



HAL
open science

Optimisation du taux de diagnostics en séquençage d'exome dans le trouble du spectre de l'autisme à l'aide d'épisignatures

Thomas Husson

► **To cite this version:**

Thomas Husson. Optimisation du taux de diagnostics en séquençage d'exome dans le trouble du spectre de l'autisme à l'aide d'épisignatures. Sciences agricoles. Normandie Université, 2022. Français. NNT : 2022NORMR072 . tel-04199453

HAL Id: tel-04199453

<https://theses.hal.science/tel-04199453v1>

Submitted on 7 Sep 2023

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

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



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité ASPECTS MOLECULAIRES ET CELLULAIRES DE LA BIOLOGIE

Préparée au sein de l'Université de Rouen Normandie

Optimisation du taux de diagnostics en séquençage d'exome dans le trouble du spectre de l'autisme à l'aide d'épisignatures

Présentée et soutenue par
THOMAS HUSSON

Thèse soutenue le 06/12/2022
devant le jury composé de

MME FREDERIQUE BONNET-BRILHAULT	PROFESSEUR DES UNIV - PRATICIEN HOSP., UNIVERSITE DE TOURS	Rapporteur du jury
MME LAURENCE OLIVIER FAIVRE	PROFESSEUR DES UNIV - PRATICIEN HOSP., CENTRE HOSPITALIER UNIVERSITAIRE DIJON- BOURGOGNE	Rapporteur du jury
MME CAROLINE DEMILY	PROFESSEUR DES UNIV - PRATICIEN HOSP., UNIVERSITE LYON 1 CLAUDE BERNARD	Membre du jury
MME CAMILLE LE CLEZIO	MAITRE DE CONFERENCES, Université de Rouen Normandie	Membre du jury
M. THOMAS SMOL	MAITRE DE CONF UNIV. - PRATICIEN HOSP., UNIVERSITE DE LILLE	Membre du jury
M. DOMINIQUE CAMPION	DIRECTEUR DE RECHERCHE, Université de Rouen Normandie	Directeur de thèse

Thèse dirigée par **DOMINIQUE CAMPION (CANCER AND BRAIN GENOMICS)**

RESUME

A ce jour, il existe plusieurs centaines de gènes pour lesquels les variations génétiques rares pathogènes sont considérées comme des facteurs de risque de Trouble du Spectre de l'Autisme (TSA). Plusieurs paradigmes expérimentaux ont concouru à la constitution de ces listes de gènes partiellement chevauchantes. Dans la mesure où il est recommandé de proposer un bilan génétique à tous les patients diagnostiqués TSA, il nous a paru nécessaire de constituer une liste de gènes associés au TSA la plus exhaustive possible afin d'établir le taux de détection de variations génétiques pathogènes dans cette population. Lors d'une première étude rétrospective nous avons séquencé 256 patients recrutés en génétique clinique. Nous avons détecté un facteur de risque génétique de TSA chez 20% d'entre eux. Ce taux de détection était plus important chez les patients présentant une déficience intellectuelle (30%) mais il restait substantiel chez les autres patients. Nous avons confirmé, dans une certaine mesure, ces données sur une cohorte prospective – afin de limiter le risque de biais de recrutement. Nous avons obtenu un taux similaire de variations nucléotidiques (autour de 10-12% dans les deux cas) mais nous avons détecté moins de variations du nombre de copie. Nous avons également montré que certaines caractéristiques cliniques du patient (notamment le retard à la marche) était associé à un plus fort taux de détection. Enfin, nous avons évalué l'intérêt (important) et les craintes (limitées) des patients et de leur famille après la consultation de génétique.

Comme habituellement dans les études de séquençage à haut débit, nous avons détecté un grand nombre de variations de signification inconnues. Dans une seconde partie de ce travail, nous nous sommes concentrés sur une stratégie d'aide à la classification de ces variations de signification inconnues à l'aide des analyses de méthylation de l'ADN sur puce. Pour plusieurs gènes, les variations pathogènes ont été montrées comme associées à des motifs aberrants de méthylation de l'ADN appelés « épisignatures ».

Nous avons effectué une évaluation indépendante des qualités diagnostiques de certaines épisignatures publiées. Pour ce faire nous avons mis en œuvre une collaboration nationale avec les services de génétique français afin de recueillir un nombre important de contrôles positifs pour 10 syndromes neurodéveloppementaux pour lesquels une épisignature a été rapportée. Nous avons montré que les performances de ces épisignatures étaient très hétérogènes et nous avons édicté des recommandations pour leur utilisation en pratique clinique.

Nous avons également essayé de détecter une épisignature pour le gène *MBD5*. Cette analyse a révélé l'ampleur des défis techniques et méthodologiques soulevés par l'analyse des épisignatures

Mots clés : variations génétiques rares, trouble du spectre de l'autisme, séquençage d'exome, variations *de novo*, analyse de méthylation l'ADN, épisignatures, biomarqueurs.

ABSTRACT

To date, there are several hundred genes for which rare pathogenic genetic variants are regarded as risk factors for Autism Spectrum Disorder (ASD). Different experimental paradigms have contributed to the constitution of these partially overlapping gene lists. Because it is recommended to offer a genetic check-up for all patients diagnosed with ASD, we felt necessary to assemble an exhaustive list of genes ASD-associated genes. Then, we aim to establish the detection rate of pathogenic variants in this population. In a first retrospective study we sequenced 256 patients recruited in our genetic department. We detected a genetic risk factor for ASD in 20% of them. This detection rate was higher in patients with intellectual disability (30%) but remained substantial in others. We confirmed these data to some extent in a prospective cohort - to attenuate recruitment bias risk. We obtained a similar rate of nucleotide variations (around 10-12% in both cases) but detected fewer copy number variations. We also showed that certain clinical characteristics of the patient (notably free walking delay) were associated with a higher detection rate. Finally, we assessed the (important) interest and (limited) fears of patients and their families after the genetic consultation.

As usual in high-throughput sequencing studies, we have detected a large number of variations of unknown significance. In a second part of this work, we focused on a strategy to help classify these variations of unknown significance using DNA methylation analysis. For several genes, pathogenic variations have been associated with aberrant DNA methylation patterns called "episignatures".

We conducted an independent evaluation of the diagnostic performances of several episignatures. A national collaboration with French geneticists allows us to collect a large number of positive controls for ten neurodevelopmental syndromes for which an episignature has been reported. We showed that the performance of these episignatures was highly heterogeneous and made recommendations for their use in clinical practice.

Finally, we tried to detect an episignature for the *MBD5* gene. This analysis revealed the extent of the technical and methodological challenges raised by the analysis of episignatures

Key word: rare genetic variants, autism spectrum disorder, whole-exome sequencing, *de novo* variants, DNA methylation analysis, episignatures, biomarkers.

REMERCIEMENTS

Je remercie tout d'abord Madame la Professeure Frédérique Bonnet-Brilhault et Madame la Professeure Laurence Faivre de me faire l'honneur de juger mon travail et d'user de leur expertise afin de produire le rapport de ma thèse.

Merci également à Madame la Professeure Caroline Demily et Monsieur le Docteur Thomas Smol d'avoir accepté d'évaluer mes travaux de thèse.

Je tiens à saluer la mémoire de Monsieur le Professeur Thierry Frebourg. Celui-ci a participé à ma formation dès la première année de médecine, puis m'a accueilli dans son laboratoire et dans son service lors de mon internat. Son enseignement a été inestimable.

Afin de produire un travail qui soit digne de votre jugement, je n'avais « *rien d'autre à faire qu'à faire pour le mieux.* » *Emile Durkheim.*

Merci tout particulièrement à mon directeur de thèse Dominique Champion et ma co-encadrante Camille Charbonnier. Dominique, merci de m'avoir enseigné ce qu'est la science. J'ai ainsi compris que « *Lorsqu'un problème est clairement et judicieusement posé, on trouve toujours des hommes capables de le résoudre* » *Léon Trotski*, et que « *La réalité ne pardonne pas une seule erreur à la théorie* » *Léon Trotski*. Camille, merci tout particulièrement pour ton soutien, ta patience et ta bienveillance. Puisque, en plus des compétences techniques, « *L'autorité morale est la qualité maîtresse de l'éducateur* » *Emile Durkheim*, alors tu es une formidable enseignante. Cette aventure m'a changé, et si c'est en bien c'est grâce à vous.

Deux de mes maîtres méritent tout particulièrement ma gratitude pour leur soutien lors de ces trois années. Gaël, tu as gagné mon respect par ta discipline

de travail et ta finesse intellectuelle et mon admiration pour avoir compris et appliqué qu'« *une pensée positive produit des millions de vibrations positives* » *John Coltrane*. Olivier, sans tes encouragements continus, tes conseils avisés et ta patience à mon égard il est certain que je n'aurai pas réussi à aller « *au-delà de me rêves* » *Simon Delestre*.

Merci à tous les membres de l'unité 1245 et de l'unité de génétique clinique pour leur agréable compagnie et leur aptitude à l'entraide. Bon courage aux futurs doctorants, dont Corentin que j'ai eu la chance d'adopter il y a de cela quelques années. Je dois noter que ce travail a grandement bénéficié de l'énergie et de la rigueur d'Anne-Claire. Alors que je ne vois que des pipettes et des tubes à essai, tu es capable de « *regarder au-delà de ce que je vois* » (adapté de *Rafiki* dans le *Roi Lion*) et d'en extraire de la connaissance.

Merci à Maud et Laëtitia, votre amitié aura été ma plus belle découverte depuis que je suis arrivé au laboratoire. Je vous remercie pour votre écoute, ainsi que pour vos conseils, y compris ceux que je n'ai pas suivis. Il se murmure que votre force de volonté aurait inspiré *Olympe de Gouges* alors qu'elle écrivait : « *quelles que soient les barrières que l'on vous oppose, il est en votre pouvoir de vous en affranchir, vous n'avez qu'à le vouloir* » et *Louise Michel* lorsqu'elle témoignait « *Chacun cherche sa route ; nous cherchons la nôtre et nous pensons que le jour où le règne de la liberté et de l'égalité sera arrivé, le genre humain sera heureux* ».

Merci aux membres de ma famille, ils m'ont toujours inspiré. A ma mère, Florence, d'avoir toujours veillé sur moi et de toujours me soutenir dans mes choix – même lorsqu'ils paraissent étranges. Je suis très heureux que tu puisses enfin dire « *le moment présent était un moment de joie* » *René Barjavel*, tu le mérites. A ma sœur, Clara : « *sur une île vivait un hobbit* » et comme *Tolkien* elle « *se sentait chez elle parmi les arbres* ». Tu ne mesures pas à quel point je suis fier d'être ton frère.

A mes meilleurs amis, Thibault, Nadir, Laure et Benji. Je vous remercie d'être qui vous êtes. Ne cessez jamais de croire que « *le sourire est la meilleure façon de faire face à tous les problèmes, d'écraser toutes les peurs, de cacher chaque douleur* », Will Smith, d'essayer « *d'imaginer Sisyphe heureux* », Albert Camus, de voir le monde « *d'une autre perspective, comme s'il était neuf* », Dr Manhattan dans *Watchmen*, et de continuer d'« *Enseigner sans imposer, donner le goût de jouer à la vie, c'est-à-dire à comprendre puis à découvrir le monde* » Henri Laborit. Je me considérerai chanceux de pouvoir côtoyer ne serait-ce que l'un d'entre vous. Ensemble, et avec tous ceux qui auraient mérité d'être cités dans ces lignes, vous faites de ma vie une fête.

Merci à Madu de s'être laissée apprivoiser et à Thor de m'avoir assisté du mieux de ses capacités pendant la rédaction de deux thèses et la préparation du concours de l'internat. « *Lorsqu'un chat accorde sa confiance à un homme, c'est sa plus belle offrande* », Charles Darwin.

Enfin et surtout, merci Fede. Tu es à la fois la meilleure des amies et l'être aimé. Partie poussine, partie bourricotte. Merveilleuse fée dont les enchantements hauts en couleurs m'empêchent de ne jamais m'ennuyer. Merci de me supporter (dans tous les sens du terme). Merci de ton amour sans conditions. L'un de tes compatriotes, il y a bien longtemps, a très joliment décrit ce que je ressens concernant notre rencontre.

« *Nel mezzo del cammin di nostra vita*

mi ritrovai per una selva oscura,

ché la diritta via era smarrita. »

...puis nous nous sommes trouvés et...

*« Io mi senti' svegliar dentro a lo core
un spirito amoroso che dormia:
e poi vidi venir da lungi Amore
allegro sì, che appena il conoscia »*

Dante Alighieri

TABLE DES MATIERES

AVANT-PROPOS	16
INTRODUCTION	21
<i>Chapitre I. Le concept d'autisme et ses évolutions</i>	22
<i>Chapitre II. Un syndrome neurodéveloppemental d'apparition précoce</i>	29
<i>Chapitre III. Une causalité multifactorielle</i>	42
EVALUATION DU SEQUENÇAGE D'EXOME DANS LE DIAGNOSTIC GENETIQUE DU TSA	61
<i>Chapitre I. Les variations génétiques rares dans le TSA</i>	62
i. L'implication de variations génétiques rares dans le TSA	62
ii. Un changement de paradigme avec le séquençage à haut débit	75
iii. Variations <i>de novo</i> et études de trios dans le TSA	86
<i>Chapitre II. Taux de détection de facteurs de risque de TSA</i>	104
i. Objectifs et résumé du travail publié	104
ii. Rare genetic susceptibility variants assessment in autism spectrum disorder: detection rate and practical use	109
<i>Chapitre III. Evaluation du bénéfice patient du séquençage d'exome dans le TSA</i>	118
INTERPRETATION DE VSI A L'AIDE D'EPISIGNATURES	141
<i>Chapitre I. Analyse de méthylation dans les syndromes neurodéveloppementaux</i>	142
i. Quel biomarqueur pour aider à classifier des VSI ?	142
ii. Caractéristiques techniques des puces Illumina	158
iii. Traitement bioinformatique des données de méthylation	164
<i>Chapitre II. Conception d'une épisignature pour le gène MBD5</i>	178

<i>Chapitre III. Evaluation des épisignatures publiées</i>	<i>194</i>
i. Objectifs et résumé du travail publié	194
<i>DISCUSSION & CONCLUSION</i>	<i>220</i>
<i>REFERENCES BIBLIOGRAPHIQUES</i>	<i>228</i>

LISTE DES FIGURES

Figure 1. Graphique en forêt des prévalences estimées et de leurs intervalles de confiance à 95% (IC95%) pour l'autisme typique (à gauche) et pour le TSA (à droite), p.27

Figure 2. Comparaison normalisée des périmètres crâniens entre patients présentant un TSA et témoins chez le fœtus et l'enfant, p.33

Figure 3. Représentation chronologique des différentes étapes de développement cérébral pour lesquels des anomalies sont suspectées chez les patients diagnostiqués avec un TSA, p.34

Figure 4. Représentation des trois principaux réseaux fonctionnels, p.36

Figure 5. Surrisque de TSA chez les apparentés de patients, en fonction de leur degré de proximité génétique, p.49

Figures 6. Explication schématique du principe des études de jumeaux, p.51

Figure 7. Etudes de jumeaux publiées dans le TSA, p.53

Figure 8. Schéma illustrant le principe des GWAS et Manhattan plot des résultats de la principale étude dans le TSA, p.56

