van Dooren, M., Willemsen, M.H., Pfundt, R., Turner, A., et al. (2009). Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. J Med Genet *46*, 598-606.
70. Rodriguez-Santiago, B., Malats, N., Rothman, N., Armengol, L., Garcia-Closas, M., Kogevinas, M., Villa, O., Hutchinson, A., Earl, J., Marenne, G., et al. (2010). Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. Am J Hum Genet *87*, 129-138.

71. Molin, A.M., Andrieux, J., Koolen, D.A., Malan, V., Carella, M., Colleaux, L., Cormier-Daire, V., David, A., de Leeuw, N., Delobel, B., et al. (2012). A novel microdeletion syndrome at 3q13.31 characterised by developmental delay, postnatal overgrowth, hypoplastic male genitals, and characteristic facial features. J Med Genet *49*, 104-109.
72. Gimelli, S., Leoni, M., Di Rocco, M., Caridi, G., Porta, S., Cuoco, C., Gimelli, G., and Tassano, E. (2013). A rare 3q13.31 microdeletion including GAP43 and LSAMP genes. Mol Cytogenet *6*, 52.
73. Gil, O.D., Zhang, L., Chen, S., Ren, Y.Q., Pimenta, A., Zanazzi, G., Hillman, D., Levitt, P., and Salzer, J.L. (2002). Complementary expression and heterophilic interactions between IgLON family members neurotrimin and LAMP. J Neurobiol *51*, 190-204.
74. Catania, E.H., Pimenta, A., and Levitt, P. (2008). Genetic deletion of Lsamp causes exaggerated behavioral activation in novel environments. Behav Brain Res *188*, 380-390.
75. Qiu, S., Champagne, D.L., Peters, M., Catania, E.H., Weeber, E.J., Levitt, P., and Pimenta, A.F. (2010). Loss of limbic system-associated membrane protein leads to reduced hippocampal mineralocorticoid receptor expression, impaired synaptic plasticity, and spatial memory deficit. Biol Psychiatry *68*, 197-204.
76. Gijsbers, A.C., den Hollander, N.S., Helderman-van de Enden, A.T., Schuurs-Hoeijmakers, J.H., Vijfhuizen, L., Bijlsma, E.K., van Haeringen, A., Hansson, K.B., Bakker, E., Breuning, M.H., et al. (2011). X-chromosome duplications in males with mental retardation: pathogenic or benign variants? Clin Genet *79*, 71-78.
77. Shimokawa, N., Haglund, K., Holter, S.M., Grabbe, C., Kirkin, V., Koibuchi, N., Schultz, C., Rozman, J., Hoeller, D., Qiu, C.H., et al. (2010). CIN85 regulates dopamine receptor endocytosis and governs behaviour in mice. Embo J *29*, 2421-2432.
78. Kowanetz, K., Husnjak, K., Holler, D., Kowanetz, M., Soubeyran, P., Hirsch, D., Schmidt, M.H., Pavelic, K., De Camilli, P., Randazzo, P.A., et al. (2004). CIN85 associates with multiple effectors controlling intracellular trafficking of epidermal growth factor receptors. Mol Biol Cell *15*, 3155-3166.
79. Kawata, A., Iida, J., Ikeda, M., Sato, Y., Mori, H., Kansaku, A., Sumita, K., Fujiwara, N., Rokukawa, C., Hamano, M., et al. (2006). CIN85 is localized at synapses and forms a complex with S-SCAM via dendrin. J Biochem *139*, 931-939.
80. Gudjonsson, T., Altmeyer, M., Savic, V., Toledo, L., Dinant, C., Grofte, M., Bartkova, J., Poulsen, M., Oka, Y., Bekker-Jensen, S., et al. (2012). TRIP12 and UBR5 suppress spreading of chromatin ubiquitylation at damaged chromosomes. Cell *150*, 697-709.
81. Poulsen, E.G., Steinhauer, C., Lees, M., Lauridsen, A.M., Ellgaard, L., and Hartmann-Petersen, R. (2012). HUWE1 and TRIP12 collaborate in degradation of ubiquitin-fusion proteins and misframed ubiquitin. PLoS One *7*, e50548.
82. Bannai, H., Fukatsu, K., Mizutani, A., Natsume, T., Iemura, S., Ikegami, T., Inoue, T., and Mikoshiba, K. (2004). An RNA-interacting protein, SYNCRIP (heterogeneous nuclear ribonuclear protein Q1/NSAP1) is a component of mRNA granule transported with inositol 1,4,5-trisphosphate receptor type 1 mRNA in neuronal dendrites. J Biol Chem *279*, 53427-53434.
83. Pak, C., Garshasbi, M., Kahrizi, K., Gross, C., Apponi, L.H., Noto, J.J., Kelly, S.M., Leung, S.W., Tzschach, A., Behjati, F., et al. (2011). Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in Drosophila and humans. Proc Natl Acad Sci U S A *108*, 12390-12395.
84. Piton, A., Michaud, J.L., Peng, H., Aradhya, S., Gauthier, J., Mottron, L., Champagne, N., Lafreniere, R.G., Hamdan, F.F., Joober, R., et al. (2008). Mutations in the calcium-related gene IL1RAPL1 are associated with autism. Hum Mol Genet *17*, 3965-3974.
85. Wang, P., Carrion, P., Qiao, Y., Tyson, C., Hrynchak, M., Calli, K., Lopez-Rangel, E., Andrieux, J., Delobel, B., Duban-Bedu, B., et al. (2013). Genotype-phenotype analysis of 18q12.1-q12.2 copy number variation in autism. Eur J Med Genet *(Epub ahead of print)*.
86. Ichida, F., Tsubata, S., Bowles, K.R., Haneda, N., Uese, K., Miyawaki, T., Dreyer, W.J., Messina, J., Li, H., Bowles, N.E., et al. (2001). Novel gene mutations in patients with left ventricular noncompaction or Barth syndrome. Circulation *103*, 1256-1263.
87. Hinton, V.J., Cyrulnik, S.E., Fee, R.J., Batchelder, A., Kiefel, J.M., Goldstein, E.M., Kaufmann, P., and De Vivo, D.C. (2009). Association of autistic spectrum disorders with dystrophinopathies. Pediatr Neurol *41*, 339-346.
88. Young, H.K., Barton, B.A., Waisbren, S., Portales Dale, L., Ryan, M.M., Webster, R.I., and North, K.N. (2008). Cognitive and psychological profile of males with Becker muscular dystrophy. J Child Neurol *23*, 155-162.
89. Pilgram, G.S., Potikanond, S., Baines, R.A., Fradkin, L.G., and Noordermeer, J.N. (2010). The roles of the dystrophin-associated glycoprotein complex at the synapse. Mol Neurobiol *41*, 1-21.
90. Grady, R.M., Grange, R.W., Lau, K.S., Maimone, M.M., Nichol, M.C., Stull, J.T., and Sanes, J.R. (1999). Role for alpha-dystrobrevin in the pathogenesis of dystrophin-dependent muscular dystrophies. Nat Cell Biol *1*, 215-220.
91. Grady, R.M., Wozniak, D.F., Ohlemiller, K.K., and Sanes, J.R. (2006). Cerebellar synaptic defects and abnormal motor behavior in mice lacking alpha- and beta-dystrobrevin. J Neurosci *26*, 2841-2851.
92. Garbett, K., Ebert, P.J., Mitchell, A., Lintas, C., Manzi, B., Mirnics, K., and Persico, A.M. (2008). Immune transcriptome alterations in the temporal cortex of subjects with autism. Neurobiol Dis *30*, 303-311.
93. Iwamoto, K., Kakiuchi, C., Bundo, M., Ikeda, K., and Kato, T. (2004). Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. Mol Psychiatry *9*, 406-416.
94. Metzinger, L., Blake, D.J., Squier, M.V., Anderson, L.V., Deconinck, A.E., Nawrotzki, R., Hilton-Jones, D., and Davies, K.E. (1997). Dystrobrevin deficiency at the sarcolemma of patients with muscular dystrophy. Hum Mol Genet *6*, 1185-1191.
95. Brauer, A.U., Savaskan, N.E., Kuhn, H., Prehn, S., Ninnemann, O., and Nitsch, R. (2003). A new phospholipid phosphatase, PRG-1, is involved in axon growth and regenerative sprouting. Nat Neurosci *6*, 572-578.
96. Willemsen, M.H., Valles, A., Kirkels, L.A., Mastebroek, M., Olde Loohuis, N., Kos, A., Wissink-Lindhout, W.M., de Brouwer, A.P., Nillesen, W.M., Pfundt, R., et al. (2011). Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. J Med Genet *48*, 810-818.

97. Carter, M.T., Nikkel, S.M., Fernandez, B.A., Marshall, C.R., Noor, A., Lionel, A.C., Prasad, A., Pinto, D., Joseph-George, A.M., Noakes, C., et al. (2011). Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder. Clin Genet *80*, 435-443.
98. Smrt, R.D., Szulwach, K.E., Pfeiffer, R.L., Li, X., Guo, W., Pathania, M., Teng, Z.Q., Luo, Y., Peng, J., Bordey, A., et al. (2010). MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. Stem Cells *28*, 1060-1070.
99. Szulwach, K.E., Li, X., Smrt, R.D., Li, Y., Luo, Y., Lin, L., Santistevan, N.J., Li, W., Zhao, X., and Jin, P. (2010). Cross talk between microRNA and epigenetic regulation in adult neurogenesis. J Cell Biol *189*, 127-141.
100. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. (2011). Genome-wide association study identifies five new schizophrenia loci. Nat Genet *43*, 969-976.
101. Cusco, I., Medrano, A., Gener, B., Vilardell, M., Gallastegui, F., Villa, O., Gonzalez, E., Rodriguez-Santiago, B., Vilella, E., Del Campo, M., et al. (2009). Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. Hum Mol Genet *18*, 1795-1804.
102. Gross, C., Nakamoto, M., Yao, X., Chan, C.B., Yim, S.Y., Ye, K., Warren, S.T., and Bassell, G.J. (2010). Excess phosphoinositide 3-kinase subunit synthesis and activity as a novel therapeutic target in fragile X syndrome. J Neurosci *30*, 10624-10638.
103. Sharma, A., Hoeffer, C.A., Takayasu, Y., Miyawaki, T., McBride, S.M., Klann, E., and Zukin, R.S. (2010). Dysregulation of mTOR signaling in fragile X syndrome. J Neurosci *30*, 694-702.
104. Lang, B., Alrahbeni, T.M., Clair, D.S., Blackwood, D.H., McCaig, C.D., and Shen, S. (2012). HDAC9 is implicated in schizophrenia and expressed specifically in post-mitotic neurons but not in adult neural stem cells. Am J Stem Cells *1*, 31-41.
105. Bachmann-Gagescu, R., Mefford, H.C., Cowan, C., Glew, G.M., Hing, A.V., Wallace, S., Bader, P.I., Hamati, A., Reitnauer, P.J., Smith, R., et al. (2010). Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. Genet Med *12*, 641-647.
106. Barge-Schaapveld, D.Q., Maas, S.M., Polstra, A., Knegt, L.C., and Hennekam, R.C. (2011). The atypical 16p11.2 deletion: a not so atypical microdeletion syndrome? Am J Med Genet A *155A*, 1066-1072.
107. Bijlsma, E.K., Gijsbers, A.C., Schuurs-Hoeijmakers, J.H., van Haeringen, A., Fransen van de Putte, D.E., Anderlid, B.M., Lundin, J., Lapunzina, P., Perez Jurado, L.A., Delle Chiaie, B., et al. (2009). Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. Eur J Med Genet *52*, 77-87.
108. Bochukova, E.G., Huang, N., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K., Saeed, S., Hamilton-Shield, J., Clayton-Smith, J., O'Rahilly, S., et al. (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. Nature *463*, 666-670.
109. Sampson, M.G., Coughlin, C.R., 2nd, Kaplan, P., Conlin, L.K., Meyers, K.E., Zackai, E.H., Spinner, N.B., and Copelovitch, L. (2010). Evidence for a recurrent microdeletion at chromosome 16p11.2 associated with congenital anomalies of the kidney and urinary tract (CAKUT) and Hirschsprung disease. Am J Med Genet A *152A*, 2618-2622.
110. Tabet, A.C., Pilorge, M., Delorme, R., Amsellem, F., Pinard, J.M., Leboyer, M., Verloes, A., Benzacken, B., and Betancur, C. (2012). Autism multiplex family with 16p11.2p12.2 microduplication syndrome in monozygotic twins and distal 16p11.2 deletion in their brother. Eur J Hum Genet *20*, 540-546.
111. Moreno-De-Luca, D., Sanders, S.J., Willsey, A.J., Mulle, J.G., Lowe, J.K., Geschwind, D.H., State, M.W., Martin, C.L., and Ledbetter, D.H. (2013). Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. Mol Psychiatry *18*, 1090-1095.
112. Guha, S., Rees, E., Darvasi, A., Ivanov, D., Ikeda, M., Bergen, S.E., Magnusson, P.K., Cormican, P., Morris, D., Gill, M., et al. (2013). Implication of a rare deletion at distal 16p11.2 in schizophrenia. JAMA Psychiatry *70*, 253-260.
113. Doche, M.E., Bochukova, E.G., Su, H.W., Pearce, L.R., Keogh, J.M., Henning, E., Cline, J.M., Saeed, S., Dale, A., Cheetham, T., et al. (2012). Human SH2B1 mutations are associated with maladaptive behaviors and obesity. J Clin Invest *122*, 4732-4736.
114. Zhang, Y., Zhu, W., Wang, Y.G., Liu, X.J., Jiao, L., Liu, X., Zhang, Z.H., Lu, C.L., and He, C. (2006). Interaction of SH2-Bbeta with RET is involved in signaling of GDNF-induced neurite outgrowth. J Cell Sci *119*, 1666-1676.
115. Maures, T.J., Chen, L., and Carter-Su, C. (2009). Nucleocytoplasmic shuttling of the adapter protein SH2B1beta (SH2-Bbeta) is required for nerve growth factor (NGF)-dependent neurite outgrowth and enhancement of expression of a subset of NGF-responsive genes. Mol Endocrinol *23*, 1077-1091.
116. Ren, D., Zhou, Y., Morris, D., Li, M., Li, Z., and Rui, L. (2007). Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. J Clin Invest *117*, 397-406.
117. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature *466*, 368-372.
118. Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., and Ragoussis, J. (2007). QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res *35*, 2013-2025.
119. Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A.C., Thiruvahindrapuram, B., Macdonald, J.R., Mills, R., et al. (2011). Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat Biotechnol *29*, 512-520.
120. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H., and Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res *17*, 1665-1674.

121. Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., Sykes, N., Pagnamenta, A.T., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet *19*, 4072-4082.
122. Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bolte, S., Bolton, P.F., Bourgeron, T., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. Hum Mol Genet *21*, 4781-4792.
123. Klei, L., Kent, B.P., Melhem, N.M., Devlin, B., and Roeder, K. (2011). GemTools: a fast and efficient approach to estimating genetic ancestry. Statistics Applications *http://arxiv.org/PS_cache/arxiv/pdf/1104/1104.1162v1.pdf*.
124. Bierut, L.J., Agrawal, A., Bucholz, K.K., Doheny, K.F., Laurie, C., Pugh, E., Fisher, S., Fox, L., Howells, W., Bertelsen, S., et al. (2010). A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A *107*, 5082-5087.
125. Newcomb, P.A., Baron, J., Cotterchio, M., Gallinger, S., Grove, J., Haile, R., Hall, D., Hopper, J.L., Jass, J., Le Marchand, L., et al. (2007). Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol Biomarkers Prev *16*, 2331-2343.
126. Figueiredo, J.C., Lewinger, J.P., Song, C., Campbell, P.T., Conti, D.V., Edlund, C.K., Duggan, D.J., Rangrej, J., Lemire, M., Hudson, T., et al. (2011). Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. Cancer Epidemiol Biomarkers Prev *20*, 758-766.
127. Fox, C.S., Liu, Y., White, C.C., Feitosa, M., Smith, A.V., Heard-Costa, N., Lohman, K., Johnson, A.D., Foster, M.C., Greenawalt, D.M., et al. (2012). Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. PLoS Genet *8*, e1002695.
128. Griswold, A.J., Ma, D., Cukier, H.N., Nations, L.D., Schmidt, M.A., Chung, R.H., Jaworski, J.M., Salyakina, D., Konidari, I., Whitehead, P.L., et al. (2012). Evaluation of copy number variations reveals novel candidate genes in autism spectrum disorder-associated pathways. Hum Mol Genet *21*, 3513-3523.
129. Leblond, C.S., Heinrich, J., Delorme, R., Proepper, C., Betancur, C., Huguet, G., Konyukh, M., Chaste, P., Ey, E., Rastam, M., et al. (2012). Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. PLoS Genet *8*, e1002521.
130. Moessner, R., Marshall, C.R., Sutcliffe, J.S., Skaug, J., Pinto, D., Vincent, J., Zwaigenbaum, L., Fernandez, B., Roberts, W., Szatmari, P., et al. (2007). Contribution of SHANK3 mutations to autism spectrum disorder. Am J Hum Genet *81*, 1289-1297.
131. Hehir-Kwa, J.Y., Rodriguez-Santiago, B., Vissers, L.E., de Leeuw, N., Pfundt, R., Buitelaar, J.K., Perez-Jurado, L.A., and Veltman, J.A. (2011). De novo copy number variants associated with intellectual disability have a paternal origin and age bias. J Med Genet *48*, 776-778.
132. Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron *70*, 863-885.
133. Froyen, G., Belet, S., Martinez, F., Santos-Reboucas, C.B., Declercq, M., Verbeeck, J., Donckers, L., Berland, S., Mayo, S., Rosello, M., et al. (2012). Copy-number gains of HUWE1 due to replication- and recombination-based rearrangements. Am J Hum Genet *91*, 252-264.
134. Berkel, S., Marshall, C.R., Weiss, B., Howe, J., Roeth, R., Moog, U., Endris, V., Roberts, W., Szatmari, P., Pinto, D., et al. (2010). Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet *42*, 489-491.
135. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., et al. (2007). Strong association of de novo copy number mutations with autism. Science *316*, 445-449.
136. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet *82*, 477-488.
137. Noor, A., Whibley, A., Marshall, C.R., Gianakopoulos, P.J., Piton, A., Carson, A.R., Orlic-Milacic, M., Lionel, A.C., Sato, D., Pinto, D., et al. (2010). Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. Sci Transl Med *2*, 49ra68.
138. Pagnamenta, A.T., Holt, R., Yusuf, M., Pinto, D., Wing, K., Betancur, C., Scherer, S.W., Volpi, E.V., and Monaco, A.P. (2011). A family with autism and rare copy number variants disrupting the Duchenne/Becker muscular dystrophy gene DMD and TRPM3. J Neurodev Disord *3*, 124-131.
139. Stankiewicz, P., Kulkarni, S., Dharmadhikari, A.V., Sampath, S., Bhatt, S.S., Shaikh, T.H., Xia, Z., Pursley, A.N., Cooper, M.L., Shinawi, M., et al. (2012). Recurrent deletions and reciprocal duplications of 10q11.21q11.23 including CHAT and SLC18A3 are likely mediated by complex low-copy repeats. Hum Mutat *33*, 165-179.
140. Luo, R., Sanders, S.J., Tian, Y., Voineagu, I., Huang, N., Chu, S.H., Klei, L., Cai, C., Ou, J., Lowe, J.K., et al. (2012). Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. Am J Hum Genet *91*, 38-55.

# Discussion

Soixante-dix ans après les premières descriptions de patients atteints d'autisme, la recherche dans le domaine de la génétique a permis d'identifier une cause chez environ 25 % des individus atteints. L'ensemble de ces causes identifiées supporte une vision de l'autisme dans laquelle de nombreux variants génétiques rares voire uniques à effet fort, jouent un rôle étiologique et expliquent, collectivement, une proportion importante de cas. Des CNVs récurrents associés à des phénomènes de pénétrance incomplète et/ou d'expressivité variable ont également été impliqués en tant que facteurs pathogènes ou de susceptibilité (**Figure 28**). Même si ce rendement étiologique représente un progrès énorme, il implique qu'une majorité des cas reste non élucidée pour le moment. L'extrême hétérogénéité clinique et génétique ainsi que la complexité de l'architecture génétique des TSA restent un frein dans l'identification des facteurs impliqués ainsi que dans la compréhension des mécanismes physiopathologiques sous-jacents.

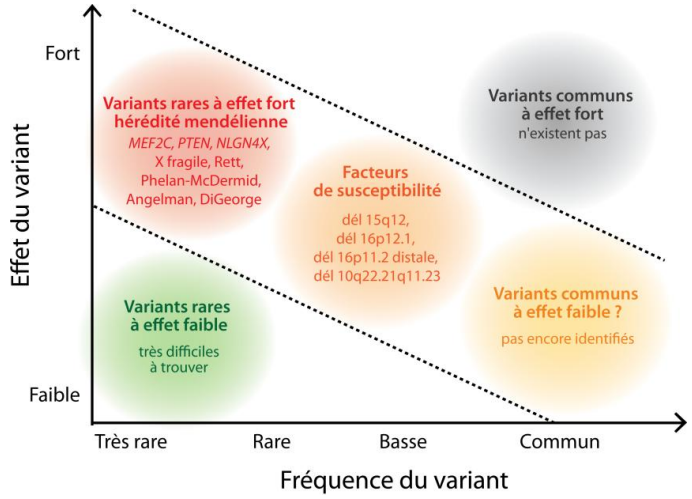



**Figure 28. Etiologie génétique des TSA : relation entre la fréquence allélique et la taille de l'effet**

Aujourd'hui, les causes génétiques connues des TSA correspondent en grande majorité à des variants rares voire uniques à effet fort, ainsi qu'à quelques CNVs à pénétrance incomplète et expressivité variable ayant un rôle de facteurs de risque. Les études d'association pangénomiques réalisées jusqu'à maintenant ont exclu l'existence de variants communs à effet fort et pour l'instant, aucun variant commun à effet faible n'a été identifié (figure adaptée de Manolio et coll., 2009).

# 1. Complexité de l'architecture génétique des TSA

## 1.1 Multitude de gènes impliqués

Le nombre de gènes impliqués dans les TSA n'a cessé d'augmenter ces dernières années. Toutes les mutations dans ces gènes sont rares et ne permettent d'expliquer qu'un nombre limité de cas. Ceci implique qu'il reste de nombreux gènes à découvrir dans l'étiologie de l'autisme. Plusieurs études ont cherché à déterminer le nombre total de gènes impliqués dans les TSA. Sanders et coll. (2011) ont estimé le nombre de régions impliquées dans les TSA à un intervalle de 130-234 en se basant sur les CNVs *de novo* identifiés dans leur cohorte de 1 124 familles à cas sporadiques de la cohorte *Simons*. En utilisant la même approche analytique sur leurs données de SNVs *de novo*, ils estiment ensuite le nombre de gènes à 1 034 (Sanders et coll., 2012a). De leur côté, O'Roak et coll. (2012b) ont estimé que si 20 à 30 % de mutations *de novo* observées chez leurs patients étaient réellement pathogéniques, le nombre de loci impliqués dans les TSA serait compris entre 384 et 821. Ces nombres ne correspondent qu'à des estimations statistiques et dépendent largement de la fraction réelle de mutations *de novo* pathogènes et des échantillons étudiés. De plus, les intervalles de confiance sont très larges et aucune notion de répartition du risque entre les différents loci n'est prise en compte. Cependant, ils donnent un ordre de grandeur du nombre de loci attendus et confirment le haut degré d'hétérogénéité observé dans les TSA.

## 1.2 Modes de transmission rares

Pour les familles à cas multiples, on suppose *a priori* qu'un même variant sera trouvé chez les atteints et que l'étiologie sera liée à des facteurs génétiques hérités. Cependant, il arrive parfois que des frères et sœurs atteints d'une même famille soient porteurs d'anomalies génétiques distinctes, apparues indépendamment et qui peuvent être *de novo*. Ce phénomène d'**hétérogénéité génétique intrafamiliale** a déjà été observé parmi des familles à cas multiples avec TSA et/ou déficience intellectuelle (Mandel et Chelly, 2004 ; Weiss et coll., 2008 ; Pinto et coll., 2010 ; Tabet et coll., 2012) et nous en avons plusieurs cas dans notre cohorte de patients.

Par ailleurs, des mutations présentes dans la lignée germinale de l'un des parents peuvent être responsables de la détection d'une mutation identique chez plusieurs enfants d'une même famille apparaissant pourtant absente des parents. Ce mécanisme appelé **mosaïcisme germinal** semble moins rare qu'attendu et a d'ailleurs permis d'expliquer l'étiologie de plusieurs familles à cas multiples avec TSA. A titre d'exemples on peut citer une famille avec un mosaïcisme germinal chez le père où les deux sœurs atteintes portent une délétion identique du gène *NRXN1* (Szatmari et coll., 2007) ; quatre familles, chacune avec deux enfants atteints, portant des délétions ou des mutations du gène *SHANK3* (Durand et coll., 2007 ; Moessner et coll., 2007 ; Gauthier et coll., 2010a ; Verhoeven et coll., 2012) et deux frères atteints avec la même mutation du gène *NLGN4X* due à un mosaïcisme maternel (Zhang et coll., 2009). Ce mécanisme est probablement sous-estimé puisqu'il ne peut être observé que dans les familles à cas multiples et passe inaperçu dans les familles à cas sporadiques. Le **mosaïcisme somatique de faible degré** chez l'un des parents peut également parfois expliquer la récurrence d'une même mutation chez plusieurs enfants d'une même famille, apparemment absente des parents (Depienne et coll., 2006 ; Jongmans et coll., 2008).

## 1.3 Expressivité variable et pénétrance incomplète

De nombreux gènes et loci ont été impliqués dans des troubles neuropsychiatriques et neurologiques distincts incluant les TSA et la déficience intellectuelle. Toutes les causes connues de TSA sont également des causes de déficience intellectuelle ; et le chevauchement étiologique entre ces deux troubles et l'épilepsie est également remarquable. Le chevauchement entre les facteurs impliqués dans ces trois troubles et ceux conférant un risque dans la schizophrénie est plus modeste, ce qui pourrait être expliqué par le fait que seule une partie de la susceptibilité à la schizophrénie est d'origine neurodéveloppementale (Hoischen et coll., 2014). En outre, plusieurs CNVs pathogènes exercent des effets pléiotropiques et sont associés à une grande variété de manifestations phénotypiques incluant les TSA, la déficience intellectuelle, l'épilepsie mais aussi la schizophrénie, les TDAH et les troubles du langage. Bien que retrouvés dans plusieurs troubles neurodéveloppementaux, certains CNVs semblent associés préférentiellement à l'un d'entre eux (**Figure 29**) (Moreno-De-Luca et coll., 2014). Ceci suggère que ces variants affectent des **processus communs** impliqués dans le développement du cerveau mais que chacun altère des **mécanismes neurologiques spécifiques**.

Par ailleurs, des mutations dans des gènes différents peuvent donner des phénotypes similaires. Ce phénomène, appelé **phénocopie**, est aussi observé dans les TSA. Par exemple, l'altération de plusieurs gènes peut donner des phénotypes similaires à ceux retrouvés chez des patients atteints du syndrome d'Angelman, regroupés sous le terme 'Angelman-like' (Tan et coll., 2014) ; de même, des gènes autres que *MECP2* sont impliqués dans des phénotypes ressemblant au syndrome de Rett, dits 'Rett-like' (Mari et coll., 2005 ; Ariani et coll., 2008 ; Zweier et coll., 2010).

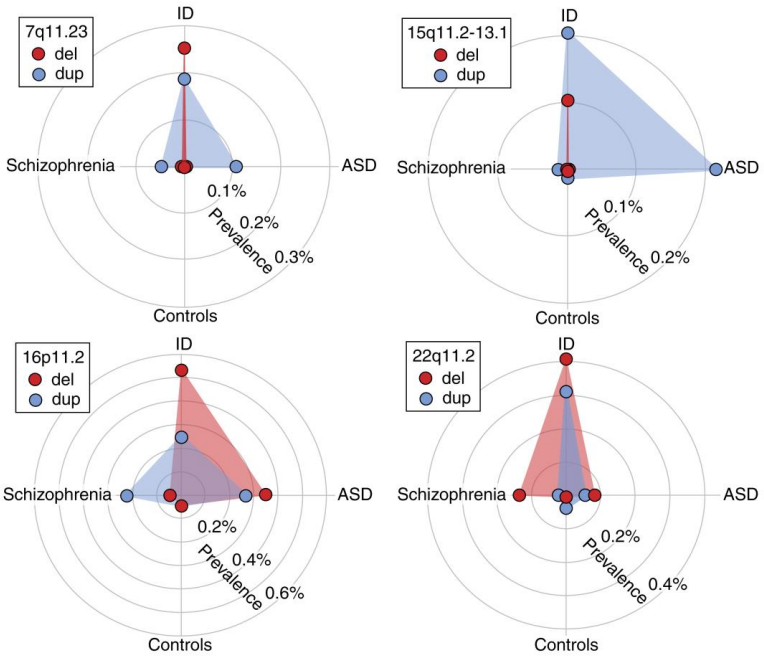



**Figure 29. Comparaison de la prévalence de quatre CNVs dans des troubles neuropsychiatriques et chez des témoins**
La prévalence de CNVs pathogènes à quatre loci est montrée pour les TSA (*ASD*), la déficience intellectuelle (*ID*), la schizophrénie (*schizophrenia*) et chez les témoins (*controls*). Certains d'entres eux confèrent un risque pour les trois troubles neurodéveloppementaux étudiés (comme les délétions 22q11.2 et les duplications 7q11.23), certains sont rarement observés dans la schizophrénie (délétions 16p11.2, duplications 22q11.2 et duplications 15q11.2-q13.1) alors que d'autres semblent être observés majoritairement dans la déficience intellectuelle (syndrome de Williams-Beuren/délétions 7q11.23 et délétions 15q11.2-q13.1) (figure tirée de Moreno-De-Luca et coll., 2014).

Plusieurs mécanismes ont été proposés pour expliquer les phénomènes de pénétrance incomplète et d'expressivité variable associés à certaines mutations :

- **Modèle oligogénique**. Selon ce modèle, c'est la présence d'une seconde mutation (en *cis* ou en *trans*) qui va moduler le phénotype associé à la première (Girirajan et coll., 2012). Si les différentes mutations touchent des gènes impliqués dans les mêmes voies cellulaires, les effets pourront être additifs ou épistatiques (Veltman et Brunner, 2010) (Figure 30).

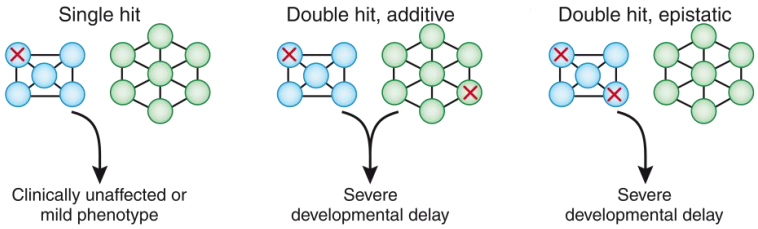



**Figure 30. Modèle oligogénique**
Une mutation (représentée par une croix rouge) entraîne l'apparition d'un phénotype léger ou n'a pas d'effet. C'est l'apparition d'une seconde mutation qui potentialisera la première et sera responsable du phénotype observé. Si les protéines mutées n'appartiennent pas aux mêmes voies biologiques, les effets observés seront généralement additifs ; si elles appartiennent à la même voie, ils seront épistatiques (figure tirée de Veltman et Brunner, 2010).

- **Révélation d'un allèle récessif**. Des délétions peuvent démasquer la présence d'une mutation récessive sur la copie homologue d'un locus donné, modulant ainsi le phénotype associé à ces délétions. Quelques rares cas de ce phénomène ont été décrits chez des patients atteints d'un

syndrome de microdélétion ayant une présentation atypique et plus sévère. La détection d'une mutation sur le second allèle d'un gène récessif appartenant à la région délétée a permis d'expliquer la variation phénotypique observée chez eux (Flipsen-ten Berg et coll., 2007 ; Bisgaard et coll., 2009 ; McDonald-McGinn et coll., 2013 ; Pebrel-Richard et coll., 2013).

- Facteurs **génétiques modificateurs**. Une autre hypothèse pouvant expliquer une variabilité phénotypique chez des patients porteurs d'une même mutation serait l'influence du contexte génétique dans lequel celle-ci s'exprime. Même si ce mécanisme a souvent été invoqué, l'identification de gènes modificateurs est difficile et dans le cas des TSA, aucun d'entre eux n'a pu être mis en évidence de façon claire. Un exemple explorant la contribution de facteurs génétiques modificateurs est celui de l'étude de deux paires de sœurs avec la même mutation dans *MECP2* et des phénotypes du syndrome de Rett discordants en termes de sévérité (Grillo et coll., 2013). Dans chaque paire, l'une des sœurs a un tableau clinique typique du syndrome de Rett avec une présentation très sévère alors que l'autre sœur présente un phénotype beaucoup plus léger, appelé 'variant Zappella', avec un langage et une capacité de marche préservés. Après avoir vérifié que le statut de l'inactivation de l'X était non biaisé chez toutes les filles atteintes, les auteurs ont étudié leurs exomes. Des variants additionnels impliquant des voies de signalisation différentes ont été détectés chez les filles avec un Rett classique et chez celles présentant le phénotype Zappella. En particulier, les sœurs avec un phénotype plus sévère présentaient des variants dans des gènes impliqués dans le stress oxydatif, le développement musculaire ainsi que des gènes impliqués dans la déficience intellectuelle et les TSA. A l'inverse, les filles moins atteintes présentent des variants impliqués dans la régulation de la fonction immunitaire qui, selon les auteurs pourraient jouer un rôle favorable dans les manifestations cliniques chez elles. Cette étude a montré le potentiel de ce type d'approche dans la recherche de facteurs génétiques modificateurs. Cependant, aucun variant n'a été identifié spécifiquement et ces résultats doivent être répliqués dans des échantillons plus larges.

  De façon intéressante, des **effets du contexte génétique** sur l'expression de certains CNVs ont été mis en évidence dans des modèles animaux. La souris modèle du syndrome de DiGeorge (délétion 22q11), présentant une haploinsuffisance du gène *TBX1*, codant un facteur de transcription, récapitule plusieurs caractéristiques phénotypiques associées au syndrome (notamment les malformations cardiaques et faciales) avec une expressivité variable et une pénétrance incomplète, également observées chez les patients. Une série de croisements a été réalisée entre cette souris et des souris mutées dans des gènes interagissant avec *TBX1* dans diverses voies biologiques et a montré une augmentation ou une diminution de la pénétrance de certaines manifestations dans leur descendance (Weischenfeldt et coll., 2013). Cette approche élégante a montré qu'il existe des mécanismes de modulation de l'expression des gènes qui pourraient être transposés chez l'humain pour expliquer l'observation de phénotypes variables chez certains patients portant le même CNV.
- **Empreinte génétique** et **facteurs épigénétiques**. La plupart des processus épigénétiques décrits concernent les modifications de l'ADN et le remodelage de la chromatine et en particulier les modifications des histones et la méthylation de l'ADN (pour revues voir Kouzarides, 2007 ; Cedar et Bergman, 2009). L'analyse de la méthylation du génome chez des paires de jumeaux concordants ou discordants pour les TSA a montré des différences de niveau de méthylation dans certaines régions génomiques entre les différents groupes de jumeaux et en particulier au sein des paires de jumeaux discordants (Wong et coll., 2014). Cette observation suggère un rôle potentiel du méthylome et plus généralement des processus épigénétiques dans l'étiologie des TSA. D'ailleurs, de nombreux gènes impliqués dans la régulation de la chromatine et dans les modifications des histones ont été impliqués dans l'autisme et la déficience intellectuelle (pour revue voir Ronan et coll., 2013). Le rôle du mécanisme d'empreinte a été clairement établi dans certaines régions génomiques comme la région 15q11-q13 impliquée dans les syndromes d'Angelman et Prader-Willi. Il n'est pas exclu que d'autres

gènes soumis à empreinte répartis sur l'ensemble du génome modulent le phénotype associé à certaines altérations génétiques.