Figure 9. Distribution des scores du quotient de spectre autistique chez des jumeaux et leurs apparentés, issus de la population générale, p.58

Figure 10. Répartition des odds-ratio en fonction du PRS pour 13 076 cas et 22 664 contrôles, p.59

Figure 11. Illustration des variations génétiques, p.65

Figure 12. Trois exemples de syndrome neurodéveloppementaux, p.66

Figure 13. Illustration du concept de variations du nombre de copies, p.70

Figure 14. Corrélation entre la localisation de la délétion et les symptômes cliniques des patients présentant une délétion de l'extrémité du bras court du chromosome 5, p.72

Figure 15. CNVs larges (> 400 kbp) et récurrents chez les sujets présentant un TSA comparés à des contrôles, p.74

Figure 16. Evolution du coût du séquençage du génome humain entre 2001 et 2021, p.77

Figure 17. Le principe de la technologie de séquençage à haut débit Illumina, p.80

Figure 18. Répartition des variations détectées dans l'étude de Karczewski et collaborateurs (2020), p.81

Figure 19. Proportion de gènes de la base de données OMIM (*Online Mendelian Inheritance in Man*) pour chaque décile de LOEUF, p.84

Figure 20. Source des variations *de novo*, p.89

Figure 21. Variations *de novo* dans le TSA, p.91

Figure 22. Comparatif du taux de variations *de novo*, p.91

Figure 23. Modèle pour la détermination de gènes avec excès de variations *de novo* dans le TSA, p.95

Figure 24. Taux comparé de variations *de novo* délétères chez les patients présentant un trouble neurodéveloppemental en ordonné et les patients présentant un TSA en abscisse, p.100

Figure 25. Intégration des différents facteurs de risque génétiques, p.102

Figure 26. Intégration des différents facteurs de risque de TSA au sein d'un pédigré, p.104

Figure 27. Diagramme de flux de l'étude, p.108

Figure 28. Expression normalisée des gènes associés au TSA, p.145

Figure 29. Caractéristiques de la méthylation de l'ADN, p.148

Figure 30. Différentes fonctions de la méthylation de l'ADN, p.149

Figure 31. Illustration théorique du concept d'épisignature pour aider à la classification de deux VSI, p.153

Figure 32. Comparaison de l'efficacité de différentes technologies d'analyse de méthylation pangénomique, p.156

Figure 33. Présence ou non d'épisignatures dans différentes conditions, p.158

Figure 34. Schéma simplifié de l'effet du traitement de l'ADN au bisulfite de sodium, p.160

Figure 35. Fonctionnement des deux types de sondes de la technologie Infinium d'Illumina, p.161

Figure 36. Distribution de la position des sondes des puces 450K et EPIC, p.162

Figure 37. Proposition d'un algorithme de traitement des données de méthylation, p.165

Figure 38. Distribution des valeurs β des sondes situées sur le chromosome Y sur un échantillon d'individus de sexe féminin, p.168

Figure 39. Trois méthodes de visualisation des données de méthylation pour la détection d'échantillons aberrants, p.171

Figure 40. Distance et classification hiérarchique entre 34 épisignatures rapportées par l'étude d'Aref-Eshghi et collaborateurs de 2020, p.177

Figure 41. Projection des deux premières composantes principales des valeurs β des sondes différenciellement méthylées des données issues du CNRGH et des contrôles appariés, p.184

Figure 42. Projection des deux premières composantes principales des valeurs β des sondes différenciellement méthylées des données issues du CNRGH, des contrôles appariés et des nouveaux cas pathogènes, p.186

Figure 43. Localisation des variations nucléotidiques et des CNVs de MBD5 incluent dans l'étude, p.188

Figure 44. Recherche d'une épisignature pour le gène MBD5 sur les sondes communes entre puces EPIC et 450K, p.190

Figure 45. Recherche d'une épisignature pour le gène MBD5 sur l'ensemble des jeux de données analysés en puce EPIC, p.191

Figure 46. Localisation du hotspot de sondes différenciellement méthylées sur le chromosome 2, p.191

LISTE DES TABLEAUX

Tableau 1. Comorbidités neuropsychiatriques chez les patients présentant un TSA, p.42

Tableau 2. Estimation de la pénétrance de CNVs récurrents, pour tout trouble du neurodéveloppement, p.75

TABLE DES ABBREVIATIONS

27K : *HumanMethylation27K BeadChip*

450K : *HumanMethylation450K BeadChip*

5-mc : 5-methylcytosine

ACMG : *American College of Medical Genetics and Genomics*

CGH-array : *Array Comparative Genomic Hybridization*, puce d'hybridation génomique comparative

CNRGH : Centre National de Recherche en Génomique Humaine

CNV : *Copy Number Variant*, variation du nombre de copies

DSM-5 : *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*

EDC : Episode Dépressif Caractérisé

EPIC : *Infinium MethylationEPIC Beadchip*

ExAC : *Exome Aggregation Database*

FDR : *False Discovery Rate*, taux de fausses découvertes

GEO : *Gene Expression Omnibus*

GnomAD : *Genome Aggregation Database*

GWAS : *Genome-Wide Association Study*, étude d'association pangénomique

IC95% : intervalle de confiance à 95%

IQR : *InterQuartile Range*, intervalle interquartile

Limma : *Linear Model for MicroArray*

LOEUF : *Loss-of-function Observed/Expected Upper bound Fraction*

MAF : *Minor Allele Frequency*, fréquence de l'allèle mineur

MBD : *Methyl-cytosine Binding Domain*, domaine de liaison aux méthylcytosines

NGS : *Next-Generation Sequencing*, séquençage de nouvelle génération

OMIM : *Online Mendelian Inheritance in Man*

OR : *Odd-Ratio*, rapport des cotes

Pb : Paire de Bases

PTV : *Protein Truncating Variant*

TADA : *Transmitted And De novo Association*

RR : Risque Relatif

SVA : *Surrogate Variable Analysis*

SVM : *Support Vector Machine*

TDAH : Trouble Déficit de l'Attention / Hyperactivité

TOC : Trouble Obsessionnel-Compulsif

TSA : Trouble du Spectre de l'Autisme

UTR : *UnTranslated Region*, région non traduite

VSI : Variations de Signification Inconnue

WES : *Whole-Exome Sequencing*, séquençage d'exome

AVANT-PROPOS

« *Prudens quaestio dimidium scientiae* », Sir Francis Bacon¹.

La méthode scientifique se fonde sur une dialectique permanente entre l'approche inductive de l'empirisme et la méthode hypothético-déductive de la démarche expérimentale. Cette approche tourne le dos au rationalisme éthéré basé sur l'exégèse des textes des Anciens, qui a longtemps prédominé, au profit d'une juste interprétation de la nature. Depuis lors, tout l'art du chercheur correspond à choisir ce sur quoi porter son attention pour réfuter ou non une théorie prédictive d'un phénomène naturel. Charge à l'intelligence collective de la communauté scientifique – somme de cerveaux Bayésiens essayant de composer avec le problème du critère² – de décerner le statut de connaissance aux théories les mieux corroborées à un instant donné. En substance, cette perspective affirme que la science véritable est la science des causes et des effets observables. Pour l'humanité, ce changement de paradigme intellectuel a été d'un apport inestimable.

¹ « *Half of science is putting forth the right questions* », référencé par Jon R. Stone dans *The Routledge Dictionary of Latin Quotations* (2005)

² « *All proofs inevitably lead to propositions which have no proof! All things are known because we want to believe in them.* » fait dire Frank Herbert à l'un de ses personnages du cycle de Dune. L'auteur illustre avec beaucoup d'acuité les difficultés à définir un point de départ de la connaissance et des artifices pratiques que l'esprit humain est contraint de mettre en place pour, malgré tout, se concevoir un savoir.

C'est pour cette raison qu'une *evidence based medicine* n'a de sens qu'associée à un effort continu pour définir les causes des symptômes, syndromes et maladies. C'est en désenchevêtrant ces liens de causalité que la médecine peut aboutir à une nosographie pertinente qui partirait des phénomènes biologiques pour aboutir à la sémiologie. C'est en reposant sur une telle nosographie, que le praticien peut fournir des diagnostics réellement individualisés à ses patients. C'est-à-dire en évitant les deux écueils que sont un réductionnisme diminuant l'humain à un vulgaire animal-machine ou à l'inverse un holisme détaché des contingences du réel. C'est, enfin, à ce prix que des thérapeutiques efficaces peuvent être conçues *a priori* pour soigner un état pathologique. Et ne plus seulement s'en remettre à la sérendipité – qui a été, jusqu'ici, l'origine première de la conception des différents psychotropes.

Dans cet effort de recherche des causes, les syndromes psychiatriques en général, et le Trouble du Spectre de l'Autisme (TSA) en particulier, sanctionnent un échec de la méthode anatomoclinique. Cette approche a, certes, fourni des informations utiles, mais est par trop réductionniste pour expliquer ces désordres touchant l'esprit humain. Il n'y a simplement pas de lésion cellulaire nécessaire et suffisante pour expliquer ces syndromes hétérogènes. Une pensée plus « complexe » est nécessaire pour appréhender l'étiologie de ces troubles.

De fait, il est très peu probable qu'à court ou moyen terme un modèle puisse intégrer les différentes strates de complexité du cerveau. C'est-à-dire qu'il soit capable d'expliquer comment les unités fonctionnelles du système nerveux s'organisent en grandes assemblées, dont la synchronisation des activités rythmiques définit des réseaux fonctionnels qui récapituleraient le fonctionnement cérébral dans toute sa complexité ;

comment ces réseaux fonctionnels permettent un traitement d'une information sensorielle à haute dimension pour aboutir à des modèles que le sujet peut stocker, manipuler et partager ; comment en émergent les heuristiques cognitivo-affectivo-comportementales décrites par les théories de l'apprentissage et les différents modèles cognitifs ; enfin, comment tout cela se cristallise chez chacun en une phénoménologie propre comportant ou non un état pathologique.

En attendant, qu'un tel modèle standard de l'esprit humain ne soit défini, si cela est seulement possible, il n'y a d'autre possibilité que de travailler avec les systèmes théoriques distincts conçus pour expliquer chacune des couches de complexité de l'esprit humain.

Y compris s'ils sont difficilement agencables, et parfois même contradictoires.

La biologie moléculaire offre une voie d'accès pour démêler, du moins partiellement, cet entrelac de causes et d'effets en partant du premier niveau d'organisation biologique. Notamment, l'étude de la génomique donne accès aussi bien à l'architecture commune du vivant qu'à l'idiosyncrasie de chaque individu, si ce n'est de chaque cellule. Les particularités génétiques du sujet permettent d'entrevoir, dans une certaine mesure, les causalités de ses troubles. Comme nous le verrons, il s'agit de causes partielles, impossible à réellement individualiser de nombreux facteurs modificateurs. Et c'est heureux, sans quoi l'Humain serait aliéné au déterminisme génétique. Je rappellerai plusieurs fois les limites de l'interprétation de ces facteurs de risque génétiques par la suite. Mais, à ce stade, il me paraît important d'insister sur leur réel pouvoir explicatif :

lorsqu'un patient (ou ses proches) reçoit une information concernant un facteur de risque génétique dont il est porteur, il a authentiquement accès à une part substantielle de l'origine de ses troubles.

Le développement des outils de génétique moléculaire sur le plan de la résolution (nucléotidique) et du débit (pangénomique) a permis une avancée majeure dans la description des facteurs de risque au TSA.

Nous avons entrepris un travail sur les apports et les limites de ces techniques dans le TSA. Cette initiative nous a paru importante considérant les particularités de ce trouble soit i) sa fréquence élevée dans la population, ii) ses limites floues et mouvantes dans le temps, iii) son spectre clinique de sévérité hétérogène et, iv) son association avec de nombreuses comorbidités développementales.

Dans un premier temps, notre objectif a été d'évaluer le taux de détection de facteurs de risques rares chez les patients TSA. Nous avons tout d'abord constitué une liste de gènes impliqués dans le TSA la plus exhaustive possible. A partir de cette liste nous avons essayé de formuler des critères de pathogénicité adaptés à ce trouble et d'estimer le taux de détection de variations génétiques pathogènes sur une cohorte rétrospective avec une réplication sur une cohorte prospective. Nous discutons également la corrélation entre le phénotype des patients et les variations génétiques identifiées. Enfin, nous avons évalué les caractéristiques cliniques associées à un plus fort rendement de détection

Dans un second temps, nous avons fait l'hypothèse que nous pourrions optimiser ce taux de détection en développant une stratégie de classification des nombreuses Variations de Signification Inconnue (VSI) détectées en séquençage à haut débit. Pour ce faire nous avons concentré nos efforts sur

le développement et l'utilisation de patterns de méthylation de l'ADN, appelés épisignatures, spécifiques de gènes impliqués dans les syndromes neurodéveloppementaux. Nous avons défini une méthodologie pour la conception d'une nouvelle épisignature, que nous avons appliqué pour le gène *MBD5*, impliqué dans un syndrome neurodéveloppemental non cliniquement reconnaissable avec surrisque de TSA. Par la suite, nous avons évalué la reproductibilité et la validité d'un premier sous-ensemble des épisignatures publiées dans la littérature.

Ces deux sections seront précédées d'une introduction générale qui portera sur les particularités cliniques, épidémiologiques et les comorbidités du TSA ainsi que sur les facteurs de risque environnementaux du trouble – qu'il nous a semblé nécessaire de détailler au vu de son caractère multifactoriel.

En sommes, comme tout travail scientifique, ce manuscrit traitera de la recherche des causes – ici génétiques – et de l'observation de leurs effets aux échelles clinique et épigénétique.

INTRODUCTION

Cette partie introductive rappellera tout d'abord l'évolution du concept d'autisme et de TSA et les caractéristiques épidémiologiques du trouble. Elle abordera par la suite son origine neurodéveloppementale très précoce. Dans un troisième chapitre, elle exposera une revue des facteurs de risque environnementaux *in-utero* décrits, et de leurs impacts somme toute limités à l'échelle individuelle. Enfin, elle conclura à une origine génétique majeure du TSA.

Chapitre I. Le concept d'autisme et ses évolutions

L'autisme infantile est un syndrome, ou plus probablement un ensemble de syndromes apparentés, décrit indépendamment par Leo Kanner en 1943 (Kanner, 1943) puis par Hans Asperger en 1944 (Wing, 1981). Les deux auteurs font référence au terme « autisme » conçu par Eugen Bleuler pour la schizophrénie dans son traité de 1911³. Pour Bleuler, ce terme décrivait, chez les patients présentant une schizophrénie, une perte du sens de la réalité au profit d'un enfermement dans une pensée hermétique et détachée du réel.

Les deux auteurs, qui s'accordent sur une origine génétique probable, décrivent cependant deux groupes de patients distincts. Là où Kanner décrit un syndrome rare et sévère, Asperger décrit lui un trouble relativement bénin et fréquent. Jusqu'à la redécouverte des travaux d'Asperger dans les années 1980 par Lorna Wing, c'est la première définition qui s'imposait. L'autisme de Kanner était donc intrinsèquement sévère et associé à une déficience intellectuelle. Ses deux symptômes cardinaux étaient la recherche de solitude (*aloneness*) et l'immuabilité (*sameness*). Par la suite, le diagnostic de syndrome d'Asperger s'est étendu aux patients ne présentant ni déficience intellectuelle ni troubles du langage. Puis différentes itérations de critères diagnostiques ont conduit à un élargissement progressif du syndrome pour inclure des formes de moins en moins sévères.

³ Eugen Bleuler. *Dementia præcox oder Gruppe der Schizophrenien* (1911) Leipzig & Wien.

En 2013, la dernière version du manuel diagnostique et statistique des troubles mentaux, du *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5), a fusionné les deux différentes entités cliniques. Elle parle de TSA, terme qui met l'accent sur le caractère hétérogène du syndrome. Toutefois, dans cet élargissement du concept clinique, la description prototypique de Kanner persiste. En effet, la dyade autistique moderne associe :

- des déficits de la communication et des interactions sociales,
- et le caractère restreint et répétitif des comportements, des intérêts ou des activités.

En sus de ce regroupement syndromique très large, le manuel propose de préciser un niveau de sévérité (de 1, le moins sévère à 3, le plus sévère) en fonction du handicap fonctionnel.

Un diagnostic clinique par un praticien expérimenté permet de poser le diagnostic de TSA avec une fiabilité satisfaisante en se basant sur ces critères explicites (Kadam *et al.*, 2021). De surcroît, plusieurs outils ont été validés dans différentes langues pour faciliter le diagnostic. *La Childhood Autism Rating Scale* (CARS), *l'Autism Diagnostic Interview-Revised* (ADI-R) et *l'Autism Diagnostic Observation Schedule* (ADOS) sont les plus fréquemment utilisés. Cette dernière échelle montre la meilleure sensibilité (Randall *et al.*, 2018) avec une spécificité équivalente aux autres. L'association ADOS-ADI-R est considérée comme le *gold standard* pour le diagnostic clinique de TSA (Falkmer *et al.*, 2013; Le Couteur *et al.*, 2008; Risi *et al.*, 2006).

Ces travaux sont rassurants quant à la fiabilité du diagnostic, c'est-à-dire essentiellement la fidélité inter-juge. Globalement, les cliniciens s'accordent sur ce qu'est un diagnostic de TSA.

Toutefois, en l'absence de validateur externe, ces considérations tautologiques ne disent rien de la validité de ce cadre nosographique. Il n'existe pas de lésion tissulaire, de signe paraclinique ou de marqueur biologique qui ferait fonction de validateur externe et encore moins qui permettrait de dépasser le syndrome pour aboutir à une maladie unitaire. La pertinence de ce regroupement syndromique lâche qu'est le TSA, agrégeant par essence des patients très hétérogènes, est discutable. Notamment, l'hypothèse implicite d'une voie physiopathologique commune entre l'autisme de Kanner et l'autisme d'Asperger reste à vérifier. Pour aller plus loin, il pourrait y avoir lieu de s'inquiéter que cet engouement anglo-saxon pour des syndromes toujours plus larges aux critères diagnostiques toujours plus larges n'aboutisse finalement à des hypothèses qui ne seraient plus réfutables. Il est vrai que la définition des critères diagnostiques n'a pas que des implications scientifiques. Le financement ou non des soins pour les patients en dépend dans beaucoup de systèmes de santé. *L'American Psychiatric Association* assume cet état de fait⁴.

En tout état de cause, et malgré les réserves que j'ai citées, nous considérerons principalement le concept de TSA dans la suite de ce travail. Par soucis de cohérence avec la littérature scientifique contemporaine.

4

https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/APA_DSM_Insurance-Implications-of-DSM-5.pdf

Dans les différentes études épidémiologiques qui ont été menées, le TSA présente une prévalence très variable. En vérité, on note une augmentation continue de la fréquence du diagnostic depuis les premières études épidémiologiques des années 1960 (Williams *et al.*, 2006; Zeidan *et al.*, 2022) (Figure 1).

Les études les plus récentes constatent une prévalence de 1 à 3% en population générale (Shenouda *et al.*, 2022; Sun *et al.*, 2019). L'élargissement des critères diagnostiques et l'amélioration des capacités de dépistage sont majoritairement en cause dans cette augmentation (King & Bearman, 2009).

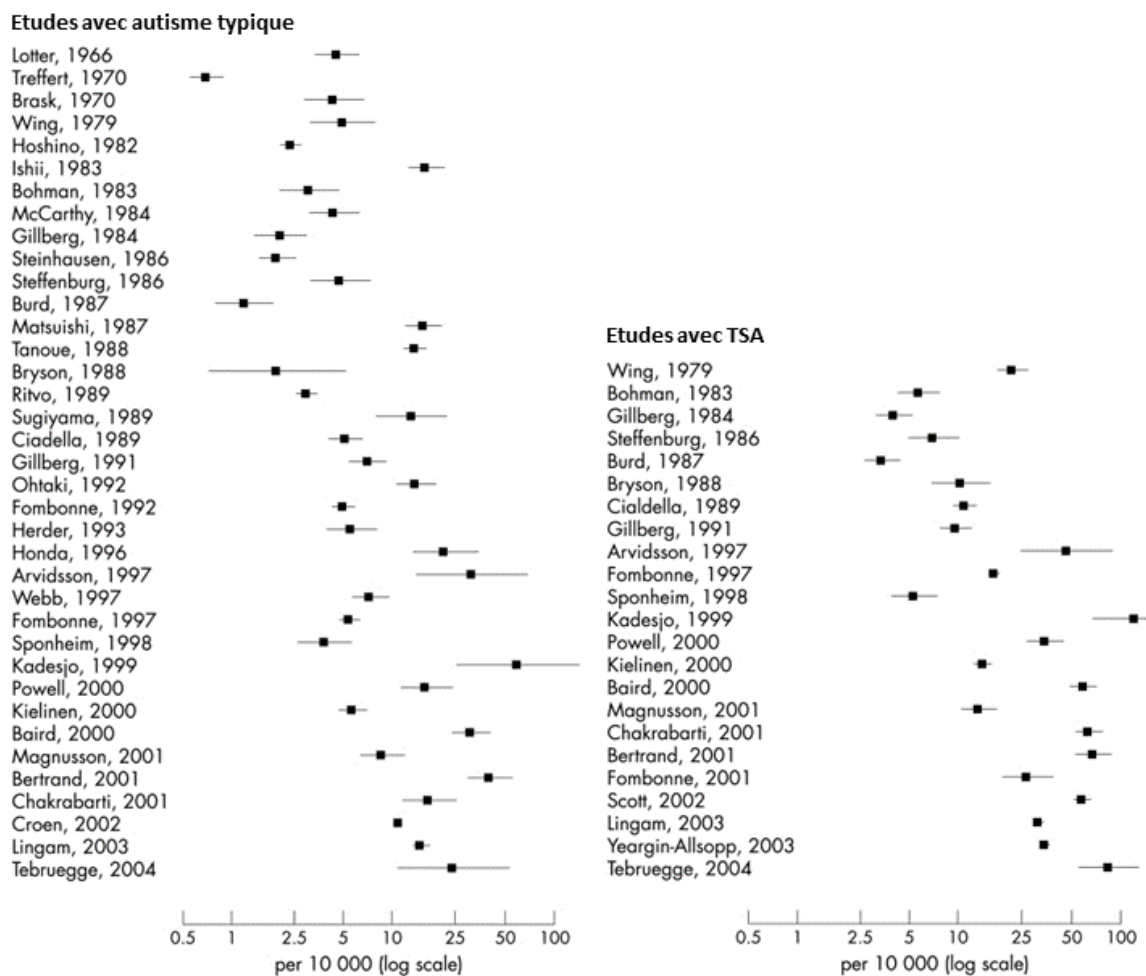


Figure 1. Graphique en forêt des prévalences estimées et de leurs intervalles de confiance à 95% (IC95%) pour l'autisme typique (à gauche) et pour le TSA (à droite). L'autisme typique est une catégorie nosographique plus resserrée – plus proche des observations de Léo Kanner – aujourd'hui abandonnée. La conception moderne de TSA comprend les anciennes catégories d'autismes typiques, atypiques, syndrome d'Asperger et trouble envahissant du développement. Quelle que soit la perspective adoptée, on remarque un accroissement progressif de la prévalence du TSA. Adapté de l'article de Williams et al., 2006.

La prédominance masculine du trouble, avec un sex-ratio autour de 3 pour 1 (Loomes et al., 2017), est une particularité épidémiologique non totalement élucidée. Pour l'expliquer, les hypothèses les plus discutées – non incompatibles entre elles – sont les suivantes :

i) Une présentation clinique particulière chez les femmes présentant un TSA. Ces jeunes patientes présenteraient des symptômes de la dyade autistique plus discrets et qui se manifesteraient plus volontiers par des troubles internalisés (sous diverses formes d'anxiété) que par des troubles du comportement (Young *et al.*, 2018). De plus, leurs intérêts restreints se dirigeraient vers des sujets plus acceptés socialement (Young *et al.*, 2018). De ce fait, leurs parents et professeurs seraient moins alertés ce qui diminuerait la probabilité d'accès au diagnostic, dont la sensibilité serait, par ailleurs, affaiblie (Young *et al.*, 2018).

ii) L'hypothèse selon laquelle l'exposition du cerveau à la testostérone lors de la période fœtale induirait une vulnérabilité au TSA. Cette hypothèse est dans la filiation de l'hypothèse du fonctionnement cognitif « hyper-mâle » de l'autisme décrit par Asperger (Baron-Cohen, 2002). Cette hypothèse a ensuite été rattachée aux différences de latéralisation cérébrale en fonction du sexe (Baron-Cohen *et al.*, 2005). Les résultats des travaux qui testent cette hypothèse via différents proxies ne sont pas consensuels et de nombreux biais diminuent leurs portées (Ferri *et al.*, 2018). La meilleure preuve est issue d'une étude danoise croisant les registres diagnostiques nationaux et une banque de liquide amniotique issu d'amniocentèses. Les auteurs ont observé un taux de stéroïdes androgéniques plus élevé dans le liquide amniotique des sujets qui seront par la suite diagnostiqués TSA par rapport à des contrôles appariés (Baron-Cohen *et al.*, 2015). Il faut noter que nous n'avons pas trouvé de réplification de ces résultats.

iii) Il est observé un excès de facteurs de risque génétiques par patiente diagnostiquée TSA par rapport aux hommes. Cette observation oriente vers un effet protecteur intrinsèque du sexe féminin vis-à-vis du TSA (Huguet *et*

al., 2013; Jacquemont *et al.*, 2014; Sanders *et al.*, 2015; Warrier *et al.*, 2022; Wigdor *et al.*, 2022).

Nous verrons dans le chapitre suivant que l'évolution clinique, la présence d'anomalies cérébrales et les comorbidités neuropsychiatriques du TSA pointent vers une origine essentiellement **neurodéveloppementale**.

Chapitre II. Un syndrome neurodéveloppemental d'apparition précoce

Un syndrome d'apparition très précoce

Quoi qu'il en soit de la validité de la nosographie actuelle, le TSA est un trouble d'apparition très précoce dans le développement. En effet, il peut être diagnostiqué de manière fiable dès 24 à 36 mois sur l'observation d'un retard, d'une stagnation voire d'une régression dans les acquisitions précoces d'aptitudes au langage ou à la socialisation et d'atypicités comportementales (Barbaro & Dissanayake, 2017; Chawarska *et al.*, 2014; Kadam *et al.*, 2021; Ozonoff *et al.*, 2015; Pham *et al.*, 2022). Avant cela, des signes d'appels sont souvent présents entre 12 et 24 mois (Miller *et al.*, 2021; Pierce *et al.*, 2019; Wieckowski *et al.*, 2021). Lorsqu'il est réalisé précocement, le diagnostic n'est remis en question que chez une faible proportion de patients. Il y a schématiquement deux situations qui expliquent qu'un patient diagnostiqué TSA dans la petite enfance ne réponde plus aux critères diagnostiques par la suite :

i) Le patient présente une forme peu sévère et passe d'un seuil léger à un seuil sub-syndromique sur une échelle diagnostique du fait des limites psychométriques de celle-ci et de la trajectoire développementale particulière du patient (Barbaro & Dissanayake, 2017; Bieleninik *et al.*, 2017; L. M. Turner & Stone, 2007).

ii) Le diagnostic de TSA est récusé en faveur d'un autre trouble neurodéveloppemental dont l'expression clinique peut ressembler au TSA dans la petite enfance. Il s'agit le plus souvent de retard de développement

évoluant vers une déficience intellectuelle sans TSA (Benedetto *et al.*, 2021; Kadam *et al.*, 2021; L. M. Turner & Stone, 2007).

Tout bien considéré, ces deux situations sont finalement marginales. Nous considérons que l'existence de formes limites d'autisme et d'un chevauchement phénotypique des syndromes neurodéveloppementaux ne remet pas en question le postulat d'une apparition du TSA dans les trois premières années de vie chez la très grande majorité des individus.

Avant la première année de vie, les données publiées concernent le devenir d'enfants issus d'une fratrie déjà touchée par le TSA. Ces enfants présentent un risque de récurrence importante d'autisme (*cf. infra*). La comparaison du développement des nourrissons qui développeront plus tard un TSA avec ceux qui auront un développement ordinaire permet la mise à jour de signes précoces, s'ils existent (Szatmari *et al.*, 2016). En l'état, ces travaux suggèrent l'émergence de symptômes de la dyade autistique entre 6 et 36 mois, avec une certaine hétérogénéité des trajectoires de développement (Szatmari *et al.*, 2016). Auparavant, il n'y a pas de différence intergroupe significative.

Le TSA est donc un syndrome dont les premiers signes sont très précoces et dont l'intensité des symptômes cardinaux évolue peu une fois déclaré (Bieleninik *et al.*, 2017; Simonoff *et al.*, 2020). Nous pouvons conclure qu'il témoigne d'un fonctionnement cérébral altéré (ou atypique pour les formes légères) de manière stable dès les phases neurodéveloppementales précoces du nourrisson voire *in-utero*.

Des anomalies cérébrales précoces et à grande échelle

Les patients diagnostiqués TSA sont connus pour présenter, en moyenne, un périmètre crânien et un volume cérébral augmentés les

premières années de vie (Redcay & Courchesne, 2005; Sacco et al., 2015). Cet élargissement cérébral prédomine sur la substance blanche mais n'épargne pas la substance grise, qu'elle soit corticale ou dans les ganglions de la base (Donovan & Basson, 2017), ni les espaces sous-arachnoïdiens (M. D. Shen et al., 2013). Certains travaux montrent même une légère augmentation du périmètre crânien dès la période in-utero (Figure 2). Cette particularité ne serait plus mesurable lors de l'adolescence (Courchesne et al., 2001; Courchesne, Campbell, et al., 2011) mais cette éventuelle « normalisation » est en trompe-l'œil. Tout d'abord, elle est remise en cause dans certaines études (Hazlett et al., 2006; Yankowitz et al., 2020). Ensuite, elle ne dit rien de la persistance d'anomalies macroscopiques régionales (Hazlett et al., 2006; van Rooij et al., 2018; H. Wang et al., 2022) ou cytoarchitecturales microscopiques (Donovan & Basson, 2017). Enfin, il ne s'agit pas d'une régression mais d'un arrêt précoce de la croissance cérébrale chez les patients TSA par rapport aux contrôles (Redcay & Courchesne, 2005). Tout bien considéré, il faudrait parler d'une avance de croissance cérébrale (Mraz et al., 2007) plutôt que d'un excès.