- **Mosaïcisme somatique**. Si la mutation survient au cours de l'embryogenèse de l'individu, elle ne sera pas observée dans toutes les cellules. Plus le stade de développement où elle apparaît est précoce, plus le nombre de cellules mutées sera important. Ainsi, selon la proportion de cellules mutées et normales, le phénotype pourra varier entre individus porteurs d'une mutation identique. Notamment, des phénomènes de mosaïcisme somatique dans le cerveau concernant aussi bien des rétrotransposons L1 que des CNVs, pourraient jouer un rôle jusqu'alors inexploré dans les pathologies neurodéveloppementales (McConnell et coll., 2013 ; Richardson et coll., 2014).
- Facteurs **environnementaux**. Même si aucune interaction directe entre gènes et environnement n'a clairement été démontrée dans les TSA, ces mécanismes pourraient potentiellement contribuer aux différences de phénotype entre les individus atteints.
- **Effets stochastiques**. Le développement cérébral est incroyablement complexe et intrinsèquement variable, même pour un phénotype donné dans un environnement donné. La nature probabiliste des processus neurodéveloppementaux au niveau cellulaire peut se traduire par des phénotypes divers au niveau de l'individu. En effet, les processus stochastiques sont inhérents au développement des organismes. Dans plusieurs modèles, surtout dans des organismes simples, il a été montré que des fluctuations de l'expression génique dues au hasard contribuaient à l'observation de phénotypes variables et de pénétrance incomplète même dans des contextes génétiques et environnementaux fixés (Losick et Desplan, 2008 ; Raj et van Oudenaarden, 2008 ; Eldar et Elowitz, 2010 ; Raj et coll., 2010).

## 1.4 Différences de susceptibilité aux altérations génétiques en fonction du sexe

***Vulnérabilité liée à l'X***

Les mutations ou réarrangements touchant des gènes du chromosome X suivent la plupart du temps un mode de transmission récessif et n'auront d'impact que chez les garçons qui ne possèdent qu'une seule copie de l'X. Même lorsque la transmission se fait selon un mode dominant, les filles peuvent tout de même se protéger par le phénomène de **biais de l'inactivation du X**. Dans des conditions physiologiques normales, l'inactivation de l'une des deux copies du chromosome X se fait au hasard dans chaque cellule et la fille est donc mosaïque pour chacun des chromosomes X paternel et maternel dans des proportions égales. Parfois, l'inactivation d'une des deux copies se fait préférentiellement et notamment lorsqu'un gène ou locus majeur est altéré sur l'une des copies, cette dernière sera sélectivement inactivée dans la majorité des cellules. A l'inverse, dans de rares cas, une maladie récessive liée à l'X peut se manifester chez une fille à cause de l'inactivation préférentielle de l'allèle normal par erreur.

Bien que les filles soient protégées la plupart du temps par leurs deux copies du chromosome X, cette diploïdie peut toutefois être un facteur de risque dans certains cas très particuliers. C'est le cas des mutations du gène *PCDH19* (protocadhérine 19) qui code une protéine d'adhésion cellulaire, responsable d'une forme d'épilepsie associée à la déficience intellectuelle et aux TSA, pour laquelle l'atteinte est limitée aux filles et dont les pères, porteurs de la mutation, ne sont pas affectés. Afin d'expliquer ce phénomène, le modèle d'**interférence cellulaire** dû à un mosaïcisme cellulaire a été proposé. Ainsi, une mutation hétérozygote de ce gène avec l'inactivation par hasard de l'un des deux chromosomes X chez les filles aura pour conséquence la présence de deux populations distinctes de cellules PCDH19-positives et PCDH19-négatives qui n'interagissent pas entre elles. Au contraire, une mutation chez un garçon entraînera l'absence totale de la protéine, non délétère et asymptomatique (Depienne et LeGuern, 2012) (**Figure 31**).

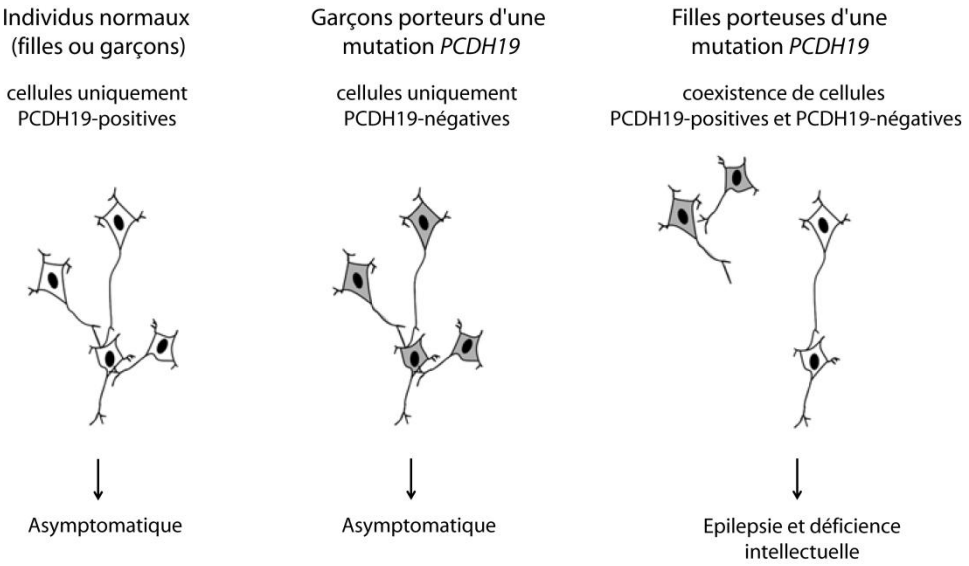



**Figure 31. Phénomène d'interférence cellulaire associé aux mutations *PCDH19***

Chez les individus normaux, il existe une population homogène de cellules PCDH19-positives qui permettent aux neurones de créer des réseaux normaux. Chez les garçons avec une perte hémizygote du gène *PCDH19*, la capacité des neurones à former des réseaux est préservée alors que chez les filles présentant une mutation hétérozygote ou chez les garçons présentant la mutation en mosaïque, la coexistence des deux populations de cellules (PCDH19-positives et PCDH19-négatives) va entraîner des défauts d'interaction cellulaire et les réseaux neuronaux formés seront anormaux (figure tirée de Depienne et coll., 2009a).

Bien que ce phénomène n'ait été décrit que dans le cas de ce gène, on peut imaginer qu'il pourrait potentiellement contribuer plus largement au risque de développer un TSA ou d'autres troubles neurodéveloppementaux.

***Effet protecteur sexe-spécifique***

Une hypothèse pour expliquer la plus forte prévalence des TSA chez les garçons serait la présence de **facteurs protecteurs chez les filles**. Dès la première étude pangénomique de CNVs dans les TSA, Sebat et coll. (2007) ont remarqué une différence entre les filles et les garçons. En effet, le sexe ratio dans leur échantillon était de 5 garçons pour une fille mais lorsque les auteurs ont considéré uniquement les porteurs de CNVs *de novo*, ce ratio était réduit à 1,8 garçons pour une fille, suggérant que les CNVs *de novo* détectés par leur méthode étaient très pénétrants et contribuaient au risque chez les garçons et les filles de façon similaire. Comme mentionné précédemment, d'autres études de CNVs dans les TSA ont rapporté une proportion plus importante de filles présentant des CNVs *de novo* par rapport aux garçons ou présentant des CNVs *de novo* plus riches en gènes (Levy et coll., 2011 ; Sanders et coll., 2011). De plus, un enrichissement en SNVs *de novo* perte de fonction chez les filles, non retrouvé chez les garçons a également été rapporté (Iossifov et coll., 2012). Ces résultats confortent l'existence d'un 'fardeau génétique' plus important chez les filles atteintes par rapport aux garçons.

Une étude plus récente, réalisée sur deux cohortes, l'une de 15 585 cas présentant un trouble neurodéveloppemental (déficience intellectuelle avec ou sans malformations associées et TSA) soumis pour un diagnostic clinique par micropuce et l'autre constituée de 762 familles avec TSA de la cohorte *Simons*, a montré qu'un nombre plus important de CNVs autosomiques 'délétères' (définis par les auteurs comme *de novo*, > 400 kb et de fréquence < 1 %) étaient détectés chez les filles comparées aux garçons (Jacquemont et coll., 2014). De plus, les filles atteintes de TSA présentaient un excès de SNVs délétères par rapport aux garçons. Enfin, les auteurs ont observé un excès de CNVs 'délétères' transmis maternellement dans la première cohorte, ainsi qu'un excès de CNVs et de SNVs hérités de la mère parmi

les familles atteintes de TSA. L'ensemble de ces résultats suggère des différences liées au sexe dans la susceptibilité aux mutations et aux CNVs non liés au chromosome X.

Les résultats de la dernière étude de *l'Autism Genome Project*, réalisée sur près de 3 000 familles avec TSA, suggèrent également la présence de facteurs protecteurs spécifiques chez les filles, puisque le sexe ratio de la cohorte (6:1) s'équilibre lorsque l'on considère les porteurs de CNVs pathogènes à pénétrance forte (2:1) et reste inchangé dans le groupe des patients porteurs de CNVs pathogènes à pénétrance incomplète et expressivité variable (6:1) (Pinto et coll., 2014). Les filles seraient donc moins susceptibles d'être affectées par des CNVs à effet modéré, caractérisés par une pénétrance incomplète et/ou une expressivité variable.

Si l'hypothèse de la présence d'effets protecteurs chez les filles, associée à un fardeau génétique plus fort chez les atteintes est vraie, on pourrait imaginer que la fratrie des filles atteintes serait plus à risque que la fratrie des garçons avec TSA. Une étude réalisée sur plus de 9 000 paires de jumeaux dizygotes a effectivement montré un risque plus élevé de présenter des caractéristiques autistiques au sein des fratries de filles atteintes, étayant ce modèle (Robinson et coll., 2013). Cependant, une étude plus récente et de taille plus grande n'a pas répliqué ces résultats (Sandin et coll., 2014). Cette discordance pourrait être expliquée par la différence des tailles d'échantillons et des méthodes diagnostiques utilisées. Néanmoins, si les filles sont plus souvent affectées par des CNVs à effet fort, on peut également imaginer que ces CNVs sont soumis à une pression de sélection négative et sont plus souvent *de novo*, ce qui implique que la probabilité de les observer dans la fratrie serait en fait faible. Bien qu'un nombre croissant d'observations soutienne l'existence d'effets protecteurs spécifiques aux filles, des études complémentaires sont nécessaires pour mieux comprendre les mécanismes sous-jacents impliqués.

## 2. Héritabilité manquante

Nos connaissances actuelles des bases génétiques des TSA nous ont permis d'identifier une étiologie chez environ 25 % des patients. Cependant, l'héritabilité de l'autisme est estimée à 50-80 % ce qui implique que beaucoup de facteurs génétiques n'ont pas encore été identifiés. A quoi correspond l'héritabilité manquante ? Plusieurs hypothèses sont envisageables.

### 2.1 Variants non détectés et de signification clinique inconnue

Parmi les nombreux variants détectés chez un individu atteint, la plupart n'a apparemment pas de rôle dans la pathologie et la difficulté est de déterminer lesquels contribuent au phénotype. A ce jour, il est clair qu'il y a un manque d'informations pour comprendre la contribution de beaucoup de CNVs détectés. Les variants de **signification clinique inconnue** (*variants of unknown significance*, **VOUS**) représentent sûrement une source majeure de facteurs étiologiques impliqués dans les TSA, non exploitée à ce jour. Des stratégies permettant de mieux comprendre leur potentiel rôle pathogène doivent être développées.

Malgré les progrès technologiques énormes réalisés ces dernières années, nos connaissances du génome restent limitées par certaines difficultés liées aux propriétés des méthodes utilisées. Par exemple, les écarts du dosage génique 3:2 étant plus difficiles à détecter qu'une réduction 2:1 par la technique de micropuces, les **duplications** sont plus difficiles à mettre en évidence que les délétions. En effet, les duplications détectées sont en moyenne de taille plus grande que les délétions et un certain nombre de petites duplications est probablement manqué. La technique de micropuces nous a permis d'identifier de nombreux syndromes de microdélétion et microduplication qui sont des évènements récurrents, les plus faciles à identifier. Pour le moment, nos connaissances des **CNVs non récurrents** restent à approfondir puisqu'ils sont associés à des fréquences plus faibles et nombre d'entre eux n'ont probablement pas encore été détectés. De plus, la recherche de facteurs génétiques étiologiques dans les TSA s'est

largement concentrée sur les régions codantes des gènes (CNVs exoniques et étude de l'exome). Cependant, plusieurs cas **d'altérations de régions génomiques non codantes** pouvant jouer un rôle dans la pathologie des individus qui les portent ont été rapportés dans la littérature (Zweier et coll., 2010 ; Bae et coll., 2014). De plus, même si les puces à ADN présentent aujourd'hui une résolution très haute, les études dans les TSA se sont concentrées sur les CNVs de taille relativement grande (> 15-30 kb) et la contribution des **petits remaniements** est encore mal appréciée. Le développement d'algorithmes permettant d'interpréter les données du séquençage de l'exome a permis de détecter certains des petits CNVs (Krumm et coll., 2013 ; Poultney et coll., 2013).

Par ailleurs, le séquençage de l'exome 'entier' ne concerne pas toutes les régions codantes puisqu'encore 8 à 10 % de ces régions ne peuvent pas être interrogées (Krumm et coll., 2014). De plus, les **indels** sont des variants difficiles à détecter (Karakoc et coll., 2012) et peu d'études ont vraiment exploré leur contribution dans les troubles neuropsychiatriques (Lemos et coll., 2012). Enfin, les variants situés dans les régions non codantes, dans les régions promotrices des gènes ou dans les éléments régulateurs situés à distance sont encore manqués.

Des progrès dans les techniques de séquençage actuelles et le développement de nouveaux algorithmes de détection pourront nous donner accès à de nouvelles ressources pour découvrir de nouveaux gènes impliqués dans les TSA.

## 2.2 Modèles oligogéniques et polygéniques

Une autre part de l'héritabilité manquante pourrait être expliquée par l'existence de modèles oligogéniques ou polygéniques. Parmi les différents modèles polygéniques proposés pour expliquer l'étiologie des TSA, le **modèle '*two-hit*'** a souvent été évoqué. Selon ce modèle oligogénique, c'est la présence de deux (ou plusieurs) variants rares à effet fort (ou modéré) qui serait responsable de l'apparition du phénotype chez l'individu. Quelques rares exemples de la littérature semblent appuyer ce modèle même si aucun n'a encore été confirmé de façon définitive. Les délétions récurrentes 16p12.1 ont été décrites dans plusieurs troubles neurodéveloppementaux incluant les TSA, la déficience intellectuelle, la schizophrénie (International Schizophrenia Consortium, 2008 ; Marshall et coll., 2008 ; Girirajan et coll., 2010 ; Pinto et coll., 2014). Dans une étude réalisée sur plus de 20 000 cas présentant divers troubles neurodéveloppementaux et près de 15 000 témoins, cette délétion est enrichie chez les cas, la plupart du temps héritée et les parents porteurs présentent souvent un phénotype neuropsychiatrique, même léger (Girirajan et coll., 2010). Ces délétions sont associées, chez un quart des patients porteurs, à la présence d'un second CNV (*second hit*). Cependant, les CNVs additionnels détectés pourraient expliquer à eux-seuls le phénotype des patients, comme par exemple une délétion terminale 22q13 impliquée dans le syndrome de Phelan-McDermid.

A l'heure actuelle, les évidences soutenant le modèle '*two-hit*' ne sont pas très convaincantes et pour pouvoir impliquer ce mécanisme, il faudrait pouvoir réunir un grand nombre de patients porteurs d'une même anomalie et rechercher systématiquement la présence de mutations ou de CNVs additionnels chez eux. En dehors du modèle 'two-hit', des **modèles polygéniques** correspondant à différentes combinaisons de variants rares, communs, *de novo*, hérités, d'effets forts ou faibles peuvent également être envisagés (**Figure 32**). Une approche alliant l'analyse des CNVs et le séquençage de l'exome entier chez un grand nombre de patients pourrait mettre en évidence des formes oligogéniques ou polygéniques de TSA (Classen et coll., 2013).

## 2.3 Contribution des variants génétiques communs

L'existence de caractéristiques autistiques chez les apparentés des individus atteints, regroupées sous le terme de '**phénotype élargi de l'autisme**', a longtemps fait penser à l'existence d'allèles communs,

présents dans la population et qui ne seraient ni nécessaires ni suffisants à l'apparition du phénotype mais qui, en combinaison les uns avec les autres, pourraient atteindre un seuil étiologique nécessaire à l'apparition de la pathologie. Les apparentés ne présenteraient que quelques uns de ces variants, expliquant les manifestations plus légères observées chez eux. De plus, des troubles sociaux de type autistique sont retrouvés dans la population générale avec des sévérités variables et une distribution normale dont l'autisme représente l'une des valeurs extrêmes (Constantino et Todd, 2003), ce qui suggère également la présence de facteurs génétiques partagés présentant un risque dans la susceptibilité aux TSA.

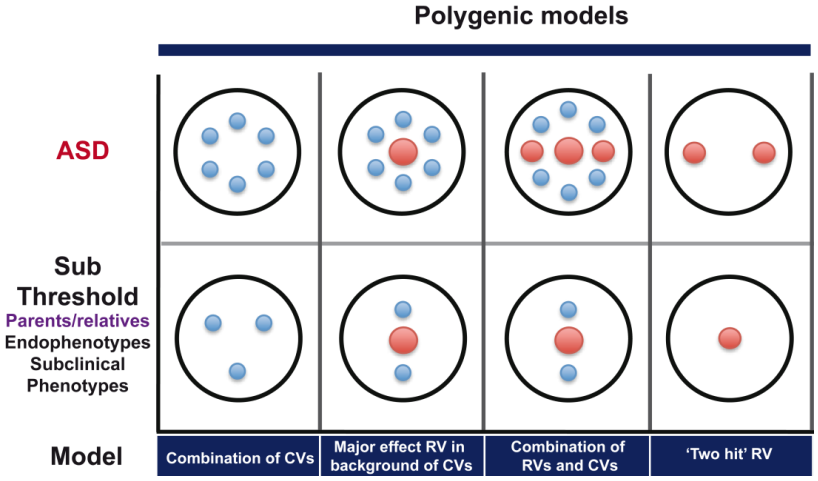



**Figure 32. Modèles polygéniques proposés dans l'étiologie des TSA**
Plusieurs modèles prenant en compte des combinaisons différentes de variants rares (RV) et de variants communs (CV) ont été proposés pour expliquer la présence de TSA chez un individu ainsi que des caractéristiques autistiques chez ses apparentés. Aujourd'hui, aucun de ces modèles n'a été clairement démontré dans les TSA (figure adaptée de Berg et Geschwind, 2012).

Une récente étude épidémiologique réalisée sur un échantillon de plus de 2 000 infirmières, leurs époux et leurs enfants, a montré une corrélation entre les habilités sociales (score de *Social Responsiveness Scale*, SRS) des parents et de leurs enfants, suggérant une contribution de facteurs génétiques fréquents (Lyall et coll., 2014) (Figure 33). De plus, les scores des enfants atteints sont encore plus élevés si les deux parents ont des scores élevés, confortant le rôle des **influences génétiques additives** qui concentreraient le risque de TSA dans les générations successives. Cette étude a également montré que les scores des parents sont corrélés entre eux, ce qui est un argument en faveur du modèle d'**appariement assortatif** (*assortative mating*), selon lequel les individus qui partagent des traits communs forment des couples et se reproduisent entre eux.

Même si un nombre croissant d'études prédisent une importante contribution des variants communs dans les TSA (Klei et coll., 2012 ; Gaugler et coll., 2014 ; Sandin et coll., 2014), aucun d'entre eux n'a été mis en évidence aujourd'hui. Cependant, les approches et stratégies utilisées dans le domaine de la génétique de l'autisme n'étaient clairement pas adaptées à la détection de tels variants. Des échantillons beaucoup plus larges que ceux disponibles aujourd'hui pour les TSA sont nécessaires afin de pouvoir espérer identifier des facteurs génétiques fréquents à effet faible dans un syndrome aussi hétérogène que l'autisme.

Comme pour d'autres pathologies, la recherche dans le domaine de la génétique de l'autisme a commencé par identifier les causes les plus évidentes, correspondant à celles qui sont associées aux effets les plus forts, notamment les variants rares et *de novo*. De nombreux facteurs étiologiques restent à découvrir incluant les variants à effets modérés, ceux qui contribuent à des formes oligogéniques et/ou

polygéniques, les variants hérités associés à des formes récessives ainsi que les variants communs associés à des effets faibles voire très faibles.

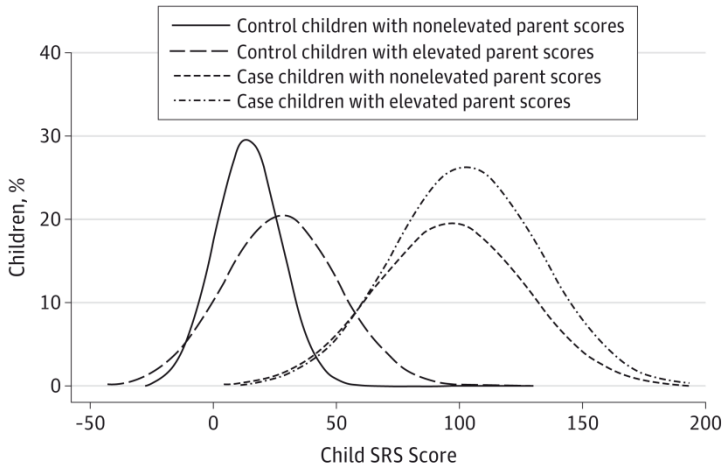



**Figure 33. Corrélation entre les habilités sociales des parents et leurs enfants**
La distribution des scores de *Social Responsiveness Scale* (SRS), mesurant les habilités d'interaction sociale, est représentée pour quatre groupes d'enfants : les enfants non atteints dont les parents ont des scores non élevés (en trait plein), les enfants non atteints dont les parents ont des scores élevés (pointillés longs), les enfants atteints de TSA dont les parents ont des scores non élevés (pointillés courts) et les enfants atteints dont les parents ont des scores élevés (pointillés et points). Les scores des enfants sont corrélés aux scores des parents (figure tirée de Lyall et coll., 2014).

## 3. Stratégies pour interpréter les variants génétiques

Les avancées technologiques dans le domaine de la génétique humaine ont été fulgurantes ces dernières années. Ces progrès ont permis la découverte d'une multitude de facteurs étiologiques et l'accumulation d'une masse de données génomiques. Rapidement, le problème n'a plus été de l'ordre de la détection des variants mais de leur interprétation et de la compréhension de leur contribution au phénotype, en particulier pour les variants de signification inconnue. L'hétérogénéité génétique des TSA étant extrême, ceci implique que certaines des mutations trouvées seront uniques à un patient. Or, la récurrence reste aujourd'hui la preuve la plus fiable de l'implication d'un gène dans une pathologie. D'ailleurs, l'implication de certains gènes liés à l'X mis en cause dans la déficience intellectuelle et les TSA par la description de mutations dans une seule famille a été récemment remise en question (Piton et coll., 2013). Il est hasardeux d'utiliser une évidence basée sur une famille unique (même de grande taille, avec beaucoup d'individus atteints) pour impliquer un gène dans une pathologie, ce qui pose un problème évident pour la recherche de facteurs impliqués dans les TSA. Pour essayer de pallier la difficulté d'observer une récurrence, une combinaison d'approches complémentaires peuvent être utilisées pour apporter des évidences additionnelles. De plus, l'identification de variants rares chez plusieurs patients nécessite des **efforts collaboratifs** comme dans le cas du consortium international de l'*Autism Genome Project*, afin d'augmenter la puissance des études en réunissant des cohortes de très grande taille.

### 3.1 Bases de données de patients et de témoins

L'existence de bases de données de patients facilite grandement l'interprétation des variants détectés. Des milliers de remaniements trouvés chez des individus atteints de troubles neurodéveloppementaux et de diverses manifestations cliniques ont été réunis dans les bases de données publiques **DECIPHER** (*DatabasE of Genomic variants and Phenotype in Humans Using Ensembl Resources*) (Firth et coll., 2009),

**ISCA** (*International Standards for Cytogenomic Arrays*) (Kaminsky et coll., 2011) et **ECARUCA** (*European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations*) (Feenstra et coll., 2006). La récurrence étant l'un des points clefs de l'analyse des CNVs, c'est en comparant un variant trouvé chez un patient avec des remaniements détectés chez d'autres individus atteints et présentant des phénotypes similaires qu'il sera parfois possible de déterminer sa pathogénicité. Puisqu'il existence un fort chevauchement étiologique entre les TSA et la déficience intellectuelle, mais aussi l'épilepsie, les données obtenues dans des cohortes de patients atteints de ces deux troubles neurodéveloppementaux sont particulièrement informatives pour l'analyse des CNVs dans les TSA. Cette stratégie peut être étendue à d'autres conditions neuropsychiatriques telles que la schizophrénie qui présente un chevauchement des facteurs étiologiques avec les TSA beaucoup plus modeste mais qui reste informatif, en particulier dans le cas de certains CNVs pathogènes à expressivité variable.

D'autres bases de données rapportent les CNVs détectés chez des individus apparemment sains. C'est par exemple le cas de la base de données des variants génomiques **DGV** (*Database of Genomic Variants*) qui contient à l'heure actuelle près de 4 millions de variants détectés chez des individus témoins de 37 études différentes. L'utilisation de ces données permet de concentrer nos analyses sur les variants rares, peu ou pas trouvés chez les témoins. Cependant, la qualité des données disponibles dans cette base est directement liée aux techniques de détection utilisées et en particulier au type de puce. Les premières études utilisaient principalement des sondes BACs qui offrent une faible résolution et surestiment souvent la taille des CNVs détectés. Plus tard, les puces à oligonucléotides ont permis d'augmenter la résolution mais également le nombre de faux-positifs, en particulier parmi les petits CNVs. La plupart des variants rapportés dans DGV n'ont pas été confirmés par une méthode indépendante et peuvent donc correspondre à de faux positifs (de Leeuw et coll., 2012). En effet, il n'est pas rare d'observer de multiples variants dans un gène ou à un locus donné, dans une seule étude de DGV, indiquant un biais de détection étude-spécifique. Idéalement, pour pouvoir comparer les CNVs entre les patients et les témoins, les données devraient être générées de la même manière ce qui implique un génotypage sur les mêmes plateformes avec le même type de puces et les mêmes méthodes analytiques (mêmes contrôles de qualité et mêmes algorithmes de détection).

## 3.2 Listes de gènes et revue de la littérature

Bien que de nombreux critères permettent de prioriser les différents variants détectés, il est souvent difficile d'interpréter un CNV trouvé chez un patient. Dans l'analyse des données AGP, nous avons recherché des CNVs pouvant expliquer la pathologie chez les patients porteurs. Pour cela, les listes de gènes développées par notre équipe se sont avérées être un outil essentiel, répertoriant :

- Les gènes et loci **impliqués dans les TSA**. Cette liste comprend à ce jour 148 gènes et 59 loci et a été créée par une revue approfondie des évidences impliquant ces gènes dans la base de données OMIM, les cas rapportés dans la littérature que ce soit dans des études de grandes cohortes ou dans la description d'un patient unique. Tous les gènes et loci de cette liste ont aussi été impliqués dans la déficience intellectuelle. La liste se veut la plus exhaustive possible et toutes les évidences sont strictement évaluées, contrairement à d'autres listes qui regroupent tous les gènes rapportés sans hiérarchiser les évidences qui leur sont associées. Par exemple la liste de gènes de l'autisme créée par SFARI (*Simons Foundation Autism Research Initiative*) est une liste dite 'inclusive' qui répertorie tous les gènes pour lesquels une analyse a suggéré qu'ils pouvaient jouer un rôle dans l'étiologie des TSA même si ces évidences ne sont pas convaincantes (études d'association non répliquées, évidences fonctionnelles isolées ou cas uniques).

- Les gènes et loci **impliqués dans la déficience intellectuelle**. Considérant le fort chevauchement étiologique existant entre les TSA et la déficience intellectuelle, les gènes impliqués dans la déficience intellectuelle mais pas encore dans les TSA sont potentiellement des causes d'autisme.
- Les gènes et loci **candidats** pour ces deux troubles neurodéveloppementaux. Cette liste contient des gènes touchés par des anomalies génétiques rares telles que des points de cassure, des délétions et des variants rapportés dans les études d'exomes et pour lesquels les évidences ne sont pas encore suffisantes pour les impliquer.

La liste des gènes impliqués dans les TSA créée par notre équipe a été publiée en 2011 (Betancur, 2011) et a été utilisée par de nombreux chercheurs dans les TSA et la schizophrénie pour identifier des variants pathogènes dans les analyses de CNVs et les études d'exome, ainsi que pour effectuer des analyses de réseaux afin d'identifier les voies biologiques impliquées dans l'autisme et de nouveaux gènes candidats (Kou et coll., 2012 ; Neale et coll., 2012b ; O'Roak et coll., 2012a ; O'Roak et coll., 2012b ; Noh et coll., 2013 ; Purcell et coll., 2014 ; Samocha et coll., 2014 ; Uddin et coll., 2014).

Parfois, la recherche bibliographique réalisée sur un gène en particulier permet de l'impliquer dans l'étiologie des TSA. C'est par exemple le cas du gène *SETD5* (*SET domain containing 5*), impliqué pour la première fois dans les TSA grâce à la découverte d'une délétion *de novo* qui n'implique que ce gène chez un patient de la cohorte AGP, mise en relation avec des évidences additionnelles de la littérature (Pinto et coll., 2014). En regroupant tous les cas décrits présentant des délétions de cette région, nous avons pu identifier 9 autres patients avec déficience intellectuelle et/ou TSA présentant des délétions 3p25.3 toutes *de novo* et définissant une région critique contenant uniquement *SETD5* (**Figure 34**). De plus, plusieurs mutations *de novo* de ce gène avaient été rapportées dans des études d'exome dans la déficience intellectuelle et les TSA. Ce gène code pour une méthyltransférase et d'autres gènes codant des méthyltransférases ont été précédemment impliqués dans les TSA (notamment *NSD1*, syndrome de Sotos et *EHMT1*, syndrome de Kleefstra).

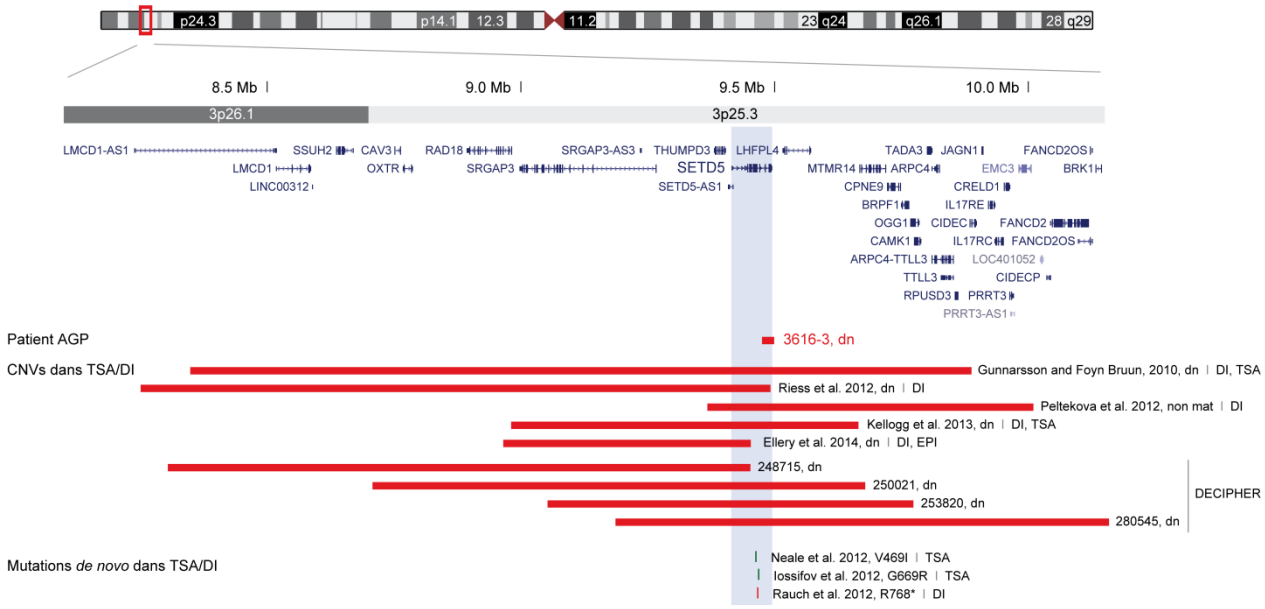



**Figure 34. Implication du gène *SETD5* dans les TSA**

Les délétions rapportées dans la région 3p25.3 sont représentées par les lignes rouges, les mutations *de novo* du gène *SETD5* sont représentées par des traits verticaux verts (faux-sens) ou rouges (non-sens). Les gènes présents dans la région sont indiqués en haut, sous la représentation schématique du chromosome 3. La région minimale de chevauchement entre les différentes délétions est montrée par la bande bleue et ne contient que le gène *SETD5* (figure adaptée de Pinto et coll., 2014). DI, déficience intellectuelle ; dn, *de novo* ; EPI, épilepsie ; mat, maternel.

Pendant la publication de notre manuscrit, une étude de séquençage ciblé de gènes candidats réalisée dans une cohorte de 996 patients atteints de déficience intellectuelle a confirmé l'implication du gène *SETD5* (Grozeva et coll., 2014). Sept mutations *de novo* perte de fonction ont été trouvées dans ce gène et cinq individus porteurs de ces mutations présentent des traits autistiques ou un TSA.

## 3.3 Outils bioinformatiques

Les études *in vitro* et *in vivo* étant très coûteuses, une analyse préliminaire *in silico* des variants détectés peut permettre de prioriser les données afin de sélectionner les variants pour lesquels l'impact délétère est le plus probable. L'**annotation des gènes** dans des bases de données répertoriant leur structure, leur fonction et leur profil d'expression facilite la sélection de candidats. En particulier, les scores d'haploinsuffisance des gènes peuvent apporter un argument en faveur de l'implication d'un gène présent dans une délétion chez un patient. Cependant, nos connaissances des gènes haploinsuffisants sont encore limitées pour le moment puisqu'environ 12 000 ont été évalués (Huang et coll., 2010) soit seulement la moitié du génome humain. La base publique **UCSC** (*University of California*, *Santa Cruz*) centralise une multitude de données concernant les gènes et a récemment intégré de nouveaux outils qui permettent de visualiser dans une région donnée la présence d'éléments régulateurs, de régions promotrices et même des données de transcriptome. En effet, de nombreux gènes présentent des isoformes tissu-spécifiques et il est essentiel de savoir lesquelles sont exprimées dans le cerveau afin d'interpréter la signification des CNVs et SNVs (Uddin et coll., 2014).

Même si les analyses de voies biologiques ne peuvent pas être utilisées pour impliquer un gène dans une pathologie, le fait qu'un gène appartienne à un réseau fonctionnel déjà mis en évidence augmente la probabilité qu'il soit lui-même impliqué. Récemment, des analyses de voies biologiques dans les TSA ont intégré les profils d'expression des isoformes de chacun des gènes pour construire des réseaux fonctionnels tissus-spécifiques (Corominas et coll., 2014). Pour une pathologie neurodéveloppementale comme les TSA, étudier les interactions des transcrits exprimés dans le cerveau va permettre de créer des réseaux fonctionnels plus pertinents et plus précis. De plus, la construction de réseaux de co-expression situant les interactions entres les différents gènes dans un **contexte spatio-temporel** dans le cortex cérébral permet également de mieux comprendre les effets de l'altération des gènes dans les TSA (Parikshak et coll., 2013 ; Willsey et coll., 2013).

## 3.4 Etudes d'expression

Aujourd'hui, la **PCR quantitative après transcription inverse** (RT-qPCR) est la méthode de choix pour tester l'impact d'un variant particulier sur l'expression d'un gène donné. Cette technique permet d'identifier les gènes touchés par une délétion dont l'expression est altérée, de révéler des effets à distance d'une délétion ou d'une duplication sur les gènes du voisinage génomique non touchés, les conséquences d'une duplication partielle d'un gène ou encore les variations d'expression liées à des CNVs touchant les régions régulatrices des gènes. Etant donné que l'accès à des échantillons cérébraux est très limité, on peut tester l'expression dans le sang des patients. En effet, bien que seuls 60 % des gènes qui sont exprimés dans le cerveau le soient aussi dans les lymphocytes (Cookson et coll., 2009), dans certains cas, l'expression d'un gène dans les lignées lymphoblastoïdes peut refléter celles des lignées neuronales. Par ailleurs, des **études d'expression pangénomiques** peuvent être utilisées pour mettre en évidence des voies biologiques dérégulées dans les TSA ou dans des sous-groupes de patients définis génétiquement (Luo et coll., 2012).

## 3.5 Analyses fonctionnelles

Des études fonctionnelles *in vitro* et/ou *in vivo* de ces variants peuvent permettre de mieux appréhender leur potentiel caractère pathogène.

Il existe de multiples modèles *in vitro* permettant d'étudier l'impact de mutations trouvées chez des patients. Le modèle doit être choisi selon la fonction du gène et peut inclure de la mutagénèse dirigée dans des cellules de mammifères ou des analyses électrophysiologiques ou immunohistochimiques par exemple. Parmi les modèles *in vitro* les plus prometteurs, on trouve les **cellules souches pluripotentes induites** (*induced pluripotent stem cells,* iPS). Il s'agit d'obtenir des cellules neuronales en culture, spécifiques d'une maladie, qui peuvent être utilisées pour étudier les bases cellulaires et moléculaires impliquées dans l'étiologie de la pathologie (**Figure 35**). Ce modèle a déjà été utilisé dans différentes maladies neurologiques incluant les TSA (Kim et coll., 2014 ; Pasca et coll., 2014).