Une des causes de cet élargissement cérébral semble être une hyperprolifération des progéniteurs neuronaux lors des deux premiers semestres de gestation, aboutissant à un excès de neurones inégalement répartis au sein des structures corticales (Courchesne, Mouton, et al., 2011; Santos et al., 2011) et sous-corticales (Schumann et al., 2009) des patients. Certains auteurs avancent que cette prolifération cellulaire excessive pourrait être la condition suffisante pour le déclenchement du TSA (Courchesne et al., 2020). Il faut tempérer ce propos en notant que l'élargissement de la substance grise ne peut pas être utilisé comme un signe paraclinique direct de TSA (Squarcina et al., 2021) et qu'il ne s'agit pas de la seule explication proposée à cet élargissement cérébral. Des

anomalies microgliales (Koyama & Ikegaya, 2015) et du système du complément (Mansur et al., 2021) ont également observé dans des astrocytes issus de cellules souches pluripotentes induites de patients diagnostiqués TSA. Ces auteurs proposent donc un défaut d'élagage des synapses glutamatergiques un peu plus tardive dans le développement. Cette hypothèse d'une dysfonction du système immunitaire inné cérébral aurait l'intérêt de faire le lien avec les facteurs de risque environnementaux pro-inflammatoires que nous détaillerons par la suite.

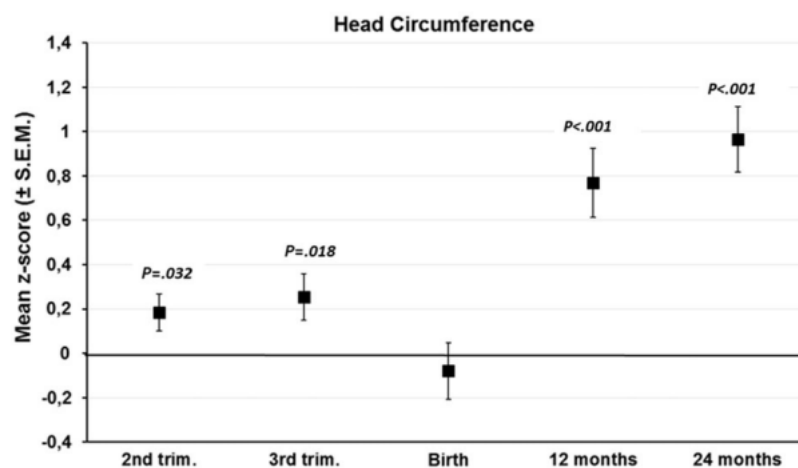


Figure 2. Comparaison normalisée des périmètres crâniens entre patients présentant un TSA et témoins chez le fœtus et l'enfant. Données obtenues à partir d'un échantillon de 61 patients avec TSA lors du 2^{ème}, 3^{ème} semestre de gestation, à la naissance puis lors des deux premières années de vie comparés aux bases de données de témoins. D'après Bonnet-Brilhault et al., 2018. On remarque que ces travaux corroborent ceux de Courchesne : le périmètre crânien est, ici aussi, normal à la naissance.

Ces éléments de survenue très précoce aboutissent à de nombreuses anomalies de neurodifférenciation qui conduisent à des atypicités dans l'activité neuronale, la proportion des différents types cellulaires, l'organisation des neurites et l'architecture fine de différentes structures cérébrales dont le détail et la discussion dépassent le cadre de ce travail (Courchesne et al., 2020; Donovan & Basson, 2017; van Rooij et al., 2018). Il suffit de dire que l'ensemble des étapes neurodéveloppementales (Figure 3) sont impactées dans le TSA. Ce modèle permet d'expliquer l'hétérogénéité clinique des patients, en fonction du degré et de la nature spécifique des atteintes à chaque étape. De plus, en décrivant des anomalies aux différentes échelles de l'organisation cérébrale, la littérature peut rendre compte de symptômes ayant trait à des niveaux d'intégration cérébrale différents. Des anomalies d'intégration sensorielle aux défauts de cognition sociale.

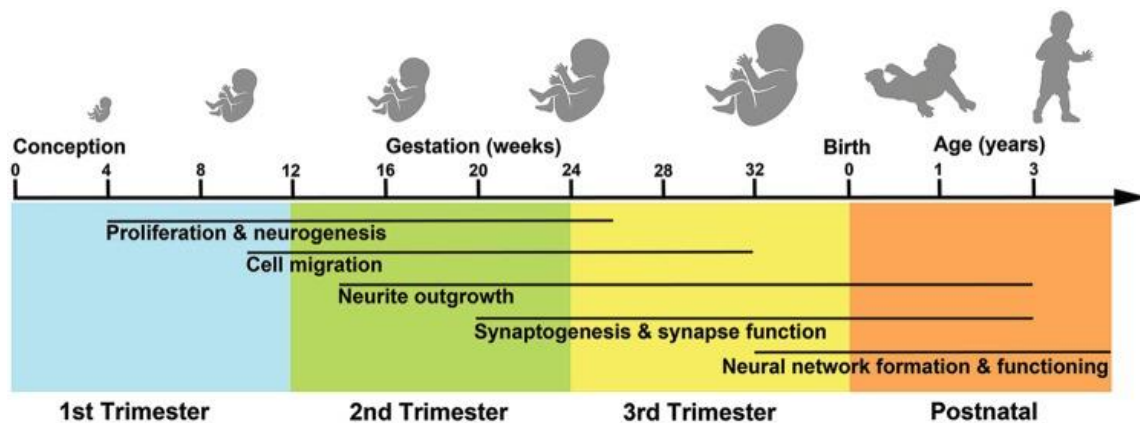


Figure 3. Représentation chronologique des différentes étapes de développement cérébral pour lesquelles des anomalies sont suspectées chez les patients diagnostiqués avec un TSA. Chez ces patients, les données de la littérature, pointent vers des altérations dans l'ensemble des étapes chronologiques du développement cérébral chez l'embryon et le fœtus. Notamment, les anomalies de prolifération neuronale évoquent une perturbation dès le premier trimestre de l'embryogénèse. D'après Courchesne et al., 2019.

Ce fonctionnement cérébral aberrant aboutit, une fois les grandes étapes du neurodéveloppement terminées, à une anomalie globale de la connectivité cérébrale à longue distance (Anderson et al., 2011; D. S. Andrews et al., 2021; Di et al., 2018; Q. Li et al., 2019; Rane et al., 2015). Les nombreux facteurs confondants et les différentes techniques d'imageries utilisées viennent obscurcir les détails mais le consensus est en faveur d'une hypoconnectivité cortico-corticale globale associé (ou compensé) par une hyperconnectivité cortico-sous-corticale (Picci et al., 2016). Le modèle initial proposait, en sus, un accroissement de la connectivité local (Belmonte et al., 2004; Just et al., 2004). Pourtant, les données d'imageries fonctionnelles (Picci et al., 2016; Rane et al., 2015), de magnétoencephalographie (Picci et al., 2016) et même de neuropathologie (McKavanagh et al., 2015) ne permettent pas de confirmer cette hypothèse d'un biais global vers un traitement local de l'information dans le cerveau des individus présentant un TSA. A vrai dire, à ce niveau local, il est plutôt observé des anomalies focales de l'organisation cytoarchitecturale chez de nombreux patients du fait d'erreurs dans la migration cellulaire (Stoner *et al.*, 2014; Wegiel *et al.*, 2010).

Ces anomalies de connectivité cortico-corticales à longue distance ont une implication clinique significative dans le TSA, car elles touchent particulièrement les structures de la cognition sociale (Sato & Uono, 2019).

Ces observations nous permettent de passer à une autre échelle de traitement de l'information : celle des réseaux fonctionnels. Ces réseaux sont définis en fonction de la synchronicité d'activation d'un ensemble de structures cérébrales dans différentes tâches cognitives ou au repos (Menon, 2011) et ils seraient particulièrement impliqués en psychopathologie humaine (Menon, 2011) (Figure 4). La coordination entre

l'activité rythmique de ces grandes populations de neurones dépend de leur interconnectivité que nous avons vu perturbés.

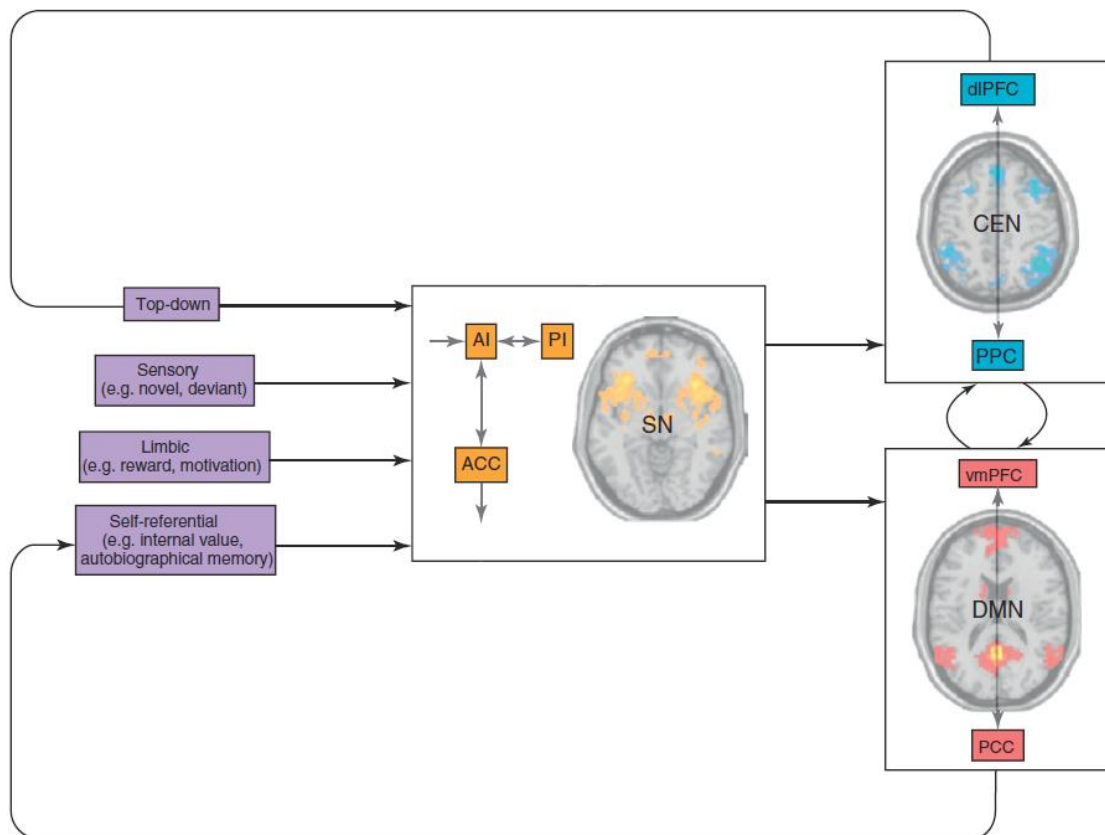


Figure 4. Représentation des trois principaux réseaux fonctionnels. Ces réseaux fonctionnels sont particulièrement impliqués dans la cognition sociale, dont la déficience est un symptôme central de TSA. CEN : Central Executive Network ; DMN : Default Mode Network ; SN : Salience Network ; AI : Anterior Insula ; PI : Posterior Insular ; ACC : Anterior Cingulate Cortex ; dIPFC : dorsolateral PreFrontal Cortex ; PPC : Posterior Parietal Cortex, vmPFC : ventromedian PreFrontal Cortex ; Posterior Cingulate Cortex. Adapté de Menon et al (2011).

Les paradigmes d'imagerie fonctionnelle permettant l'étude de ces réseaux neuronaux dans le TSA sont très complexes et de nombreux biais existent. Il est difficile de rentrer dans le détail des structures cérébrales impliquées, de même que sur le lien avec les signes cliniques sans expertise

dans le domaine. Je me contenterais d'observer qu'un grand nombre d'études montrent des anomalies de ces réseaux fonctionnels et de leurs interconnexions chez les patients présentant un TSA, au point que ces altérations sont considérées comme centrales dans la clinique du TSA (Abbott et al., 2016; Chen et al., 2021; Lynch et al., 2013; Moradimanesh et al., 2021; Mougá et al., 2022; Nomi et al., 2019; Nomi & Uddin, 2015; Padmanabhan et al., 2017; Uddin, 2015; Yerys et al., 2015).

Des troubles de la latéralisation cérébrale

Un exemple de connectivité à longue distance concerne la transmission d'information entre les hémisphères cérébraux *via* le corps calleux et les commissures. Ainsi, le cerveau tire parti d'une spécialisation fonctionnelle de chaque hémisphère dans le traitement de l'information qui pourrait dériver d'une différence dans la structure des minicolonnes corticales, les unités élémentaires de traitement de l'information (Chance, 2014). Ceci se reflète dans la latéralisation de nombreuses fonctions cognitives, c'est-à-dire l'activation préférentielle d'un hémisphère cérébral pour une tâche donnée, aussi appelé dominance hémisphérique (Karolis et al., 2019).

Comme attendu au vu des conclusions du chapitre précédent, la connectivité interhémisphérique est également perturbée dans le TSA (Q. Li et al., 2019; Yao et al., 2021). Chez ces patients, il semble exister un défaut de partage de l'information, particulièrement entre régions homologues.

Sans qu'il soit possible de trancher entre cause, conséquence ou contingence, cela fait le lien avec le deuxième stigmatisme cérébral macroscopique de TSA. Outre la macrocéphalie relative précoce que nous avons citée, on observe également une atténuation globale de l'asymétrie

cérébrale (Floris *et al.*, 2021; Postema *et al.*, 2019) dont l'ampleur est corrélée aux symptômes cardinaux de TSA. Cette observation semble d'autant moins triviale, qu'à l'échelle clinique les patients diagnostiqués TSA sont connus pour présenter une altération de plusieurs fonctions latéralisées. C'est le cas du langage (Kleinhans *et al.*, 2008; Lindell & Hudry, 2013), de la latéralité manuelle⁵ (Lindell & Hudry, 2013; Markou *et al.*, 2017) et de l'exploration visuo-spatiale (English *et al.*, 2017; Liu *et al.*, 2021).

L'image globale des altérations neuropathologiques et neurophysiologiques dans le TSA n'est pas encore très nette. De plus, il n'est pas encore possible d'inférer un lien de causalité entre ces différentes anomalies. Néanmoins, il semble se dégager les éléments suivants : des anomalies du développement cérébral *in-utero* induisent une croissance cérébrale trop précoce. Celle-ci aboutit à des anomalies cytoarchitecturales diffuses – notamment la spécialisation fonctionnelle différentielle des structures homologues entre les hémisphères – et une altération des voies de connectivité à longue distance. Ces altérations entraînent des anomalies de fonctionnement des réseaux fonctionnels impliqués dans le traitement de l'information à haut niveau.

⁵ 15% des patients présentant un TSA sont gaucher et 44% ambidextres, comparé à respectivement 5 et 16% pour les sujets normotypiques.

Les relations entre ces différents éléments et le lien avec l'émergence des symptômes de la dyade autistique reste à préciser. Il n'est pas certain non plus qu'elles soient présentes chez tous les cas de TSA.

De fréquentes comorbidités neuropsychiatriques

Comme nous l'avons vu, le TSA est marqué par une atteinte plus ou moins sévère des différents stades de développement neuronal *in-utero* aboutissant à un fonctionnement cérébral atypique. Celui-ci s'exprime très précocement par la dyade autistique, *i.e.* trouble de la cognition sociale et caractère restreint et stéréotypé des activités. Au-delà de ce regroupement syndromique spécifique, les patients atteints de TSA présentent de nombreuses comorbidités neuropsychiatriques. La première concerne les fonctions intellectuelles des patients. Il est bien établi qu'il existe trois populations distinctes de patients : intellectuellement déficient sans TSA, TSA sans déficience et TSA avec déficience (Matson & Shoemaker, 2009).

La cooccurrence fréquente du TSA avec la déficience intellectuelle suppose le lien entre certaines formes de syndrome autistique et une dysfonction cérébrale généralisée. L'ampleur de cette cooccurrence est sujette à débat en raison de la difficulté clinique à séparer les deux entités chez bon nombre d'individus. Il existe effectivement quasi-systématiquement une certaine altération des capacités d'interaction sociale et une rigidité comportementale chez les patients présentant une altération cognitive, en fonction de l'ampleur de cette dernière (Matson & Shoemaker, 2009). Chez ces patients, il est recommandé de poser ou non le diagnostic de TSA en fonction de l'ampleur des difficultés d'interaction sociale vis-à-vis du niveau de développement général, sans qu'il n'existe de

ligne de conduite autre que l'évaluation subjective du clinicien⁶. Il est notable que, aux Etats-Unis, l'augmentation de la prévalence de TSA de ces dernières décennies est corrélée à une diminution du nombre de diagnostic de déficience intellectuelle (Shattuck, 2006). Cela permet de supposer qu'une partie de l'augmentation de la prévalence du TSA est liée à un basculement diagnostique de toute une population de patients de la déficience intellectuelle vers le TSA.

Ainsi, une zone grise apparaît à l'intersection de ces trois catégories qui rend leur quantification relative difficile. En fonction des critères diagnostiques retenus, de l'organisation de l'offre de soin, et du faisceau d'indices utilisé pour différencier le TSA de la déficience intellectuelle le niveau de cooccurrence change de manière drastique. Au point d'aboutir à la constatation absurde qu'entre 0 et 100% des patients présentant un TSA souffrent également d'une déficience intellectuelle, selon différentes équipes de recherche (Bougeard *et al.*, 2021; Fombonne, 2003). Une cooccurrence entre 20 et 50% est néanmoins retrouvée dans une grande majorité des études (Bougeard *et al.*, 2021; Fombonne, 2003).

D'une manière plus générale, un grand nombre de comorbidités psychiatriques sont surreprésentées chez ces patients témoignant de dysfonctions cérébrales variées à expression plus ou moins précoces (tableau 1). Il est possible d'ajouter à celle-ci une augmentation du risque

⁶ American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*

d'infirmités motrices cérébrales avec un *odd-ratio* (OR) à 8.75 IC_{95%} : [4,6–16,7] (Pan *et al.*, 2021).

Ces comorbidités neuropsychiatriques sont si fréquentes qu'un diagnostic de TSA pur fait exception. La présence de troubles cognitifs, particulièrement attentionnels, ou de manifestations anxieuses est la norme. Au-delà de ses singularités cliniques, le TSA se place donc dans le champ d'une altération globale du fonctionnement cérébral.

Comorbidité	Prévalence chez les patients atteints de TSA [IC_{95%}]*	Prévalence en population générale
EDC ¹	11% [9%-13%]	4,7%
Epilepsie ²	12.1% [IQR : 15.4%]	0.6%
Schizophrénie ¹	4% [3%-5%]	0.75%
TDAH ¹	28% [25%-32%]	7,2%
TOC ¹	9% [7%-10%]	0.7%
Troubles anxieux ¹	20% [17%-23%]	7,3%
Trouble bipolaire ¹	5% [3%-6%]	1,06% type 1 1,57% type 2
Trouble des conduites ¹	12% [10%-15%]	8,9%
Trouble du rythme circardien ¹	13% [9%-17%]	3,7%

Tableau 1. Comorbidités neuropsychiatriques chez les patients présentant un TSA. *sauf précision contraire. ¹Lai et al., 2019 ; ²Lukmanji et al., 2019 ; TDAH : Trouble Déficit de l'Attention / Hyperactivités ; EDC : Episode Dépressif Caractérisé ; TOC : Trouble Obsessionnel-Compulsif ; IQR : Intervalle Interquartile. Les patients diagnostiqués avec TSA présentent systématiquement une prévalence plus élevée des pathologies psychiatriques et de l'épilepsie que la population générale.

Chapitre III. Une causalité multifactorielle

L'évolution de la nosographie vers la notion de **spectre** autistique rend compte de la diversité des phénotypes et de la sévérité du syndrome. Cette catégorisation prend comme prémisse l'hypothèse d'une voie physiopathologique (au moins partiellement) commune entre les différentes formes du trouble. Néanmoins, quels que soient les mécanismes conduisant à ce regroupement de symptômes, rien n'indique qu'une cause commune y aboutisse chez l'ensemble des patients. Effectivement, aucun facteur nécessaire et suffisant n'est actuellement identifié pour qu'un patient déclare un TSA. Nous ne décrivons que des facteurs de risque plus ou moins pénétrants (Antaki et al., 2022; Bölte et al., 2019; Grove et al., 2019; Myers et al., 2020; Warrier et al., 2022). De plus, comme ceux-ci ne sont pas indépendants, il est souvent difficile de démêler la causalité de l'association statistique. Nous décrivons d'abord les facteurs environnementaux puis les éléments en faveur d'une origine partiellement génétique du TSA.

Revue des facteurs de risque environnementaux identifiés pour le TSA

Les facteurs de risque environnementaux de TSA peuvent se regrouper en plusieurs catégories. Tout d'abord et majoritairement, des facteurs toxiques *in-utero* ont été décrits. Différentes complications obstétricales ont également été associées à un risque plus élevé de TSA chez l'enfant. Enfin, nous évoquerons la possible existence de facteurs de risque postnataux précoces.

Concernant les facteurs toxiques, la société industrielle expose la femme enceinte et donc son fœtus à un grand nombre de xénobiotiques dont certains sont neurotoxiques.

Les thérapeutiques médicamenteuses sont une cause d'exposition volontaire fréquente, avec un bénéfice attendu. La plupart de ceux-ci n'ont pas été spécifiquement étudiés comme facteur de risque de TSA (Bölte *et al.*, 2019). Plus généralement, les études ont porté sur leur aspect tératogène et/ou sur le devenir neurodéveloppemental global des enfants. La seule molécule pour laquelle un surrisque de TSA est démontrée est le valproate de sodium. Cet antiépileptique induit une augmentation très importante des complications malformatives, des déficiences intellectuelles et, également, de TSA (OR 17,3, IC_{95%} [2,4–217,6]) (Veroniki *et al.*, 2017). Ce risque n'est pas retrouvé pour les autres antiépileptiques concernant le TSA.

La prise d'antidépresseurs sérotoninergiques pendant la grossesse est également associée à une augmentation modeste de survenue de TSA chez l'enfant. Toutefois, l'association statistique est de même ampleur si l'on considère les antécédents de prise d'antidépresseur **avant** et non **pendant** la grossesse chez la mère, la causalité est donc très peu probable (Leshem *et al.*, 2021).

L'exposition à divers polluants, surtout les particules fines, et pesticides pendant la grossesse serait également un facteur de risque très modéré de TSA (Chun *et al.*, 2020; von Ehrenstein *et al.*, 2019).

En dehors de ces molécules, la majeure partie des facteurs de risque démontrés concernent l'environnement fœtal. Ils peuvent être décomposés en trois groupes : i) l'exposition fœtale aux androgènes, ii) l'activation du système immunitaire de la mère et iii) le risque métabolique de celle-ci.

Comme nous l'avons évoqué, le sex-ratio du TSA montre une prédominance masculine qui serait partiellement expliquée par un effet de

la testostérone sur le cerveau en développement. La présence d'un syndrome des ovaires polykystiques, qui induit une production excessive d'androgènes, chez la mère (OR : 1,6, IC_{95%} : [1,3–1,9]) a été montrée comme associée à un accroissement des cas de TSA chez l'enfant sur une étude de registres nationaux suédois (Cesta *et al.*, 2020). Sur ces mêmes registres, l'hirsutisme – marqueur d'imprégnation à la testostérone – est également associée au TSA (OR : 1,6, IC_{95%} : [0,9-2,8]) (Lee *et al.*, 2017).

Un autre sujet de conjonctures est le lien entre une dérégulation du système immunitaire et la survenue de pathologies psychiatriques. Dans le cas du TSA, il a été montré que la survenue d'une infection, lors de la grossesse, par différents virus pathogènes (comme la rubéole (Hutton, 2016) ou l'influenza (Atladóttir *et al.*, 2012)) serait un facteur de risque. Plus généralement, un faisceau d'arguments est en faveur d'un effet spécifique de l'activation du système immunitaire sur le neurodéveloppement. En effet, dans les travaux d'Atladóttir et collaborateurs, tout épisode de fièvre prolongé lors de la grossesse était associé à un surrisque de TSA (aHR⁷ : 3,2, IC_{95%} : [1,8–5,6]) alors que d'autres travaux montraient que c'était le cas pour toute infection maternelle ayant nécessité une hospitalisation (OR = 1,4, IC_{95%} : [1,3-1,5]) (Lee *et al.*, 2015). Enfin, plusieurs modèles animaux d'activation immunitaire lors de la grossesse montrent des phénotypes isomorphes avec le TSA dans la descendance (Careaga *et al.*, 2017), en considérant une certaine prudence quant aux possibilités d'inférences à partir de modèles animaux de pathologies psychiatriques (Bey & Jiang, 2014).

⁷ aHR : *adjusted Hazard Ratio*

D'autres facteurs toxiques *in-utero* pourraient amener un plus grand risque de TSA chez l'enfant. L'obésité (OR : 1,5, IC_{95%} : [1,3-1,7]) (Sanchez *et al.*, 2018), le diabète (Risque Relatif (RR) : 1,6, IC_{95%} : [1,3-1,9]) (Wan *et al.*, 2018) et l'hypertension artérielle (OR : 1,3, IC_{95%} : [1,1-1,6]) (Maher *et al.*, 2018) maternels seraient tous trois associés à une faible augmentation du risque de TSA. Les méta-analyses citées incluent des effectifs importants avec une hétérogénéité modérée et un risque de biais faible. Le niveau de preuve semble donc satisfaisant. Comme pour l'activation immunitaire, la taille d'effet est faible mais il s'agit de conditions fréquentes lors des grossesses qui peuvent donc avoir un impact significatif sur le risque de TSA à l'échelle de la population. Ces trois conditions sont interdépendantes et s'intègrent dans le syndrome métabolique – connu pour son association avec la maladie athéromateuse – avec l'hypertriglycéridémie et l'hypercholestérolémie. A notre connaissance, seule une étude évoque une association entre perturbation du bilan lipidique maternel et surrisque de TSA (Park *et al.*, 2021).

Les liens établis entre le syndrome métabolique et la physiopathologie du TSA sont pour l'instant ténus :

- L'hyperglycémie intra-utérine pourrait avoir un effet nocif sur le développement du cerveau fœtal, particulièrement avant 26 semaines de gestation (Wan *et al.*, 2018).
- L'hypertension artérielle maternelle entraînerait une dysfonction placentaire et donc *in-fine* une hypoperfusion du cerveau fœtal (Maher *et al.*, 2018).
- Une étude sur une fratrie discordante concernant l'obésité maternelle lors de la grossesse ne montre pas d'effet de ce facteur sur le

risque de TSA (Gardner *et al.*, 2015). Les auteurs concluaient qu'il était peu probable que l'obésité soit directement causale du surrisque de TSA.

Il est notable que le sexe masculin soit, tout comme le syndrome métabolique, un facteur de risque de maladie athéromateuse du fait de la répartition androïde des graisses. Ce syndrome marqué par différents événements cardio-vasculaires comprend une inflammation systémique chronique de bas grade (Libby, 2021). Il est donc possible de formuler l'hypothèse que ces différents facteurs endogènes *in-utero* convergent vers la neuro-inflammation avec un impact sur le développement cérébral fœtal.

Une dernière période de la vie fœtale a été étudiée comme pourvoyeuse de facteurs de risque de TSA. Il s'agit de l'accouchement. En effet, de nombreux facteurs périnataux ont été associés au TSA.

La prématurité, les anomalies de la présentation, la réalisation d'une césarienne ou de faibles paramètres de naissance (Gardener *et al.*, 2011; C. Wang *et al.*, 2017) font parties des facteurs les plus mieux répliqués. Ces différents éléments interdépendants sont associés à une souffrance cérébrale hypoxiques chez le fœtus (Gardener *et al.*, 2011). Parmi les nombreuses voies physiologiques impactées par cette hypoxie cérébrale néo-natale, on note encore une fois une dysrégulation de la réponse inflammatoire (Gardener *et al.*, 2011). Ces événements qui se compliquent d'un impact délétère à long terme pour le développement cérébral (B. Wang *et al.*, 2021) ne semblent pas spécifiques au TSA mais communs aux pathologies neuropsychiatriques et neurodéveloppementales au sens large (Bölte *et al.*, 2019). Il pourrait donc s'agir d'un facteur de souffrance cérébrale aspécifique et non d'un facteur de risque *per se* de TSA.

Ainsi, l'essentiel des facteurs de risque environnementaux de TSA ont été établis dans un contexte d'exposition *in-utero*. L'hypothèse d'un effet de la vaccination du nourrisson et les théories psychogéniques de l'autisme ont été démontrées comme fausses au-delà de tout doute raisonnable (Taylor et al., 2014). Malheureusement, leur impact négatif sur l'opinion publique a limité les études sur les facteurs de risque postnataux (Bölte et al., 2019). Des travaux récents ont associé divers facteurs de faible niveau socio-économique et de mauvaise santé maternelle au TSA sur des cohortes prospectives (Jo et al., 2019; Pham et al., 2022). Ces facteurs démographiques sont très fortement associés aux facteurs de risque *in-utero* déjà cités. Les études transversales d'où ils sont issus ne permettent pas de les considérer comme des facteurs étiologiques de TSA. En revanche, ces études mettent en lumière les populations pour lesquelles l'exposition aux facteurs de risque *in-utero* de TSA est démesurément présente.

Deux autres facteurs de risque sont liés à la temporalité de la grossesse et sont plus difficiles à classer. L'âge des parents au moment de la conception augmente le risque de TSA d'une manière âge-dépendante. En pratique, chaque augmentation de 10 ans de l'âge maternel est associée à une augmentation de 18% du risque de TSA (OR = 1,2, IC_{95%} : [1,1-1,3]) alors que celui-ci est de 21% pour 10 ans d'âge paternel (OR = 1,2, IC_{95%} : [1,18-1,24]) (Wu et al., 2017). Nous verrons par la suite que l'âge des parents au moment de la conception peut induire la présence d'un certain type de facteurs de risque génétique de TSA chez les enfants. Le second concerne l'intervalle entre les grossesses. Une deuxième grossesse à intervalle court (défini par un intervalle inférieur à 12 ou 24 mois selon les études) après la précédente fait porter un risque plus élevé de TSA chez l'enfant issu de celle-ci, par rapport à un intervalle long (> 60-72 mois) (Cheslack Postava & Winter, 2015; Kozuki & Walker, 2013). Certains

auteurs expliquent ce phénomène par l'effet de la déplétion en nutriments essentiels lors de la précédente grossesse (Bölte *et al.*, 2019).