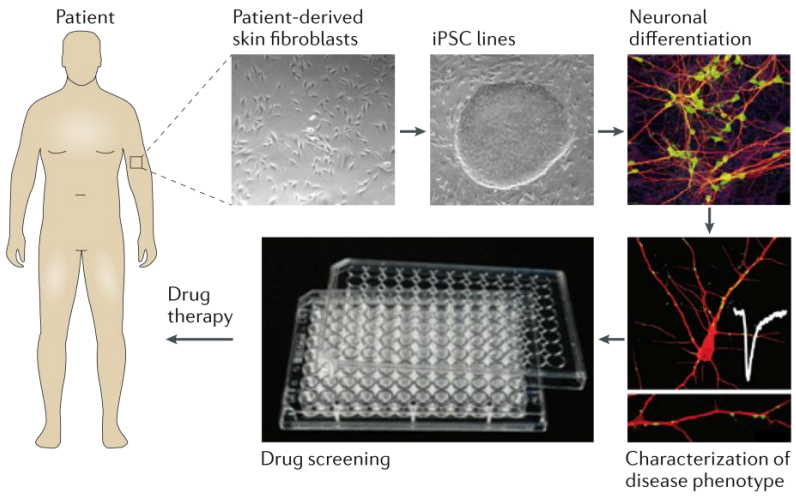



**Figure 35. Génération de cellules souches pluripotentes induites pour étudier la physiopathologie des TSA**

Des cellules d'un patient sont prélevées (la plupart du temps, des lymphoblastes ou des fibroblastes), reprogrammées en cellules souches pluripotentes et dérivées dans un type cellulaire particulier en l'occurrence des neurones dans le cas des TSA. Les cellules ainsi obtenues permettent de définir les phénotypes associés et de tester un grand nombre de composés pharmacologiques dans l'espoir de trouver une thérapie applicable aux patients (figure tirée de Ghosh et coll., 2013).

Les cellules iPS ont été notamment utilisées pour modéliser le syndrome de Rett (Marchetto et coll., 2010), le syndrome de l'X fragile (Urbach et coll., 2010), le syndrome de Phelan-McDermid (Shcheglovitov et coll., 2013) et le syndrome de Timothy (Pasca et coll., 2011). Grâce à ces approches, notre connaissance des caractéristiques moléculaires et cellulaires associées à certaines mutations s'est largement enrichie. En outre, cette méthode peut être utilisée pour réaliser un criblage à grande échelle de molécules pharmaceutiques et ainsi offrir une perspective thérapeutique. Cependant, elle reste marginale pour le moment puisqu'elle représente un niveau technique et un coût très élevés. Une autre application intéressante des cellules iPS dans des patients porteurs de syndromes de microdélétion est la possibilité d'identifier des phénotypes cellulaires caractéristiques du syndrome (synaptiques, cardiaques ou osseux par exemple) et d'essayer de les corriger en ré-exprimant les gènes délétés, afin d'identifier quels sont ceux responsables des phénotypes observés chez les patients.

La souris est l'animal le plus souvent utilisé pour modéliser *in vivo* les mutations trouvées chez l'humain. En plus d'être physiologiquement très similaire à l'homme, la plupart des gènes humains ont un

homologue chez la souris et il y a une importante conservation de synténie[3] entre ces deux espèces, ce qui peut s'avérer particulièrement intéressant pour l'étude de syndromes de microdélétion à gènes contigus. Il existe de **nombreux modèles murins** pour des syndromes génétiques associés aux TSA qui essaient de récapituler les atteintes observées chez les patients. C'est par exemple le cas de formes monogéniques comme le syndrome de Rett et le syndrome de duplication *MECP2*, le syndrome de Phelan-McDermid et le syndrome de l'X fragile pour lesquels plusieurs modèles de souris transgéniques ont été créés. Il existe également des modèles murins de divers syndromes de microdélétion/microduplication tels que les syndromes de DiGeorge (délétion 22q11.2), de Smith-Magenis et Ptocki-Lupski (délétion et duplication 17p11.1), d'Angelman/Prader-Willi et de duplication 15q11-q13 et de microdélétion/microduplication 16p11.2 (pour revue voir Weischenfeldt et coll., 2013). La caractérisation de ces animaux par différents tests comportementaux (interaction sociale, stéréotypies, apprentissage, mémoire à court ou long terme par exemple) peut montrer des troubles et déficits semblables à ceux observés chez l'humain. Ces souris ont ainsi permis de réaliser une caractérisation phénotypique plus approfondie au niveau cellulaire et moléculaire contribuant considérablement à l'accumulation de connaissances sur ces syndromes et dans certains cas au développement de thérapies.

D'autres modèles animaux ont également été utilisés pour étudier certaines pathologies du développement. Par exemple, le poisson zèbre a été utilisé pour l'étude du système nerveux central et de ses fonctions (pour revues voir Kabashi et coll., 2010 ; Burne et coll., 2011). En effet, ce modèle bien qu'étant invertébré, possède une architecture neuronale très similaire à celle de l'homme. De plus, la transparence des embryons et des larves en cours de développement fait de cet animal un outil simple et pratique dans l'étude des mécanismes à la base des maladies neurodéveloppementales. Par exemple, ce modèle a été utilisé dans l'étude des délétions 16p11.2 pour essayer de trouver des gènes responsables des différentes manifestations associées au CNV chez les patients (Blaker-Lee et coll., 2012 ; Golzio et coll., 2012). La région délétée de façon récurrente contient une trentaine de gènes et les auteurs ont injecté des oligonucléotides antisens dirigés contre plusieurs gènes candidats dans les larves en développement afin d'observer les phénotypes associés à l'altération de chacun des gènes. Les résultats montrent que la plupart des gènes de cet intervalle génomique sont impliqués dans le développement cérébral (Blaker-Lee et coll., 2012). De façon intéressante, Golzio et coll. (2012) ont impliqué *KCTD13* dans les phénotypes de micro- et macrocéphalie liés respectivement à une sous-expression et à une surexpression du gène. Les auteurs proposent une contribution majeure de ce gène aux phénotypes neurologiques associés aux CNVs 16p11.2 chez l'humain.

Utiliser une combinaison d'approches variées permet parfois d'interpréter plus facilement les variants génétiques en réunissant un nombre important d'évidences en faveur du caractère pathogène des anomalies détectées. Cependant, l'implication d'un gène doit être évaluée en premier lieu sur la base d'évidences génétiques fortes et non sur l'observation d'un phénotype anormal associé à un variant donné dans un modèle animal ou *in vitro*. Les évidences génétiques constituent les arguments les plus fiables pour impliquer un gène dans une pathologie, les études bioinformatiques et fonctionnelles complémentaires permettent de les appuyer.

## 4. Intérêt du diagnostic étiologique chez les patients atteints de TSA

***Conseil génétique et prise en charge adaptée***

L'analyse par micropuces a révolutionné le diagnostic moléculaire dans les TSA et le retard du développement et est maintenant recommandée comme test de première intention dans la recherche

[3] La synténie correspond à la présence de plusieurs loci sur un même chromosome ; on parle de conservation de synténie lorsque l'ordre des loci sur les chromosomes est le même entre deux espèces.

d'étiologie dans ces pathologies (Miller et coll., 2010). Notre analyse des CNVs dans le cadre de l'étude AGP nous a permis de réaliser 82 diagnostics étiologiques, correspondant à 3,3 % de la cohorte (**Figure 36**) (Pinto et coll., 2014).

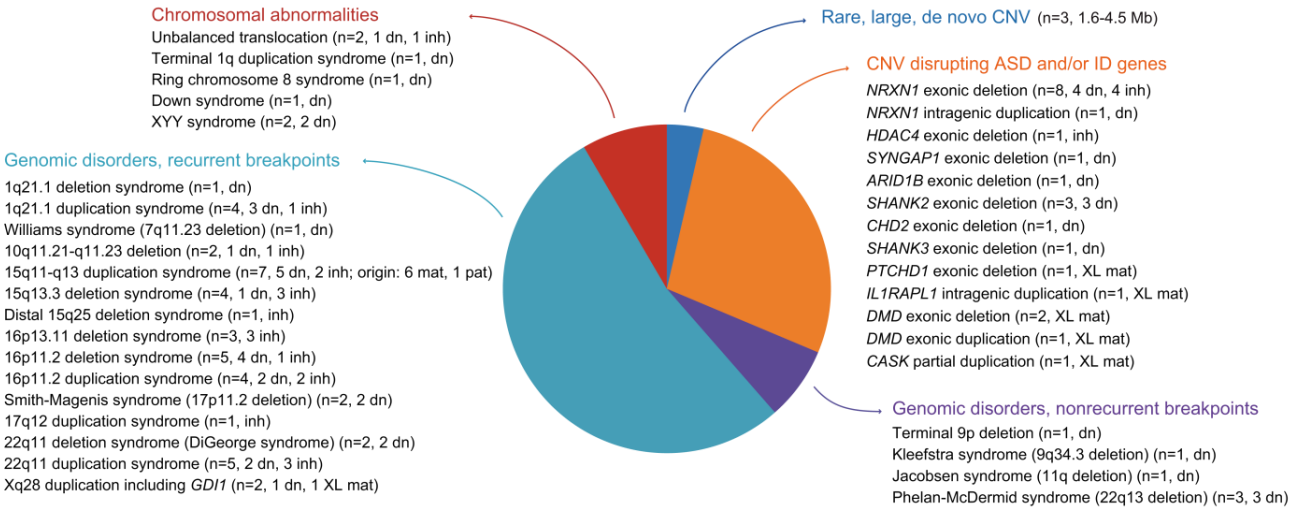



**Figure 36. CNVs pathogènes identifiés chez 2 446 patients atteints de TSA de l'*Autism Genome Project***

Grâce à notre analyse, 84 CNVs pathogènes ont été découverts chez 82 patients de la cohorte, représentant 38 étiologies différentes. Parmi les CNVs pathogènes identifiés, on trouve des anomalies chromosomiques, des CNVs récurrents et non récurrents correspondant à des troubles génomiques décrits, des CNVs de grande taille, rares et *de novo* et des CNVs touchant des gènes impliqués dans la TSA et/ou la déficience intellectuelle. dn, *de novo* ; inh ; hérité ; mat ; maternel ; pat ; paternel (figure tirée de Pinto et coll. 2014).

Identifier l'étiologie associée au phénotype chez un individu est primordial autant pour le patient et pour la famille, que pour la communauté scientifique. En faisant un diagnostic étiologique, on offre aux parents la possibilité de bénéficier d'un **conseil génétique adapté.** Les études épidémiologiques dans les TSA ont rapporté un taux de récurrence élevé (18,7 %) pour la fratrie des enfants atteints (Ozonoff et coll., 2011). Connaître l'étiologie permet d'évaluer le risque de récurrence spécifique de la famille (qui peut être similaire à celui de la population générale dans le cas d'anomalies *de novo*) au lieu de se baser sur ce chiffre de récurrence globale. Un diagnostic étiologique permet également de prévenir des récurrences en recherchant les porteurs parmi les membres de la famille et en réalisant un dépistage prénatal. Pour le patient, le diagnostic génétique permet d'anticiper et de prendre en charge les comorbidités médicales et comportementales. Par exemple, deux de nos patients sont atteints du syndrome de DiGeorge qui est associé à la présence de malformations cardiaques congénitales et à une hypocalcémie, un suivi particulier pourra donc être mis en place pour eux. Pour les parents, le diagnostic soulage l'anxiété et l'incertitude, permet d'arrêter les tests diagnostiques parfois coûteux et invasifs et de mieux comprendre les comorbidités et le pronostic associé à l'anomalie de leur enfant. Par ailleurs, connaître l'étiologie peut permettre aux parents de s'investir dans des associations de familles spécifiques à la maladie et leur donne la possibilité de participer à des programmes de recherche et de faire bénéficier leur enfant de thérapies actuellement en cours d'évaluation. Quatre patients de notre cohorte présentent un syndrome de Phelan-McDermid pour lequel des essais cliniques sont en cours. Enfin, pour la communauté scientifique, chaque découverte étiologique permet d'augmenter les connaissances quant à la fonction des gènes impliqués, les réseaux fonctionnels et voies biologiques touchés afin de mieux appréhender les corrélations génotype-phénotype et les mécanismes physiopathologiques sous-jacents. A terme, une meilleure compréhension des processus neurodéveloppementaux impliqués conduira au développement de thérapies adaptées.

***Perspectives thérapeutiques***

L'identification de voies biologiques dérégulées dans les TSA grâce aux études génétiques a fournit des cibles potentielles pour le développement de composés visant à rétablir les déséquilibres observés. Les modèles *in vitro* ou *in vivo* permettent de tester les effets des différents composés candidats au traitement. Ainsi, des thérapies ciblées pour améliorer les manifestations de certains syndromes sont en plein essor. Des réversions de certaines des manifestations cliniques ont été observées à l'âge adulte chez les souris modèles de plusieurs syndromes (pour revue voir Ehninger et coll., 2008b), ce qui offre un espoir inespéré pour les patients et leurs familles. Ces résultats suggèrent que des processus physiopathologiques du cerveau mature pourraient contribuer aux phénotypes neurologiques, indiquant que la fenêtre d'intervention thérapeutique dans les maladies neurodéveloppementales telles que la déficience intellectuelle et les TSA serait en fait plus longue qu'imaginée.

L'étude princeps ayant montré la possibilité de réversion du phénotype à l'âge adulte a été réalisée dans un modèle de souris du syndrome de Rett pour lequel l'expression du gène *MECP2* est conditionnelle. Ces souris présentaient les manifestations phénotypiques classiquement associées au syndrome et la réexpression de *MECP2* dans les neurones de souris adultes a permis l'amélioration significative des défauts cellulaires et comportementaux (Guy et coll., 2007). D'autres stratégies ont montré leur efficacité pour améliorer les manifestations dues à des déficits dans certaines voies biologiques. Le traitement des souris déficientes pour *PTEN* (Zhou et coll., 2009) ou des souris modèles de la sclérose tubéreuse de Bourneville (*TSC1* et *TSC2*) (Ehninger et coll., 2008a) par un inhibiteur de la voie mTOR (rapamycine), suractivée dans ces deux pathologies, a permis d'améliorer des anomalies au niveau cellulaire et les déficits comportementaux, réduit la macrocéphalie et l'épilepsie et prolongé la survie (pour revue voir Ehninger et Silva, 2011). Cette molécule est maintenant utilisée en clinique. Dans le cas du syndrome de Phelan-McDermid (délétion 22q13, *SHANK3*), le facteur de croissance IGF1 (*insulin growth factor 1*) a permis de corriger certains déficits électrophysiologiques et comportementaux des souris mutantes (Bozdagi et coll., 2013). Des essais cliniques sur des patients atteints du syndrome sont actuellement en cours à l'Université Mount Sinaï à New York.

Les progrès réalisés ces dernières années dans le domaine de la génétique des TSA sont extraordinaires et ont apporté des résultats remarquables et parfois inattendus. Cependant, la contribution des variants communs, celle des facteurs non génétiques quels qu'ils soient, le rôle du contexte génétique et environnemental dans l'expression d'un génotype donné sont autant de zones d'ombre qu'il reste à explorer. Les efforts collaboratifs et le partage des données doivent continuer pour mettre en place de nouvelles approches et de nouvelles stratégies qui nous permettront d'accroître nos connaissances et de mieux comprendre l'architecture génétique de cette pathologie complexe.

# Références

# A


Abdelmoity AT, LePichon JB, Nyp SS, Soden SE, Daniel CA, Yu S (2012) 15q11.2 proximal imbalances associated with a diverse array of neuropsychiatric disorders and mild dysmorphic features. *J Dev Behav Pediatr* 33, 570-576

Adegbola A, Gao H, Sommer S, Browning M (2008) A novel mutation in JARID1C/SMCX in a patient with autism spectrum disorder (ASD). *Am J Med Genet A* 146A, 505-511

Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407, 513-516

American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (Washington, DC,)

Andari E, Duhamel JR, Zalla T, Herbrecht E, Leboyer M, Sirigu A (2010) Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. *Proc Natl Acad Sci U S A* 107, 4389-4394

Andersen S, Laurberg P, Wu C, Olsen J (2014) Attention deficit hyperactivity disorder and autism spectrum disorder in children born to mothers with thyroid dysfunction: a Danish nationwide cohort study. *BJOG* (Epub ahead of print)

Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Sykes N, Pagnamenta AT et al. (2010) A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 19, 4072-4082

Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G, Bolshakova N, Bolte S, Bolton PF, Bourgeron T et al. (2012) Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* 21, 4781-4792

Antshel KM, Aneja A, Strunge L, Peebles J, Fremont WP, Stallone K, Abdulsabur N, Higgins AM, Shprintzen RJ, Kates WR (2007) Autistic spectrum disorders in velo-cardio facial syndrome (22q11.2 deletion). *J Autism Dev Disord* 37, 1776-1786

Archer HL, Evans J, Edwards S, Colley J, Newbury-Ecob R, O'Callaghan F, Huyton M, O'Regan M, Tolmie J, Sampson J et al. (2006) CDKL5 mutations cause infantile spasms, early onset seizures, and severe mental retardation in female patients. *J Med Genet* 43, 729-734

Ariani F, Hayek G, Rondinella D, Artuso R, Mencarelli MA, Spanhol-Rosseto A, Pollazzon M, Buoni S, Spiga O, Ricciardi S et al. (2008) FOXG1 is responsible for the congenital variant of Rett syndrome. *Am J Hum Genet* 83, 89-93

Ascano M, Jr., Mukherjee N, Bandaru P, Miller JB, Nusbaum JD, Corcoran DL, Langlois C, Munschauer M, Dewell S, Hafner M et al. (2012) FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* 492, 382-386

Asperger H (1944) Die "Autistischen Psychopathen" im Kindesalter. *Archiv fur Psychiatrie und Nervenkrankheiten* 117, 76-136

August GJ, Stewart MA, Tsai L (1981) The incidence of cognitive disabilities in the siblings of autistic children. *Br J Psychiatry* 138, 416-422

Autism and Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators (2014) Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2010. *MMWR Surveill Summ* 63, 1-21


# B


Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S, Bader PI, Hamati A, Reitnauer PJ, Smith R et al. (2010) Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genet Med* 12, 641-647

Bae BI, Tietjen I, Atabay KD, Evrony GD, Johnson MB, Asare E, Wang PP, Murayama AY, Im K, Lisgo SN et al. (2014) Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. *Science* 343, 764-768

Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M (1995) Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 25, 63-77

Baird TD, August GJ (1985) Familial heterogeneity in infantile autism. *J Autism Dev Disord* 15, 315-321

Balemans MC, Kasri NN, Kopanitsa MV, Afinowi NO, Ramakers G, Peters TA, Beynon AJ, Janssen SM, van Summeren RC, Eeftens JM et al. (2013) Hippocampal dysfunction in the Euchromatin histone methyltransferase 1 heterozygous knockout mouse model for Kleefstra syndrome. *Hum Mol Genet* 22, 852-866

Ballif BC, Theisen A, Coppinger J, Gowans GC, Hersh JH, Madan-Khetarpal S, Schmidt KR, Tervo R, Escobar LF, Friedrich CA et al. (2008) Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Mol Cytogenet* 1, 8

Barbosa AC, Kim MS, Ertunc M, Adachi M, Nelson ED, McAnally J, Richardson JA, Kavalali ET, Monteggia LM, Bassel-Duby R et al. (2008) MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proc Natl Acad Sci U S A* 105, 9391-9396

Barcia G, Chemaly N, Gobin S, Milh M, Van Bogaert P, Barnerias C, Kaminska A, Dulac O, Desguerre I, Cormier V et al. (2014) Early epileptic encephalopathies associated with STXBP1 mutations: Could we better delineate the phenotype? *Eur J Med Genet* 57, 15-20

Barge-Schaapveld DQ, Maas SM, Polstra A, Knegt LC, Hennekam RC (2011) The atypical 16p11.2 deletion: a not so atypical microdeletion syndrome? *Am J Med Genet A* 155A, 1066-1072

Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a "theory of mind"? *Cognition* 21, 37-46

Baron-Cohen S (2002) The extreme male brain theory of autism. *Trends Cogn Sci* 6, 248-254

Bayes A, van de Lagemaat LN, Collins MO, Croning MD, Whittle IR, Choudhary JS, Grant SG (2011) Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat Neurosci* 14, 19-21

Ben-Shachar S, Lanpher B, German JR, Qasaymeh M, Potocki L, Nagamani SC, Franco LM, Malphrus A, Bottenfield GW, Spence JE et al. (2009) Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet* 46, 382-388

Bena F, Bruno DL, Eriksson M, van Ravenswaaij-Arts C, Stark Z, Dijkhuizen T, Gerkes E, Gimelli S, Ganesamoorthy D, Thuresson AC et al. (2013) Molecular and clinical characterization of 25 individuals with exonic deletions of NRXN1 and comprehensive review of the literature. *Am J Med Genet B Neuropsychiatr Genet* 162B, 388-403

Berg AT, Plioplys S (2012) Epilepsy and autism: is there a special relationship? *Epilepsy Behav* 23, 193-198

Berg JM, Geschwind DH (2012) Autism genetics: searching for specificity and convergence. *Genome Biol* 13, 247

Berg JS, Brunetti-Pierri N, Peters SU, Kang SH, Fong CT, Salamone J, Freedenberg D, Hannig VL, Prock LA, Miller DT et al. (2007) Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams-Beuren syndrome region. *Genet Med* 9, 427-441

Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, Endris V, Roberts W, Szatmari P, Pinto D et al. (2010) Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat Genet* 42, 489-491

Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, Witherspoon K, Gerdts J, Baker C, Vulto-van Silfhout AT et al. (2014) Disruptive CHD8 Mutations Define a Subtype of Autism Early in Development. *Cell* 158, 263-276

Berryer MH, Hamdan FF, Klitten LL, Moller RS, Carmant L, Schwartzentruber J, Patry L, Dobrzeniecka S, Rochefort D, Neugnot-Cerioli M et al. (2013) Mutations in SYNGAP1 cause intellectual disability, autism, and a specific form of epilepsy by inducing haploinsufficiency. *Hum Mutat* 34, 385-394

Betancur C, Sakurai T, Buxbaum JD (2009) The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci* 32, 402-412

Betancur C (2011) Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* 1380, 42-77

Betancur C, Coleman M (2013) Etiological heterogeneity in autism spectrum disorders: role of rare variants, In *The Neuroscience of Autism Spectrum Disorders*, pp. 113-144

Bisgaard AM, Kirchhoff M, Nielsen JE, Kibaek M, Lund A, Schwartz M, Christensen E (2009) Chromosomal deletion unmasking a recessive disease: 22q13 deletion syndrome and metachromatic leukodystrophy. *Clin Genet* 75, 175-179

Bishop DV, Jacobs PA, Lachlan K, Wellesley D, Barnicoat A, Boyd PA, Fryer A, Middlemiss P, Smithson S, Metcalfe K et al. (2011) Autism, language and communication in children with sex chromosome trisomies. *Arch Dis Child* 96, 954-959

Blaker-Lee A, Gupta S, McCammon JM, De Rienzo G, Sive H (2012) Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes. *Dis Model Mech* 5, 834-851

Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, Saeed S, Hamilton-Shield J, Clayton-Smith J, O'Rahilly S et al. (2010) Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 463, 666-670

Boido D, Farisello P, Cesca F, Ferrea E, Valtorta F, Benfenati F, Baldelli P (2010) Cortico-hippocampal hyperexcitability in synapsin I/II/III knockout mice: age-dependency and response to the antiepileptic drug levetiracetam. *Neuroscience* 171, 268-283

Bolton P, Macdonald H, Pickles A, Rios P, Goode S, Crowson M, Bailey A, Rutter M (1994) A case-control family history study of autism. *J Child Psychol Psychiatry* 35, 877-900

Bonaglia MC, Giorda R, Borgatti R, Felisari G, Gagliardi C, Selicorni A, Zuffardi O (2001) Disruption of the ProSAP2 gene in a t(12;22)(q24.1;q13.3) is associated with the 22q13.3 deletion syndrome. *Am J Hum Genet* 69, 261-268

Bonaglia MC, Giorda R, Mani E, Aceti G, Anderlid BM, Baroncini A, Pramparo T, Zuffardi O (2006) Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3 deletion syndrome. *J Med Genet* 43, 822-828

Bonaglia MC, Giorda R, Beri S, De Agostini C, Novara F, Fichera M, Grillo L, Galesi O, Vetro A, Ciccone R et al. (2011) Molecular mechanisms generating and stabilizing terminal 22q13 deletions in 44 subjects with Phelan/McDermid syndrome. *PLoS Genet* 7, e1002173

Bonati MT, Russo S, Finelli P, Valsecchi MR, Cogliati F, Cavalleri F, Roberts W, Elia M, Larizza L (2007) Evaluation of autism traits in Angelman syndrome: a resource to unfold autism genes. *Neurogenetics* 8, 169-178

Bondy A, Frost L (2001) The Picture Exchange Communication System. *Behav Modif* 25, 725-744

Bozdagi O, Tavassoli T, Buxbaum JD (2013) Insulin-like growth factor-1 rescues synaptic and motor deficits in a mouse model of autism and developmental delay. *Mol Autism* 4, 9

Brugha TS, McManus S, Bankart J, Scott F, Purdon S, Smith J, Bebbington P, Jenkins R, Meltzer H (2011) Epidemiology of autism spectrum disorders in adults in the community in England. *Arch Gen Psychiatry* 68, 459-465

Burne T, Scott E, van Swinderen B, Hilliard M, Reinhard J, Claudianos C, Eyles D, McGrath J (2011) Big ideas for small brains: what can psychiatry learn from worms, flies, bees and fish? *Mol Psychiatry* 16, 7-16

Burnside RD, Pasion R, Mikhail FM, Carroll AJ, Robin NH, Youngs EL, Gadi IK, Keitges E, Jaswaney VL, Papenhausen PR et al. (2011) Microdeletion/microduplication of proximal 15q11.2 between BP1 and BP2: a susceptibility region for neurological dysfunction including developmental and language delay. *Hum Genet* 130, 517-528

Butler MG, Dasouki MJ, Zhou XP, Talebizadeh Z, Brown M, Takahashi TN, Miles JH, Wang CH, Stratton R, Pilarski R et al. (2005) Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J Med Genet* 42, 318-321

Buxbaum JD, Cai G, Chaste P, Nygren G, Goldsmith J, Reichert J, Anckarsater H, Rastam M, Smith CJ, Silverman JM et al. (2007) Mutation screening of the PTEN gene in patients with autism spectrum disorders and macrocephaly. *Am J Med Genet B Neuropsychiatr Genet* 144B, 484-491

Buxbaum JD, Daly MJ, Devlin B, Lehner T, Roeder K, State MW (2012) The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron* 76, 1052-1056

# C

Camacho A, Simon R, Sanz R, Vinuela A, Martinez-Salio A, Mateos F (2012) Cognitive and behavioral profile in females with epilepsy with PDCH19 mutation: two novel mutations and review of the literature. *Epilepsy Behav* 24, 134-137

Canitano R (2007) Epilepsy in autism spectrum disorders. *Eur Child Adolesc Psychiatry* 16, 61-66

Carney RM, Wolpert CM, Ravan SA, Shahbazian M, Ashley-Koch A, Cuccaro ML, Vance JM, Pericak-Vance MA (2003) Identification of MeCP2 mutations in a series of females with autistic disorder. *Pediatr Neurol* 28, 205-211

Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, Cook J, Khan A, Dorschner MO, Weaver M, Calvert S et al. (2013) Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nat Genet* 45, 825-830

Casey JP, Magalhaes T, Conroy JM, Regan R, Shah N, Anney R, Shields DC, Abrahams BS, Almeida J, Bacchelli E et al. (2012) A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum Genet* 131, 565-579

Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* 10, 295-304

Centers for Disease Control (CDC) (2012) Prevalence of autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *MMWR Surveill Summ* 61, 1-19

Cesca F, Baldelli P, Valtorta F, Benfenati F (2010) The synapsins: key actors of synapse function and plasticity. *Prog Neurobiol* 91, 313-348

Chaste P, Sanders SJ, Mohan KN, Klei L, Song Y, Murtha MT, Hus V, Lowe JK, Willsey AJ, Moreno-De-Luca D et al. (2014) Modest impact on risk for autism spectrum disorder of rare copy number variants at 15q11.2, specifically breakpoints 1 to 2. *Autism Res* 7, 355-362

Chess S (1971) Autism in children with congenital rubella. *J Autism Child Schizophr* 1, 33-47

Chess S, Fernandez P, Korn S (1978) Behavioral consequences of congenital rubella. *J Pediatr* 93, 699-703
Ching MS, Shen Y, Tan WH, Jeste SS, Morrow EM, Chen X, Mukaddes NM, Yoo SY, Hanson E, Hundley R et al. (2010) Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am J Med Genet B Neuropsychiatr Genet* 153B, 937-947
Chiyonobu T, Hayashi S, Kobayashi K, Morimoto M, Miyanomae Y, Nishimura A, Nishimoto A, Ito C, Imoto I, Sugimoto T et al. (2007) Partial tandem duplication of GRIA3 in a male with mental retardation. *Am J Med Genet A* 143A, 1448-1455
Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 106, 19096-19101
Christensen J, Gronborg TK, Sorensen MJ, Schendel D, Parner ET, Pedersen LH, Vestergaard M (2013) Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA* 309, 1696-1703
Classen CF, Riehmer V, Landwehr C, Kosfeld A, Heilmann S, Scholz C, Kabisch S, Engels H, Tierling S, Zivicnjak M et al. (2013) Dissecting the genotype in syndromic intellectual disability using whole exome sequencing in addition to genome-wide copy number analysis. *Hum Genet* 132, 825-841
Clement JP, Aceti M, Creson TK, Ozkan ED, Shi Y, Reish NJ, Almonte AG, Miller BH, Wiltgen BJ, Miller CA et al. (2012) Pathogenic SYNGAP1 mutations impair cognitive development by disrupting maturation of dendritic spine synapses. *Cell* 151, 709-723
Clifford S, Dissanayake C, Bui QM, Huggins R, Taylor AK, Loesch DZ (2007) Autism spectrum phenotype in males and females with fragile X full mutation and premutation. *J Autism Dev Disord* 37, 738-747
Cohen D, Lazar G, Couvert P, Desportes V, Lippe D, Mazet P, Heron D (2002) MECP2 mutation in a boy with language disorder and schizophrenia. *Am J Psychiatry* 159, 148-149
Constantino JN, Todd RD (2003) Autistic traits in the general population: a twin study. *Arch Gen Psychiatry* 60, 524-530
Constantino JN, Todd RD (2005) Intergenerational transmission of subthreshold autistic traits in the general population. *Biol Psychiatry* 57, 655-660
Constantino JN, Todorov A, Hilton C, Law P, Zhang Y, Molloy E, Fitzgerald R, Geschwind D (2013) Autism recurrence in half siblings: strong support for genetic mechanisms of transmission in ASD. *Mol Psychiatry* 18, 137-138
Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. *Nat Rev Genet* 10, 184-194
Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V et al. (2011) A copy number variation morbidity map of developmental delay. *Nat Genet* 43, 838-846
Corominas R, Yang X, Lin GN, Kang S, Shen Y, Ghamsari L, Broly M, Rodriguez M, Tam S, Trigg SA et al. (2014) Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat Commun* 5, 3650
Costain G, Lionel AC, Merico D, Forsythe P, Russell K, Lowther C, Yuen T, Husted J, Stavropoulos DJ, Speevak M et al. (2013) Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. *Hum Mol Genet* 22, 4485-4501
Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* 1, 40-47
Curatolo P, Bombardieri R, Jozwiak S (2008) Tuberous sclerosis. *Lancet* 372, 657-668

# D

Darnell JC, Van Driesche SJ, Zhang C, Hung KY, Mele A, Fraser CE, Stone EF, Chen C, Fak JJ, Chi SW et al. (2011) FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247-261
de Kovel CG, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, Kluck C, Muhle H, von Spiczak S, Ostertag P et al. (2010) Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain* 133, 23-32
de la Chapelle A, Herva R, Koivisto M, Aula P (1981) A deletion in chromosome 22 can cause DiGeorge syndrome. *Hum Genet* 57, 253-256
de Leeuw N, Dijkhuizen T, Hehir-Kwa JY, Carter NP, Feuk L, Firth HV, Kuhn RM, Ledbetter DH, Martin CL, van Ravenswaaij-Arts CM et al. (2012) Diagnostic interpretation of array data using public databases and internet sources. *Hum Mutat* 33, 930-940
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 et al. (2012) Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 367, 1921-1929
Depienne C, Arzimanoglou A, Trouillard O, Fedirko E, Baulac S, Saint-Martin C, Ruberg M, Dravet C, Nabbout R, Baulac M et al. (2006) Parental mosaicism can cause recurrent transmission of SCN1A mutations associated with severe myoclonic epilepsy of infancy. *Hum Mutat* 27, 389
Depienne C, Heron D, Betancur C, Benyahia B, Trouillard O, Bouteiller D, Verloes A, LeGuern E, Leboyer M, Brice A (2007) Autism, language delay and mental retardation in a patient with 7q11 duplication. *J Med Genet* 44, 452-458
Depienne C, Bouteiller D, Keren B, Cheuret E, Poirier K, Trouillard O, Benyahia B, Quelin C, Carpentier W, Julia S et al. (2009a) Sporadic infantile epileptic encephalopathy caused by mutations in PCDH19 resembles Dravet syndrome but mainly affects females. *PLoS Genet* 5, e1000381
Depienne C, Moreno-De-Luca D, Heron D, Bouteiller D, Gennetier A, Delorme R, Chaste P, Siffroi JP, Chantot-Bastaraud S, Benyahia B et al. (2009b) Screening for genomic rearrangements and methylation abnormalities of the 15q11-q13 region in autism spectrum disorders. *Biol Psychiatry* 66, 349-359
Depienne C, LeGuern E (2012) PCDH19-related infantile epileptic encephalopathy: an unusual X-linked inheritance disorder. *Hum Mutat* 33, 627-634
Devlin B, Melhem N, Roeder K (2011) Do common variants play a role in risk for autism? Evidence and theoretical musings. *Brain Res* 1380, 78-84
Doche ME, Bochukova EG, Su HW, Pearce LR, Keogh JM, Henning E, Cline JM, Saeed S, Dale A, Cheetham T et al. (2012) Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *J Clin Invest* 122, 4732-4736
Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, Nygren G, Rastam M, Gillberg IC, Anckarsater H et al. (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* 39, 25-27

# E

Ehninger D, Han S, Shilyansky C, Zhou Y, Li W, Kwiatkowski DJ, Ramesh V, Silva AJ (2008a) Reversal of learning deficits in a Tsc2+/- mouse model of tuberous sclerosis. *Nat Med* 14, 843-848
Ehninger D, Li W, Fox K, Stryker MP, Silva AJ (2008b) Reversing neurodevelopmental disorders in adults. *Neuron* 60, 950-960

Ehninger D, Silva AJ (2011) Rapamycin for treating Tuberous sclerosis and Autism spectrum disorders. *Trends Mol Med* 17, 78-87
Ekstrom AB, Hakenas-Plate L, Samuelsson L, Tulinius M, Wentz E (2008) Autism spectrum conditions in myotonic dystrophy type 1: a study on 57 individuals with congenital and childhood forms. *Am J Med Genet B Neuropsychiatr Genet* 147B, 918-926
Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467, 167-173
Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, Milh M, Kortum F, Fritsch A, Pientka FK et al. (2010) Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat Genet* 42, 1021-1026

# F

Fassio A, Patry L, Congia S, Onofri F, Piton A, Gauthier J, Pozzi D, Messa M, Defranchi E, Fadda M et al. (2011) SYN1 loss-of-function mutations in autism and partial epilepsy cause impaired synaptic function. *Hum Mol Genet* 20, 2297-2307
Feenstra I, Fang J, Koolen DA, Siezen A, Evans C, Winter RM, Lees MM, Riegel M, de Vries BB, Van Ravenswaaij CM et al. (2006) European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA); an online database for rare chromosome abnormalities. *Eur J Med Genet* 49, 279-291
Feenstra I, Hanemaaijer N, Sikkema-Raddatz B, Yntema H, Dijkhuizen T, Lugtenberg D, Verheij J, Green A, Hordijk R, Reardon W et al. (2011) Balanced into array: genome-wide array analysis in 54 patients with an apparently balanced de novo chromosome rearrangement and a meta-analysis. *Eur J Hum Genet* 19, 1152-1160
Fehr S, Wilson M, Downs J, Williams S, Murgia A, Sartori S, Vecchi M, Ho G, Polli R, Psoni S et al. (2013) The CDKL5 disorder is an independent clinical entity associated with early-onset encephalopathy. *Eur J Hum Genet* 21, 266-273
Fernandez BA, Roberts W, Chung B, Weksberg R, Meyn S, Szatmari P, Joseph-George AM, Mackay S, Whitten K, Noble B et al. (2010) Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. *J Med Genet* 47, 195-203
Fine SE, Weissman A, Gerdes M, Pinto-Martin J, Zackai EH, McDonald-McGinn DM, Emanuel BS (2005) Autism spectrum disorders and symptoms in children with molecularly confirmed 22q11.2 deletion syndrome. *J Autism Dev Disord* 35, 461-470
Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP (2009) DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 84, 524-533
Fischbach GD, Lord C (2010) The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* 68, 192-195
Flipsen-ten Berg K, van Hasselt PM, Eleveld MJ, van der Wijst SE, Hol FA, de Vroede MA, Beemer FA, Hochstenbach PF, Poot M (2007) Unmasking of a hemizygous WFS1 gene mutation by a chromosome 4p deletion of 8.3 Mb in a patient with Wolf-Hirschhorn syndrome. *Eur J Hum Genet* 15, 1132-1138
Folstein S, Rutter M (1977) Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry* 18, 297-321
Fombonne E (2009) Epidemiology of pervasive developmental disorders. *Pediatr Res* 65, 591-598
Fombonne E, Quirke S, Hagen A (2011) Epidemiology of pervasive developmental disorders, In Autism Spectrum Disorders, Amaral DG, Dawson G, and Geschwind DH (eds). *Oxford University Press*, 90-111
Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM et al. (2014) De novo mutations in schizophrenia implicate synaptic networks. *Nature* 506, 179-184

# G

Gamsiz ED, Viscidi EW, Frederick AM, Nagpal S, Sanders SJ, Murtha MT, Schmidt M, Triche EW, Geschwind DH, State MW et al. (2013) Intellectual disability is associated with increased runs of homozygosity in simplex autism. *Am J Hum Genet* 93, 103-109
Garcia CC, Blair HJ, Seager M, Coulthard A, Tennant S, Buddles M, Curtis A, Goodship JA (2004) Identification of a mutation in synapsin I, a synaptic vesicle protein, in a family with epilepsy. *J Med Genet* 41, 183-186
Gardener H, Spiegelman D, Buka SL (2011) Perinatal and neonatal risk factors for autism: a comprehensive meta-analysis. *Pediatrics* 128, 344-355
Garg S, Lehtonen A, Huson SM, Emsley R, Trump D, Evans DG, Green J (2013) Autism and other psychiatric comorbidity in neurofibromatosis type 1: evidence from a population-based study. *Dev Med Child Neurol* 55, 139-145
Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, Mahajan M, Manaa D, Pawitan Y, Reichert J et al. (2014) Most genetic risk for autism resides with common variation. *Nat Genet* 46, 881-885
Gauthier J, Champagne N, Lafreniere RG, Xiong L, Spiegelman D, Brustein E, Lapointe M, Peng H, Cote M, Noreau A et al. (2010a) De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. *Proc Natl Acad Sci USA* 107, 7863-7868
Gauthier J, Champagne N, Lafreniere RG, Xiong L, Spiegelman D, Brustein E, Lapointe M, Peng H, Cote M, Noreau A et al. (2010b) De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. *Proc Natl Acad Sci U S A* 107, 7863-7868
Ghosh A, Michalon A, Lindemann L, Fontoura P, Santarelli L (2013) Drug discovery for autism spectrum disorder: challenges and opportunities. *Nat Rev Drug Discov* 12, 777-790
Giannandrea M, Bianchi V, Mignogna ML, Sirri A, Carrabino S, D'Elia E, Vecellio M, Russo S, Cogliati F, Larizza L et al. (2010) Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. *Am J Hum Genet* 86, 185-195
Gillberg C, Cederlund M, Lamberg K, Zeijlon L (2006) Brief report: "the autism epidemic". The registered prevalence of autism in a Swedish urban area. *J Autism Dev Disord* 36, 429-435
Gillott A, Furniss F, Walter A (2001) Anxiety in high-functioning children with autism. *Autism* 5, 277-286
Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D (2011) Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 70, 898-907
Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, Vives L, Walsh T, McCarthy SE, Baker C et al. (2010) A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 42, 203-209
Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, Filipink RA, McConnell JS, Angle B, Meschino WS et al. (2012) Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med* 367, 1321-1331

Goldstein S, Schwebach AJ (2004) The comorbidity of Pervasive Developmental Disorder and Attention Deficit Hyperactivity Disorder: results of a retrospective chart review. *J Autism Dev Disord* 34, 329-339

Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S, Reymond A, Sun M, Sawa A, Gusella JF et al. (2012) KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 485, 363-367

Gong X, Jiang YW, Zhang X, An Y, Zhang J, Wu Y, Wang J, Sun Y, Liu Y, Gao X et al. (2012) High proportion of 22q13 deletions and SHANK3 mutations in Chinese patients with intellectual disability. *PLoS One* 7, e34739

Gratten J, Visscher PM, Mowry BJ, Wray NR (2013) Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease. *Nat Genet* 45, 234-238

Gratten J, Wray NR, Keller MC, Visscher PM (2014) Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat Neurosci* 17, 782-790

Grillo E, Lo Rizzo C, Bianciardi L, Bizzarri V, Baldassarri M, Spiga O, Furini S, De Felice C, Signorini C, Leoncini S et al. (2013) Revealing the complexity of a monogenic disease: Rett syndrome exome sequencing. *PLoS One* 8, e56599

Grozeva D, Carss K, Spasic-Boskovic O, Parker MJ, Archer H, Firth HV, Park SM, Canham N, Holder SE, Wilson M et al. (2014) De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. *Am J Hum Genet* 94, 618-624

Gu W, Zhang F, Lupski JR (2008) Mechanisms for human genomic rearrangements. *Pathogenetics* 1, 4

Guastella AJ, Einfeld SL, Gray KM, Rinehart NJ, Tonge BJ, Lambert TJ, Hickie IB (2010) Intranasal oxytocin improves emotion recognition for youth with autism spectrum disorders. *Biol Psychiatry* 67, 692-694

Guha S, Rees E, Darvasi A, Ivanov D, Ikeda M, Bergen SE, Magnusson PK, Cormican P, Morris D, Gill M et al. (2013) Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA Psychiatry* 70, 253-260

Guinchat V, Thorsen P, Laurent C, Cans C, Bodeau N, Cohen D (2012) Pre-, peri- and neonatal risk factors for autism. *Acta Obstet Gynecol Scand* 91, 287-300

Guy J, Gan J, Selfridge J, Cobb S, Bird A (2007) Reversal of neurological defects in a mouse model of Rett syndrome. *Science* 315, 1143-1147

# H

Haas RH (2010) Autism and mitochondrial disease. *Dev Disabil Res Rev* 16, 144-153

Hackett A, Tarpey PS, Licata A, Cox J, Whibley A, Boyle J, Rogers C, Grigg J, Partington M, Stevenson RE et al. (2010) CASK mutations are frequent in males and cause X-linked nystagmus and variable XLMR phenotypes. *Eur J Hum Genet* 18, 544-552

Hagerman R, Hoem G, Hagerman P (2010) Fragile X and autism: Intertwined at the molecular level leading to targeted treatments. *Mol Autism* 1, 12

Halgren C, Kjaergaard S, Bak M, Hansen C, El-Schich Z, Anderson C, Henriksen K, Hjalgrim H, Kirchhoff M, Bijlsma E et al. (2011) Corpus callosum abnormalities, intellectual disability, speech impairment, and autism in patients with haploinsufficiency of ARID1B. *Clin Genet*

Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, Miller J, Fedele A, Collins J, Smith K et al. (2011) Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry* 68, 1095-1102

Hamdan FF, Daoud H, Piton A, Gauthier J, Dobrzeniecka S, Krebs MO, Joober R, Lacaille JC, Nadeau A, Milunsky JM et al. (2011a) De novo SYNGAP1 mutations in nonsyndromic intellectual disability and autism. *Biol Psychiatry* 69, 898-901

Hamdan FF, Gauthier J, Araki Y, Lin DT, Yoshizawa Y, Higashi K, Park AR, Spiegelman D, Dobrzeniecka S, Piton A et al. (2011b) Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am J Hum Genet* 88, 306-316

Hamdan FF, Gauthier J, Dobrzeniecka S, Lortie A, Mottron L, Vanasse M, D'Anjou G, Lacaille JC, Rouleau GA, Michaud JL (2011c) Intellectual disability without epilepsy associated with STXBP1 disruption. *Eur J Hum Genet* 19, 607-609

Hannes FD, Sharp AJ, Mefford HC, de Ravel T, Ruivenkamp CA, Breuning MH, Fryns JP, Devriendt K, Van Buggenhout G, Vogels A et al. (2009) Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J Med Genet* 46, 223-232

Hanson E, Bernier R, Porche K, Jackson FI, Goin-Kochel RP, Snyder LG, Snow AV, Wallace AS, Campe KL, Zhang Y et al. (2014) The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population. *Biol Psychiatry* (Epub ahead of print)

Harrington JW, Rosen L, Garnecho A, Patrick PA (2006) Parental perceptions and use of complementary and alternative medicine practices for children with autistic spectrum disorders in private practice. *J Dev Behav Pediatr* 27, S156-161

He X, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS, Schellenberg GD, Gibbs RA, Daly MJ, Buxbaum JD et al. (2013) Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet* 9, e1003671

Heinzen EL, Radtke RA, Urban TJ, Cavalleri GL, Depondt C, Need AC, Walley NM, Nicoletti P, Ge D, Catarino CB et al. (2010) Rare deletions at 16p13.11 predispose to a diverse spectrum of sporadic epilepsy syndromes. *Am J Hum Genet* 86, 707-718

Heron B, Mikaeloff Y, Froissart R, Caridade G, Maire I, Caillaud C, Levade T, Chabrol B, Feillet F, Ogier H et al. (2011) Incidence and natural history of mucopolysaccharidosis type III in France and comparison with United Kingdom and Greece. *Am J Med Genet A* 155A, 58-68

Hinton VJ, Cyrulnik SE, Fee RJ, Batchelder A, Kiefel JM, Goldstein EM, Kaufmann P, De Vivo DC (2009) Association of autistic spectrum disorders with dystrophinopathies. *Pediatr Neurol* 41, 339-346

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. *Genet Med* 4, 45-61

Hoffmann TJ, Windham GC, Anderson M, Croen LA, Grether JK, Risch N (2014) Evidence of Reproductive Stoppage in Families With Autism Spectrum Disorder: A Large, Population-Based Cohort Study. *JAMA Psychiatry* 71, 943-951

Hogart A, Wu D, LaSalle JM, Schanen NC (2010) The comorbidity of autism with the genomic disorders of chromosome 15q11.2-q13. *Neurobiol Dis* 38, 181-191

Hoischen A, Krumm N, Eichler EE (2014) Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nat Neurosci* 17, 764-772

Hood RL, Lines MA, Nikkel SM, Schwartzentruber J, Beaulieu C, Nowaczyk MJ, Allanson J, Kim CA, Wieczorek D, Moilanen JS et al. (2012) Mutations in SRCAP, encoding SNF2-related CREBBP activator protein, cause Floating-Harbor syndrome. *Am J Hum Genet* 90, 308-313

Hoppman-Chaney N, Wain K, Seger PR, Superneau DW, Hodge JC (2013) Identification of single gene deletions at 15q13.3: further evidence that CHRNA7 causes the 15q13.3 microdeletion syndrome phenotype. *Clin Genet* 83, 345-351

Howlin P, Karpf J, Turk J (2005) Behavioural characteristics and autistic features in individuals with Cohen Syndrome. *Eur Child Adolesc Psychiatry* 14, 57-64

Huang N, Lee I, Marcotte EM, Hurles ME (2010) Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 6, e1001154

Hultman CM, Sandin S, Levine SZ, Lichtenstein P, Reichenberg A (2011) Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. *Mol Psychiatry* 16, 1203-1212

Hynes K, Tarpey P, Dibbens LM, Bayly MA, Berkovic SF, Smith R, Raisi ZA, Turner SJ, Brown NJ, Desai TD et al. (2010) Epilepsy and mental retardation limited to females with PCDH19 mutations can present de novo or in single generation families. *J Med Genet* 47, 211-216

# I

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36, 949-951

Ingason A, Kirov G, Giegling I, Hansen T, Isles AR, Jakobsen KD, Kristinsson KT, le Roux L, Gustafsson O, Craddock N et al. (2011a) Maternally derived microduplications at 15q11-q13: implication of imprinted genes in psychotic illness. *Am J Psychiatry* 168, 408-417

Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietilainen OP, Buizer-Voskamp JE, Strengman E, Francks C, Muglia P et al. (2011b) Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry* 16, 17-25

International Molecular Genetic Study of Autism Consortium (1998) A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Hum Mol Genet* 7, 571-578

International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455, 237-241

Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A et al. (2012) De novo gene disruptions in children on the autistic spectrum. *Neuron* 74, 285-299

# J

Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, Lyonnet S, Amiel J, Le Merrer M, Heron D, de Blois MC et al. (2006) Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J Med Genet* 43, 843-849

Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsesia A, Beckmann ND et al. (2011) Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 478, 97-102

Jacquemont S, Coe BP, Hersch M, Duyzend MH, Krumm N, Bergmann S, Beckmann JS, Rosenfeld JA, Eichler EE (2014) A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. *Am J Hum Genet* 94, 415-425

Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C et al. (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 34, 27-29

Jeste SS, Sahin M, Bolton P, Ploubidis GB, Humphrey A (2008) Characterization of autism in young children with tuberous sclerosis complex. *J Child Neurol* 23, 520-525

Jiang YH, Yuen RK, Jin X, Wang M, Chen N, Wu X, Ju J, Mei J, Shi Y, He M et al. (2013) Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am J Hum Genet* 93, 249-263

Johansson M, Rastam M, Billstedt E, Danielsson S, Stromland K, Miller M, Gillberg C (2006) Autism spectrum disorders and underlying brain pathology in CHARGE association. *Dev Med Child Neurol* 48, 40-50

Jongmans MC, Hoefsloot LH, van der Donk KP, Admiraal RJ, Magee A, van de Laar I, Hendriks Y, Verheij JB, Walpole I, Brunner HG et al. (2008) Familial CHARGE syndrome and the CHD7 gene: a recurrent missense mutation, intrafamilial recurrence and variability. *Am J Med Genet A* 146A, 43-50

# K

Kabashi E, Champagne N, Brustein E, Drapeau P (2010) In the swim of things: recent insights to neurogenetic disorders from zebrafish. *Trends Genet* 26, 373-381

Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, Moreno-De-Luca D, Moreno-De-Luca A, Mulle JG, Warren ST et al. (2011) An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med* 13, 777-784

Kanne SM, Abbacchi AM, Constantino JN (2009) Multi-informant ratings of psychiatric symptom severity in children with autism spectrum disorders: the importance of environmental context. *J Autism Dev Disord* 39, 856-864

Kanner L (1943) Autistic disturbance of affective contact. *Nervous Child* 2, 217-250

Karakoc E, Alkan C, O'Roak BJ, Dennis MY, Vives L, Mark K, Rieder MJ, Nickerson DA, Eichler EE (2012) Detection of structural variants and indels within exome data. *Nat Methods* 9, 176-178

Kielinen M, Rantala H, Timonen E, Linna SL, Moilanen I (2004) Associated medical disorders and disabilities in children with autistic disorder: a population-based study. *Autism* 8, 49-60

Kim DS, Ross PJ, Zaslavsky K, Ellis J (2014) Optimizing neuronal differentiation from induced pluripotent stem cells to model ASD. *Front Cell Neurosci* 8, 109

Kim MS, Akhtar MW, Adachi M, Mahgoub M, Bassel-Duby R, Kavalali ET, Olson EN, Monteggia LM (2012) An essential role for histone deacetylase 4 in synaptic plasticity and memory formation. *J Neurosci* 32, 10879-10886

Kim SH, Lord C (2012) New autism diagnostic interview-revised algorithms for toddlers and young preschoolers from 12 to 47 months of age. *J Autism Dev Disord* 42, 82-93

King M, Bearman P (2009) Diagnostic change and the increased prevalence of autism. *Int J Epidemiol* 38, 1224-1234

Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, Moran J, Chambert K, Toncheva D, Georgieva L et al. (2012) De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* 17, 142-153

Kleefstra T, van Zelst-Stams WA, Nillesen WM, Cormier-Daire V, Houge G, Foulds N, van Dooren M, Willemsen MH, Pfundt R, Turner A et al. (2009) Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. *J Med Genet* 46, 598-606

Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey AJ, Moreno-De-Luca D, Yu TW, Fombonne E, Geschwind D et al. (2012) Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism* 3, 9

Klein-Tasman BP, Phillips KD, Lord C, Mervis CB, Gallo FJ (2009) Overlap with the autism spectrum in young children with Williams syndrome. *J Dev Behav Pediatr* 30, 289-299

Kolevzon A (2013) Current Trends in the Pharmacological Treatment of Autism Disorders, In *The Neuroscience of Autism Spectrum Disorders*, pp. 85-101

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Wong WS et al. (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488, 471-475

Kou Y, Betancur C, Xu H, Buxbaum JD, Ma'ayan A (2012) Network- and attribute-based classifiers can prioritize genes and pathways for autism spectrum disorders and intellectual disability. *Am J Med Genet C Semin Med Genet* 160C, 130-142

Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128, 693-705

Krumm N, O'Roak BJ, Karakoc E, Mohajeri K, Nelson B, Vives L, Jacquemont S, Munson J, Bernier R, Eichler EE (2013) Transmission disequilibrium of small CNVs in simplex autism. *Am J Hum Genet* 93, 595-606

Krumm N, O'Roak BJ, Shendure J, Eichler EE (2014) A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci* 37, 95-105

Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, Gilliam TC, Nowak NJ, Cook EH, Jr., Dobyns WB et al. (2008) Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet* 17, 628-638

# L

Laje G, Morse R, Richter W, Ball J, Pao M, Smith AC (2010) Autism spectrum features in Smith-Magenis syndrome. *Am J Med Genet C Semin Med Genet* 154C, 456-462

Laumonnier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, Moizard MP, Raynaud M, Ronce N, Lemonnier E, Calvas P et al. (2004) X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet* 74, 552-557

Laumonnier F, Cuthbert PC, Grant SG (2007) The role of neuronal complexes in human X-linked brain diseases. *Am J Hum Genet* 80, 205-220

Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G, Konyukh M, Chaste P, Ey E, Rastam M et al. (2012) Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet* 8, e1002521

Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS et al. (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45, 984-994

Lemos RR, Souza MB, Oliveira JR (2012) Exploring the implications of INDELs in neuropsychiatric genetics: challenges and perspectives. *J Mol Neurosci* 47, 419-424

Leonard H, Wen X (2002) The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev* 8, 117-134

Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, Zhang N, Mowry BJ, Olincy A, Amin F et al. (2011) Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry* 168, 302-316

Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K et al. (2011) Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 70, 886-897

Lichtenstein P, Carlstrom E, Rastam M, Gillberg C, Anckarsater H (2010) The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *Am J Psychiatry* 167, 1357-1363

Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG, Neale BM, Kirby A, Ruderfer DM, Fromer M et al. (2013) Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 77, 235-242

Lin EI, Jeyifous O, Green WN (2013) CASK regulates SAP97 conformation and its interactions with AMPA and NMDA receptors. *J Neurosci* 33, 12067-12076

Longo N, Ardon O, Vanzo R, Schwartz E, Pasquali M (2011) Disorders of creatine transport and metabolism. *Am J Med Genet C Semin Med Genet* 157, 72-78

Lord C, Rutter M, Le Couteur A (1994) Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24, 659-685

Lord C, Risi S, Lambrecht L, Cook EH, Jr., Leventhal BL, DiLavore PC, Pickles A, Rutter M (2000) The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 30, 205-223

Losh M, Childress D, Lam K, Piven J (2008) Defining key features of the broad autism phenotype: a comparison across parents of multiple- and single-incidence autism families. *Am J Med Genet B Neuropsychiatr Genet* 147B, 424-433

Losick R, Desplan C (2008) Stochasticity and cell fate. *Science* 320, 65-68

Lotter V (1966) Epidemiology of autistic conditions in young children. *Social Psychiatry* 1, 124-135

Lovaas OI (1987) Behavioral treatment and normal educational and intellectual functioning in young autistic children. *J Consult Clin Psychol* 55, 3-9

Lowther C, Costain G, Stavropoulos DJ, Melvin R, Silversides CK, Andrade DM, So J, Faghfoury H, Lionel AC, Marshall CR et al. (2014) Delineating the 15q13.3 microdeletion phenotype: a case series and comprehensive review of the literature. *Genet Med* (Epub ahead of print)

Lundstrom S, Chang Z, Rastam M, Gillberg C, Larsson H, Anckarsater H, Lichtenstein P (2012) Autism spectrum disorders and autistic like traits: similar etiology in the extreme end and the normal variation. *Arch Gen Psychiatry* 69, 46-52

Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N, Chu SH, Klei L, Cai C, Ou J, Lowe JK et al. (2012) Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. *Am J Hum Genet* 91, 38-55

Lyall K, Constantino JN, Weisskopf MG, Roberts AL, Ascherio A, Santangelo SL (2014) Parental social responsiveness and risk of autism spectrum disorder in offspring. *JAMA Psychiatry* 71, 936-942

# M


Mabb AM, Judson MC, Zylka MJ, Philpot BD (2011) Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. *Trends Neurosci* 34, 293-303

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB et al. (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823-828

Malow B, Adkins KW, McGrew SG, Wang L, Goldman SE, Fawkes D, Burnette C (2012) Melatonin for sleep in children with autism: a controlled trial examining dose, tolerability, and outcomes. *J Autism Dev Disord* 42, 1729-1737; author reply 1738

Mandel JL, Chelly J (2004) Monogenic X-linked mental retardation: is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations. *Eur J Hum Genet* 12, 689-693

Marchetto MC, Carromeu C, Acab A, Yu D, Yeo GW, Mu Y, Chen G, Gage FH, Muotri AR (2010) A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* 143, 527-539

Mari F, Azimonti S, Bertani I, Bolognese F, Colombo E, Caselli R, Scala E, Longo I, Grosso S, Pescucci C et al. (2005) CDKL5 belongs to the same molecular pathway of MeCP2 and it is responsible for the early-onset seizure variant of Rett syndrome. *Hum Mol Genet* 14, 1935-1946

Marini C, Scheffer IE, Nabbout R, Mei D, Cox K, Dibbens LM, McMahon JM, Iona X, Carpintero RS, Elia M et al. (2009) SCN1A duplications and deletions detected in Dravet syndrome: Implications for molecular diagnosis. *Epilepsia* 50, 1670-1678

Marini C, Mei D, Parmeggiani L, Norci V, Calado E, Ferrari A, Moreira A, Pisano T, Specchio N, Vigevano F et al. (2010) Protocadherin 19 mutations in girls with infantile-onset epilepsy. *Neurology* 75, 646-653

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y et al. (2008) Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82, 477-488

Masurel-Paulet A, Andrieux J, Callier P, Cuisset JM, Le Caignec C, Holder M, Thauvin-Robinet C, Doray B, Flori E, Alex-Cordier MP et al. (2010) Delineation of 15q13.3 microdeletions. *Clin Genet* 78, 149-161

McBride KL, Varga EA, Pastore MT, Prior TW, Manickam K, Atkin JF, Herman GE (2010) Confirmation study of PTEN mutations among individuals with autism or developmental delays/mental retardation and macrocephaly. *Autism Res* 3, 137-141

McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, Perkins DO, Dickel DE, Kusenda M, Krastoshevsky O et al. (2009) Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* 41, 1223-1227

McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C, Shumilina S, Lasken RS, Vermeesch JR, Hall IM et al. (2013) Mosaic copy number variation in human neurons. *Science* 342, 632-637

McDonald-McGinn DM, Fahiminiya S, Revil T, Nowakowska BA, Suhl J, Bailey A, Mlynarski E, Lynch DR, Yan AC, Bilaniuk LT et al. (2013) Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and contribute to atypical findings in patients with 22q11.2DS. *J Med Genet* 50, 80-90

McGrath LM, Yu D, Marshall C, Davis LK, Thiruvahindrapuram B, Li B, Cappi C, Gerber G, Wolf A, Schroeder FA et al. (2014) Copy number variation in obsessive-compulsive disorder and tourette syndrome: a cross-disorder study. *J Am Acad Child Adolesc Psychiatry* 53, 910-919

Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, Huang S, Maloney VK, Crolla JA, Baralle D et al. (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359, 1685-1699

Mefford HC (2009) Genotype to phenotype-discovery and characterization of novel genomic disorders in a "genotype-first" era. *Genet Med* 11, 836-842

Mesibov GB, Shea V (2010) The TEACCH program in the era of evidence-based practice. *J Autism Dev Disord* 40, 570-579

Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, Jian M, Liu G, Greer D, Bhandari A et al. (2012) Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151, 1431-1442

Milh M, Villeneuve N, Chouchane M, Kaminska A, Laroche C, Barthez MA, Gitiaux C, Bartoli C, Borges-Correia A, Cacciagli P et al. (2011) Epileptic and nonepileptic features in patients with early onset epileptic encephalopathy and STXBP1 mutations. *Epilepsia* 52, 1828-1834

Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, Bridgemohan C, Cox GF, Dickinson H, Gentile J, Harris DJ et al. (2009) Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J Med Genet* 46, 242-248

Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, Church DM, Crolla JA, Eichler EE, Epstein CJ et al. (2010) Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 86, 749-764

Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, Zwaigenbaum L, Fernandez B, Roberts W, Szatmari P et al. (2007) Contribution of SHANK3 mutations to autism spectrum disorder. *Am J Hum Genet* 81, 1289-1297

Moreno-De-Luca D, Mulle JG, Kaminsky EB, Sanders SJ, Myers SM, Adam MP, Pakula AT, Eisenhauer NJ, Uhas K, Weik L et al. (2010) Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am J Hum Genet* 87, 618-630

Moreno-De-Luca D, Sanders SJ, Willsey AJ, Mulle JG, Lowe JK, Geschwind DH, State MW, Martin CL, Ledbetter DH (2013) Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. *Mol Psychiatry* 18, 1090-1095

Moreno-De-Luca D, Moreno-De-Luca A, Cubells JF, Sanders SJ (2014) Cross-disorder comparison of four neuropsychiatric CNV loci. *Curr Genet Med Rep* 2, 151-161

Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, Mukaddes NM, Balkhy S, Gascon G, Hashmi A et al. (2008) Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321, 218-223

Moss J, Howlin P, Magiati I, Oliver C (2012) Characteristics of autism spectrum disorder in Cornelia de Lange syndrome. *J Child Psychol Psychiatry* 53, 883-891

Mouridsen SE, Rich B, Isager T (1999) Epilepsy in disintegrative psychosis and infantile autism: a long-term validation study. *Dev Med Child Neurol* 41, 110-114

Mouridsen SE, Rich B, Isager T (2011) A longitudinal study of epilepsy and other central nervous system diseases in individuals with and without a history of infantile autism. *Brain Dev* 33, 361-366

# N


Na ES, Nelson ED, Kavalali ET, Monteggia LM (2013) The impact of MeCP2 loss- or gain-of-function on synaptic plasticity. *Neuropsychopharmacology* 38, 212-219

Najm J, Horn D, Wimplinger I, Golden JA, Chizhikov VV, Sudi J, Christian SL, Ullmann R, Kuechler A, Haas CA et al. (2008) Mutations of CASK cause an X-linked brain malformation phenotype with microcephaly and hypoplasia of the brainstem and cerebellum. *Nat Genet* 40, 1065-1067

Nakanishi M, Deardorff MA, Clark D, Levy SE, Krantz I, Pipan M (2012) Investigation of autistic features among individuals with mild to moderate Cornelia de Lange syndrome. *Am J Med Genet A* 158A, 1841-1847

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V et al. (2012a) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485, 242-245

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V et al. (2012b) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485, 242-245

Neul JL (2012) The relationship of Rett syndrome and MECP2 disorders to autism. *Dialogues Clin Neurosci* 14, 253-262

Niklasson L, Rasmussen P, Oskarsdottir S, Gillberg C (2009) Autism, ADHD, mental retardation and behavior problems in 100 individuals with 22q11 deletion syndrome. *Res Dev Disabil* 30, 763-773

Noh HJ, Ponting CP, Boulding HC, Meader S, Betancur C, Buxbaum JD, Pinto D, Marshall CR, Lionel AC, Scherer SW et al. (2013) Network topologies and convergent aetiologies arising from deletions and duplications observed in individuals with autism. *PLoS Genet* 9, e1003523

Novara F, Beri S, Giorda R, Ortibus E, Nageshappa S, Darra F, Bernardina BD, Zuffardi O, Van Esch H (2010) Refining the phenotype associated with MEF2C haploinsufficiency. *Clin Genet* 78, 471-477

Numis AL, Major P, Montenegro MA, Muzykewicz DA, Pulsifer MB, Thiele EA (2011) Identification of risk factors for autism spectrum disorders in tuberous sclerosis complex. *Neurology* 76, 981-987


# O


O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C et al. (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 43, 585-589

O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K et al. (2012a) Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338, 1619-1622

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD et al. (2012b) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485, 246-250

Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, Bryson S, Carver LJ, Constantino JN, Dobkins K et al. (2011) Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* 128, e488-495


# P


Pagnamenta AT, Holt R, Yusuf M, Pinto D, Wing K, Betancur C, Scherer SW, Volpi EV, Monaco AP (2011) A family with autism and rare copy number variants disrupting the Duchenne/Becker muscular dystrophy gene DMD and TRPM3. *J Neurodev Disord* 3, 124-131

Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, Horvath S, Geschwind DH (2013) Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 155, 1008-1021

Pasca SP, Portmann T, Voineagu I, Yazawa M, Shcheglovitov A, Pasca AM, Cord B, Palmer TD, Chikahisa S, Nishino S et al. (2011) Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nat Med* 17, 1657-1662

Pasca SP, Panagiotakos G, Dolmetsch RE (2014) Generating human neurons in vitro and using them to understand neuropsychiatric disease. *Annu Rev Neurosci* 37, 479-501

Pebrel-Richard C, Debost-Legrand A, Eymard-Pierre E, Greze V, Kemeny S, Gay-Bellile M, Gouas L, Tchirkov A, Vago P, Goumy C et al. (2013) An unusual clinical severity of 16p11.2 deletion syndrome caused by unmasked recessive mutation of CLN3. *Eur J Hum Genet* (Epub ahead of print)

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466, 368-372

Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, Xu X, Ziman R, Wang Z et al. (2014) Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 94, 677-694

Piton A, Michaud JL, Peng H, Aradhya S, Gauthier J, Mottron L, Champagne N, Lafreniere RG, Hamdan FF, Joober R et al. (2008) Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Hum Mol Genet* 17, 3965-3974

Piton A, Redin C, Mandel JL (2013) XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am J Hum Genet* 93, 368-383

Poultney CS, Goldberg AP, Drapeau E, Kou Y, Harony-Nicolas H, Kajiwara Y, De Rubeis S, Durand S, Stevens C, Rehnstrom K et al. (2013) Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. *Am J Hum Genet* 93, 607-619

Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748-752

Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, O'Dushlaine C, Chambert K, Bergen SE, Kahler A et al. (2014) A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 506, 185-190


# Q-R


Quintero-Rivera F, Sharifi-Hannauer P, Martinez-Agosto JA (2010) Autistic and psychiatric findings associated with the 3q29 microdeletion syndrome: case report and review. *Am J Med Genet A* 152A, 2459-2467

Rabbani B, Tekin M, Mahdieh N (2014) The promise of whole-exome sequencing in medical genetics. *J Hum Genet* 59, 5-15

Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216-226

Raj A, Rifkin SA, Andersen E, van Oudenaarden A (2010) Variability in gene expression underlies incomplete penetrance. *Nature* 463, 913-918

Ramalingam A, Zhou XG, Fiedler SD, Brawner SJ, Joyce JM, Liu HY, Yu S (2011) 16p13.11 duplication is a risk factor for a wide spectrum of neuropsychiatric disorders. *J Hum Genet* 56, 541-544

Ramocki MB, Tavyev YJ, Peters SU (2010) The MECP2 duplication syndrome. *Am J Med Genet A* 152A, 1079-1088

Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, Albrecht B, Bartholdi D, Beygo J, Di Donato N et al. (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 380, 1674-1682

Reddy KS (2005) Cytogenetic abnormalities and fragile-X syndrome in Autism Spectrum Disorder. *BMC Med Genet* 6, 3

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W et al. (2006) Global variation in copy number in the human genome. *Nature* 444, 444-454

Reinthaler EM, Lal D, Lebon S, Hildebrand MS, Dahl HH, Regan BM, Feucht M, Steinbock H, Neophytou B, Ronen GM et al. (2014) 16p11.2 600 kb Duplications confer risk for typical and atypical Rolandic epilepsy. *Hum Mol Genet* (Epub ahead of print)

Richardson SR, Morell S, Faulkner GJ (2014) L1 Retrotransposons and Somatic Mosaicism in the Brain. *Annu Rev Genet* (Epub ahead of print)

Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M et al. (2013) Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* 45, 1150-1159

Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, Cook EH, Jr., Leventhal BL, Pickles A (2006) Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry* 45, 1094-1103

Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM (1985) Concordance for the syndrome of autism in 40 pairs of afflicted twins. *Am J Psychiatry* 142, 74-77

Ritvo ER, Jorde LB, Mason-Brothers A, Freeman BJ, Pingree C, Jones MB, McMahon WM, Petersen PB, Jenson WR, Mo A (1989) The UCLA-University of Utah epidemiologic survey of autism: recurrence risk estimates and genetic counseling. *Am J Psychiatry* 146, 1032-1036

Rivet TT, Matson JL (2011) Review of gender differences in core symptomatology in autism spectrum disorders. *Res Autism Spect Dis* 5, 957-976

Robinson EB, Lichtenstein P, Anckarsater H, Happe F, Ronald A (2013) Examining and interpreting the female protective effect against autistic behavior. *Proc Natl Acad Sci U S A* 110, 5258-5262

Ronan JL, Wu W, Crabtree GR (2013) From neural development to cognition: unexpected roles for chromatin. *Nat Rev Genet* 14, 347-359

Ronemus M, Iossifov I, Levy D, Wigler M (2014) The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* 15, 133-141

Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA (2009) Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch Pediatr Adolesc Med* 163, 907-914

Rosenfeld JA, Coppinger J, Bejjani BA, Girirajan S, Eichler EE, Shaffer LG, Ballif BC (2010) Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11.2 microdeletions and microduplications. *J Neurodev Disord* 2, 26-38

Ross JL, Roeltgen DP, Kushner H, Zinn AR, Reiss A, Bardsley MZ, McCauley E, Tartaglia N (2012) Behavioral and social phenotypes in boys with 47,XYY syndrome or 47,XXY Klinefelter syndrome. *Pediatrics* 129, 769-778

Rujescu D, Ingason A, Cichon S, Pietilainen OP, Barnes MR, Toulopoulou T, Picchioni M, Vassos E, Ettinger U, Bramon E et al. (2009) Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* 18, 988-996

Russo S, Marchi M, Cogliati F, Bonati MT, Pintaudi M, Veneselli E, Saletti V, Balestrini M, Ben-Zeev B, Larizza L (2009) Novel mutations in the CDKL5 gene, predicted effects and associated phenotypes. *Neurogenetics* 10, 241-250

# S

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928-933

Saghatelyan AK, Nikonenko AG, Sun M, Rolf B, Putthoff P, Kutsche M, Bartsch U, Dityatev A, Schachner M (2004) Reduced GABAergic transmission and number of hippocampal perisomatic inhibitory synapses in juvenile mice deficient in the neural cell adhesion molecule L1. *Mol Cell Neurosci* 26, 191-203

Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnstrom K, Mallick S, Kirby A et al. (2014) A framework for the interpretation of de novo mutation in human disease. *Nat Genet* (Epub ahead of print)

Sampson MG, Coughlin CR, 2nd, Kaplan P, Conlin LK, Meyers KE, Zackai EH, Spinner NB, Copelovitch L (2010) Evidence for a recurrent microdeletion at chromosome 16p11.2 associated with congenital anomalies of the kidney and urinary tract (CAKUT) and Hirschsprung disease. *Am J Med Genet A* 152A, 2618-2622

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA et al. (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70, 863-885

Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL et al. (2012a) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485, 237-241

Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, Dilullo NM, Parikshak NN, Stein JL et al. (2012b) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485, 237-241

Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A (2014) The familial risk of autism. *JAMA* 311, 1770-1777

Sando R, 3rd, Gounko N, Pieraut S, Liao L, Yates J, 3rd, Maximov A (2012) HDAC4 governs a transcriptional program essential for synaptic plasticity and memory. *Cell* 151, 821-834

Santen GW, Aten E, Sun Y, Almomani R, Gilissen C, Nielsen M, Kant SG, Snoeck IN, Peeters EA, Hilhorst-Hofstee Y et al. (2012) Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat Genet* 44, 379-380