Facteurs génétiques et TSA

Une agrégation familiale des cas de TSA a été notée dans de nombreux travaux. Par exemple, une étude sur la population suédoise a montré un risque relatif, pour un apparenté de sujet diagnostiqué TSA, s'échelonnant de RR : 2, IC_{95%} : [1,8-2,2] pour un cousin à RR : 10,3, IC_{95%} : [9,4-10,3] pour la fratrie (voir Figure 5) (Sandin *et al.*, 2014). Néanmoins, ce surrisque proportionnel à la proximité génétique des individus n'est pas suffisant pour impliquer une causalité génétique du fait de l'importance de l'environnement partagé.

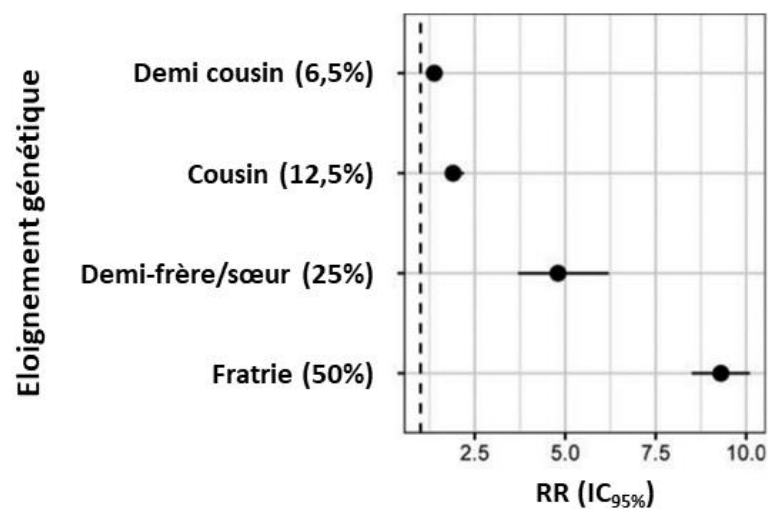
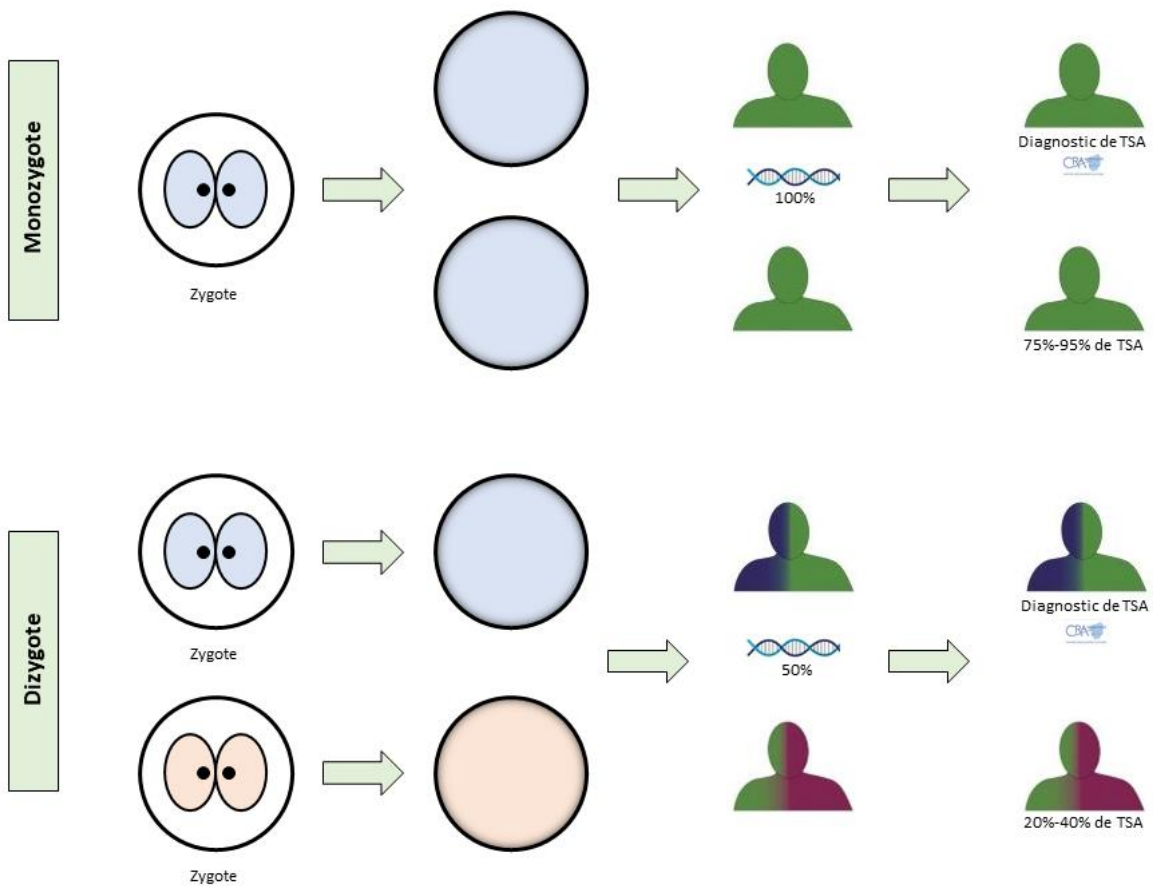


Figure 5. Surrisque de TSA chez les apparentés de patients, en fonction de leur degré de proximité génétique. Entre parenthèse, la proportion du génome partagé. Plus la proximité génétique des individus est importante, plus le risque relatif de récurrence (ajusté sur les âges) de TSA est important. D'après Havdahl *et al.*, 2021.

En réalité, la meilleure preuve provient des études de jumeaux. Ces études prennent en considération des paires de jumeaux monozygotes ou dizygotes, dont l'un des deux est diagnostiqué TSA. Les jumeaux présentent une correspondance dans leur patrimoine génétique de 100% (monozygotes) ou 50% (dizygotes). Plus le facteur génétique est important dans l'étiologie du trouble, plus les monozygotes seront concordants sur le diagnostic de TSA par rapport aux dizygotes (Figure 6). Ce paradigme considère que les enfants partagent un environnement similaire dans une même proportion quelle que soit leur zygote.



Figures 6. Explication schématique du principe des études de jumeaux. Les jumeaux monozygotes proviennent de la division d'un zygote unique. Ils partagent l'ensemble de leur patrimoine génétique. A l'inverse les jumeaux dizygotes proviennent de deux zygotes différents, et ils partagent la moitié de leur patrimoine génétique, comme tous membres d'une fratrie. En partant du postulat que l'exposition aux facteurs environnementaux ne varie pas en fonction du statut mono ou dizygote, alors on peut conclure que plus la concordance du phénotype TSA est importante chez les monozygotes par rapport aux dizygotes, plus le déterminisme génétique du trouble est important.

Toutes ces études montrent une concordance plus importante entre les jumeaux monozygotes qu'entre les jumeaux dizygotes ce qui établit l'implication de facteurs génétiques dans l'étiologie du trouble. En considérant les études publiées (Figure 7), on observe que la concordance, si elle est importante, n'est jamais complète ce qui est en accord avec l'existence des facteurs des risque environnementaux que nous avons détaillé précédemment.

Ces études de jumeaux permettent le calcul de l'héritabilité du trouble. Ce concept statistique a été, à l'origine, défini afin de départager la part de variabilité du trait capturée par la variabilité génétique par rapport à celle issue de la variabilité environnementale dans la population. Elle s'entend pour un caractère quantitatif (le quotient intellectuel, par exemple) (Tenesa & Haley, 2013). Son calcul nécessite de poser l'hypothèse que les variations de ce trait seraient issues de la somme d'une multitude de facteurs environnementaux et génétiques indépendants portant chacun un effet minime (Tenesa & Haley, 2013). Il est entendu qu'aucune des sous-parties de cette hypothèse ne va de soi en pathologie humaine en général, et dans le TSA en particulier. De plus, pour l'étude d'un trait binaire – malade ou non – une hypothèse supplémentaire est nécessaire : le modèle considère qu'il existe une dimension « susceptibilité » sous-jacente au trouble, non observable cliniquement, chez tout individu. La déclaration de la maladie chez un individu donné dépendrait du dépassement d'un effet seuil sur cette dimension « susceptibilité » (ou *liability*) (Tenesa & Haley, 2013). Un tel trait autistique est effectivement décrit en population générale à partir de questionnaires psychométriques (Constantino & Todd, 2003; Hoekstra *et al.*, 2007), et est héritable (Hoekstra *et al.*, 2007). Toutefois, il n'est pas certain que les facteurs explicatifs de cette dimension soient les mêmes que

ceux qui expliquent le syndrome TSA chez l'intégralité ou même la majorité des patients.

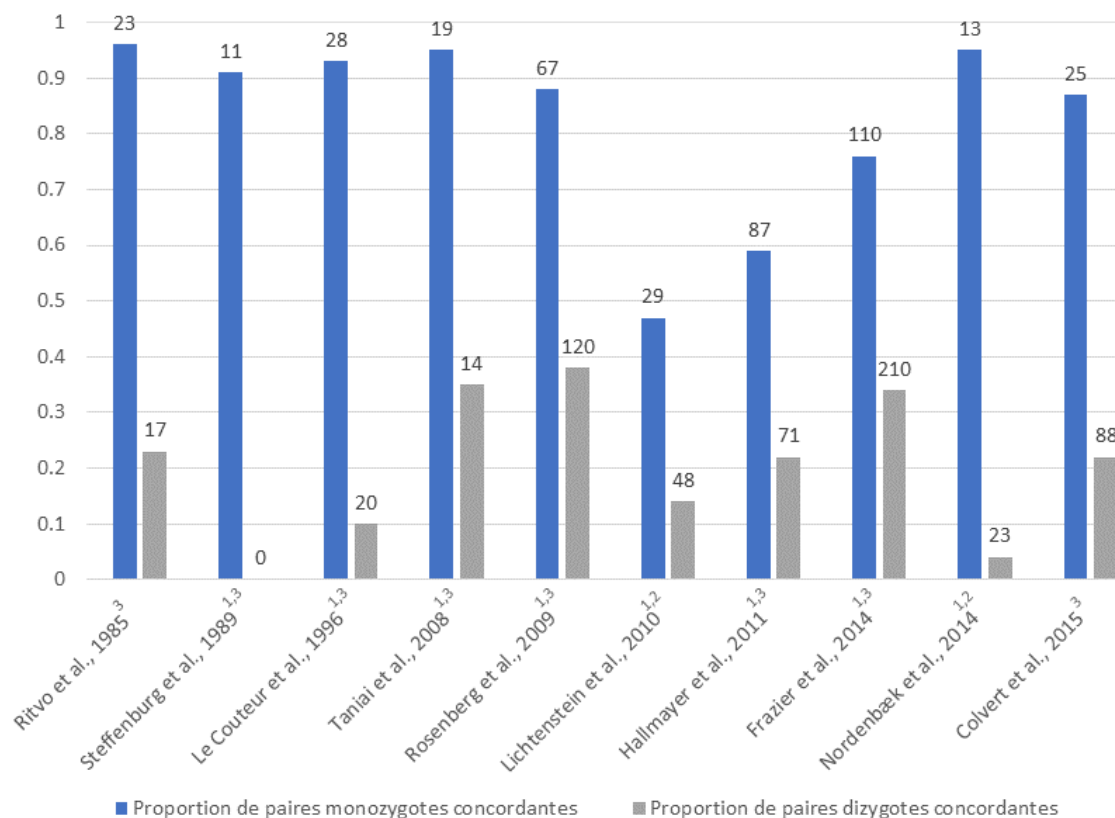


Figure 7. Etudes de jumeaux publiées dans le TSA ; ¹résultats pour les jumeaux de même sexe ; ²Etude prospective ou sur registre ; ³Paires de jumeaux inclus car l'un au moins présente un diagnostic de TSA. Dans l'ensemble des études publiées, les jumeaux monozygotes sont plus souvent concordants pour le TSA que les dizygotiques. L'ampleur de la concordance est variable. Elle est influencée par les critères diagnostiques utilisés, par le mode de recrutement des paires de jumeaux et par la méthode pour établir la zygote. En méta-analysant les données brutes (disponibles dans l'article de Tick et collaborateurs (2016)) on obtient 80% de concordance pour les monozygotes, 18% pour les dizygotiques et 10% au sein de la fratrie.

Malgré ses limitations, ce concept permet de donner une idée du poids respectif des facteurs environnementaux et génétiques qui influent la variabilité d'un trouble dans une population donnée. Une méta-analyse regroupant la plupart des études citées plus haut calcule une héritabilité entre 60 et 90% en fonction des critères diagnostiques utilisés et de la prévalence du TSA dans la population (Tick *et al.*, 2016). Il s'agit de l'héritabilité la plus importante parmi les pathologies neuropsychiatriques.

Risque génétique additif porté par les polymorphismes génétiques

A l'échelle du génome, il existe un très grand nombre de variations génétiques fréquentes ou polymorphismes, c'est-à-dire avec une fréquence de l'allèle mineur $> 0,05$ (MAF, *Minor Allele Frequency*) dans la population (Karczewski *et al.*, 2020). Elles sont particulièrement représentées dans les régions intergéniques et introniques où la pression de sélection est faible.

L'association entre ces polymorphismes et le TSA, postulée dans les études d'héritabilité, a été recherchée. C'est dans cette perspective « *common disease, common variant* » qu'ont été menées les études d'association pangénomiques ou GWAS (*Genome-Wide Association Studies*). Dans ces travaux, plusieurs centaines de milliers de polymorphismes génétiques sont testés simultanément le long du génome et la fréquence de chacun d'eux est comparée entre les sujets TSA et des témoins. Une méthodologie statistique complexe est nécessaire pour tenir compte de l'origine ethnique des individus et de la liaison génétique entre les variations au sein d'un locus (Tam *et al.*, 2019). Pour l'interprétation de ces études il faut garder à l'esprit les éléments suivants (Tam *et al.*, 2019) :

- Chaque polymorphisme porte un risque très faible à l'échelle et les individus non atteints en portent tous un certain nombre.
- Au vu du grand nombre de polymorphismes testés, un seuil de significativité pangénomique défini à 5.10^{-8} est nécessaire dans chaque test d'association. On en déduit qu'une taille d'échantillon de plusieurs dizaines de milliers de patients et de témoins est nécessaire pour atteindre une puissance statistique acceptable.
- Chaque *hit* de GWAS associe un locus du génome au trouble. Le risque en lui-même peut être porté par le polymorphisme testé ou le plus souvent par une autre variation avec lequel il se trouve en déséquilibre de liaison au sein du locus.
- Ces loci sont localisés le plus souvent en dehors des séquences codantes. Le gène le plus proche est généralement rapporté mais il n'y a pas de preuve qu'il soit effectivement impacté : dans la cellule, le génome prend une structure tridimensionnelle rapprochant des séquences linéairement éloignées (Maurano *et al.*, 2012). Le lien des variations identifiées en GWAS avec leur impact biologique est donc difficile à définir.

La principale étude GWAS publiée dans le TSA a nécessité l'inclusion de plus de 20 000 sujets porteurs d'un TSA et 170 000 contrôles. Les auteurs ont établi l'existence de 5 loci significativement associés au TSA (Grove *et al.*, 2019), dont la localisation est visible sur la Figure 8.