Sasaki M, Lange J, Keeney S (2010) Genome destabilization by homologous recombination in the germ line. *Nat Rev Mol Cell Biol* 11, 182-195

Schaaf CP, Boone PM, Sampath S, Williams C, Bader PI, Mueller JM, Shchelochkov OA, Brown CW, Crawford HP, Phalen JA et al. (2012) Phenotypic spectrum and genotype-phenotype correlations of NRXN1 exon deletions. *Eur J Hum Genet* 20, 1240-1247

Schaefer GB, Mendelsohn NJ (2013) Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. *Genet Med* 15, 399-407

Schafer MK, Altevogt P (2010) L1CAM malfunction in the nervous system and human carcinomas. *Cell Mol Life Sci* 67, 2425-2437
Schizophrenia Psychiatric Genome-Wide Association Study Consortium (2011) Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 43, 969-976
Schorry EK, Keddache M, Lanphear N, Rubinstein JH, Srodulski S, Fletcher D, Blough-Pfau RI, Grabowski GA (2008) Genotype-phenotype correlations in Rubinstein-Taybi syndrome. *Am J Med Genet A* 146A, 2512-2519
Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M et al. (2004) Large-scale copy number polymorphism in the human genome. *Science* 305, 525-528
Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J et al. (2007) Strong association of de novo copy number mutations with autism. *Science* 316, 445-449
Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC et al. (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 38, 1038-1042
Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R et al. (2008) A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* 40, 322-328
Shattuck PT (2006) The contribution of diagnostic substitution to the growing administrative prevalence of autism in US special education. *Pediatrics* 117, 1028-1037
Shcheglovitov A, Shcheglovitova O, Yazawa M, Portmann T, Shu R, Sebastiano V, Krawisz A, Froehlich W, Bernstein JA, Hallmayer JF et al. (2013) SHANK3 and IGF1 restore synaptic deficits in neurons from 22q13 deletion syndrome patients. *Nature* 503, 267-271
Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, Caronna EB, Miller KJ, Frazier JA, Silverstein I, Picker J et al. (2010) Clinical genetic testing for patients with autism spectrum disorders. *Pediatrics* 125, e727-735
Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudbridge F, Holmans PA, Whittemore AS, Mowry BJ et al. (2009) Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 460, 753-757
Shinawi M, Schaaf CP, Bhatt SS, Xia Z, Patel A, Cheung SW, Lanpher B, Nagl S, Herding HS, Nevinny-Stickel C et al. (2009) A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes. *Nat Genet* 41, 1269-1271
Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA, Probst FJ, Craigen WJ, Graham BH, Pursley A et al. (2010) Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J Med Genet* 47, 332-341
Sidorov MS, Auerbach BD, Bear MF (2013) Fragile X mental retardation protein and synaptic plasticity. *Mol Brain* 6, 15
Sikora DM, Pettit-Kekel K, Penfield J, Merkens LS, Steiner RD (2006) The near universal presence of autism spectrum disorders in children with Smith-Lemli-Opitz syndrome. *Am J Med Genet A* 140, 1511-1518
Simonati A, Boaretto F, Vettori A, Dabrilli P, Criscuolo L, Rizzuto N, Mostacciuolo ML (2006) A novel missense mutation in the L1CAM gene in a boy with L1 disease. *Neurol Sci* 27, 114-117
Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G (2008) Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry* 47, 921-929
Snow AV, Lecavalier L, Houts C (2009) The structure of the Autism Diagnostic Interview-Revised: diagnostic and phenotypic implications. *J Child Psychol Psychiatry* 50, 734-742
Somerville MJ, Mervis CB, Young EJ, Seo EJ, del Campo M, Bamforth S, Peregrine E, Loo W, Lilley M, Perez-Jurado LA et al. (2005) Severe expressive-language delay related to duplication of the Williams-Beuren locus. *N Engl J Med* 353, 1694-1701
Soorya L, Kolevzon A, Zweifach J, Lim T, Dobry Y, Schwartz L, Frank Y, Wang AT, Cai G, Parkhomenko E et al. (2013) Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. *Mol Autism* 4, 18
Sousa SB, Abdul-Rahman OA, Bottani A, Cormier-Daire V, Fryer A, Gillessen-Kaesbach G, Horn D, Josifova D, Kuechler A, Lees M et al. (2009) Nicolaides-Baraitser syndrome: Delineation of the phenotype. *Am J Med Genet A* 149A, 1628-1640
Spiegel EK, Colman RF, Patterson D (2006) Adenylosuccinate lyase deficiency. *Mol Genet Metab* 89, 19-31
Splawski I, Timothy KW, Sharpe LM, Decher N, Kumar P, Bloise R, Napolitano C, Schwartz PJ, Joseph RM, Condouris K et al. (2004) Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell* 119, 19-31
Spooren W, Lindemann L, Ghosh A, Santarelli L (2012) Synapse dysfunction in autism: a molecular medicine approach to drug discovery in neurodevelopmental disorders. *Trends Pharmacol Sci* 33, 669-684
Stankiewicz P, Kulkarni S, Dharmadhikari AV, Sampath S, Bhatt SS, Shaikh TH, Xia Z, Pursley AN, Cooper ML, Shinawi M et al. (2012) Recurrent deletions and reciprocal duplications of 10q11.21q11.23 including CHAT and SLC18A3 are likely mediated by complex low-copy repeats. *Hum Mutat* 33, 165-179
State MW (2010) Another piece of the autism puzzle. *Nat Genet* 42, 478-479
Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE et al. (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* 455, 232-236
Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietilainen OP, Mors O, Mortensen PB et al. (2009) Common variants conferring risk of schizophrenia. *Nature* 460, 744-747
Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, Bohman M (1989) A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *J Child Psychol Psychiatry* 30, 405-416
Stevens SA, Nash K, Koren G, Rovet J (2013) Autism characteristics in children with fetal alcohol spectrum disorders. *Child Neuropsychol* 19, 579-587
Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, Stephan DA, Morton DH (2006) Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *N Engl J Med* 354, 1370-1377
Su SC, Tsai LH (2011) Cyclin-dependent kinases in brain development and disease. *Annu Rev Cell Dev Biol* 27, 465-491
Sudhof TC (2008) Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* 455, 903-911
Sweeten TL, Posey DJ, McDougle CJ (2004) Brief report: autistic disorder in three children with cytomegalovirus infection. *J Autism Dev Disord* 34, 583-586
Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L et al. (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39, 319-328

# T


Tabet AC, Pilorge M, Delorme R, Amsellem F, Pinard JM, Leboyer M, Verloes A, Benzacken B, Betancur C (2012) Autism multiplex family with 16p11.2p12.2 microduplication syndrome in monozygotic twins and distal 16p11.2 deletion in their brother. *Eur J Hum Genet* 20, 540-546

Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM et al. (2012) Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 149, 525-537

Tan WH, Bird LM, Thibert RL, Williams CA (2014) If not Angelman, what is it? A review of Angelman-like syndromes. *Am J Med Genet A* 164A, 975-992

Taniai H, Nishiyama T, Miyachi T, Imaeda M, Sumi S (2008) Genetic influences on the broad spectrum of autism: study of proband-ascertained twins. *Am J Med Genet B Neuropsychiatr Genet* 147B, 844-849

Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C, Dicks E, Menzies A et al. (2009) A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet* 41, 535-543

Thomas NS, Sharp AJ, Browne CE, Skuse D, Hardie C, Dennis NR (1999) Xp deletions associated with autism in three females. *Hum Genet* 104, 43-48

Tierney E, Nwokoro NA, Porter FD, Freund LS, Ghuman JK, Kelley RI (2001) Behavior phenotype in the RSH/Smith-Lemli-Opitz syndrome. *Am J Med Genet* 98, 191-200

Tordjman S, Anderson GM, Botbol M, Toutain A, Sarda P, Carlier M, Saugier-Veber P, Baumann C, Cohen D, Lagneaux C et al. (2012) Autistic disorder in patients with Williams-Beuren syndrome: a reconsideration of the Williams-Beuren syndrome phenotype. *PLoS One* 7, e30778

Treadwell-Deering DE, Powell MP, Potocki L (2010) Cognitive and behavioral characterization of the Potocki-Lupski syndrome (duplication 17p11.2). *J Dev Behav Pediatr* 31, 137-143

Tropeano M, Ahn JW, Dobson RJ, Breen G, Rucker J, Dixit A, Pal DK, McGuffin P, Farmer A, White PS et al. (2013) Male-biased autosomal effect of 16p13.11 copy number variation in neurodevelopmental disorders. *PLoS One* 8, e61365

Tsai L, Stewart MA, August G (1981) Implication of sex differences in the familial transmission of infantile autism. *J Autism Dev Disord* 11, 165-173

Tuchman R, Rapin I (2002) Epilepsy in autism. *Lancet Neurol* 1, 352-358

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D et al. (2005) Fine-scale structural variation of the human genome. *Nat Genet* 37, 727-732


# U


Uddin M, Tammimies K, Pellecchia G, Alipanahi B, Hu P, Wang Z, Pinto D, Lau L, Nalpathamkalam T, Marshall CR et al. (2014) Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. *Nat Genet* 46, 742-747

Ullmann R, Turner G, Kirchhoff M, Chen W, Tonge B, Rosenberg C, Field M, Vianna-Morgante AM, Christie L, Krepischi-Santos AC et al. (2007) Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat* 28, 674-682

Urbach A, Bar-Nur O, Daley GQ, Benvenisty N (2010) Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells. *Cell Stem Cell* 6, 407-411


# V


Valnegri P, Montrasio C, Brambilla D, Ko J, Passafaro M, Sala C (2011) The X-linked intellectual disability protein IL1RAPL1 regulates excitatory synapse formation by binding PTPdelta and RhoGAP2. *Hum Mol Genet* 20, 4797-4809

van Bon BW, Mefford HC, Menten B, Koolen DA, Sharp AJ, Nillesen WM, Innis JW, de Ravel TJ, Mercer CL, Fichera M et al. (2009) Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. *J Med Genet* 46, 511-523

Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, Romano C, Delle Chiaie B, Mortier G, Menten B et al. (2009) Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome. *Eur J Med Genet* 52, 94-100

Van Houdt JK, Nowakowska BA, Sousa SB, van Schaik BD, Seuntjens E, Avonce N, Sifrim A, Abdul-Rahman OA, van den Boogaard MJ, Bottani A et al. (2012) Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat Genet* 44, 445-449, S441

van Kuilenburg AB, Meijer J, Mul AN, Hennekam RC, Hoovers JM, de Die-Smulders CE, Weber P, Mori AC, Bierau J, Fowler B et al. (2009) Analysis of severely affected patients with dihydropyrimidine dehydrogenase deficiency reveals large intragenic rearrangements of DPYD and a de novo interstitial deletion del(1)(p13.3p21.3). *Hum Genet* 125, 581-590

Veenstra-Vanderweele J, Christian SL, Cook EH, Jr. (2004) Autism as a paradigmatic complex genetic disorder. *Annu Rev Genomics Hum Genet* 5, 379-405

Veltman JA, Brunner HG (2010) Understanding variable expressivity in microdeletion syndromes. *Nat Genet* 42, 192-193

Veltman MW, Thompson RJ, Craig EE, Dennis NR, Roberts SE, Moore V, Brown JA, Bolton PF (2005) A paternally inherited duplication in the Prader-Willi/Angelman syndrome critical region: a case and family study. *J Autism Dev Disord* 35, 117-127

Verhoeven WM, Egger JI, Willemsen MH, de Leijer GJ, Kleefstra T (2012) Phelan-McDermid syndrome in two adult brothers: atypical bipolar disorder as its psychopathological phenotype? *Neuropsychiatr Dis Treat* 8, 175-179

Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M et al. (2010) A de novo paradigm for mental retardation. *Nat Genet* 42, 1109-1112

Vorstman JA, Morcus ME, Duijff SN, Klaassen PW, Heineman-de Boer JA, Beemer FA, Swaab H, Kahn RS, van Engeland H (2006) The 22q11.2 deletion in children: high rate of autistic disorders and early onset of psychotic symptoms. *J Am Acad Child Adolesc Psychiatry* 45, 1104-1113

# W


Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, Falchi M, Chen F, Andrieux J, Lobbens S et al. (2010) A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 463, 671-675

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280, 1077-1082

Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PM et al. (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459, 528-533

Wassink TH, Piven J, Patil SR (2001) Chromosomal abnormalities in a clinic sample of individuals with autistic disorder. *Psychiatr Genet* 11, 57-63

Weischenfeldt J, Symmons O, Spitz F, Korbel JO (2013) Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* 14, 125-138

Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MA, Green T et al. (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358, 667-675

Weiss LA, Arking DE, Gene Discovery Project of Johns Hopkins & the Autism Consortium, Daly MJ, Chakravarti A (2009) A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461, 802-808

Werling DM, Geschwind DH (2013) Sex differences in autism spectrum disorders. *Curr Opin Neurol* 26, 146-153

Wheelwright S, Baron-Cohen S, Goldenfeld N, Delaney J, Fine D, Smith R, Weil L, Wakabayashi A (2006) Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Res* 1079, 47-56

White SM, Morgan A, Da Costa A, Lacombe D, Knight SJ, Houlston R, Whiteford ML, Newbury-Ecob RA, Hurst JA (2010) The phenotype of Floating-Harbor syndrome in 10 patients. *Am J Med Genet A* 152A, 821-829

Wieczorek D, Bogershausen N, Beleggia F, Steiner-Haldenstatt S, Pohl E, Li Y, Milz E, Martin M, Thiele H, Altmuller J et al. (2013) A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. *Hum Mol Genet* 22, 5121-5135

Willatt L, Cox J, Barber J, Cabanas ED, Collins A, Donnai D, FitzPatrick DR, Maher E, Martin H, Parnau J et al. (2005) 3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome. *Am J Hum Genet* 77, 154-160

Willemsen MH, Vulto-van Silfhout AT, Nillesen WM, Wissink-Lindhout WM, van Bokhoven H, Philip N, Berry-Kravis EM, Kini U, van Ravenswaaij-Arts CM, Delle Chiaie B et al. (2012) Update on Kleefstra Syndrome. *Mol Syndromol* 2, 202-212

Williams JG, Higgins JP, Brayne CE (2006) Systematic review of prevalence studies of autism spectrum disorders. *Arch Dis Child* 91, 8-15

Williams NM, Franke B, Mick E, Anney RJ, Freitag CM, Gill M, Thapar A, O'Donovan MC, Owen MJ, Holmans P et al. (2012) Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: the role of rare variants and duplications at 15q13.3. *Am J Psychiatry* 169, 195-204

Williams SR, Aldred MA, Der Kaloustian VM, Halal F, Gowans G, McLeod DR, Zondag S, Toriello HV, Magenis RE, Elsea SH (2010) Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *Am J Hum Genet* 87, 219-228

Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA et al. (2013) Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155, 997-1007

Wimmer H, Perner J (1983) Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103-128

Wing L, Gould J (1979) Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *J Autism Dev Disord* 9, 11-29

Wing L (1981) Asperger's syndrome: a clinical account. *Psychol Med* 11, 115-129

Wong CC, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC, Plomin R, Mill J (2014) Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry* 19, 495-503

Wu Y, Arai AC, Rumbaugh G, Srivastava AK, Turner G, Hayashi T, Suzuki E, Jiang Y, Zhang L, Rodriguez J et al. (2007) Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. *Proc Natl Acad Sci U S A* 104, 18163-18168


# Y-Z


Yamashita Y, Fujimoto C, Nakajima E, Isagai T, Matsuishi T (2003) Possible association between congenital cytomegalovirus infection and autistic disorder. *J Autism Dev Disord* 33, 455-459

Young HK, Barton BA, Waisbren S, Portales Dale L, Ryan MM, Webster RI, North KN (2008) Cognitive and psychological profile of males with Becker muscular dystrophy. *J Child Neurol* 23, 155-162

Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, Schmitz-Abe K, Harmin DA, Adli M, Malik AN et al. (2013) Using whole-exome sequencing to identify inherited causes of autism. *Neuron* 77, 259-273

Zafeiriou DI, Ververi A, Dafoulis V, Kalyva E, Vargiami E (2013) Autism spectrum disorders: the quest for genetic syndromes. *Am J Med Genet B Neuropsychiatr Genet* 162B, 327-366

Zhang C, Milunsky JM, Newton S, Ko J, Zhao G, Maher TA, Tager-Flusberg H, Bolliger MF, Carter AS, Boucard AA et al. (2009) A neuroligin-4 missense mutation associated with autism impairs neuroligin-4 folding and endoplasmic reticulum export. *J Neurosci* 29, 10843-10854

Zhou J, Blundell J, Ogawa S, Kwon CH, Zhang W, Sinton C, Powell CM, Parada LF (2009) Pharmacological inhibition of mTORC1 suppresses anatomical, cellular, and behavioral abnormalities in neural-specific Pten knock-out mice. *J Neurosci* 29, 1773-1783

Zhubi A, Cook EH, Guidotti A, Grayson DR (2014) Epigenetic mechanisms in autism spectrum disorder. *Int Rev Neurobiol* 115, 203-244

Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, Bijlsma EK, Oortveld MA, Ekici AB, Reis A et al. (2009) CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. *Am J Hum Genet* 85, 655-666

Zweier M, Gregor A, Zweier C, Engels H, Sticht H, Wohlleber E, Bijlsma EK, Holder SE, Zenker M, Rossier E 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. *Hum Mutat* 31, 722-733

# Annexes

# Liste des publications


Pinto D, **Delaby E**, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, Xu X, Ziman R, Wang Z, Vorstman JA, Thompson A, Regan R, Pilorge M, Pellecchia G, Pagnamenta AT, Oliveira B, Marshall CR, Magalhaes TR, Lowe JK, Howe JL, Griswold AJ, Gilbert J, Duketis E, Dombroski BA, De Jonge MV, Cuccaro M, Crawford EL, Correia CT, Conroy J, Conceição IC, Chiocchetti AG, Casey JP, Cai G, Cabrol C, Bolshakova N, Bacchelli E, Anney R, Gallinger S, Cotterchio M, Casey G, Zwaigenbaum L, Wittemeyer K, Wing K, Wallace S, van Engeland H, Tryfon A, Thomson S, Soorya L, Rogé B, Roberts W, Poustka F, Mouga S, Minshew N, McInnes LA, McGrew SG, Lord C, Leboyer M, Le Couteur AS, Kolevzon A, Jiménez González P, Jacob S, Holt R, Guter S, Green J, Green A, Gillberg C, Fernandez BA, Duque F, Delorme R, Dawson G, Chaste P, Café C, Brennan S, Bourgeron T, Bolton PF, Bölte S, Bernier R, Baird G, Bailey AJ, Anagnostou E, Almeida J, Wijsman EM, Vieland VJ, Vicente AM, Schellenberg GD, Pericak-Vance M, Paterson AD, Parr JR, Oliveira G, Nurnberger JI, Monaco AP, Maestrini E, Klauck SM, Hakonarson H, Haines JL, Geschwind DH, Freitag CM, Folstein SE, Ennis S, Coon H, Battaglia A, Szatmari P, Sutcliffe JS, Hallmayer J, Gill M, Cook EH, Buxbaum JD, Devlin B, Gallagher L, Betancur C, Scherer SW. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* (2014) 94(5), 677-94 **(Article 1)**

Celestino-Soper PB, Violante S, Crawford EL, Luo R, Lionel AC, **Delaby E**, Cai G, Sadikovic B, Lee K, Lo C, Gao K, Person RE, Moss TJ, German JR, Huang N, Shinawi M, Treadwell-Deering D, Szatmari P, Roberts W, Fernandez B, Schroer RJ, Stevenson RE, Buxbaum JD, Betancur C, Scherer SW, Sanders SJ, Geschwind DH, Sutcliffe JS, Hurles ME, Wanders RJ, Shaw CA, Leal SM, Cook EH Jr, Goin-Kochel RP, Vaz FM, Beaudet AL. A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism. *Proc Natl Acad Sci USA* (2012) 109(21), 7974-81 **(Annexe, Article 2)**

Tabet AC, Verloes A, Pilorge M, **Delaby E**, Delorme R, Nygren G, Devillard F, Gérard M, Passemard S, Jacquette A, Delahaye A, Perrin L, Dupont C, Aboura A, Bitoun P, Coleman M, Leboyer M, Gillberg C, Benzacken B, Betancur C. Complex nature of apparently balanced chromosomal rearrangements in patients with autism spectrum disorder (*soumis pour publication*) **(Annexe, Article 3)**

Article 3

# *A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism*

Celestino-Soper PB, Violante S, Crawford EL, Luo R, Lionel AC, **Delaby E**, Cai G, Sadikovic B, Lee K, Lo C, Gao K, Person RE, Moss TJ, German JR, Huang N, Shinawi M, Treadwell-Deering D, Szatmari P, Roberts W, Fernandez B, Schroer RJ, Stevenson RE, Buxbaum JD, Betancur C, Scherer SW, Sanders SJ, Geschwind DH, Sutcliffe JS, Hurles ME, Wanders RJ, Shaw CA, Leal SM, Cook EH Jr, Goin-Kochel RP, Vaz FM, Beaudet AL

*Proc Natl Acad Sci USA* (2012) 109(21), 7974-81

# A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism

Patrícia B. S. Celestino-Soper[a,1], Sara Violante[b,c,1], Emily L. Crawford[d], Rui Luo[e], Anath C. Lionel[f], Elsa Delaby[g], Guiqing Cai[h], Bekim Sadikovic[a], Kwanghyuk Lee[a], Charlene Lo[a], Kun Gao[e], Richard E. Person[a], Timothy J. Moss[a], Jennifer R. German[a], Ni Huang[i], Marwan Shinawi[a,j,2], Diane Treadwell-Deering[j,k], Peter Szatmari[l], Wendy Roberts[m], Bridget Fernandez[n], Richard J. Schroer[o], Roger E. Stevenson[o], Joseph D. Buxbaum[h], Catalina Betancur[g], Stephen W. Scherer[f,m], Stephan J. Sanders[p], Daniel H. Geschwind[e], James S. Sutcliffe[d], Matthew E. Hurles[i], Ronald J. A. Wanders[b], Chad A. Shaw[a], Suzanne M. Leal[a], Edwin H. Cook, Jr.[q], Robin P. Goin-Kochel[a,j,r], Frédéric M. Vaz[b,1], and Arthur L. Beaudet[a,j,r,1,3]

Departments of [a]Molecular and Human Genetics, [k]Psychiatry, and [r]Pediatrics, Baylor College of Medicine, Houston, TX 77030; [j]Texas Children's Hospital, Houston, TX 77030; [b]Laboratory Genetic Metabolic Disease, Departments of Clinical Chemistry and Pediatrics, Academic Medical Center, University of Amsterdam, 1105 AZ, Amsterdam, The Netherlands; [c]Metabolism and Genetics Group, Research Institute for Medicines and Pharmaceutical Sciences (iMed.UL), Faculdade de Farmácia, Universidade de Lisboa, 1649-003 Lisbon, Portugal; [d]Department of Molecular Physiology and Biophysics, Center for Molecular Neuroscience, Vanderbilt University, Nashville, TN 37232; [e]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095; [f]Centre for Applied Genomics and Program in Genetics and Genome Biology, [m]Hospital for Sick Children, Toronto, ON, Canada M5G 1X8; [g]Institut National de la Santé et de la Recherche Médicale U952, Centre National de la Recherche Scientifique Unité Mixte de Recherche 7224, and Université Pierre et Marie Curie, University of Paris 6, Paris 94010, France; [h]Seaver Autism Center for Research and Treatment, Department of Psychiatry, and Friedman Brain Institute, Mount Sinai School of Medicine, New York, NY 10029; [i]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; [l]The Offord Centre for Child Studies, McMaster Children's Hospital and Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada L8S 4L8; [n]Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St. John's, NF, Canada A1B 3V6; [o]Greenwood Genetic Center, Greenwood, SC 29646; [p]Program on Neurogenetics, Child Study Center and Departments of Psychiatry and Genetics, Yale University School of Medicine, New Haven, CT 06520; and [q]Institute for Juvenile Research, Department of Psychiatry, University of Illinois, Chicago, IL 60608

This article is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2011.

Edited by Helen H. Hobbs, University of Texas Southwestern Medical Center, Dallas, TX, and approved March 27, 2012 (received for review December 20, 2011)

**We recently reported a deletion of exon 2 of the trimethyllysine hydroxylase epsilon (*TMLHE*) gene in a proband with autism. *TMLHE* maps to the X chromosome and encodes the first enzyme in carnitine biosynthesis, 6-*N*-trimethyllysine dioxygenase. Deletion of exon 2 of *TMLHE* causes enzyme deficiency, resulting in increased substrate concentration (6-*N*-trimethyllysine) and decreased product levels (3-hydroxy-6-*N*-trimethyllysine and γ-butyrobetaine) in plasma and urine. *TMLHE* deficiency is common in control males (24 in 8,787 or 1 in 366) and was not significantly increased in frequency in probands from simplex autism families (9 in 2,904 or 1 in 323). However, it was 2.82-fold more frequent in probands from male-male multiplex autism families compared with controls (7 in 909 or 1 in 130; *P* = 0.023). Additionally, six of seven autistic male siblings of probands in male-male multiplex families had the deletion, suggesting that *TMLHE* deficiency is a risk factor for autism (metaanalysis Z-score = 2.90 and *P* = 0.0037), although with low penetrance (2–4%). These data suggest that dysregulation of carnitine metabolism may be important in nondysmorphic autism; that abnormalities of carnitine intake, loss, transport, or synthesis may be important in a larger fraction of nondysmorphic autism cases; and that the carnitine pathway may provide a novel target for therapy or prevention of autism.**

The role of carnitine in biology and disease has been studied for decades (1, 2). Carnitine has been proposed to be a conditionally essential nutrient, and even termed vitamin $B_T$. Carnitine content of foods varies widely, being very low in fruits, vegetables, and grains; intermediate in milk products, eggs, chicken, and fish; and very high in red meats. The proportion of carnitine derived from the diet varies widely in humans, being quite low in vegetarians and especially low with a vegan diet that excludes dairy products and eggs. In contrast, about 75% of carnitine is derived from the diet in meat eaters. Carnitine homeostasis in humans (Fig. 1) is maintained by a modest rate of endogenous synthesis, absorption from dietary sources, and efficient tubular reabsorption by the kidney. Apart from the dietary intake, carnitine is synthesized in humans in kidney, liver, and brain from protein-derived 6-*N*-trimethyllysine (TML) via 3-hydroxy-6-*N*-trimethyllysine (HTML), 4-*N*-trimethylaminobutyraldehyde (TMABA), and 4-*N*-trimethylaminobutyric acid [γ-butyrobetaine (γBB)] (3). Renal resorption plays an important role in carnitine metabolism, with considerable excretion if carnitine intake is abundant, but there is extremely efficient resorption if body stores of carnitine are low. Carnitine is present as free carnitine and as acylcarnitines, of which the latter reflect the cellular acyl-CoA ester pool. Up to 99% of carnitine is intracellular and is essential for mitochondrial function, where its role is to enable transport of fatty acids into mitochondria, where β-oxidation takes place (3); it is also involved in the transport of peroxisomal oxidation products to the mitochondria.

Carnitine deficiency can develop secondary to dietary inadequacy or as an adverse effect of medical treatments. Although humans can synthesize carnitine, nutritional deficiency can occur, as when infants were fed early preparations of soy formulas that were deficient in carnitine (4). Similarly, deficiency can arise with

Author contributions: P.B.S.C.-S., S.V., S.J.S., R.J.A.W., C.A.S., F.M.V., and A.L.B. designed research; P.B.S.C.-S., S.V., E.L.C., R.L., A.C.L., B.S., K.L., C.L., R.E.P., T.J.M., J.R.G., N.H., S.J.S., C.A.S., F.M.V., and A.L.B. performed research; K.G., C.A.S., and S.M.L. contributed new reagents/analytic tools; P.B.S.C.-S., S.V., E.L.C., R.L., A.C.L., E.D., G.C., B.S., K.L., R.E.P., T.J.M., N.H., M.S., D.T.-D., P.S., W.R., B.F., R.J.S., R.E.S., J.D.B., C.B., S.W.S., S.J.S., D.H.G., J.S.S., M.E.H., R.J.A.W., C.A.S., S.M.L., E.H.C., R.P.G.-K., F.M.V., and A.L.B. analyzed data; and P.B.S.C.-S., S.V., F.M.V., and A.L.B. wrote the paper.

Conflict of interest statement: Multiple authors are based in the Department of Molecular and Human Genetics at Baylor College of Medicine, which offers extensive genetic laboratory testing, and Baylor College of Medicine derives revenue from this activity. The department offers biochemical and molecular diagnostic testing for trimethyllysine hydroxylase, epsilon gene deficiency. P.B.S.C.-S., S.V., F.M.V., and A.L.B. have filed a patent related to some of the work reported.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]P.B.S.C.-S., S.V., F.M.V., and A.L.B. contributed equally to this work.

[2]Present address: Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110.

[3]To whom correspondence should be addressed. E-mail: abeaudet@bcm.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1120210109/-/DCSupplemental.

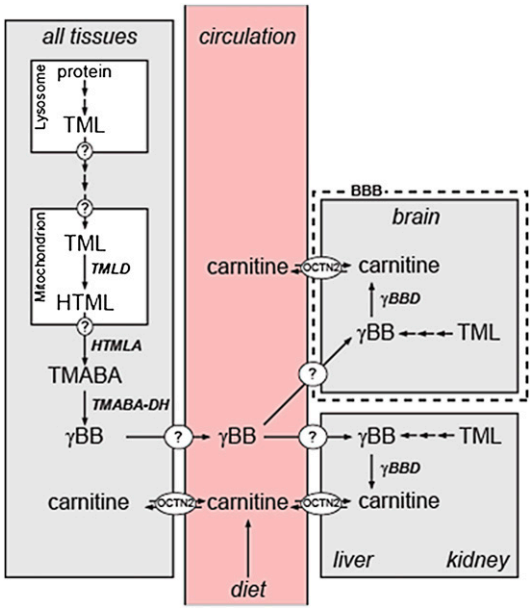



**Fig. 1.** Carnitine biosynthesis and homeostasis in humans. Carnitine is synthesized in four enzymatic steps. After release of TML by lysosomal protein degradation, this compound is hydroxylated by TMLD, producing HTML. HTML is cleaved by HTML aldolase (HTMLA) into TMABA and glycine. Subsequently, TMABA is oxidized by TMABA-dehydroxygenase (TMABA-DH) to form 4-*N*-trimethylaminobutyrate, also named γ-butyrobetaine (γBB). Finally, γBB is hydroxylated by γBBD, yielding L-carnitine. Because TMLD is located in mitochondria, TML needs to be transported out of the lysosome and across the inner mitochondrial membrane into the mitochondrial matrix by means of transporters, which are unknown at this time. Depending on the subcellular localization of the HTLMA (also uncertain, likely the cytosol), HTML or γBB needs to be transported back to the cytosol (where γBBD is located). In cells that do not contain γBBD, γBB is exported from the cell and imported into tissues (liver, kidney, and brain in humans) that do express γBBD by means of at least one specific transporter, presumably *SLC6A13*. Carnitine is transported by OCTN2 and other lower affinity transporters (not shown).

parenteral alimentation in neonates (5). Carnitine deficiency can also occur secondary to administration of pivalate-conjugated antibiotics or valproic acid (2). Various disease processes and medical interventions, such as renal tubular disorders and chronic hemodialysis, respectively, can also be associated with carnitine deficiency.

There are primary and secondary genetic forms of carnitine deficiency (6). Secondary deficiency is caused by various fatty oxidation defects and organic acidemias that lead to carnitine deficiency through urinary loss of acylcarnitines that accumulate related to the enzyme deficiency. Primary systemic carnitine deficiency is caused by biallelic loss-of-function mutations in the *SLC22A5* gene that encodes the plasma membrane organic cation transporter-2 (OCTN2). OCTN2 deficiency is characterized by excessive urinary loss of carnitine, leading to systemic deficiency with associated skeletal myopathy, cardiomyopathy, fatty liver, and hypoglycemia. Although the possibility of a primary systemic carnitine deficiency caused by a defect in carnitine biosynthesis was postulated long ago, no primary disorders of carnitine biosynthesis have been described until now (7).

Administration of carnitine is the centerpiece of therapy for systemic carnitine deficiency, and it is beneficial in some genetic forms of secondary carnitine deficiency. Administration of carnitine and acetylcarnitine has been explored as an antioxidant and for treatment of many disorders, including diabetic peripheral neuropathy (8), heart failure (9), and mitochondrial disorders.

Recently, we discovered a deletion of the 6-*N*-trimethyllysine dioxygenase (TMLD) gene [also known as trimethyllysine hydroxylase, epsilon (*TMLHE*)] while studying probands with autism, raising the question of whether there might be an association of autism with *TMLHE* mutations (10). *TMLHE* maps to the long arm of the X chromosome near the boundary of the pseudoautosomal region and encodes TMLD. TMLD is the first enzyme of the carnitine biosynthesis pathway (3) and is localized in mitochondria (11).

The etiology of severe, dysmorphic autism with a male/female ratio of 3.2:1 (12) has become increasingly well defined as often attributable to de novo mutations or recent mutations transmitted for a few generations. These mutations include large copy number variants (CNVs), which are detectable by chromosomal microarray analysis in up to 25% of the most severe cases with phenotypes that include major intellectual disability, which restricts reproduction (13). De novo point mutations are also being discovered as causes of autism using next-generation sequencing of genomic DNA (14). Disease-causing CNVs are found in ∼10% of patients with intermediate phenotypes, often with less severe intellectual disability (15, 16). In these cases, penetrance may be incomplete and the phenotype can be highly variable, with diagnoses of intellectual disability, autism, schizophrenia, and idiopathic epilepsy seen with the same CNV (17–19). These examples typify single-locus conditions, perhaps with genetic and nongenetic modifier effects. Autism spectrum disorders and related neurocognitive phenotypes blend into even more complex genotype-phenotype relationships with evidence for two-hit or two-locus pathogenesis (20). At the milder end of the autism spectrum are patients who often have speech, have an intelligence quotient (IQ) ranging from low to the normal range, and are nondysmorphic. This milder population, some of whom meet diagnostic criteria for Asperger syndrome, can display up to an 8:1 male/female ratio (21, 22), and will be referred to herein as having nondysmorphic autism (NDA). This includes the milder portion of the autism spectrum, but patients who have NDA can have severe cognitive and behavioral phenotypes. The etiology of NDA remains almost completely unknown, but the extreme sex ratio may provide a clue as to its etiology.

In this paper, we show that *TMLHE* deficiency is a very common inborn error of metabolism in males and suggest that it may be significantly more frequent in autistic male-male sib pairs than in controls.

## Results

**Deletions of Exon 2 Are Heterogeneous and Common in Autistic and Healthy Males.** Given the discovery of a deletion of exon 2 in *TMLHE* in a male simplex proband with autism (10), we examined the frequency of *TMLHE* mutations in autism and control populations. We studied simplex families primarily from the Simons Simplex Collection (SSC) and multiplex families primarily from the Autism Genetic Resource Exchange (AGRE) collection, and we recruited multiple collaborators to study additional families (Table 1). We have now identified a total of 16 male autism probands, six affected male siblings of probands, and 24 healthy adult males with deletions of exon 2, indicating that this is a relatively common CNV (Table 1). For 28 of the 29 deletions characterized more thoroughly, size ranged from 5.7 to 15.9 kb and only exon 2 was deleted; one additional deletion of 59.6 kb removed exons 2–6 (Fig. 2 *A* and *B* and Table S1). Based on position, size, and sequence, there was a minimum of 14 different deletion junctions among 29 unrelated families. Sequencing of the breakpoints of many deletions showed that almost all junctions occurred in long interspersed elements and short interspersed elements in the introns flanking exon 2 (Fig. 2*B* and Table S2), as has been seen in other loci (23). For all SSC samples (probands, heterozygous mothers, and healthy fathers), the deletions were present in both DNA extracted directly from blood and DNA extracted from lymphoblastoid cell lines (LCLs).

**Table 1. Sources of male autism and control samples and methods of testing**

| | No. | Deletion |
|---|---|---|
| Simplex autism | | |
| SSC | 1,887 | 6* |
| SCAP | 80 | 0 |
| Houston | 24 | 0 |
| Toronto | 328 | 3 |
| Paris | 333 | 0 |
| New York | 252 | 0 |
| Totals | 2,904 | 9 |
| Male probands from male-male sibling pairs† | | |
| AGRE + NIMH | 752 | 7 |
| Toronto | 93 | 0 |
| Paris | 53 | 0 |
| New York | 11 | 0 |
| Totals | 909 | 7 |
| Male probands from male-female sibling pairs† | | |
| Paris | 38 | 0 |
| New York | 5 | 0 |
| Controls | | |
| SSC, NIMH, and AGRE autism fathers | 2,197 | 7 |
| NIMH controls | 897 | 3 |
| BPR | 49 | 1 |
| Houston fathers | 36 | 0 |
| Multiplex fathers | 615 | 0 |
| WTCCC | 3,018 | 9 |
| Toronto | 1,975 | 4 |
| Totals | 8,787 | 24 |

SCAP, South Carolina Autism Project; WTCCC, Wellcome Trust Case–Control Consortium.

*Screening for the deletion and confirmation were performed as described in *Materials and Methods*.

†Excludes affected sibling.

In addition, we identified an extremely common intronic deletion (Fig. 2*A*) that appeared on the basis of comparative genomic hybridization (CGH) array to be indistinguishable in all cases. This is equivalent to Database of Genomic Variants numbers 115349 and 104572, which appear to be identical, and/or to number 97130, which is very similar. The intronic deletion was present in 74% of 93 autism male probands and 71% of 48 control males examined, and it should not be misinterpreted as causing enzyme deficiency. This intronic deletion was present on 24 of 29 chromosomes from unrelated families with exon 2 deletion of *TMLHE* (Table S1).

Genomic sequencing of exons for *TMLHE* is complicated by the presence of two pseudoexons (7aP and 8aP) that are highly homologous to exons 7a and 8a, and are imbedded in a large inverted repeat downstream of *TMLHE* (24) (Fig. 2*C*). In addition, there are two alternative exons 7b and 8b, which are located between the two inverted repeats and have sequences unrelated to 7a and 8a. Sequencing of exons 1–8 of *TMLHE* from genomic DNA in 536 SSC autism male probands, 98 AGRE probands, and 443 National Institute of Mental Health (NIMH) controls identified very few point mutations (Table S3). In addition to exon 2 deletion, sequencing in a multiplex AGRE family (AU 0177) identified an arginine-to-glutamine change in codon 241 (R241Q) in the mother and unaffected half-brother of the two autistic males (Fig. S1). More recently, we have been able to study plasma from 156 male SSC probands for carnitine biosynthesis metabolites. We identified one male (Table S3) with biochemical abnormalities similar to those described below, and sequencing identified an R70H mutation likely causing *TMLHE* deficiency.

**Exon 2 Deletion Results in Loss of TMLD Activity and Absence of TMLD Protein.** The functional effect of deletion of exon 2 of *TMLHE* was examined as TMLD enzyme activity based on its role in carnitine biosynthesis. Cultured LCLs from males with deletion of exon 2 had low or undetectable TMLD enzyme activity, and heterozygous mothers had reduced activity compared with healthy males (Fig. 3*A*). Results from family AU 0177 were complex, with the two affected brothers with deletion of exon 2 having very low or undetectable enzyme activity but the unaffected mother and half-brother showing low activity but higher than that of the affected brothers (Fig. 3 *A* and *B*). The mother is a compound heterozygote and transmitted the R241Q mutation to the unaffected half-brother of the siblings with autism. Cells with exon 2 deletion also lacked immunodetectable protein by Western blot analysis (Fig. 3*C*). Many of the control and autism cell lines in Fig. 3*D* have the intron 1 deletion, and many do not.

Analysis of RNA from LCLs using RT-PCR revealed low levels of skipping of exon 2 in most samples and a stable transcript with complete absence of exon 2 in cells from males with deletion of this exon (Fig. S2*A*). Thus, nonsense-mediated decay is not prominent for the exon 2 deletion transcript. This was confirmed using a quantitative RT-PCR assay, which showed normal levels of transcript for the exon 5/6 junction in all samples but complete absence of the exon 1/2 junction in deletion samples (Fig. S2*B*).

**Diagnostic Metabolite Abnormalities in Plasma and Urine.** The two affected brothers from AGRE family AU 0177 had a normal facial appearance in childhood and were otherwise nondysmorphic. They both had normal plasma free carnitine levels (33 and 34 μmol/L, normal = 22–65 μmol/L) at the recent ages of 15 and 17 y. In urine of the affected brothers, HTML and γBB were undetectable and the excretion of TML was threefold that of controls (Fig. 4*A*). Plasma from the two brothers and from 5 SSC probands showed a significant increase in TML, complete absence of HTML, and severely reduced levels of γBB, except for one case (Fig. 4*B*). The (HTML + γBB)/TML ratio was very low in patients compared with control plasma and may be an excellent index of TMLD activity (Fig. 4*C*). These data indicate that *TMLHE* deficiency represents a unique inborn error of carnitine biosynthesis. To search for other evidence of a common *TMLHE* deficiency and for other defects in carnitine biosynthesis, urine from 29 SSC probands who did not have any known *TMLHE* mutations was studied and did not reveal any abnormal carnitine metabolites.

**Sex Ratio in NDA Is Not Caused by a Common Inherited Mutation in *TMLHE*.** It was important to determine whether there was a common mutation or epigenetic mechanism causing TMLD enzyme deficiency, and perhaps explaining the male predominance in some forms of autism. To address the possibility of a common but difficult to detect inherited mutation, we analyzed SNP data from Illumina arrays on 411 AGRE families; this revealed no evidence of linkage for *TMLHE*, which is the most telomeric gene on Xq that is not on Y, or for *VAMP7*, which is nearby but across the pseudoautosomal boundary and present on both the X and Y chromosomes. For the *TMLHE* region, a maximum nonparametric linkage score of 1.25 and logarithm of odds (LOD) score of 0.34 were observed for markers located within and flanking the *TMLHE* gene. For the *VAMP7* gene region, a maximum nonparametric linkage score of 0.76 and LOD score of 0.15 were observed. Thus, there was no evidence for linkage at either locus. This is not surprising, given the extensive genetic heterogeneity in autism.

***TMLHE* Deficiency Likely Is a Risk Factor for Autism.** Because deletion of exon 2 was more common than any other mutations detectable by genomic sequencing, and because it was associated with loss of enzyme activity, it was expedient to analyze a large series of autism cases and controls for exon 2 deletion. A PCR assay

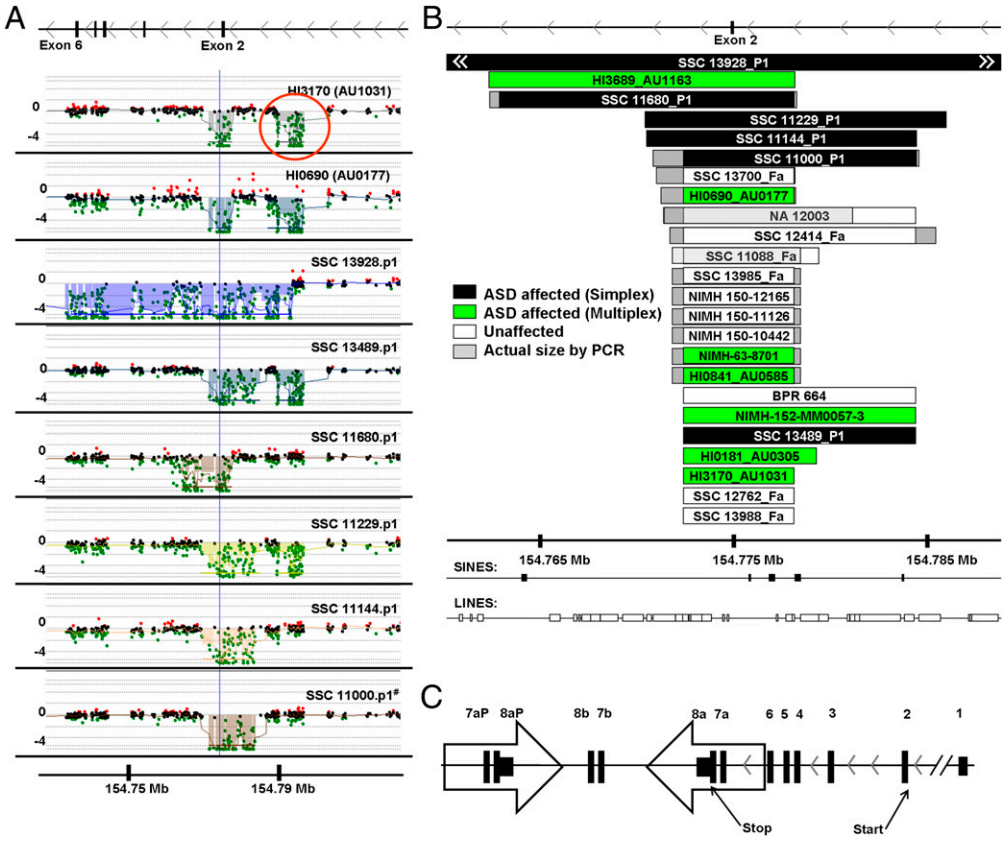



**Fig. 2.** Exon 2-containing deletions found in *TMLHE* and structure of *TMLHE*. (*A*) Array CGH showing eight exonic deletions in *TMLHE*, with the relative location of *TMLHE* exons 2–6 aligned above the array CGH plots (GRCh37/hg19 assembly; http://genome.ucsc.edu). The horizontal axis shows chromosome position, and the vertical axis shows the $\log_2$ ratio of array signal. Semitransparent filled boxes on CGH plots highlight the regions of aberration; all samples have deletions involving exon 2, and most have a separate deletion in intron 1 (red circle). All samples are autism probands with identifiers found in Tables S1, S4, and S5. (*B*) Twenty-four unrelated individuals with deletions involving exon 2 of *TMLHE* are mapped. Deletion coordinates were determined by array CGH unless specified by PCR assay. White arrowheads indicate the continuation of the deletion for SSC 13928.p1. NA 12003 is an unaffected individual whose deletion is published (38) and was better characterized in this study. BPR 664 is an unaffected individual. (*C*) Diagram of gene structure. Large open arrows represent near-identical inverted repeats. Fa, father; P1, proband; HI, AGRE individuals; #, individual first described by Celestino-Soper et al. (10).

was designed with primers slightly outside the boundaries of exon 2 to give a product of 538 bp in normal males but no product for males with deletion of exon 2. An example of the PCR assay with an internal control product is shown in Fig. S3. If a sample failed to give a PCR product for exon 2, the presence or absence of the deletion was then confirmed using CGH custom array with densely spaced oligonucleotides interrogating the *TMLHE* region, as shown in Fig. 2*A*. For PCR analysis, we focused entirely on males because the assay did not reliably detect the deletion in heterozygous females.

Using the PCR assay, we tested simplex male probands primarily from the SSC with lesser numbers of probands from the South Carolina Autism Project and from probands from Houston, TX. We also tested male controls, including SSC fathers of autism probands, NIMH controls, and Baylor Polymorphism Resource (BPR) controls. With the collaboration of the laboratories of two of the authors (J.S.S. and D.H.G.), we tested multiplex probands (here, multiplex refers to male-male sibling pairs both affected with autism) from the AGRE, NIMH, and Nashville collections. We subsequently developed collaborations with the laboratories of four of the authors (S.W.S., C.B., J.D.B., and M.E.H.) to expand the data for exon 2 deletion in autism male probands and control males. Some collaborating laboratories used quantitative PCR (qPCR) or existing Affymetrix 6.0 array data as the primary test for deletion of exon 2, as specified in Table 1. All deletion probands were validated, and approximate coordinates were determined using the custom *TMLHE* array. Deletions in control males from the laboratories of S.W.S. and M.E.H. were not validated.

Comparison of the data for male probands from simplex families (9 in 2,904 or 1 in 323 deleted) with all controls (24 in 8,787 or 1 in 366 deleted) did not provide evidence for an association ($P = 0.44$) (Table 1). One SSC proband (11680.p1) also had deletion of chromosome 16p11.2 and was eliminated from these calculations and from the phenotypic data. We hypothesized for numerous reasons reviewed in the discussion that the frequency of exon 2 deletion might be substantially higher in probands from male-male affected sib-pairs. The frequency of exon 2 deletion was 2.85-fold higher in these multiplex probands (7 in 909 or 1 in 130) compared with all male controls, with a $P$ value of 0.023 (Table 1). For each multiplex family, only one affected male (identified as the proband) was tested initially. All genotypes were consistent with the X-linked inheritance; all 17 mothers of probands with deletion of exon 2 were tested and were heterozygous for the deletion, confirming that the deletions were not cell culture artifacts.

We next examined the affected male siblings of the multiplex male probands and found that six of the seven had the same deletion as the proband. Based on analysis using the Transmission Disequilibrium Test (TDT), the probability of obtaining this result, if there were no association, is 0.012. Using metaanalysis, we calculated a Stouffer's $z$-statistic to combine the data for multiplex probands compared with control males and data from the TDT. We obtained a Z-score of 2.90 and a $P$ value of 0.0037 using Stouffer's method, suggesting that *TMLHE* deficiency is a risk factor for autism. If the data from the simplex families are included in the metaanalysis, the Z-score is –2.81 and the $P$ value is 0.0051, which is only slightly higher than the $P$ value of 0.0037.

**Cognitive Function of *TMLHE*-Deficient Males with Autism Varies Widely.** Significant phenotypic information was available for seven SSC probands, two of three Canadian Genetic probands, seven SSC unaffected fathers, all seven multiplex probands and six affected brothers, and 3 NIMH control males with deletion of exon 2 of *TMLHE* (Tables S4 and S5). The levels of cognitive and language functioning varied considerably across patients. The full-scale IQ of autistic males with deletion of exon 2 ranged from

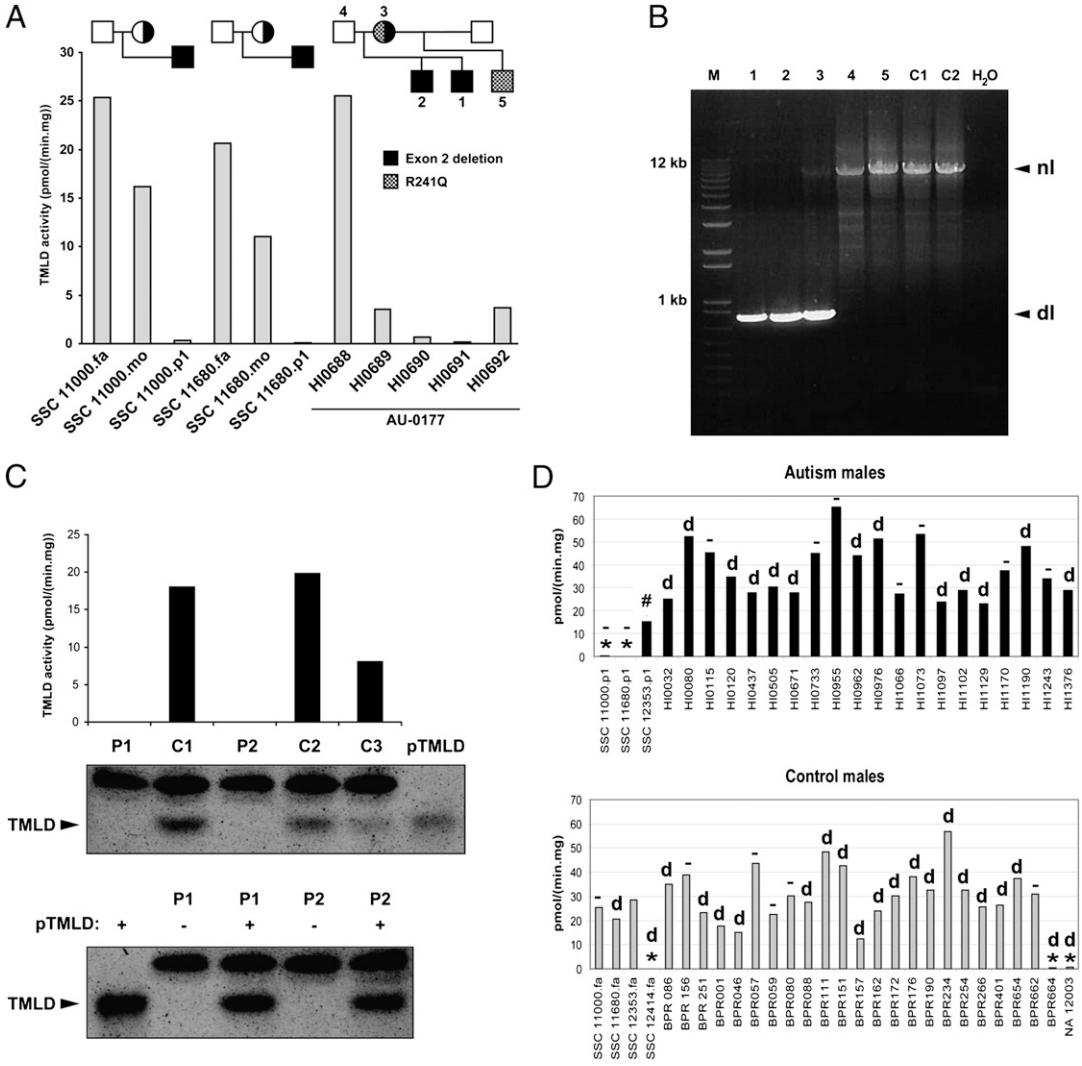



**Fig. 3.** Genetic and enzymatic characterization of hemizygous deletion of exon 2. (*A*) TMLD activity measured in lymphoblast homogenates of three families with exon 2 deletion. (*B*) PCR assay results for the AU 0177 family showing the deletion in the two affected brothers (1, 2) and in the mother (3) but not in the father (4), unaffected maternal half-brother (5), or unaffected controls (C1 and C2). There is bias of amplification in the mother, such that the normal band is faint. dl, deletion; nl, normal. (*C*) (*Upper*) TMLD activity and Western blot analysis of 2 individuals with exon 2 deletion (P1, HI0690; P2, BPR664) and three controls (C1–C3). Purified TMLD (pTMLD) is used as a positive control. (*Lower*) Western blot analysis of 2 individuals (P1 and P2) with (+) or without (–) addition of pTMLD, showing the complete absence of protein in cases of exon 2 deletion and confirmation of the identity of the immunoreactive material as TMLD. The upper band in the Western blot is an irrelevant protein. (*D*) (*Upper*) TMLD activity measured in lymphoblast homogenates from several autism males. *,*TMLHE* exon 2 deletion; #, E287K; d, deletion in intron 1 in 13 individuals; –, no deletion in intron 1 in 9 individuals. SSC 12353.p1 was not tested for the presence of intron 1 deletion. (*Lower*) TMLD activity measured in lymphoblast homogenates from male controls. BPR indicates local unaffected controls, and NA 12003 is an unaffected individual. SSC 12353.fa was not tested for presence of intron 1 deletion. There was no apparent correlation of the level of enzyme activity with the presence or absence of the intronic deletion. For *A*, *C*, and *D*, assays were run in duplicate and the average is plotted without error bars. fa, father; mo, mother; p1, proband.

38 to 143; 5 of 21 with available data were in the range of intellectual disability, and 3 of 21 were reported as untestable. One proband had seizures. For six of six cases in which information was available, patients were described as nondysmorphic. With respect to the controls, two of the seven SSC fathers had at least one domain with an elevated broader autism phenotype score based on the self-report Broad Autism Phenotype Questionaire, but the Social Responsiveness Scale rating by significant other and Family History Interview-Interviewee Impression scores were not consistent with the broader autism phenotype.

**Assuming a True Association, the Penetrance for Autism in *TMLHE* Deficiency Would Be Very Low.** The majority of males with an exon 2 deletion in the US and UK populations are expected to be phenotypically "normal" as adults. *TMLHE* deficiency was present in slightly less than 1% of probands from male-male affected sibling pairs; thus, it would be present in substantially less than 1% of all cases of autism. If we assume an overall frequency of 1 in 100 for autism, with a 4:1 male/female ratio, a frequency of 1 in 350 for *TMLHE* deficiency in normal males, and a frequency of *TMLHE* deficiency of 1 in 250 or 1 in 150 in males with autism, the penetrance would calculate at 2.2% or 3.6%, respectively (*SI Materials and Methods* and Table S6).

## Discussion

*TMLHE* deficiency is a previously undescribed inborn error of metabolism discovered about 100 y after Garrod described such conditions in his 1908 Croonian Lectures to the Royal College of

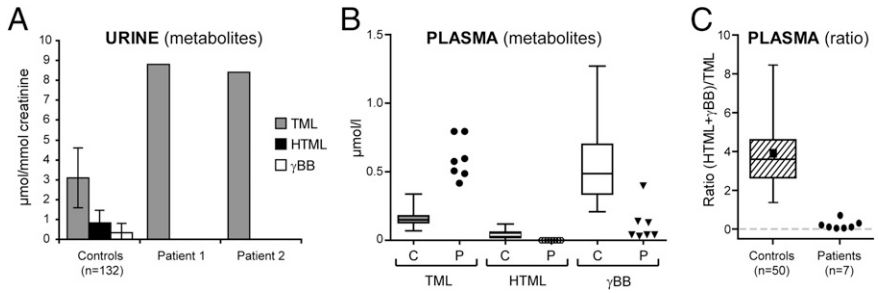



**Fig. 4.** Increased TML and decreased HTML and γBB in patients with *TMLHE* exon 2 deletions. (*A*) Bar diagram of concentrations (mean ± SD) of carnitine biosynthesis intermediates in urine of two patients with exon 2 deletion (1, HI0690; 2, HI0691) compared with controls. (*B*) Box and whisker plot of carnitine biosynthesis intermediates in plasma of seven patients with exon 2 deletion (HI0690, HI0691, SSC 13928.p1, SSC 13489.p1, SSC 11000.p1, SSC 11229.p1, and SSC 11680.p1) compared with controls. (*C*) Box and whisker plot showing the diagnostic potential of the (HTML + γBB)/HTML ratio as an indicator of *TMLHE* deficiency. All seven patients have a very low ratio. The black square represents the mean of the controls, and whiskers show the minimum and maximum values of the control group.

Surgeons. The frequency of *TMLHE* deficiency is startling, at ~1 in 350 control males of European descent, making it at least 20-fold more frequent than phenylketonuria in males. The enzyme deficiency and metabolite changes in plasma and urine are typical for an inborn error of metabolism. *TMLHE* appears to be a gene in which deletions are much more common than point mutations. There is precedent for this at the *DMD*, *PMP22*, *UBE3A*, and other loci causing Duchenne muscular dystrophy, hereditary neuropathy with liability to pressure palsies, Angelman syndrome, and other phenotypes, respectively. These biases are usually explained, in part, by genome architecture, as is likely the case for *TMLHE*.

It might be of some concern that we did not detect any indication that *TMLHE* deficiency is a risk factor for simplex autism. However, simplex and multiplex groups of autism families are significantly different, with an expectation of higher rates of de novo mutations in simplex compared with multiplex families and a higher rate of inherited mutations in multiplex vs. simplex families. In addition, shared genetic modifiers and shared environment are potential factors in multiplex families. Given these differences between simplex and multiplex families, the apparent low penetrance of *TMLHE* deficiency for autism, and the modest sample size, the negative result for simplex families is not surprising. Assuming that the association with multiplex autism is replicated, we would expect that there would be a significant association with simplex autism males with a much larger sample size, because all multiplex families are initially simplex before the birth of a second affected sibling. We would argue that it is not appropriate in this circumstance of low penetrance to combine the probabilities from simplex and multiplex families for an association of *TMLHE* deficiency with autism, because they are significantly different samples. However, if one were to do so, the failure to detect an association in a sample of simplex families of this size, given the necessarily low penetrance, has relatively weak statistical significance and does not detract substantially from the *P* value of 0.0037 observed with male-male multiplex families. We conclude that *TMLHE* deficiency is likely to be a weak risk factor for autism, but replication studies are needed, particularly those focusing on male-male multiplex families. The data make it clear that *TMLHE* deficiency is neither necessary nor sufficient to cause autism. With roughly 4 million births per year in the United States, this would equate to about 5,600 deficient males born per year, which, in turn, would equate to 168 males with *TMLHE* deficiency and autism assuming a 3% penetrance.

One might ask whether carnitine metabolism plays a broader role in the etiology of NDA. One possibility is that *TMLHE* deficiency is entirely benign, as is generally believed to be the case for pentosuria and histidinemia. Alternatively, *TMLHE* deficiency could mediate harmful effects either through toxic accumulation of TML or through deficiency of downstream metabolites, including HTML, TMABA, γBB, or carnitine. All these are possible, but we believe that the most attractive hypothesis at this time is that there is an increased risk for autism, and that this risk is modified by dietary intake of carnitine from birth through the first few years of life. Carnitine intake of the pregnant or nursing mother could also be important. There are extensive reports of mitochondrial abnormalities in autism, as reviewed recently (25), and some of the mitochondrial dysfunction could be secondary to carnitine deficiency. There are reports of low plasma carnitine in autism (26–29), but these reports have not prompted intensive investigations into a possible role of carnitine deficiency in autism and further studies are needed.

Another hypothesis could be that other genetic abnormalities involving the carnitine pathway might confer a risk for autism. Features of autism generally are not reported in children with systemic carnitine deficiency, although cases of autism with carnitine deficiency have been reported (26). Given the neurological basis of autism and the prominent expression of *TMLHE* in hippocampal neurons and Purkinje cells, one possibility would be that symptoms of autism might be secondary to carnitine deficiency in the brain. If that were the case, the pathophysiology of systemic carnitine deficiency would be very different from *TMLHE* deficiency. The former has low plasma carnitine, but ability to synthesize carnitine in the brain and elsewhere is intact. In the latter, plasma carnitine may be normal or low-normal based on dietary intake, but neurons are unable to synthesize carnitine and become completely dependent on transfer across the blood–brain barrier. If carnitine deficiency in the brain was deleterious, dietary deficiency, excess renal losses, disorders of transport (especially across the blood–brain barrier), and defects in synthesis might be risk factors for autism. Relatively little is known about transport across the blood–brain barrier, but this transport may be a limiting factor, because the concentration of carnitine in cerebrospinal fluid is 10- to 15-fold lower than in plasma (30, 31). As shown in Fig. 1, not all tissues are capable of complete carnitine biosynthesis because of the differential expression of the last enzyme, γBB dioxygenase (γBBD), which is only expressed in kidney, liver, and brain in humans. After degradation of proteins that contain TML residues, TML is converted to γBB, which is then transported to the tissues that express γBBD and converted into carnitine. The plasma membrane γBB transporter likely is encoded by the *SLC6A13* gene, which is known as a betaine/GABA transporter and has recently been suggested to function in carnitine biosynthesis as the liver γBB transporter (32). Transport of either or both carnitine and γBB across the blood–brain barrier could be important.

One important question is whether the association with autism is valid and can be replicated in future studies. Although

*TMLHE* deficiency was discovered by a genome-wide molecular analysis, a *P* value of genome-wide significance is not needed here, because this is a simple test of the hypothesis that a newly discovered inborn error of metabolism is a risk factor for autism or not. The metaanalysis *P* value of 0.0037, indicating an association with multiplex autism, suggests that the data indicating an association are unlikely to have occurred by chance. If penetrance for autism is influenced by carnitine intake during infancy, the risk for autism associated with *TMLHE* deficiency may be greater in countries with a high frequency of vegetarian diets and lower meat or beef intake. China, India, and South Korea are all countries where some studies of the incidence of autism are available (22, 33, 34) and there is a more vegetarian diet and/or much lower beef intake.

Two clinical investigations are of immediate interest and are being initiated. One is studying carnitine metabolites in cerebrospinal fluid of infants with autism with and without *TMLHE* deficiency near the age of onset, and the other is treating very young infants with autism with and without *TMLHE* deficiency with carnitine or γBB supplementation. Whether increased carnitine intake before onset of autism might prevent the development of symptoms would require a more complex study. There is a recent report of a trial of carnitine supplementation in autism suggesting clinical improvement (35), but the study included some patients up to 10 y of age who might be unlikely to respond; it would be desirable to have data from very young patients, preferably nondysmorphic, with and without *TMLHE* deficiency. The data reported here suggest that *TMLHE* deficiency is a risk factor for NDA and that carnitine metabolism could be a target for therapeutic intervention in this and related disorders.

## Materials and Methods

**Human Subjects and DNA.** All work with direct involvement of human subjects was approved by the relevant institutional review boards or equivalents, and informed consent was obtained from all subjects. For SSC, AGRE, and NIMH samples, DNA derived from LCLs was obtained from the Rutgers University Cell and DNA Repository. Additional information is provided in *SI Materials and Methods*. The numbers of simplex probands, multiplex male-male sib pairs, and controls from various sources are specified in Table 1. Detailed information for cell culture and for identification of deleted probands and controls is given in *SI Materials and Methods*. For the screening of samples included in Table 1, we used PCR assays (Fig. S2), except for SSC samples, where Illumina arrays (36) were also used; for Toronto and the Wellcome Trust Case–Control Consortium, where Affymetrix 6.0 arrays were used; and for Paris and New York, where qPCR assays were used. All deletions in patients with autism were confirmed using custom arrays for the *TMLHE* region (Fig. 2*A*).

PCR and Sanger dideoxy-sequencing of *TMLHE* exons 1–8 was performed for 536 SSC male probands, 98 affected AGRE males from male-male multiplex families (brothers or half-brothers with the same mother), and 443 NIMH male controls (primers provided in Table S7).

**CGH Array.** All arrays used in this study were designed and analyzed based on University of California, Santa Cruz (UCSC) Genome Browser hg18 (National Center for Biotechnology Information Build 36, March 2006). The coordinates found in tables and figures are converted to hg19 (Genome Reference Consortium: human GRCh Build 37, February 2009). An Agilent CGH custom array of design ID 028249 was used to confirm *TMLHE* deletions originally found by PCR assay or those that were detected by the 1M Illumina SNP array through a collaborative study of SSC families (36). The custom array design is available on the Agilent's eArray website (www.agilent.com/genomics/earray). Analysis of CNVs was done using Agilent's DNA Analytics software (version 4.0.76) with the following settings: aberration algorithm ADM-2, a minimum of three consecutive probes per region, and a minimum absolute average $\log_2$ ratio of 0.25 for any given region.

The protocol for DNA digestion, labeling, purification, and hybridization to the arrays followed the manufacturer's instructions with some modifications, as described previously (37). Genomic DNA (800 ng) from the SSC individual and from a single male reference was used in the digestion. Each slide was scanned into an image file using the Agilent G2565 DNA Microarray Scanner at a 3-μm scan resolution. Each image file was quantified using Agilent Feature Extraction software (version 10.7.3.1). The Agilent custom-focused validation files were uploaded into the DNA Analytics software for analysis.

**Enzyme Assays and Metabolite Determinations.** All individuals tested for TMLD enzyme activity were assayed for the presence or absence of exon 2 of *TMLHE* by PCR or CGH array. These included BPR controls, AGRE and SSC individuals, and a Centre d'Étude du Polymorphisme Humain control (NA12003) (38). TML was obtained from Sigma–Aldrich. [$^2H_9$]TML and [$^2H_3$]γBB were synthesized as described previously (11). [$^2H_9$]HTML was prepared enzymatically by incubating [$^2H_9$]TML with *Neurospora crassa* TLMD, heterologously expressed in *Saccharomyces cerevisiae* as described previously (39). The resulting mixture of [$^2H_9$]HTML and [$^2H_9$]TML was applied to Amicon Ultra 30-kDa filters (Millipore), and the deproteinized filtrate was used as an internal standard for TML and HTML. All other reagents were of analytical grade.

Lymphoblast pellets were homogenized in 10 mM Mops buffer containing 0.9% (wt/vol) NaCl, 10% (wt/vol) glycerol, and 5 mM DTT (pH 7.4). The protein concentration was determined by the method of Bradford (40) using human serum albumin as a standard. For measurement of TMLD and γBBD activities, the reaction mixture consisted of 20 mM potassium phosphate buffer containing 50 mM KCl, 3 mM 2-oxoglutarate, 10 mM sodium ascorbate, 0.5 mM DTT, 0.5 mM ammonium iron sulfate, 2.5 mg/mL BSA, 2 mM TML, and 0.2 mM [$^2H_3$] γBB at pH 7.4, with a final volume of 250 μL. The reaction was started by adding 50 μL of homogenate (target final protein concentration of 0.2 mg/mL for lymphoblast homogenates) to the reaction mixture and was incubated at 37 °C for 30 min. The reaction was terminated by the addition of $ZnCl_2$ to a final concentration of 1 mM, and the reaction mixtures were placed on ice. The $ZnCl_2$ solution also contained the following internal standards: 50 pmol of [$^2H_9$]HTML, 140 pmol of [$^2H_9$]TML, 140 pmol of [$^2H_3$]γ-BB, and 550 pmol of [$^2H_3$]carnitine. Subsequently, the reaction mixture was loaded onto an Amicon Ultra 30-kDa filter and centrifuged at 14,000 × *g* for 20 min to separate the metabolites (TML, HTML, γ-BB, and carnitine) from the enzymes and remove most of the proteins. The filtrate (100 μL) was derivatized with methylchloroformate, and the produced HTML was quantified using ion-pair ultra performance liquid chromatography (UPLC)-tandem MS essentially as previously described (11).

For determination of carnitine biosynthesis metabolites in plasma and urine, internal standards were added to each homogenate and derivatization was performed as described above. Plasma samples were deproteinized using an Amicon Ultra 30-kDa filter. Urine samples were directly derivatized, and TML, HTML, carnitine, and γ-BB were quantified using ion-pair UPLC-tandem MS as previously described (11). For immunoblot analysis, a Multiphor II Nova Blot electrophoretic transfer unit (Amersham Pharmacia Biotech) was used to transfer proteins onto a Protran nitrocellulose membrane (Whatman) as described by the manufacturer. After blocking of nonspecific binding sites with 50 g/L Protifar (Nutricia) and 10 g/L BSA in PBS with Tween 20 (1 g/L) for 1 h, the membrane was incubated for 2 h in the same buffer without Protifar with 1:3,000 dilution of rabbit polyclonal antibodies raised against human recombinant TMLD fused to maltose-binding protein (41). Detection was performed with IRDye 800-conjugated goat anti-rabbit antibody (LI-COR Biosciences) according to the manufacturer's instructions. Membranes were then dried and scanned using the Odyssey Infrared Imaging System (LI-COR Biosciences).

**TDT, Metaanalysis, and Penetrance Calculations.** After 7 of 909 probands from multiplex male-male families were identified, the *P* value favoring a risk relationship of *TMLHE* deficiency to autism was 0.022 based on a one-sided Fisher's exact test. This outcome could have occurred by chance or could have occurred because there is indeed a risk relationship. If the result occurred by chance, 3.5 of the seven siblings would be expected statistically to have the deletion, if all mothers are assumed to be carriers. If the result reflects a true risk relationship, a higher proportion, but not necessarily all, of the autistic siblings should have the deletion. Seven of the eight siblings carried the deletion. Statistical analysis of the sibling data was performed by implementing the TDT (42–45). Because the X chromosome is being analyzed, only transmissions from the mother are informative. We examined whether or not the deletion had been transmitted to the affected male sibling from his mother and applied McNemar's $\chi^2$ (46) to the resulting 2 × 2 table to guard against significant results attributable to population substructure/admixture. To combine the results from the comparison of multiplex probands to controls ($P = 0.023$) and from the TDT analysis of affected male siblings ($P = 0.012$), metaanalysis was performed using Stouffer's method (47). The metaanalysis resulted in a Z-score of 2.90 and a *P* value of 0.0037.

For estimating penetrance, we used a hypothetical population of 2 million individuals at risk with an equal number of males and females, as shown in Table S6. We assumed a frequency of autism of 1 in 100 with a 4:1 male/female ratio. We assumed that 1 in 350 normal males carried deletion of exon 2 of *TMLHE*. We then calculated the penetrance assuming that either 1 in 250 or 1 in 150 males with autism carries the deletion.

**ACKNOWLEDGMENTS.** We thank all the families and the clinicians at the participating Simons Foundation Autism Research Initiative (SFARI) SSC sites and collaborators within the SSC Genetic Consortium; the principal investigators of the SSC Genetic Consortium are listed elsewhere (36). We also thank many other clinicians and families from around the world for making samples available. We are grateful for the participating AGRE Consortium families and all the resources provided by AGRE. We especially acknowledge access to the Illumina AGRE genotype data submitted by Dr. Hakon Hakonarson at the Center for Applied Genomics at The Children's Hospital of Philadelphia. We thank Dr. Joachim Hallmayer for help in recruiting Autism Genome Project investigators. This study makes use of data generated by the Wellcome Trust Case–Control Consortium. A full list of the investigators who contributed to the generation of the data is available at www.wtccc.org.uk. J.S.S. acknowledges the assistance of Melissa Potter in the Computational Genomics Core for managing family, sample, genetic, and phenotypic data for samples examined at Vanderbilt University. A.L.B. thanks Dr. Karen Wang for help in working directly with a family, Drs. James Lupski and Huda Zoghbi for helpful discussions and critical reading of the manuscript, and Dr. John Belmont and Gladys Zapata for providing LCLs and DNA from the BPR collection. F.M.V. acknowledges assistance from the Metabolite/Mass Spectroscopy section of the laboratory Genetic Metabolic Diseases for technical assistance and L. IJlst for helpful discussions. C.B. acknowledges all the families and the clinicians participating in the Paris Autism Research International Sibpair study, particularly Christopher Gillberg, Marion Leboyer, Mary Coleman, Maria Rastam, Gudrun Nygren, and Richard Delorme. The AGRE is a program of Autism Speaks and is supported, in part, by Grant 1U24MH081810 from the National Institute of Mental Health (to Clara M. Lajonchere). Part of this work was supported by Grant SFARI 124827 from the Simons Foundation (to the investigators of the SSC Genetic Consortium) and Grant HD-37283 (to A.L.B) and Grant P30HD-0240640 from the National Institutes of Health. Part of this work was financially supported by the Fundação para a Ciência e Tecnologia, Lisbon, Portugal, by Grant SFRH/BD/38074/2007 (to. S.V.). Part of this work was supported by National Institutes of Health Grants R01 MH061009 and R01 NS049261 (to J.S.S.). Funding for part of this work was provided by the Wellcome Trust under Award 076113 and by Grant 077014/Z/05/Z. Funding for the Paris Autism Research International Sibpair study was provided, in part, by the Institut National de la Santé et de la Recherche Médicale, Fondation de France, Fondation Orange, Fondation pour la Recherche Médicale, Assistance Publique–Hôpitaux de Paris, and the Swedish Science Council.