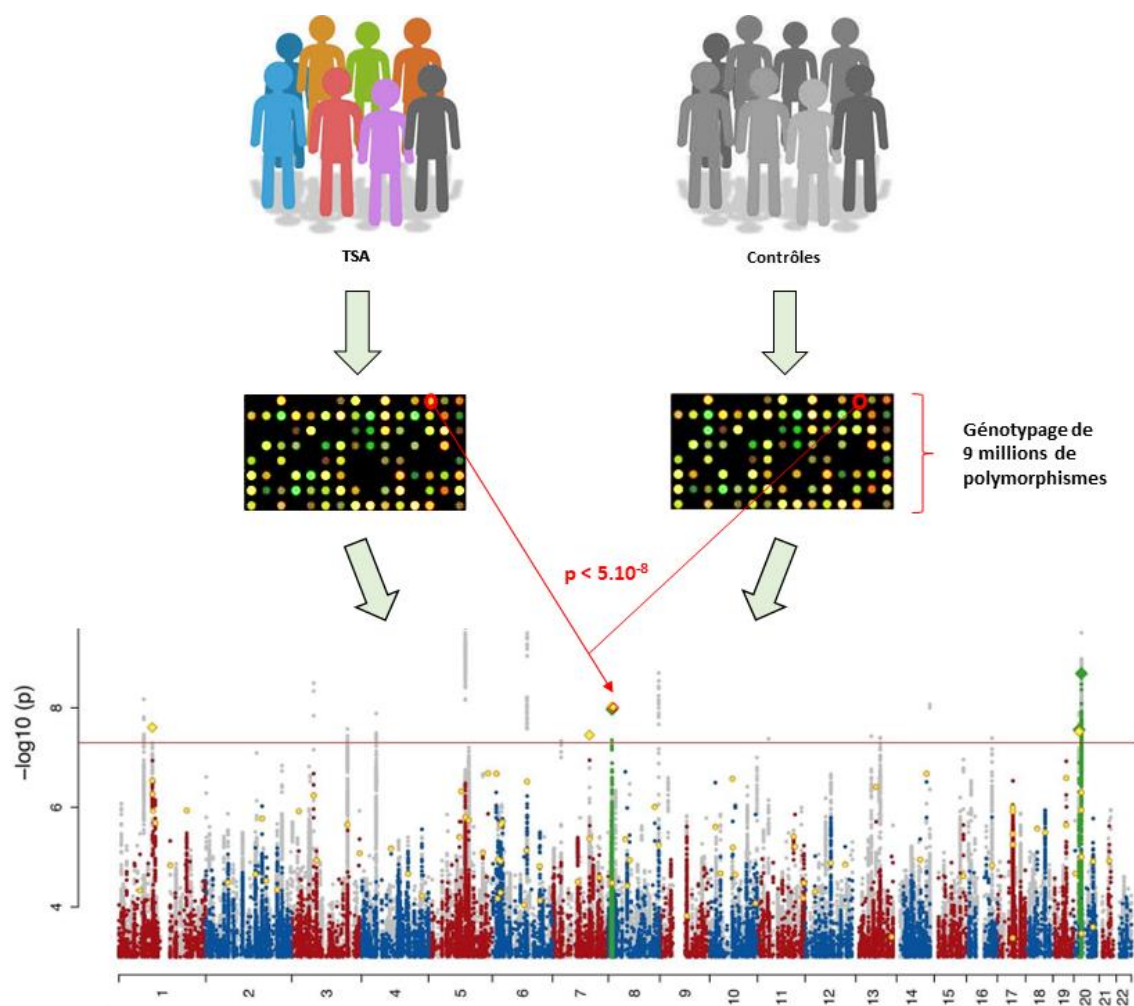


Figure 8. Schéma illustrant le principe des GWAS et Manhattan plot des résultats de la principale étude dans le TSA. La fréquence de chaque polymorphisme étudié est comparée entre cas et témoins. Lorsque l'un de ceux-ci passe le seuil de significativité, tenant compte des comparaisons multiples, alors le loci en regard est associé au trouble. Les nombreux biais possible, techniques ou de stratification de population, imposent des analyses post-hoc pour fiabiliser les résultats. Ici, elles ne confirment qu'une partie des associations. Les autres sont représentées en grisé. Adapté de Grove et al., 2019. (Grove et al., 2019)

Pour aller plus loin dans la quantification du risque porté par ces polymorphismes et dans la mesure où : i) c'est la sommation du risque qu'ils portent qui explique leur risque à l'échelle individuelle, ii) la plupart

des loci réellement associés au TSA ne passent pas le seuil de significativité pour des raisons de puissance statistique, les auteurs de *GWAS* ont proposé le calcul d'un score de risque polygénique (PRS, Polygenic Risk Score). Ce score est composé, pour chaque individu, de la somme des allèles à risque dont il est porteur, pondérés par leurs OR. Il est testé pour différents seuils de p-valeur de manière à obtenir le meilleur rapport signal/bruit (Choi *et al.*, 2020).

Le calcul du PRS (et avant cela l'idée d'héritabilité) nécessite de poser l'hypothèse d'une origine dimensionnelle du TSA. Cette conceptualisation s'oppose à une logique catégorielle binaire (malade ou non), plus opérante en médecine mais rarement fidèle aux événements biologiques.

Comme nous l'avons expliqué concernant l'héritabilité, une dimension « trait autistique » existe bien en population générale. Celui-ci entraîne des variations de la symptomatologie autistique infra-clinique qui s'expriment comme des particularités dans la personnalité de l'individu. Lorsque qu'un certain seuil (Figure 9) est dépassé l'impact est suffisant pour parler de TSA : la logique catégorielle est rejointe.

L'intensité de ce trait autistique, non pathologique en dehors des extrêmes, est distribué selon une loi gaussienne dans la population (Constantino & Todd, 2003; Hoekstra *et al.*, 2007). On remarque qu'il est plus important chez les apparentés de patients présentant un TSA. De plus, il est bien héritable tout comme la catégorie morbide TSA. Deux observations qui orienteraient vers une voie neurologique (au moins partiellement) commune avec le syndrome TSA de la nosographie classique (Hoekstra *et al.*, 2007).

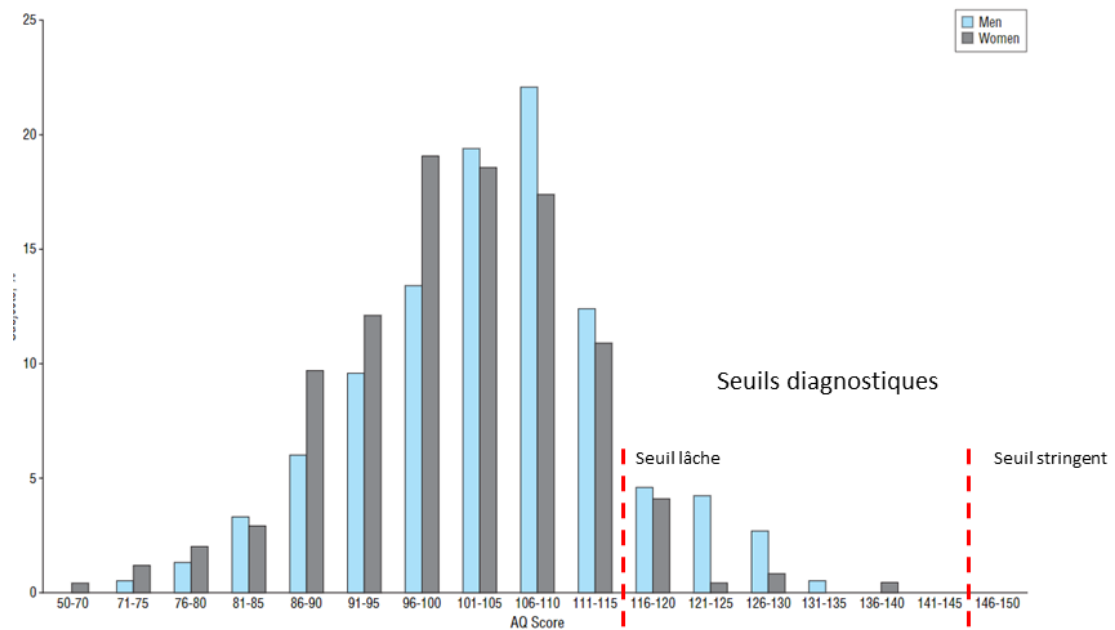


Figure 9. Distribution des scores du quotient de spectre autistique chez des jumeaux et leurs apparentés, issus de la population générale. Adapté de Hoekstra et al., 2007. Deux seuils diagnostiques sont représentés, issus d'autres travaux de la même équipe (Hoekstra et al., 2008) : un seuil lâche correspondant au trouble envahissant du développement non spécifique de la classification CIM-10 et un seuil stringent issu de la classification du DSM-IV.

Les critiques proposées quant au modèle de l'héritabilité sont, ici aussi, valables. Le modèle sous-jacent sur lequel ces calculs se basent suppose un risque génétique additif, porté par les polymorphismes. Ces facteurs génétiques additifs de faible impact permettent le calcul d'une variable continue appelée liability (que l'on suppose être plus ou moins équivalente au « trait » autistique décrit en population générale). Ils pourraient s'additionner avec des facteurs environnementaux de faible impact mais il n'existerait ni effets multiplicatifs ni facteurs de risque à effet fort. La maladie (ou le syndrome) « TSA » surviendrait du fait d'un effet seuil de liability. Enfin, l'ensemble des formes de TSA serait lié à une position extrême sur cette distribution (Gaussienne) de liability. En somme, cette

modélisation ne tient que grâce à une série de postulats (soulignés) tous très discutables.

En plus de ces critiques, il faut souligner que le PRS n'a d'intérêt qu'en termes de recherche. Il ne doit pas être utilisé comme un test diagnostique. En effet, même en comparant des individus aux extrêmes de la distribution, l'OR de présenter un TSA ne se situe au maximum qu'entre 2.5 et 3 (Figure 10). Ainsi, ses capacités prédictives sont bien trop limitées.

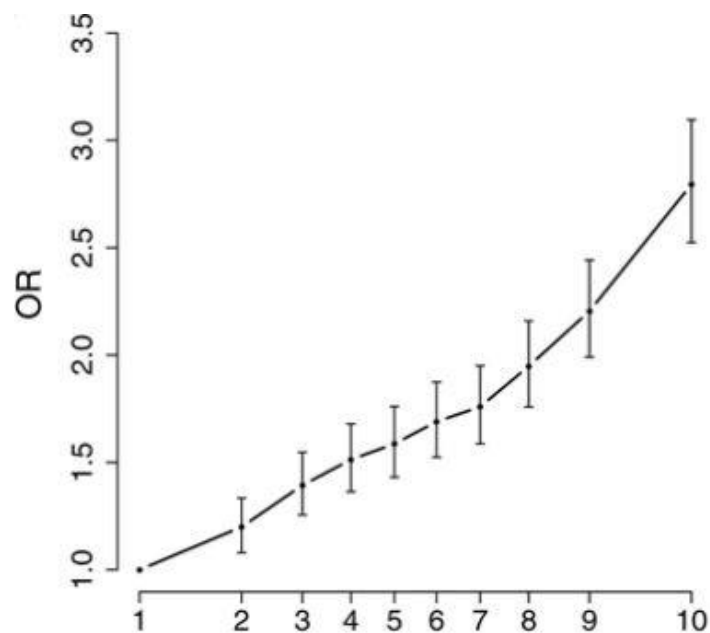


Figure 10. Répartition des odds-ratio en fonction du PRS pour 13 076 cas et 22 664 contrôles. Les intervalles de confiance à 95% des odds-ratio sont représentés pour chacun des déciles. Le PRS représenté est calculé à partir des polymorphismes associés au test avec une p-valeur $p < 0.1$. Même en considérant deux individus chacun à un extrême de la distribution de PRS, ce dernier n'a qu'une valeur prédictive limitée. D'après Grove et al., 2019.

C'est à l'échelle de la population que ces polymorphismes – par leur fréquence par définition élevée – expliquent une part potentiellement substantielle du risque de TSA (Gaugler *et al.*, 2014).

L'ensemble des travaux que nous avons cités dans ce chapitre permet de conclure que le TSA est un trouble issu d'altérations cérébrales précoces et étendues. Le déterminisme de celles-ci est complexe et inclus des facteurs environnementaux et génétiques.

Les facteurs de risque environnementaux et génétiques que nous avons décrit doivent être considérés à l'échelle de la population entière en raison de leur faible impact individuel ($OR \leq 2$). D'autant plus qu'un certain nombre d'entre eux sont très probablement des *proxies* marquant le risque mais non directement impliqué dans l'étiologie du trouble. Ainsi, aucun⁸ de ces facteurs de risque n'est pertinent pour expliquer le déterminisme du trouble à l'échelle individuelle. Le risque qu'ils induisent n'y est pas cliniquement significatif à l'échelle individuelle. En effet, si nous prenons une prévalence du TSA de 2% et un facteur de risque induisant un OR de 2 en faveur du TSA, le risque de TSA n'est que de $2 \times 2\% = 4\%$ chez un individu porteur. Ce facteur de risque ne représenterait qu'une part très minoritaire dans le déterminisme du trouble et n'est pertinent qu'à une échelle populationnelle.

⁸A l'exception de l'exposition du fœtus au valproate de sodium qui provoque un syndrome neurodéveloppemental complexe dont le TSA est l'une des manifestations phénotypiques possibles.

Bien sûr, si ces facteurs de risque présentaient un effet multiplicatif, leur distribution singulière chez un individu pourrait aboutir à un risque fort chez un individu donné mais il n'existe pas de démonstration en ce sens à notre connaissance.

Nous verrons dans la partie suivante, qu'à l'inverse, certaines variations génétiques rares sont associées à un risque suffisamment élevé pour expliquer une part significative de la causalité du trouble chez un individu porteur.

EVALUATION DU SEQUENÇAGE D'EXOME DANS LE DIAGNOSTIC GENETIQUE DU TSA

Dans un premier chapitre, nous détaillerons les différentes approches qui ont concouru à impliquer des gènes dans l'origine du TSA chez certains patients *via* les variations génétiques rares. Nous le concluons en évoquant l'association entre ces variations rares et les polymorphismes pour concourir au risque de TSA.

Le chapitre suivant sera consacré à la détermination d'une liste de gène de TSA et à l'évaluation de son rendement diagnostique en séquençage d'exome sur une cohorte rétrospective de patients TSA. Ce travail a été validé par une publication scientifique. Le troisième chapitre concernera la réplication de ce travail sur une cohorte prospective avec évaluation du bénéfice patient, ce travail sera soumis pour publication en 2023 après réception des dernières données et réalisation des dernières analyses.

Chapitre I. Les variations génétiques rares dans le TSA

i. L'implication de variations génétiques rares dans le TSA

La diversité du génome humain

Les premiers travaux de séquençage extensif des 3.1 milliards de paires de bases (pb) du génome humain ont été publiés en 2001 (Lander *et al.*, 2001; Venter *et al.*, 2001) avant d'aboutir à une séquence quasi-complète de référence en 2004 (International Human Genome Sequencing Consortium, 2004). La séquence de référence d'*Homo sapiens* a été depuis plusieurs fois amendée par le Genome Reference Consortium⁹ et les portions les plus difficiles d'accès n'ont été totalement séquencées que tout récemment (Nurk *et al.*, 2022).