1. Conference Proceedings (2004) Carnitine. The science behind a conditionally essential nutrient. Proceedings of a conference. March 25–26, 2004. Bethesda, MD. *Ann N Y Acad Sci* 1033:1–197.
2. Flanagan JL, Simmons PA, Vehige J, Willcox MD, Garrett Q (2010) Role of carnitine in disease. *Nutr Metab (Lond)* 7:30.
3. Vaz FM, Wanders RJ (2002) Carnitine biosynthesis in mammals. *Biochem J* 361: 417–429.
4. Slonim AE, et al. (1981) Dietary-dependent carnitine deficiency as a cause of nonketotic hypoglycemia in an infant. *J Pediatr* 99:551–555.
5. Schmidt-Sommerfeld E, Penn D (1990) Carnitine and total parenteral nutrition of the neonate. *Biol Neonate* 58(Suppl 1):81–88.
6. Longo N, Amat di San Filippo C, Pasquali M (2006) Disorders of carnitine transport and the carnitine cycle. *Am J Med Genet C Semin Med Genet* 142C(2):77–85.
7. Strijbis K, Vaz FM, Distel B (2010) Enzymology of the carnitine biosynthesis pathway. *IUBMB Life* 62:357–362.
8. Evans JD, Jacobs TF, Evans EW (2008) Role of acetyl-L-carnitine in the treatment of diabetic peripheral neuropathy. *Ann Pharmacother* 42:1686–1691.
9. Sarma S, Gheorghiade M (2010) Nutritional assessment and support of the patient with acute heart failure. *Curr Opin Crit Care* 16:413–418.
10. Celestino-Soper PBS, et al. (2011) Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. *Hum Mol Genet* 20:4360–4370.
11. Vaz FM, et al. (2002) Analysis of carnitine biosynthesis metabolites in urine by HPLC-electrospray tandem mass spectrometry. *Clin Chem* 48:826–834.
12. Miles JH, et al. (2008) Development and validation of a measure of dysmorphology: Useful for autism subgroup classification. *Am J Med Genet A* 146A:1101–1116.
13. Jacquemont M-L, et al. (2006) Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J Med Genet* 43:843–849.
14. O'Roak BJ, et al. (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 43:585–589.
15. Sebat J, et al. (2007) Strong association of de novo copy number mutations with autism. *Science* 316:445–449.
16. Marshall CR, et al. (2008) Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82:477–488.
17. van Bon BW, et al. (2009) Further delineation of the 15q13 microdeletion and duplication syndromes: A clinical spectrum varying from non-pathogenic to a severe outcome. *J Med Genet* 46:511–523.
18. Sebat J, Levy DL, McCarthy SE (2009) Rare structural variants in schizophrenia: One disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet* 25: 528–535.
19. Shinawi M, et al. (2009) A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes. *Nat Genet* 41:1269–1271.
20. Girirajan S, et al. (2010) A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 42:203–209.
21. Scott FJ, Baron-Cohen S, Bolton P, Brayne C (2002) Brief report: Prevalence of autism spectrum conditions in children aged 5-11 years in Cambridgeshire, UK. *Autism* 6: 231–237.
22. Kalra V, Seth R, Sapra S (2005) Autism—Experiences in a tertiary care hospital. *Indian J Pediatr* 72:227–230.
23. Boone PM, et al. (2011) Alu-specific microhomology-mediated deletion of the final exon of SPAST in three unrelated subjects with hereditary spastic paraplegia. *Genet Med* 13:582–592.
24. Monfregola J, et al. (2007) Functional characterization of the TMLH gene: Promoter analysis, in situ hybridization, identification and mapping of alternative splicing variants. *Gene* 395(1–2):86–97.
25. Rossignol DA, Frye RE (2012) Mitochondrial dysfunction in autism spectrum disorders: A systematic review and meta-analysis. *Mol Psychiatry* 17:290–314.
26. Gargus JJ, Imtiaz F (2008) Mitochondrial energy-deficient endophenotype in autism. *American Journal of Biochemistry and Biotechnology* 4(2):198–207.
27. Mostafa GA, El-Gamal HA, El-Wakkad ASE, El-Shorbagy OE, Hamza MM (2005) Polyunsaturated fatty acids, carnitine and lactate as biological markers of brain energy in autistic children. *International Journal of Child Neuropsychiatry* 2(2):179–188.
28. Lombard J (1998) Autism: A mitochondrial disorder? *Med Hypotheses* 50:497–500.
29. Filipek PA, Juranek J, Nguyen MT, Cummings C, Gargus JJ (2004) Relative carnitine deficiency in autism. *J Autism Dev Disord* 34:615–623.
30. Rubio JC, et al. (1998) Cerebrospinal fluid carnitine levels in patients with Alzheimer's disease. *J Neurol Sci* 155(2):192–195.
31. Shinawi M, Gruener N, Lerner A (1998) CSF levels of carnitine in children with meningitis, neurologic disorders, acute gastroenteritis, and seizure. *Neurology* 50: 1869–1871.
32. Fujita M, et al. (2009) Hepatic uptake of gamma-butyrobetaine, a precursor of carnitine biosynthesis, in rats. *Am J Physiol Gastrointest Liver Physiol* 297:G681–G686.
33. Wong VC, Hui SL (2008) Epidemiological study of autism spectrum disorder in China. *J Child Neurol* 23(1):67–72.
34. Kim YS, et al. (2011) Prevalence of autism spectrum disorders in a total population sample. *Am J Psychiatry* 168:904–912.
35. Geier DA, et al. (2011) A prospective double-blind, randomized clinical trial of levocarnitine to treat autism spectrum disorders. *Med Sci Monit* 17:PI15–PI23.
36. Sanders SJ, et al. (2011) Multiple recurrent *de novo* copy number variations (CNVs), including duplications of the 7q11.23 Williams-Beuren syndrome region, are strongly associated with autism. *Neuron* 70:863–885.
37. Ou Z, et al. (2008) Bacterial artificial chromosome-emulation oligonucleotide arrays for targeted clinical array-comparative genomic hybridization analyses. *Genet Med* 10:278–289.
38. McCarroll SA, et al.; International HapMap Consortium (2006) Common deletion polymorphisms in the human genome. *Nat Genet* 38:86–92.
39. Swiegers JH, Vaz FM, Pretorius IS, Wanders RJ, Bauer FF (2002) Carnitine biosynthesis in Neurospora crassa: Identification of a cDNA coding for epsilon-N-trimethyllysine hydroxylase and its functional expression in Saccharomyces cerevisiae. *FEMS Microbiol Lett* 210(1):19–23.
40. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248–254.
41. Vaz FM, Ofman R, Westinga K, Back JW, Wanders RJ (2001) Molecular and Biochemical Characterization of Rat epsilon-N-Trimethyllysine Hydroxylase, the First Enzyme of Carnitine Biosynthesis. *J Biol Chem* 276:33512–33517.
42. Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516.
43. Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: History, subdivision, and admixture. *Am J Hum Genet* 57:455–464.
44. McGinnis RE, Ewens WJ, Spielman RS (1995) The TDT reveals linkage and linkage disequilibrium in a rare disease. *Genet Epidemiol* 12:637–640.
45. Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 59:983–989.
46. McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12(2):153–157.
47. Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RM, Jr. (1949) *Adjustment During Army Life* (Princeton Univ Press, Princeton).

Article 4

# *Complex nature of apparently balanced chromosomal rearrangements in patients with autism spectrum disorder*

Tabet AC, Verloes A, Pilorge M, **Delaby E**, Delorme R, Nygren G, Devillard F, Gérard M, Passemard S, Jacquette A, Delahaye A, Perrin L, Dupont C, Aboura A, Bitoun P, Coleman M, Leboyer M, Gillberg C, Benzacken B, Betancur C

(*soumis pour publication*)

*Article soumis pour publication*

# Complex nature of apparently balanced chromosomal rearrangements in patients with autism spectrum disorder

Anne-Claude Tabet[1,2,3,4], Alain Verloes[1,5], Marion Pilorge[2,3,4], Elsa Delaby[2,3,4], Richard Delorme[6,7], Gudrun Nygren[8], Françoise Devillard[9], Marion Gérard[1], Sandrine Passemard[10], Aurelia Jacquette[11], Andrée Delahaye[5,12], Laurence Perrin[1], Céline Dupont[1], Azzedine Aboura[1], Pierre Bitoun[13], Mary Coleman[14], Marion Leboyer[7,15,16,17], Christopher Gillberg[8], Brigitte Benzacken[1,5,13], Catalina Betancur[2,3,4]

[1] AP-HP, Robert Debré University Hospital, Department of Genetics, Paris, France
[2] INSERM, U1130, Paris, France
[3] CNRS, UMR 8246, Paris, France
[4] Sorbonne Universités, UPMC Univ Paris 6, Neuroscience Paris Seine, Paris, France
[5] INSERM U676, Paris, France
[6] AP-HP, Robert Debré University Hospital, Department of Child and Adolescent Psychiatry, Paris, France
[7] Fondation Fondamental, Créteil, France
[8] Gillberg Neuropsychiatry Centre, University of Gothenburg, Göteborg, Sweden
[9] Grenoble University Hospital, Department of Genetics, Grenoble, France
[10] AP-HP, Robert Debré University Hospital, Neurology Unit, Paris, France
[11] AP-HP, Pitié-Salpêtrière University Hospital, Medical Genetics Unit, Paris, France
[12] AP-HP, Jean-Verdier Hospital, Cytogenetics Unit, Bondy, France
[13] AP-HP, Jean Verdier Hospital, Medical Genetics Unit, Bondy, France
[14] Foundation for Autism Research, Sarasota, Florida, USA
[15] AP-HP, Henri Mondor-Albert Chenevier Hospital, Department of Psychiatry, Créteil, France
[16] INSERM U955, Institut Mondor de Recherche Biomédicale, Psychiatric Genetics, Créteil, France
[17] University Paris-Est Créteil, Faculty of Medicine, Créteil, France

**Running title:** Apparently balanced chromosomal rearrangements in autism

**Corresponding author:** Dr. Anne-Claude Tabet, Cytogenetics Unit, Robert Debré Hospital, 48 Boulevard Sérurier, 75019 Paris, France. Phone: +33 1 40 03 57 10; fax: +33 1 40 03 53 19; e-mail: anne-claude.tabet@rdb.aphp.fr

## Abstract

Apparently balanced chromosomal rearrangements can be associated with an abnormal phenotype, including intellectual disability and autism spectrum disorders (ASD). Genome-wide microarrays reveal cryptic genomic imbalances related or not to the breakpoints in about 25%-50% of patients with an abnormal phenotype carrying microscopically balanced chromosomal rearrangements. Here we performed single nucleotide polymorphism array analysis of 18 patients with ASD carrying balanced chromosomal abnormalities (nine *de novo*, seven inherited, and two of unknown inheritance). We detected clinically significant *de novo* copy number variants in 4/18 patients (22%), including 3/9 with *de novo* rearrangements and 1/7 with an inherited abnormality. The sizes ranged from 3.3 to 4.9 Mb; three were related to the breakpoint regions and one occurred elsewhere in the genome. Of note, we report the first patient with autism carrying a duplication of the Wolf-Hirschhorn syndrome critical region, contributing to delineate the phenotype of this rare genomic disorder. The patient has a chromosome 4p inverted duplication deletion, with a 0.5 Mb deletion of terminal 4p and a 4.3 Mb duplication of 4p16.2p16.3. In addition, we identified a rare *de novo* deletion of unknown significance on a chromosome unrelated to the initial rearrangement and disrupting a single gene, *RFX3*. These findings underscore the utility of SNP arrays for investigating apparently balanced chromosomal abnormalities in subjects with ASD or related neurodevelopmental disorders in both clinical and research settings.

**Key words:** autism spectrum disorder; balanced chromosomal rearrangement, CNV; deletion; duplication; karyotype; inversion; translocation; SNP microarray

## Introduction

Balanced chromosome abnormalities, including translocations and inversions, are structural rearrangements of genetic material with no overall gain or loss detected with conventional karyotyping. Apparently balanced chromosomal rearrangements (ABCR) have an estimated frequency of 0.5% in newborns [1], including 14% *de novo*. About 6% of *de novo* ABCR detected at amniocentesis are associated with an abnormal phenotype, including intellectual disability (ID) and multiple congenital anomalies [2]. Because of the insufficient discrimination power of conventional cytogenetics, in which genomic imbalances smaller than 5-10 megabases (Mb) are usually not detected, the most common explanation for the clinical abnormalities is cryptic loss or gain of genomic material at or in the vicinity of the breakpoint. When there is no loss or gain of DNA sequences, the rearrangement can disrupt a dosage-sensitive gene, separate a gene from its *cis* regulatory elements, or generate a functional chimeric gene.

Studies using DNA microarray technologies have demonstrated submicroscopic anomalies related or not to the breakpoint in 46% (range 31%-100%) of patients with an abnormal phenotype carrying a *de novo* ABCR [3-10]. Inherited ABCR in patients with an abnormal phenotype have been studied less often, but they can also be associated with cryptic imbalances at the breakpoint or elsewhere in the genome, with a combined frequency of 25% (6/24) in three studies [4, 6, 8]. Imbalances are more frequent in complex rearrangements involving more than two breakpoints and in patients with a complex phenotype [7]. In contrast, genomic imbalances are unlikely to be detected in phenotypically normal carriers of apparently balanced translocations [5]. Taken together, these findings indicate that a significant proportion of ABCR in phenotypically abnormal patients are in fact associated with genomic imbalances and that these rearrangements should be systematically investigated by high-resolution microarrays independently of their *de novo* or inherited origin. The phenotypes of the patients included in previous studies were very heterogeneous, including developmental delay, ID of variable severity, multiple congenital anomalies and/or autism spectrum disorders (ASD).

ASD is an etiologically heterogeneous neurodevelopmental disorder characterized by impairments in social communication and by restricted interests and stereotyped behaviors. Hundreds of rare variants, including chromosomal abnormalities, copy number variants (CNVs) and single nucleotide variants have been implicated in ASD [11, 12]. However, for about 80% of cases, the underlying genetic determinants remain unknown. The frequency of structural chromosomal imbalances detected by conventional cytogenetics in autism varies between 2-6% [13, 14], including karyotypically balanced chromosomal abnormalities. To date, only isolated cases with ABCR and ASD have been studied by array technology [15, 16]. Here we report a series of 18 patients with ASD carrying *de novo* or inherited ABCR studied by single nucleotide polymorphism (SNP) arrays to identify cryptic CNVs implicated in abnormal neurodevelopment.

## Materials and methods

### Subjects

Patients with ASD and ABCR were ascertained through two sources: 1) 12 patients from the PARIS (Paris Autism Research International Study) family dataset [17], and 2) six patients referred by the network of French cytogeneticists. A summary of the clinical and cytogenetic data of the patients is shown in Table 1. Fourteen patients had ID. Based on the presence of facial dysmorphism and/or malformations, six subjects were considered to have syndromic ASD (Patients 3, 4, 5, 9, 10, and 15). Sixteen patients fulfilled DSM-IV criteria for autistic disorder, while two subjects (Patients 9 and 13) had a previous diagnosis of ASD but could not be formally evaluated for ASD for this study. The Autism Diagnostic Interview-Revised was performed in 14 subjects; in one individual, the Diagnostic Interview for Social and Communication Disorders, tenth revision (DISCO-10) was used instead. In addition, five subjects were assessed with the

Autism Diagnostic Observation Schedule (ADOS). Fragile X molecular testing and metabolic screening were normal in all individuals. All parents were phenotypically normal, including those carrying a rearrangement. Informed consent was obtained from all families participating in the study.

### Conventional cytogenetic analysis

Karyotype (G-banded and/or R-banded, 450 band level) showing an ABCR was available for all patients and in most cases their parents before inclusion. The sample included ten autosomal reciprocal translocations, one X-autosome translocation, three rare inversions and four common pericentric inversions of chromosomes 2, 5, 9, and Y, usually considered as cytogenetic variants (Table 1). Nine chromosomal rearrangements appeared *de novo*, four were maternally inherited, three were paternal and in two cases, the inheritance could not be defined because DNA samples from the fathers were not available.

### Whole-genome SNP array

SNP array analysis was performed in all affected individuals and their parents using the Human CNV370-Duo DNA Analysis BeadChip (Illumina, San Diego, CA, USA). In three cases, parental DNA was not available (Patients 8, 9, and 18) and in one case, only the mother was available (Patient 17). The array contains more than 370,000 markers, with a mean resolution of 30 kb. Genomic DNA was processed according to the Infinium II assay manual. SNP copy numbers (log R Ratio) and B allele frequencies were assessed using the Bead Studio software version 3.2, with the CNV partition algorithm, V 1.3.2 or 2.4.4 (Illumina). All genomic coordinates are based on GRCh37/hg19.

### FISH analyses

Whenever possible, fluorescence in situ hybridization (FISH) analysis was carried out using specific bacterial artificial chromosomes (BACs) to confirm CNVs larger than 1 Mb. BAC clones were hybridized on metaphases spreads of patients and their parents, in order to confirm the inheritance.

### Quantitative PCR (qPCR)

Real time qPCR was used to confirm and map CNVs shorter than 1 Mb. For Patient 2, qPCR was used to confirm a larger CNV because metaphase spreads were not available. We used the Universal Probe Library (UPL) system (Roche, Indianapolis, USA) and a LightCycler 480 real time PCR system (Roche), as described previously [18].

## Results

The microarray analysis detected clinically significant CNVs in 4/18 patients (22%), including 3/9 with *de novo* ABCR and 1/7 with an inherited abnormality (Table 2). Patients 1 and 4 had both a deletion and a duplication in the chromosome implicated in the rearrangement. Patient 2 had a deletion at one of the translocation breakpoints, whereas Patient 3 carried a deletion on a chromosome unrelated to the known rearrangement. The size of the pathogenic CNVs ranged from 3.31 to 4.93 Mb. In addition, Patient 5 carried a *de novo* deletion on a chromosome unrelated to the initial rearrangement and disrupting a single gene, *RFX3*. In the remaining 13 patients, no clinically relevant rearrangement was detected.

### Patient 1

The patient is a 2 year 9 month old boy referred for evaluation because of developmental delay and ASD features. He was born at term after an uneventful pregnancy, with birth weight 5280 g (+4.3 SD), probably due to latent diabetes in mother. An inguinal hernia was corrected at the age of two months. There was no delay of motor milestones or language development. He was diagnosed with autism after formal testing at an autism referral center. He had a borderline IQ of 69. There was no history of seizures and his EEG was normal. His physical and neurological exams were normal except for mild strabismus; in

particular he has no dysmorphic features. At the age of 4 years, he was +2 SD to +3 SD in weight and length. Conventional karyotyping showed a *de novo* translocation between the telomeric region of the short arm of chromosome 4 and the short arm of an acrocentric chromosome. His karyotype was: 46,XY,der(4)t(4;acro p)(p16.3;acro p)dn. SNP array analysis revealed a 541 kb deletion in 4p16.3 (17,764-558,839, hg19) and a 4.24 Mb duplication in 4p16.2p16.3 (577,581-4,812,859) in the proband (Figure 1A), absent from the parents. Both rearrangements were confirmed by qPCR and FISH. The deletion was confirmed by FISH with the 4p subtelomeric probe RP11-2H3. Using probes RP11-296G16 (4p16.3) and RP11-265012 (4p16.2), the FISH findings were consistent with an inverted duplication (data not shown), indicating a chromosome 4p inverted duplication deletion. SNP genotyping indicated that both rearrangements occurred on the paternal chromosome. The deletion includes 8 RefSeq genes; the duplication encompasses 70 genes and encompasses the Wolf-Hirschhorn critical region.

### Patient 2

The patient is a 5 year-old boy presenting with non-syndromic autism and mild ID. He is the second child of healthy parents; antenatal and postnatal periods were uneventful. His language was delayed; he said his first words at 36 months, and did not use sentences. He was diagnosed with autism following evaluation at the autism unit of a university hospital; he met criteria for autism according to the ADI-R. He suffered from recurrent otitis media in childhood and eczema. No dysmorphic features were noticed, except for long fingers and toes, and numerous secondary palmar creases. Physical and neurological exams were normal. Karyotyping showed a *de novo*, apparently balanced reciprocal translocation, 46,XY,t(5;18)(q11.2;p11.2)dn. ish t(5;18)(WCP 5+;WCP 5+,D18Z1+). The array revealed a 4.22 Mb deletion of chromosome 18p11.22p11.31 (5,408,997-9,625,750) (Figure 1B) encompassing 20 genes and originating on the paternally inherited chromosome. The deletion was confirmed by qPCR. Both parents had a normal array profile and normal gene dosage by qPCR.

### Patient 3

This boy is the first child of non-consanguineous healthy European parents with unremarkable family history. Increased nuchal translucency was observed at 12 weeks of gestation. The amniocentesis karyotype was considered normal. At 22 weeks, an ultrasound revealed bilateral ectopic testes and dilated colon, and intrauterine growth retardation was noted at 32 and 36 weeks. He was delivered by cesarean section at 37 weeks because of fetal heart deceleration. Birth weight was 1850 g (<5th centile), length 43 cm (<5th centile) and occipitofrontal circumference (OFC) 32.5 cm (<25th centile). As a newborn, he had feeding difficulties, axial hypotonia, jitteriness and uncoordinated movements. At ten months, his weight and length were below -2 SD, and his OFC decreased at -3 SD. His language development and motor milestones were delayed; he was able to stand up with support at 42 months and did not walk or talk at 4 years. Dysmorphic features included high forehead, horizontal eyebrows, upward slanting palpebral fissures, bulbous nose, smooth philtrum, thin upper lip, and posteriorly rotated ears. Brain MRI showed ventricular enlargement and increased frontal subarachnoid spaces. The EEG, visceral and heart ultrasounds were normal. At 42 months, his growth parameters remained at -3 SD. He was referred to a child psychiatric unit at the age of 4 years, where he received a diagnosis of autism based on DSM-IV criteria, and confirmed by the ADI-R and ADOS. Evaluation with the Brunet-Lezine Revised Scale showed a developmental age of 13 months, indicating severe cognitive impairment. The Vineland Adaptive Behavior Scales showed developmental ages between 9 months (language) and 18 months (socialization). High resolution karyotype revealed a *de novo* pericentric inversion of chromosome 11: 46,XY,inv(11)(p14q23.2)dn. The SNP array detected no imbalance on chromosome 11 but showed a *de novo* 7q21.3q22.1 deletion of 4.93 Mb (97,043,362-101,977,945) (Figure 1C), encompassing over 100 genes and occurring on the paternal allele. The deletion was confirmed by FISH with BACs targeting 7q22.1 (RP11-44M6 and DBACA-20A02, Integragen, France), and excluded in the parents.

### Patient 4

This 14 year-old male was reported previously [15]. Briefly, he had autism, ID, insulin-dependent diabetes, growth hormone deficiency, and mild dysmorphic features. Standard karyotype showed an apparently balanced paracentric inversion of the long arm of chromosome 2 inherited from his healthy mother: 46,XY,inv(2)(q14.2q37.3)mat. Array analysis showed a 3.31 Mb terminal deletion (239,735,269-243,044,147) at the distal breakpoint of the inversion, and a 2q14.1q14.2 duplication spanning 4.23 Mb (117,072,756-121,304,548) at the proximal breakpoint (Figure 1D). Both imbalances occurred on the maternally derived chromosome and were confirmed by qPCR. FISH analysis revealed that the duplicated material was located at the telomeric end of chromosome 2, distal to the inverted region. Both parents had normal SNP array profiles. The terminal deletion results in the 2q37 deletion syndrome, also known as brachydactyly-mental retardation syndrome. Brachydactyly, reported in ~50% of affected individuals, was not observed clinically in Patient 4.

### Patient 5

The patient is a female referred at the age of 4 years 8 months because of developmental delay, autistic behavior, and speech delay, and diagnosed with autism after formal testing. She was born at term after an uneventful pregnancy, with birth parameters on the normal range. She walked at 22 months. At 3 years she exhibited overgrowth, with a height +3 SD and macrocephaly (+3 SD). Assessment of cognitive functioning at 6 years indicated severe ID. When re-evaluated at 11 years she met criteria for autism according to the ADI-R, and the Vineland Adaptive Behavior Scales indicated developmental ages between 14 months (communication and socialization) and 37 months (motor skills). At the age of 13 years, all her growth parameters were increased (weight +3.5 SD, height +2 SD, OFC +4.2 SD), and she remained non-verbal. She had mild dysmorphic features, including downslanting palpebral fissures, and short philtrum; her hands and feet were large, and a café-au-lait spot was present on a toe. The neurological exam was normal. A brain MRI performed at 9 years showed a Chiari malformation type 1 and mild hyperintensities of the white matter. The standard karyotype was 46,XX,inv(2)(p11q13)dn, which is considered a frequent polymorphic cytogenetic variant. SNP array analysis revealed a *de novo* intragenic deletion of 179 kb in the *RFX3* gene at 9p24.2 (3,344,203-3,523,286). The deletion occurred on the paternal allele. qPCR confirmed a deletion of exons 2-4 of *RFX3* in the patient and was normal in both parents.

## Discussion

### Cryptic abnormalities in ABCR

We identified *de novo* genomic imbalances in 5/18 patients with ASD and ABCR, of which four are clinically significant (4/18, 22%) and one of unknown significance. This is in accordance with a previous study showing that 24% of patients with abnormal phenotype and a *de novo* two-point chromosome rearrangement have a clinically significant cryptic genomic imbalance [7]. Three imbalances were related to the breakpoint regions and two deletions occurred on unrelated chromosomes. In two patients, the SNP array analysis identified a terminal deletion and a duplication occurring on the same chromosome involved in the ABCR. Four out of five ABCR with a genomic imbalance occurred *de novo*: two paracentric inversions, one pericentric inversion of chromosome 2 considered as a chromosomal variant, and two reciprocal translocations. Two of the five patients with a *de novo* imbalance have a non-syndromic form of ASD. Our results are concordant with previous studies reporting that ABCR often hide more complex rearrangements regardless of the type of ABCR, particularly when they appear *de novo* [3, 7], and when they are associated with an abnormal phenotype [5, 6, 8]. Of the five *de novo* imbalances, three originated on the paternally inherited chromosome. This is in agreement with previous studies showing that male gametogenesis is more susceptible to rearrangements.

Four of our patients carry common inversions considered variants without phenotypic effects. In all cases, the inversion was indeed balanced, confirming that common pericentric inversions of chromosomes 2, 5, 9 and Y are not implicated in abnormal phenotype by imbalances at the breakpoints, even when they occur *de novo* [19-21]. In one case (Patient 5), the pericentric inversion of chromosome 2 appeared balanced but a *de novo* deletion occurred on an unrelated chromosome.

## Imbalances at the breakpoint region

### Duplication of the Wolf-Hirschhorn region

Patient 1, carrying a translocation (4;acro p), has two imbalances, a 541 kb subtelomeric 4p16.3 deletion, and a directly adjacent 4.29 Mb inverted duplication. The 4p16.3 deletion overlaps two previously reported 4p terminal deletions between 200 and 400 kb in phenotypically normal individuals, and can be considered to be benign [22, 23]. The phenotype of our patient is likely related to the proximal duplication, which encompasses the Wolf-Hirschhorn critical region (WHSCR), with two proposed loci, WHSCR1 and WHSCR2 (Figure 2). Partial 4p trisomy, including the WHSCR, is associated with variable clinical manifestations depending on the size of the duplicated segment. Clinical features include growth retardation, ID, dysmorphic features, heart and renal abnormalities. Most cases reported thus far were identified cytogenetically, and are therefore very large [24-26]. Smaller molecularly defined duplications overlapping the WHSCR have been reported in six patients (Figure 2). Roselló et al. [27] described a boy with a 1.1 Mb 4p16.3 duplication encompassing the WHSCR, presenting with ID, dysmorphic features, language delay, hyperactivity, aggressive behavior, and non-specific anomalies on MRI. The subject also has a terminal 4p16.3 deletion of 1.3 Mb, interpreted as benign. A terminal 2.3 Mb duplication of 4p16.3 was reported in a girl with mild phenotype, including speech and language delay, borderline normal IQ, mild hypertelorism, and normal height with advanced bone age [28]. The duplication resulted from a reciprocal translocation and was associated with a 314 kb deletion of 17q25.3, considered benign. A 5 Mb terminal 4p16.2 duplication very similar to the one observed in our patient was identified in two individuals from the same family, a 2-month-old boy and his paternal uncle with mild ID, after the older sister with typical WHS features was found to carry a 4p deletion due to a translocation [29]. The two affected subjects shared some characteristics, including facial features, spaced nipples and small penis. Similarly, a 3 Mb terminal 4p16.3 duplication was recently described in a three-generation family with macrocephaly, tall stature, speech delay, mild ID, and minor dysmorphic features [30]. Finally, two smaller interstitial duplications overlapping the WHSCR have been reported. An interstitial 4p16.3 duplication of 560 kb encompassing both WHSCR1 and WHSCR2 was reported in a subject with profound ID, dysmorphic features, hand malformation, glaucoma, hypotonia, and seizures [31]. This subject also carried a pericentric inversion between 4p16.3 and 4q22; given the severity of his features, it is possible that the inversion contributes to his phenotype by disrupting a gene at the breakpoint [31]. A 506 kb duplication in 4p16.3 involving only WHSCR2 was reported in a 16-month-old boy with some characteristics of the trisomy 4p syndrome, including psychomotor delay, prominent glabella, low-set ears, and short neck [32]. Compared to the patient reported by Hannes et al., the only common features observed in these individuals were developmental delay and seizures. Our patient presented with mild ID and autism, without malformations or seizures. ASD was not mentioned in any of the subjects with 4p16.3 duplications including the WHSCR or in patients with cytogenetically detectable duplications. ASD has been reported in at least two patients with Wolf-Hirschhorn syndrome [33, 34], suggesting that overexpressed genes within this region could be responsible for ID and ASD in our patient.

Inverted duplications of chromosome 4 with terminal deletions have been reported in subjects with abnormal phenotypes [26, 35, 36]. The majority of rearrangements with terminal deletion associated with a duplication in the same chromosome arm correspond to inverted duplication deletions (inv dup del) [37]. Three mechanisms have been proposed to explain such rearrangements. All involve an intermediate

dicentric chromosome which breaks during meiotic division to produce a monocentric duplicated and deleted chromosome. The first mechanism occurs because of a parental paracentric inversion: during meiosis pairing, both homologues create an inversion loop; a crossing-over in the loop leads to the formation of a dicentric chromosome and a reciprocal acentric chromosome. A subsequent breakage of the dicentric allows the formation of a monocentric chromosome with the inverted-duplicated-deleted chromosome. The second mechanism occurs through non-allelic homologous recombination between inverted segmental duplications in the same arm. For both mechanisms, the inverted duplicated region and the deleted region appear separated by a single copy region. The third mechanism, which is probably the most frequent, involves U-type exchange between the sister chromatids; in this case the duplicated and deleted regions are adjacent and not separated by a single copy region [38, 39]. In the patient reported here, there is no single copy region between the deletion and duplication, suggesting an U-type exchange mechanism. Moreover, as a telomere capture mechanism is frequently used for stabilizing the broken chromosome ends [39], we suggest that the translocation between the short arm of an acrocentric chromosome and the derivative chromosome 4 was generated to stabilize it.

### 18p11.22p11.31 deletion

In Patient 2, with a *de novo* translocation (5;18), we identified an interstitial deletion of 4.21 Mb in the 18p11.22p11.31 region. The 18p deletion syndrome, associated with terminal deletions of variable size, is characterized by typical facial features, growth retardation and ID [40]. Our patient does not have any facial dysmorphic features or growth retardation. Pure interstitial deletions in this region have not been reported in the literature so far. The DECIPHER and ISCA databases include several patients with partially overlapping interstitial deletions of variable sizes; although several are inherited from a normal parent, two are *de novo*, a 9 Mb deletion in a patient with autism (290283), and a 3 Mb deletion (286638, no phenotype information). The deletion in Patient 2 overlaps at least 20 known genes; only one has a disease-associated OMIM entry, *NDUFV2*, involved in autosomal recessive mitochondrial complex I deficiency characterized by early-onset cardiomyopathy and encephalopathy. We speculate that haploinsufficiency of one or more genes included in the deletion could be implicated in the ASD phenotype. Two genes in the region have a high haploinsufficiency score, *EPB41L3* and *ANKRD12,* both highly expressed in brain*. EPB41L3* encodes band 4.1-like protein 3 (also known as protein 4.1B), which interacts with the synaptic cell adhesion molecule 1 (SynCAM1) to recruit NMDA receptors to synapses [41], and also acts as tumor suppressor. *ANKRD12* (ankyrin repeat domain 12) is a member of the ankyrin repeats-containing cofactor family, which can inhibit the transcriptional activity of nuclear receptors through the recruitment of histone deacetylases.

### 2q37 terminal deletion

Patient 4 carries a maternal inv(2)(q14q36) with a terminal deletion 2q37.3 and a duplication 2q14.2 at the inversion breakpoints, which arose from a recombination event between the normal and inverted homologues 2 on maternal meiosis 1, followed by breakage of a dicentric chromosome [15]. The deletion involves the region implicated in the 2q37 deletion syndrome, frequently associated with ASD [34, 42]. His phenotype is similar to that reported in patients with terminal 2q deletions, and includes autism, ID, language delay, growth retardation, and mild facial dysmorphism. Haploinsufficiency of the *HDAC4* gene is responsible for brachymetaphalangy and ID [43], but the contribution of one or more distal genes cannot be excluded, since individuals with distal deletions not including *HDAC4* have been reported with ID, ASD and seizures. The 2q14.1q14.2 duplication includes 20 genes; no similar imbalances are reported in DECIPHER, ISCA or DGV databases so its clinical significance is unknown.

### Imbalances on unrelated chromosomes

In Patient 3, the *de novo* pericentric chromosome 11 inversion was not associated with a genomic imbalance but the SNP array analysis revealed a 7q21.3q22.1 *de novo* deletion of 4.9 Mb. Among the deleted genes, only *AP1S1* has been implicated in a neurological phenotype, a recessive disorder of copper metabolism. No similar deletions have been reported in the literature although several cases in DECIPHER and ISCA have smaller or larger deletions. Two inherited apparently balanced rearrangements mapping to this region have been reported in autism: a pericentric inversion (7q22.1q31.1) in two siblings with autism [44] and a translocation involving chromosome 7q22.1, just distal to *NPTX2*, in a subject with autism and ID [45].

Patient 5 has a pericentric inversion of chromosome 2, which is considered a benign cytogenetic variant. No CNV was observed in chromosome 2 but the SNP array identified a *de novo* 9p24.2 deletion interrupting the *RFX3* gene. A ~440 kb deletion including *RFX3* was identified in one individual in a schizophrenia cohort [46]. Following–up on this finding, Sahoo et al. [47] analyzed 38,779 individuals referred for microarray testing and identified seven deletions encompassing the *RFX3* gene, with sizes ranging from 177 kb to 3.5 Mb. Two deletions were *de novo*, three inherited, and three of unknown origin. The clinical features included autism, ID, and behavioral problems. In addition, one DECIPHER patient (248290) has a partial deletion of *RFX3* but she also carries two large deletions in other chromosomes (no inheritance or phenotype information available). A similar deletion was reported in a patient from ISCA (nssv580652) with cataract and global developmental delay (unknown inheritance). DGV lists only one coding exon deletion (nsv892090), identified among 6533 Asians from the general population. Distal deletions of 9p of varying sizes are characterized by ID, trigonocephaly, dysmorphic facial features, and genital abnormalities; autism or autistic-like behavior are also common [48]. Two critical regions have been proposed for the 9p deletion syndrome, between 11 Mb and 16 Mb, and the first 2 Mb of 9pter [48], neither of which includes *RFX3* (located at 3.2 Mb). *RFX3* encodes a transcription factor required for ciliogenesis in mammals [49]. Many ciliopathy genes, such as those involved in Joubert syndrome, can present with neurodevelopmental phenotypes, including ID and ASD. *Rfx3*-/- mice show several hallmarks of ciliopathies, including left-right asymmetry defects, hydrocephalus and corpus callosum agenesis [50, 51]. The phenotype of heterozygous mice has not been reported. Although the clinical significance of the *RFX3* deletion identified in our patient is unknown, the results on *Rfx3* mutant mice and the presence of rare *RFX3* deletions in patients with ID or abnormal behavior suggest that further studies are warranted on the contribution of *RFX3* loss of function to abnormal brain development.

### Conclusion

In this study, we used genome-wide microarray analysis to elucidate the molecular complexity of microscopically balanced chromosome rearrangements in a series of 18 patients with ASD. We detected imbalances related to the initial ABCR but also CNVs occurring elsewhere in the genome. We report the first case of a duplication encompassing the Wolf-Hirschhorn syndrome critical region in a patient with autism. In addition, we identified clinically relevant *de novo* deletions in regions that have not been implicated in autism so far. This work provides additional evidence that ABCR are often more complex than they appear at first and can hide cryptic rearrangements involved in the abnormal phenotype regardless of the type and inheritance of the rearrangement. Thus, ABCR associated with abnormal phenotype have to be analyzed with genome-wide molecular cytogenetic techniques in order to exclude imbalances either within or outside the initial breakpoints. In the patients where a genomic imbalance at the breakpoints was not detected by array analysis, higher resolution techniques based on next-generation sequencing [52, 53] might be necessary to explore the rearrangement at the nucleotide level.

## Electronic databases

Database of Genomic Variants (DGV), http://dgv.tcag.ca/dgv/app/home
DECIPHER, http://decipher.sanger.ac.uk
International Standards for Cytogenomic Arrays (ISCA) Consortium, https://www.iscaconsortium.org
UCSC Genome Browser, http://genome.ucsc.edu


## Acknowledgements

We are grateful to the patients and families who participated in this study. We thank the Clinical Investigation Center of the Robert Debré Hospital and the DNA and cell bank of the Pitié-Salpêtrière Hospital. This research was supported by Fondation de France, INSERM, and Institut de France. M.P. and E.D. were supported by a PhD fellowship from the French Ministry of Research.


## References


1. Jacobs PA, Browne C, Gregson N, Joyce C, White H: Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. *J Med Genet* 1992, 29(2):103-108.
2. Warburton D: De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet* 1991, 49(5):995-1013.
3. De Gregori M, Ciccone R, Magini P, Pramparo T, Gimelli S, Messa J, Novara F, Vetro A, Rossi E, Maraschio P *et al*: Cryptic deletions are a common finding in "balanced" reciprocal and complex chromosome rearrangements: a study of 59 patients. *J Med Genet* 2007, 44(12):750-762.
4. Sismani C, Kitsiou-Tzeli S, Ioannides M, Christodoulou C, Anastasiadou V, Stylianidou G, Papadopoulou E, Kanavakis E, Kosmaidou-Aravidou Z, Patsalis PC: Cryptic genomic imbalances in patients with de novo or familial apparently balanced translocations and abnormal phenotype. *Mol Cytogenet* 2008, 1:15.
5. Baptista J, Mercer C, Prigmore E, Gribble SM, Carter NP, Maloney V, Thomas NS, Jacobs PA, Crolla JA: Breakpoint mapping and array CGH in translocations: comparison of a phenotypically normal and an abnormal cohort. *Am J Hum Genet* 2008, 82(4):927-936.
6. Schluth-Bolard C, Delobel B, Sanlaville D, Boute O, Cuisset JM, Sukno S, Labalme A, Duban-Bedu B, Plessis G, Jaillard S *et al*: Cryptic genomic imbalances in de novo and inherited apparently balanced chromosomal rearrangements: array CGH study of 47 unrelated cases. *Eur J Med Genet* 2009, 52(5):291-296.
7. Feenstra I, Hanemaaijer N, Sikkema-Raddatz B, Yntema H, Dijkhuizen T, Lugtenberg D, Verheij J, Green A, Hordijk R, Reardon W *et al*: Balanced into array: genome-wide array analysis in 54 patients with an apparently balanced de novo chromosome rearrangement and a meta-analysis. *Eur J Hum Genet* 2011, 19(11):1152-1160.
8. Gijsbers AC, Bosch CA, Dauwerse JG, Giromus O, Hansson K, Hilhorst-Hofstee Y, Kriek M, van Haeringen A, Bijlsma EK, Bakker E *et al*: Additional cryptic CNVs in mentally retarded patients with apparently balanced karyotypes. *Eur J Med Genet* 2010, 53(5):227-233.
9. Gribble SM, Prigmore E, Burford DC, Porter KM, Ng BL, Douglas EJ, Fiegler H, Carr P, Kalaitzopoulos D, Clegg S *et al*: The complex nature of constitutional de novo apparently balanced translocations in patients presenting with abnormal phenotypes. *J Med Genet* 2005, 42(1):8-16.
10. Higgins AW, Alkuraya FS, Bosco AF, Brown KK, Bruns GA, Donovan DJ, Eisenman R, Fan Y, Farra CG, Ferguson HL *et al*: Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am J Hum Genet* 2008, 82(3):712-722.
11. Betancur C: Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* 2011, 1380:42-77.
12. Murdoch JD, State MW: Recent developments in the genetics of autism spectrum disorders. *Curr Opin Genet Dev* 2013, 23(3):310-315.
13. Kielinen M, Rantala H, Timonen E, Linna SL, Moilanen I: Associated medical disorders and disabilities in children with autistic disorder: a population-based study. *Autism* 2004, 8(1):49-60.
14. Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, Caronna EB, Miller KJ, Frazier JA, Silverstein I, Picker J *et al*: Clinical genetic testing for patients with autism spectrum disorders. *Pediatrics* 2010, 125(4):e727-735.
15. Devillard F, Guinchat V, Moreno-De-Luca D, Tabet AC, Gruchy N, Guillem P, Nguyen Morel MA, Leporrier N, Leboyer M, Jouk PS *et al*: Paracentric inversion of chromosome 2 associated with cryptic duplication of 2q14 and deletion of 2q37 in a patient with autism. *Am J Med Genet A* 2010, 152A(9):2346-2354.
16. Haldeman-Englert CR, Chapman KA, Kruger H, Geiger EA, McDonald-McGinn DM, Rappaport E, Zackai EH, Spinner NB, Shaikh TH: A de novo 8.8-Mb deletion of 21q21.1-q21.3 in an autistic male with a complex rearrangement involving chromosomes 6, 10, and 21. *Am J Med Genet A* 2010, 152A(1):196-202.
17. Depienne C, Moreno-De-Luca D, Heron D, Bouteiller D, Gennetier A, Delorme R, Chaste P, Siffroi JP, Chantot-Bastaraud S, Benyahia B *et al*: Screening for genomic rearrangements and methylation abnormalities of the 15q11-q13 region in autism spectrum disorders. *Biol Psychiatry* 2009, 66(4):349-359.

18. Tabet AC, Aboura A, Gerard M, Pilorge M, Dupont C, Gadisseux JF, Hervy N, Pipiras E, Delahaye A, Kanafani S *et al*: Molecular characterization of a de novo 6q24.2q25.3 duplication interrupting UTRN in a patient with arthrogryposis. *Am J Med Genet A* 2010, 152A(7):1781-1788.
19. Groupe de Cytogeneticiens Francais: Pericentric inversions in man. A French collaborative study. *Ann Genet* 1986, 29(3):129-168.
20. Hsu LY, Benn PA, Tannenbaum HL, Perlis TE, Carlson AD: Chromosomal polymorphisms of 1, 9, 16, and Y in 4 major ethnic groups: a large prenatal study. *Am J Med Genet* 1987, 26(1):95-101.
21. Thomas NS, Bryant V, Maloney V, Cockwell AE, Jacobs PA: Investigation of the origins of human autosomal inversions. *Hum Genet* 2008, 123(6):607-616.
22. Van Buggenhout G, Melotte C, Dutta B, Froyen G, Van Hummelen P, Marynen P, Matthijs G, de Ravel T, Devriendt K, Fryns JP *et al*: Mild Wolf-Hirschhorn syndrome: micro-array CGH analysis of atypical 4p16.3 deletions enables refinement of the genotype-phenotype map. *J Med Genet* 2004, 41(9):691-698.
23. Ravnan JB, Tepperberg JH, Papenhausen P, Lamb AN, Hedrick J, Eash D, Ledbetter DH, Martin CL: Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. *J Med Genet* 2006, 43(6):478-489.
24. Patel SV, Dagnew H, Parekh AJ, Koenig E, Conte RA, Macera MJ, Verma RS: Clinical manifestations of trisomy 4p syndrome. *Eur J Pediatr* 1995, 154(6):425-431.
25. Gerard-Blanluet M, Romana S, Munier C, Le Lorc'h M, Kanafani S, Sinico M, Touboul C, Levaillant JM, Haddad B, Lopez N *et al*: Classical West "syndrome" phenotype with a subtelomeric 4p trisomy. *Am J Med Genet A* 2004, 130A(3):299-302.
26. Cotter PD, Kaffe S, Li L, Gershin IF, Hirschhorn K: Loss of subtelomeric sequence associated with a terminal inversion duplication of the short arm of chromosome 4. *Am J Med Genet* 2001, 102(1):76-80.
27. Roselló M, Monfort S, Orellana C, Ferrer-Bolufer I, Quiroga R, Oltra S, Martinez F: Submicroscopic duplication of the Wolf-Hirschhorn critical region with a 4p terminal deletion. *Cytogenet Genome Res* 2009, 125(2):103-108.
28. Carmany EP, Bawle EV: Microduplication of 4p16.3 due to an unbalanced translocation resulting in a mild phenotype. *Am J Med Genet A* 2011, 155A(4):819-824.
29. Wu L, Meng D, Zhou Z, Du J, Long Z, Liang D: A family with partial duplication/deletion 4p due to a balanced t (4; 15) (p16.2; p11.2) translocation. *Am J Med Genet A* 2011, 155A(3):656-659.
30. Schonewolf-Greulich B, Ravn K, Hamborg-Petersen B, Brondum-Nielsen K, Tumer Z: Segregation of a 4p16.3 duplication with a characteristic appearance, macrocephaly, speech delay and mild intellectual disability in a 3-generation family. *Am J Med Genet A* 2013, 161(9):2358-2362.
31. Hannes F, Drozniewska M, Vermeesch JR, Haus O: Duplication of the Wolf-Hirschhorn syndrome critical region causes neurodevelopmental delay. *Eur J Med Genet* 2010, 53(3):136-140.
32. Cyr AB, Nimmakayalu M, Longmuir SQ, Patil SR, Keppler-Noreuil KM, Shchelochkov OA: A novel 4p16.3 microduplication distal to WHSC1 and WHSC2 characterized by oligonucleotide array with new phenotypic features. *Am J Med Genet A* 2011, 155A(9):2224-2228.
33. Sogaard M, Tumer Z, Hjalgrim H, Hahnemann J, Friis B, Ledaal P, Pedersen VF, Baekgaard P, Tommerup N, Cingoz S *et al*: Subtelomeric study of 132 patients with mental retardation reveals 9 chromosomal anomalies and contributes to the delineation of submicroscopic deletions of 1pter, 2qter, 4pter, 5qter and 9qter. *BMC Med Genet* 2005, 6:21.
34. Fisch GS, Grossfeld P, Falk R, Battaglia A, Youngblom J, Simensen R: Cognitive-behavioral features of Wolf-Hirschhorn syndrome and other subtelomeric microdeletions. *Am J Med Genet C Semin Med Genet* 2010, 154C(4):417-426.
35. Paskulin GA, Riegel M, Cotter PD, Kiss A, Rosa RF, Zen PR, Mombach R, Graziadio C: Inv dup del(4)(:p13-->p16.3::p16.3->qter) in a girl without typical manifestations of Wolf-Hirschhorn syndrome. *Am J Med Genet A* 2009, 149A(6):1302-1307.
36. Kondoh Y, Toma T, Ohashi H, Harada N, Yoshiura K, Ohta T, Kishino T, Niikawa N, Matsumoto N: Inv dup del(4)(:p14 --> p16.3::p16.3 --> qter) with manifestations of partial duplication 4p and Wolf-Hirschhorn syndrome. *Am J Med Genet A* 2003, 120A(1):123-126.
37. Zuffardi O, Bonaglia M, Ciccone R, Giorda R: Inverted duplications deletions: underdiagnosed rearrangements?? *Clin Genet* 2009, 75(6):505-513.
38. Rowe LR, Lee JY, Rector L, Kaminsky EB, Brothman AR, Martin CL, South ST: U-type exchange is the most frequent mechanism for inverted duplication with terminal deletion rearrangements. *J Med Genet* 2009, 46(10):694-702.
39. Yu S, Graf WD: Telomere capture as a frequent mechanism for stabilization of the terminal chromosomal deletion associated with inverted duplication. *Cytogenet Genome Res* 2010, 129(4):265-274.
40. Wester U, Bondeson ML, Edeby C, Anneren G: Clinical and molecular characterization of individuals with 18p deletion: a genotype-phenotype correlation. *Am J Med Genet A* 2006, 140(11):1164-1171.
41. Hoy JL, Constable JR, Vicini S, Fu Z, Washbourne P: SynCAM1 recruits NMDA receptors via protein 4.1B. *Mol Cell Neurosci* 2009, 42(4):466-483.
42. Leroy C, Landais E, Briault S, David A, Tassy O, Gruchy N, Delobel B, Gregoire MJ, Leheup B, Taine L *et al*: The 2q37-deletion syndrome: an update of the clinical spectrum including overweight, brachydactyly and behavioural features in 14 new patients. *Eur J Hum Genet* 2013, 21(6):602-612.
43. Williams SR, Aldred MA, Der Kaloustian VM, Halal F, Gowans G, McLeod DR, Zondag S, Toriello HV, Magenis RE, Elsea SH: Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *Am J Hum Genet* 2010, 87(2):219-228.

44. Cukier HN, Skaar DA, Rayner-Evans MY, Konidari I, Whitehead PL, Jaworski JM, Cuccaro ML, Pericak-Vance MA, Gilbert JR: Identification of chromosome 7 inversion breakpoints in an autistic family narrows candidate region for autism susceptibility. *Autism Res* 2009, 2(5):258-266.
45. Vincent JB, Choufani S, Horike S, Stachowiak B, Li M, Dill FJ, Marshall C, Hrynchak M, Pewsey E, Ukadike KC *et al*: A translocation t(6;7)(p11-p12;q22) associated with autism and mental retardation: localization and identification of candidate genes at the breakpoints. *Psychiatr Genet* 2008, 18(3):101-109.
46. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A *et al*: Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008, 320(5875):539-543.
47. Sahoo T, Theisen A, Rosenfeld JA, Lamb AN, Ravnan JB, Schultz RA, Torchia BS, Neill N, Casci I, Bejjani BA *et al*: Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. *Genet Med* 2011, 13(10):868-880.
48. Hauge X, Raca G, Cooper S, May K, Spiro R, Adam M, Martin CL: Detailed characterization of, and clinical correlations in, 10 patients with distal deletions of chromosome 9p. *Genet Med* 2008, 10(8):599-611.
49. Bonnafe E, Touka M, AitLounis A, Baas D, Barras E, Ucla C, Moreau A, Flamant F, Dubruille R, Couble P *et al*: The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. *Mol Cell Biol* 2004, 24(10):4417-4427.
50. Baas D, Meiniel A, Benadiba C, Bonnafe E, Meiniel O, Reith W, Durand B: A deficiency in RFX3 causes hydrocephalus associated with abnormal differentiation of ependymal cells. *Eur J Neurosci* 2006, 24(4):1020-1030.
51. Benadiba C, Magnani D, Niquille M, Morle L, Valloton D, Nawabi H, Ait-Lounis A, Otsmane B, Reith W, Theil T *et al*: The ciliogenic transcription factor RFX3 regulates early midline distribution of guidepost neurons required for corpus callosum development. *PLoS Genet* 2012, 8(3):e1002606.
52. Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM *et al*: Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 2012, 149(3):525-537.
53. Schluth-Bolard C, Labalme A, Cordier MP, Till M, Nadeau G, Tevissen H, Lesca G, Boutry-Kryza N, Rossignol S, Rocas D *et al*: Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *J Med Genet* 2013, 50(3):144-150.
54. Bitoun P, Philippe C, Cherif M, Mulcahy MT, Gilgenkrantz S: Incontinentia pigmenti (type 1) and X;5 translocation. *Ann Genet* 1992, 35(1):51-54.

**Table 1. Clinical and cytogenetic characteristics of patients**

| Patient | Initial karyotype | Sex | Age | ASD | Cognitive level | Language | Birth and early development | Body measures | Dysmorphic features | Other | Brain MRI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1* | 46,XY,der(4)t(4;acrop)(p16.3;acro p)dn | M | 4 y | Autism (DISCO) | VIQ 67, PIQ 78, FSIQ 69 (WPPSI-R, 34 mo) | Functional language | Born at 42 wk, W +4.3 SD (gestational diabetes). Normal motor milestones | 4 y: W & H +2-3 SD | No dysmorphic features, mild strabismus | Inguinal hernia | Not done |
| 2* | 46,XY,t(5;18)(q12;p11.2)dn | M | 5 y | Autism (ADI-R) | DQ 52 (PEP-R, 40 mo) | 1st words 36 mo, no sentences | Born at 38 wk, W –0.6 SD, H –1 SD, OFC mean. Walked at 17 mo | 5 y: W –1 SD, H +1 SD, OFC –0.4 SD | No dysmorphic features, slender habitus, long fingers and toes, numerous secondary palmar creases | Frequent otitis, eczema | Small bilateral insular hypersignal suggesting myelinisation delay (at 4 y) |
| 3** | 46,XY,inv(11)(p14q23.2)dn | M | 4 y | Autism (ADI-R, ADOS) | DQ 13 mo (Brunet Lezine-R, 4 y) | Non verbal | Born at 37 wk, W –3.5 SD, H –3.5 SD, OFC –2.5 SD. Delayed motor development, unable to walk at 4 y | 42 mo: W, H & OFC –3 SD | Mild dysmorphic features, high forehead, horizontal eyebrows, bulbous nose, thin upper lip | Aggressiveness, anxiety, stereotypies, head banging, bilateral cryptorchidism, axial hypotonia, hyperlaxity | Supratentorial ventricular enlargement, increased subarachnoid spaces |
| 4* | 46,XY,inv(2)(q14q36)mat (reported previously [15]) | M | 14 y | Autism (ADI-R) | VIQ 46, PIQ 50, FSIQ 46 (WISC-III, 14 y) | 1st words 30 mo, 1st sentences 60 mo | Born at 39 wk, W mean, H –0.5 SD, OFC –1.1 SD. Sat at 9 mo, walked at 14 mo | 12 y: W +0.5 SD, H –0.5 SD, OFC mean | Mild dysmorphic features, frontal bossing, flattened nasal bridge, deep-set eyes, downslanting palpebral fissures, thin upper lip | Hyperactivity, head banging, anxiety, asthma, insulin-dependent diabetes, growth hormone deficit | Normal (at 3 y) |
| 5* | 46,XX,inv(2)(p12q14.1)dn (benign cytogenetic variant) | F | 13 y | Autism (ADI-R) | DQ 30 (PEP-R, 6 y) | Non verbal | Born at t39 wk, W +0.5 SD, H mean, OFC +1.9 SD. Sat at normal age, walked at 22 mo | 13 y: W +3.5 SD, H +2 SD, macrocephaly (OFC +4.2 SD) | Mild dysmorphic features, upslanted palpebral fissures, short philtrum, large hands and feet, one café-au-lait spot | Stereotypies, self-injurious behavior, sleep disturbance, one episode of febrile seizure at 31 mo, normal EEG | Chiari type I malformation, mild white matter hypersintensities (at 9 y) |
| 6* | 46,XX,inv(9)(p11q13)dn (benign cytogenetic variant) | F | 7 y | Autism (ADI-R) | Intellectual disability | Isolated words | Born at term, W +0.5 SD, H +0.9 SD. Walked at 12 mo | 7 y 2 mo: W +3.8 SD, H +4.1 SD), OFC +1.7 SD | No dysmorphic features | Precocious puberty at 7 y, bone age 8 y, normal hormone levels, pulmonar stenosis, moderate systolic murmur | Normal (at 7 y) |
| 7* | 46,XY,inv(5)(q13q34)dn | M | 18 y | Autism (ADI-R) | VIQ 93, PIQ 80, FSIQ 86 (WISC-III, 18 y) | No delay, functional language | Born at 40 wk, mean W, H and OFC. Sat at 9 mo, walked at 20 mo | 18 y: W –1.4 SD, H mean, OFC +1.6 SD | No dysmorphic features, mild clubbing | Refraction error | Myelinisation delay (at 18 y) |
| 8** | 46,XY,t(9;19)(p12;q13.4)dn | M | 7 y | Autism (ADI-R, ADOS) | VIQ 83, PIQ 80, FSIQ 79 (WPPSI, 5 y 5 mo) | 1st words 24 mo, 1st sentences 36 mo, functional language | Born at term, W +1 SD, OFC +0.3 SD. Sat at 7 mo, walked at 11 mo | NA | No dysmorphic features | ADHD | Normal |
| 9** | 46,XX,t(X;5)(p11.2;q35.2)dn (reported previously [54]) | F | 5 y | ASD‡ | Normal intelligence | Non verbal | NA | NA | Non specific dysmorphic features | Incontinentia pigmenti type 1 (hypomelanosis of Ito), pigmented lesions on the legs, no epilepsy, abnormal EEG | Not done |
| 10* | 46,XY,t(20;21)(q11.2;q21)dn | M | 12 y | Autism (ADI-R) | DQ 20 (PEP-R, 12 y) | Non verbal | Born at 41 wk, W –1.7 SD, H –0.5 SD, OFC –1.1 SD. Sat at 9 mo, walked at 18 mo | 12 y: W –1.8 SD, H mean, microcephaly (–3.6 SD) | Dysmorphic features, low forehead, thick eyebrows, long nose, short philtrum, right single transverse palmar crease | Strabismus, epilepsy (onset at 9 mo) | Not done |
| 11* | 46,XY,t(9;16)(q3.2;p1.2)mat | M | 18 y | Autism (ADI-R) | VIQ 46, PIQ 46, FSIQ 40 (WISC-III, 14 y) | Language delay, 1st phrases 5 y, functional language | Born at term, W +1.3 SD, H +1 SD, OFC +2.1 SD. Sat at 8 mo, walked at 14 mo | 18 y: W +3 SD, H mean, OFC +2.3 SD | No dysmorphic features, short neck, brachymetatarsia of 4th and 5th rays | Hyperactivity in childhood, strabismus, hypermetropia | Not done |
| 12* | 46,XY,t(3;8)(q13.2;p23)mat | M | 5 y | Autism (DSM-IV, CARS) | VIQ 79, PIQ 74, FSIQ 77 (Stanford-Binet IV, 3 y 9 mo) | 1st words 26 mo, 1st sentences 36 mo, few sentences at 5 y, dysarthric speech | Normal pregnancy, delivery and early development; walked at 13 mo | 4 y: W, H and OFC within normal limits | No dysmorphic facial features, long ring finger, brachymesophalangia V, sandal gap | Severe ADHD, pica, self-injurious behavior, partial complex epilepsy (onset at 4 y, currently seizure free on medication), chronic ear infections, chronic diarrhea, food allergies | Not done |
| 13** | 46,XY, t(2;20)(q13;q13.33)mat | M | 22 y | ASD‡ | Intellectual disability | Language delay, uses only a few words, answers with signs | Born at term, W –0.8 SD, H –1 SD, OFC –1.1 SD. Sat at 11 mo, walked at 2 y | 22 y: W +1.4 SD, H mean, OFC –1 SD | No dysmorphic facial features, bilateral clinodactyly 5th finger, several café au lait spots | Hyperactivity, aggressiveness, inappropriate laughter, hand stereotypies, unilateral strabismus, seizures at 7 mo, no recurrence after stopping treatment | Normal (at 1 y) |
| 14* | 46,XX,inv(5)(p13.3q13.3)pat (benign cytogenetic variant) | F | 8 y | Autism (ADI-R) | DQ 55 (PEP-R, 5 y) | 1st words 18 mo, 1st sentences before 3 y; echolalic language | Dizygotic twin pregnancy, born at 27 wk, W 880 g. Walked at 18 mo | 8 y: W mean, H +1 SD, OFC +1 SD | No dysmorphic features, long face, flat feet, tuberous angioma scar on scapula | Insensitivity to pain | Non specific white substance hypersignal (at 5 y) |

**Table 1. Clinical and cytogenetic characteristics of patients *(continued)***

| Patient | Initial karyotype | Sex | Age | ASD | Cognitive level | Language | Birth and early development | Body measures | Dysmorphic features | Other | Brain MRI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 15* | 46,X,inv(Y)(p11q11)pat (benign cytogenetic variant) | M | 14 y | Autism (ADI-R) | DQ 15 (PEP-R, 14 y) | Non verbal | Born at 38 wk, W –2.1 SD, H –1.5 SD, OFC –2 SD. Sat at 9 mo, walked at 13 mo | 14 y: W +2 SD, H +1 SD, OFC mean | Synophris, abnormal dental implantation (delayed tooth loss) | | Normal (at 14 y) |
| 16** | 46,XY,t(2;13)(p23;q14)pat | M | 7 y | Autism (ADI-R, ADOS) | PIQ <1$^{st}$ %ile (Raven, 7 y 5 mo) | 1$^{st}$ words 36 mo, 1$^{st}$ sentences 48 mo, functional language | Born at term, W –1.2 SD, H +0.5 SD, OFC +0.7 SD. Sat at 20 mo, walked at 36 mo | NA | No dysmorphic features | ADHD | Normal |
| 17* | 46,XY,t(4;9)(p13;p23) (not maternal, father not tested) | M | 30 y | Autism (ADI-R, ADOS) | DQ <20 (PEP-R, 16 y) | Non verbal | Born at term, W +0.7 SD, H +1 SD. Feeding difficulties, hypotonia, growth delay (–3 SD). Walked at 4 y | 30 y: W –1 SD, H +1.1 SD, OFC –0.5 SD | No dysmorphic features except for enophtalmia, low-set eyebrows, large ears, narrow hands | Hyperactivity, severe sleep disturbance, insensitivity to pain, gastric and esophageal ulcers, abnormal EEG at 1 y (temporal lobe focus), no epilepsy | Not done (encephalography at 6 mo: global ventricular dilatation) |
| 18** | 46,XY,t(2;13)(q22;q31) (not maternal, father not tested) | M | 4 y | Autism (ADI-R, ADOS) | IQ <1$^{th}$ %ile (Raven, 4 y 8 mo) | Non verbal | Born post-term at 47 wk, W –0.8 SD, H –0.8 SD, OFC –0.6 SD. Sat at 20 mo, walked at 36 mo | 4 y: W –0.5 SD, H –0.5 SD, OFC –2 SD | No dysmorphic features, brachyplagiocephaly | ADHD, aggressiveness, anxiety | Normal |

* Patients from the PARIS cohort; ** patients referred by the network of French cytogeneticists; ‡ ASD was not formally evaluated in these two patients. Abbreviations: ADHD, attention deficit-hyperactivity disorder; ASD, autism spectrum disorder; CARS, Childhood Autism Rating Scale; DISCO, Diagnostic Interview for Social and Communication Disorders; DQ, developmental quotient; FSIQ, full scale IQ; H, height; IQ, intellectual quotient; mo, months; MRI, magnetic resonance imaging; NA, not available; OFC, occipitofrontal circumference; PEP-R, Psychoeducational Profile Revised; PIQ, performance IQ; SD, standard deviation; VIQ, verbal IQ; W, weight; WISC-III, Wechsler Intelligence Scale for Children, third edition; wk, weeks; WPPSI-R, Wechsler Preschool and Primary Scale of Intelligence Revised; y, years

**Table 2. *De novo* abnormalities detected by SNP array**

| Patient | Initial karyotype | Imbalance | Chromosomal region | Start | End | Size (bp) | Parental origin |
|---|---|---|---|---|---|---|---|
| **Clinically significant abnormalities** | | | | | | | |
| 1[a] | 46,XY,der(4)t(4;acrop)(p16.3;acro p)dn | Deletion | 4p16.3 | 17,764 | 558,839 | 541,076 | Paternal |
| | | Duplication | 4p16.2p16.3 | 577,581 | 4,812,859 | 4,235,279 | Paternal |
| 2 | 46,XY,t(5;18)(q12;p11.2)dn | Deletion | 18p11.22p11.31 | 5,408,997 | 9,625,750 | 4,216,754 | Maternal |
| 3 | 46,XY,inv(11)(q14q23.2)dn | Deletion | 7q21.3q22.1 | 97,043,362 | 101,977,945 | 4,934,584 | Paternal |
| 4[a] | 46,XY,inv(2)(q14q36)mat | Deletion | 2q37.3 | 239,735,269 | 243,044,147 | 3,308,879 | Maternal |
| | | Duplication | 2q14.1q14.2 | 117,072,756 | 121,304,548 | 4,231,793 | Maternal |
| **Variant of unknown clinical significance** | | | | | | | |
| 5 | 46,XX,inv(2)(p12q14.1)dn | Deletion | 9p24.2 | 3,334,203 | 3,523,286 | 179,084 | Paternal |

[a]In Patient 1, the 4p duplication is pathogenic; the deletion is considered benign. In Patient 4, the 2q37 deletion is pathogenic; the clinical significance of the duplication is unknown.

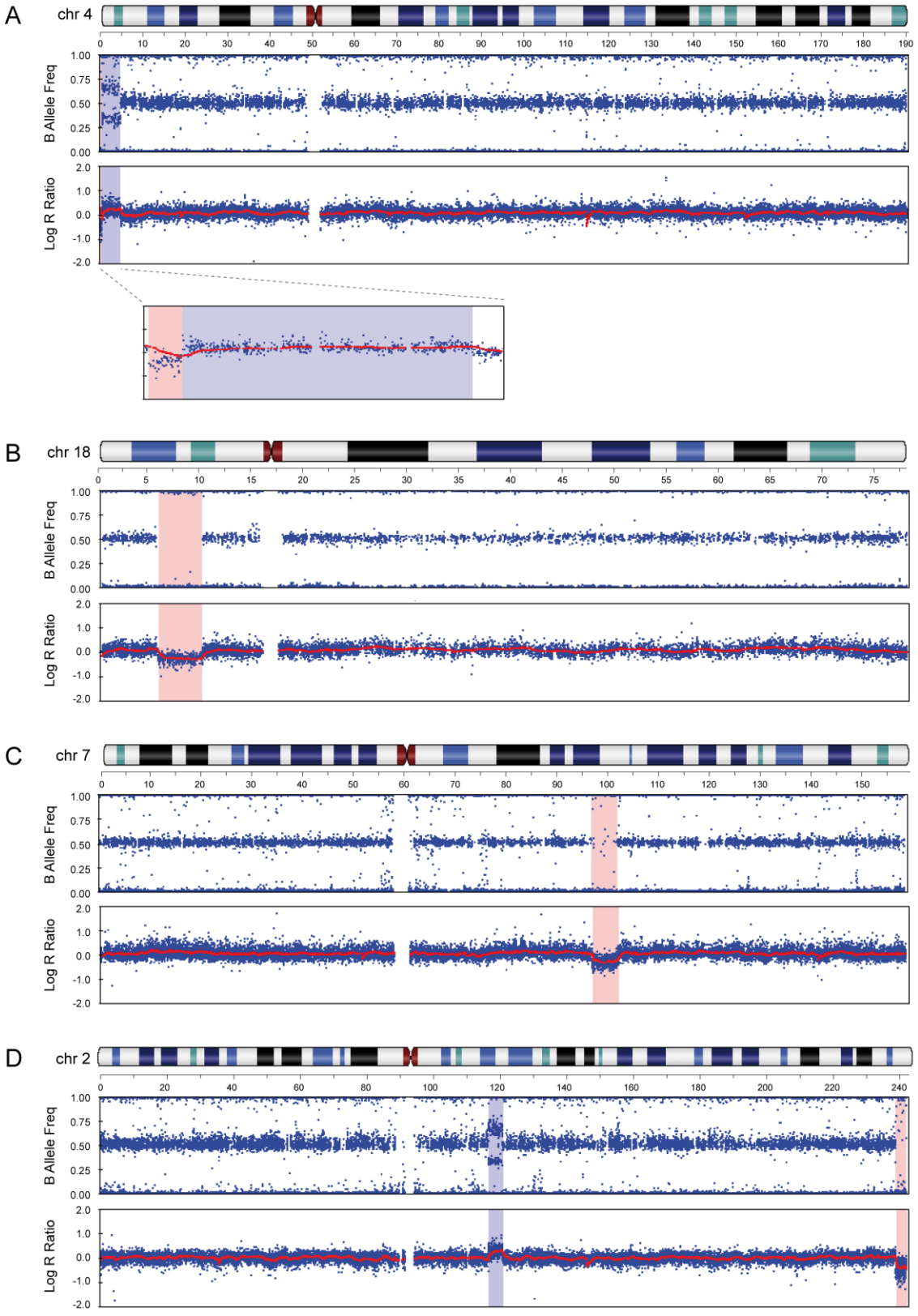



**Figure 1.** SNP arrays of patients with clinical significant findings. **A.** Patient 1 (46,XY,der(4)t(4;acro p)dn) has a copy number loss of 541 kb in 4p16.3, and a copy number gain of 4.29 Mb in 4p16.2p16.3. The distal deleted and duplicated 4p segments are shown in detail below. **B**. Patient 2 (t(5;18)(q11.2;p11.2)dn) has deletion of 4.21 Mb in 18p11.31p11.22. **C.** Patient 3 (46,XY,inv(11)(p14q23.2)dn) had no imbalance on chromosome 11 but the SNP array revealed a *de novo* deletion of 4.3 Mb in 7q21.3q22.1. **D.** Patient 4 (46,XY,inv(2)(q14.2q37.3)mat) has a maternally inherited paracentric inversion of chromosome 2q, with a 2q14.1q14.2 duplication of 4.2 Mb at the proximal breakpoint, and a 2q37.3 deletion of 3.5 Mb extending to the telomere.

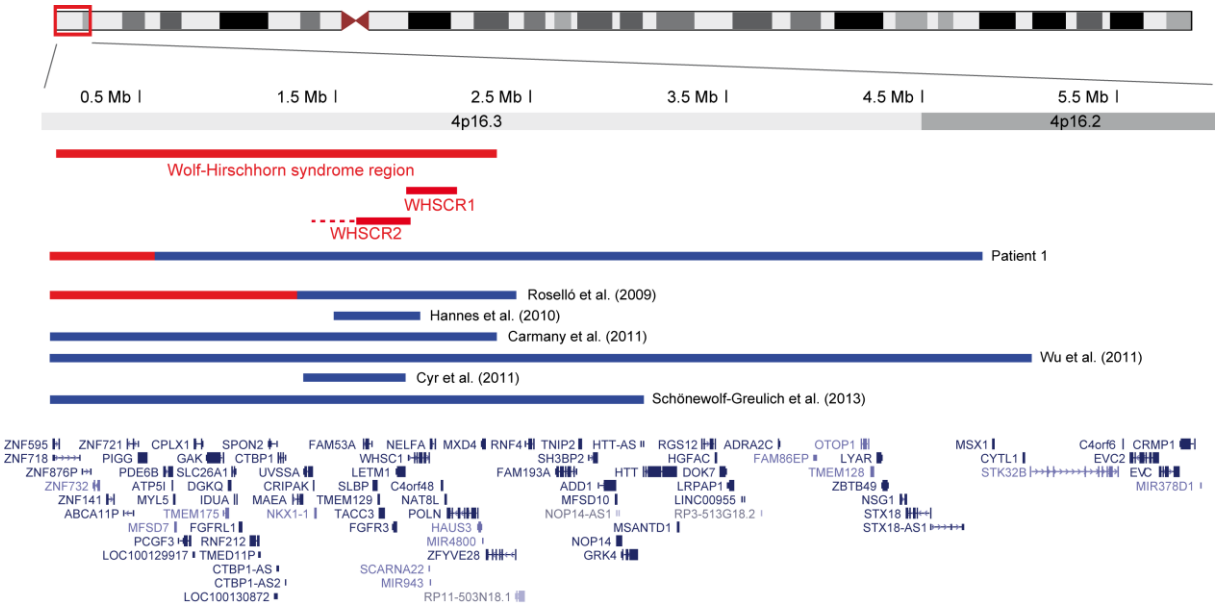



**Figure 2**. Duplication of the Wolf-Hirschhorn region in Patient 1. Map of chromosome 4 (1-6,000,000, hg19) showing the rearrangement detected in Patient 1 and previously reported overlapping rearrangements. The region commonly deleted in Wolf-Hirschhorn syndrome and the two proposed critical regions (WHSCR1 and WHSCR2) are represented. Horizontal blue lines represent duplications and red lines indicate deletions. RefSeq genes that lie in this region are indicated at the bottom of the map (UCSC, hg19).