Une partie minoritaire de l'ADN comprend les gènes codant pour des protéines ou d'autres éléments de régulation de l'expression génique (comme les ARNs non codants). Seuls les exons (1,5% du génome) sont finalement exprimés sous forme de protéine.

Une particularité remarquable du génome humain est son extrême variabilité : chaque individu porte plusieurs millions de variations génétiques par rapport à la séquence de référence (Lappalainen *et al.*, 2019). Les deux principaux types de variations génétiques étudiées, les

⁹ <https://www.ncbi.nlm.nih.gov/grc>

variations de la séquence nucléotidique et les variations du nombre de copies sont représentés dans la Figure 11. A l'échelle de la population, 99% de ces variations sont rares, c'est-à-dire avec une fréquence de l'allèle $\leq 0,01$ (Karczewski et al., 2020).

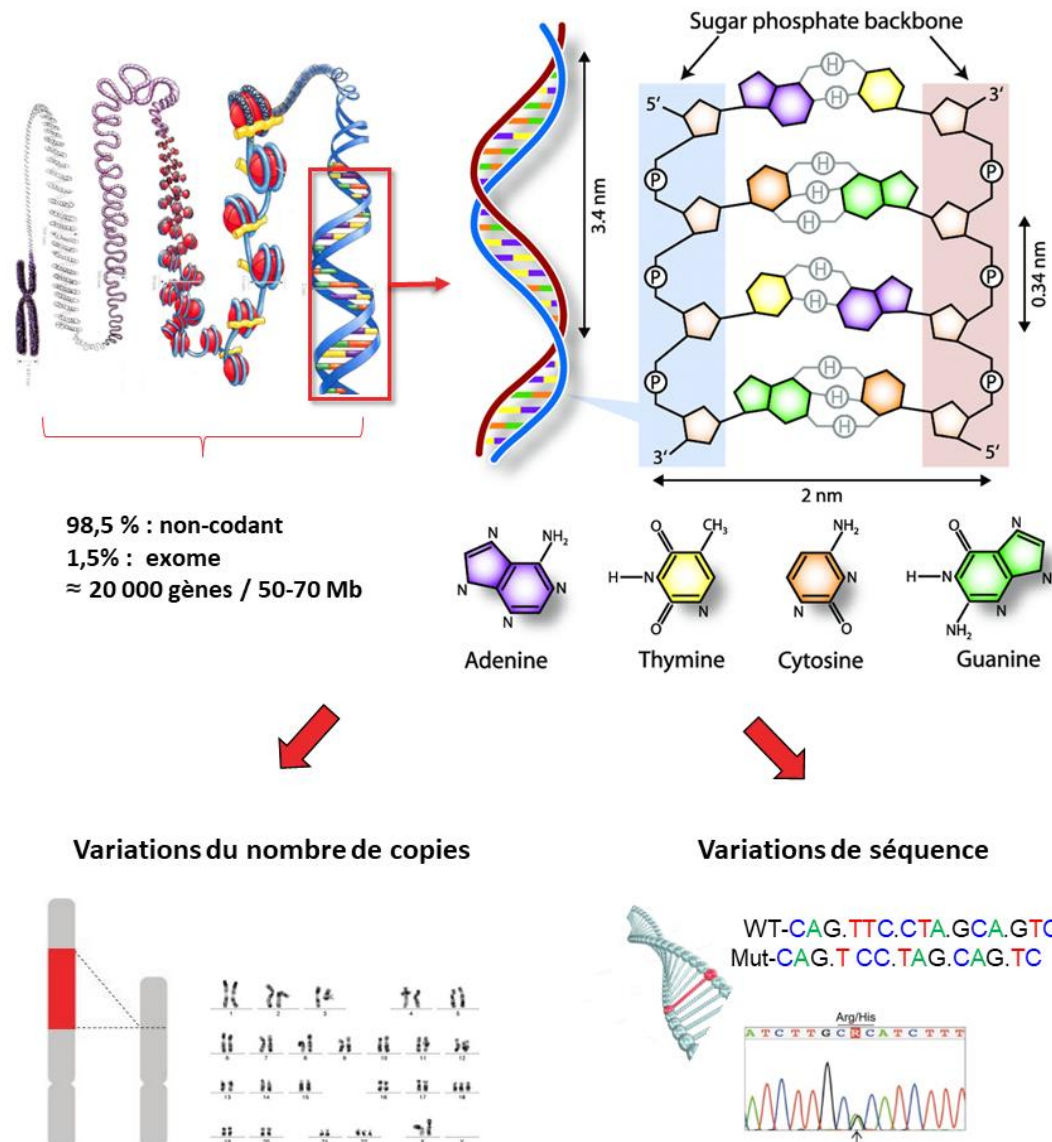


Figure 11. Illustration des variations génétiques. Les variations de la séquence nucléotidique comprennent d'une part les variations ponctuelles et les petites insertions et délétions (indels) et d'autre part les variations du nombre de copies qui sont des remaniements de plus grande taille. Adapté de Gauthier, 2007.

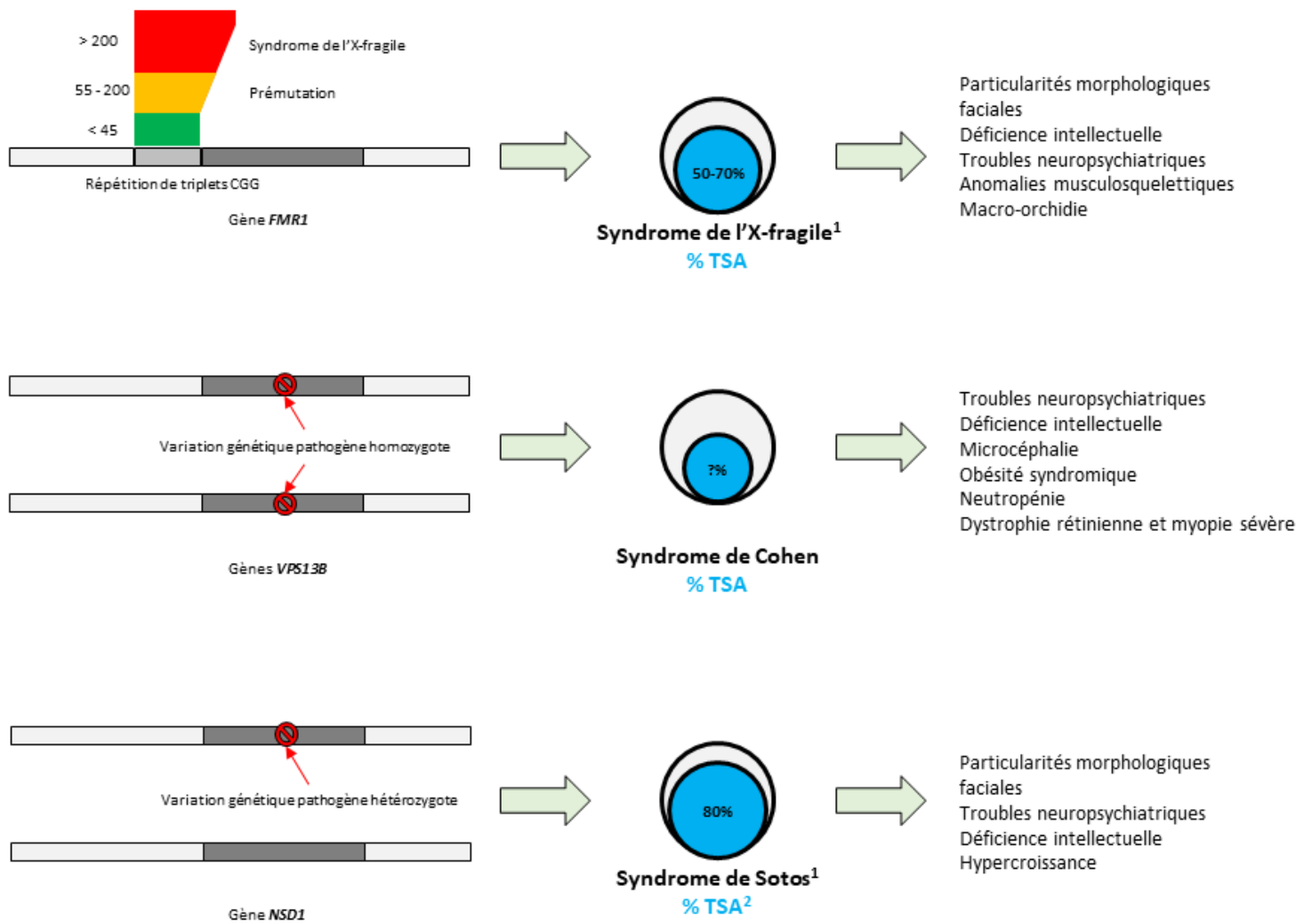


Figure 12. Trois exemples de syndrome neurodéveloppementaux. *Syndrôme de l'X fragile (*FMR1*) : le gène contient une séquence de répétitions de triplets CGG. Le nombre de triplets peut augmenter au fil des générations par erreur lors de la méiose. La maladie se déclare lorsque le nombre de triplets dépasse le nombre de 200. Il s'agit d'une maladie de déterminisme non mendélien liée à l'X et pour laquelle la pénétrance est réduite chez les sujets de sexe féminin et pour laquelle il existe d'autres phénotypes en présence d'une prémutation (allèles avec 55 à <200 répétitions entraînant un gain de fonction toxique et ainsi une autre pathologie de pénétrance incomplète). Le syndrome de l'X-fragile est le premier syndrome neurodéveloppemental à avoir été associé à un surrisque de TSA (Ciaccio et al., 2017). Syndrôme de Cohen (*VPS13B*) : il s'agit d'une maladie autosomique récessive, l'inactivation des deux copies du gène est nécessaire pour que la maladie soit présente. L'augmentation du risque de TSA n'est à notre connaissance pas quantifié. Syndrôme de Sotos (*NSD1*) : il s'agit d'une maladie autosomique dominante, l'inactivation d'une copie du gène est suffisante pour que la maladie soit présente. ¹Phénotypes issus de l'encyclopédie en ligne GeneReviews. ²selon (Lane et al., 2017).*

La description de formes syndromiques de TSA à partir des années 2000

Les années 2000 ont vu la description d'un grand nombre de corrélations clinico-biologiques entre des syndromes neurodéveloppementaux cliniquement reconnaissables et l'altération de gènes spécifiques. Un grand nombre de ces syndromes présentent le TSA dans leur spectre clinique (Betancur, 2011) avec une prévalence plus ou moins importante.

Ces syndromes sont des maladies monogénique, le plus souvent à hérédité Mendélienne classique. Dans ce cas, une variation génétique pathogène dans un gène spécifique entraîne la maladie correspondante. Différents types de transmissions (autosomique dominant, récessif ou lié à l'X¹⁰) et différents types de variations pathogènes (maladies à expansion de triplets, variations du nombre de copies et variations nucléotidiques) peuvent être impliqués (Figure 12).

Ce modèle idéal doit néanmoins être quelque peu modéré. En réalité, ces gènes présentent une expressivité très variable : les individus touchés par une même variation pathogène dans un gène donné présentent fréquemment un phénotype différent.

La neurofibromatose de type 1 est un bon exemple de cette problématique d'expressivité variable. Une variation génétique pathogène dans le gène *NF1* induit ce syndrome selon un mode de transmission

¹⁰ Les syndromes récessifs liés à l'X, s'exprimant uniquement ou majoritairement (selon les cas) chez les garçons n'expliquent pas de manière décisive le *sex-ratio* dans le TSA : ils n'expliqueraient qu'environ 2% des cas chez l'homme (Lim *et al.*, 2013).

autosomique dominant. En pratique clinique, il s'agit d'une maladie multi-systémique avec des manifestations cutanées, vasculaires, ostéoarticulaires, tumorales et neurodéveloppementales (J. M. Friedman, 2022). Sur le plan neurodéveloppemental, le risque de déficience intellectuelle est multiplié par deux (Al-Farsi et al., 2022) et un diagnostic de TSA ou de TDAH est posé chez respectivement 25% (Chisholm et al., 2022) et 30-50% (Vogel et al., 2017) des patients. Tous les états intermédiaires de handicap peuvent se rencontrer entre des patients ne présentant que des manifestations cutanées peu ou pas morbides, d'autres au pronostic vital engagé par une évolution tumorale maligne de la maladie et enfin d'autres avec un handicap cognitif plus marqué.

L'approche classique pour le diagnostic de ces syndromes neurodéveloppementaux est le séquençage ciblé du ou des gènes incriminés lorsque le clinicien constate un point d'appel clinique. Celui-ci est le plus souvent une association connue de malformations d'organe, la reconnaissance d'une dysmorphie faciale spécifique ou plus rarement des particularités comportementales. Cette approche peut cependant être mise en défaut pour les patients qui présentent une forme légère. De plus, il existe des chevauchements syndromiques fréquents entre des syndromes associés à des gènes différents. Par exemple, le syndrome de Rett est causé par une perte de fonction du gène *MECP2* mais au moins 17 gènes entraînent des présentations cliniques similaires et souvent indiscernables (Vidal et al., 2019).

A mesure que de nouveaux syndromes neurodéveloppementaux sont décrits, aux phénotypes moins spécifiques ou reconnaissables, cette stratégie est de plus en plus mise en défaut. Au quatrième trimestre 2021,

la base de données SFARI Gene¹¹ de la *Simons Foundation Autism Research Initiative* (Abrahams *et al.*, 2013) référençait 239 gènes impliqués dans des syndromes neurodéveloppementaux avec un surrisque de TSA. La plupart ne sont pas cliniquement reconnaissables.

Un changement d'échelle est donc nécessaire pour démêler cette question clinique complexe.

Etudes pangénomiques des variations du nombre de copies

Les variations du nombre de copies (*Copy Number Variants, CNVs*) sont des variations de ploïdie d'au moins 50 pb sur un locus donné d'un chromosome (Collins *et al.*, 2020). Cette limite a été fixée arbitrairement pour les différencier des indels, c'est-à-dire les petites insertions et délétions, qui ont longtemps été détectées par des technologies différentes. Il s'agit de remaniements dont la taille peut s'étendre à l'échelle de la gigabase (Collins *et al.*, 2020). Les CNVs sont beaucoup moins nombreux que les variations ponctuelles puisque le génome d'un individu ne comporte qu'un nombre médian d'environ 3500 délétions et 720 duplications (Collins *et al.*, 2020). En revanche, la taille importante de ces variations génétiques – et donc le nombre de gènes qu'ils peuvent potentiellement impliquer – en font l'un des déterminants importants de la diversité humaine et des phénomènes évolutifs (Pös *et al.*, 2021).

Le mécanisme le plus souvent impliqué dans ces remaniements structuraux est la recombinaison homologue non allélique (Pös *et al.*, 2021).

¹¹ <https://gene.sfari.org/>

Elle est déclenchée par des séquences répétées appelées duplications segmentaires. C'est le cas, par exemple, de la délétion 22q11.2 que nous avons déjà citée. Dans ce cas, les points de cassures sont répétitifs sur un locus donné (Figure 13) (Pös *et al.*, 2021). Dans environ 5% des cas, d'autres mécanismes de recombinaison génétique sont impliqués et les bornes du CNV sont variables (Pös *et al.*, 2021).

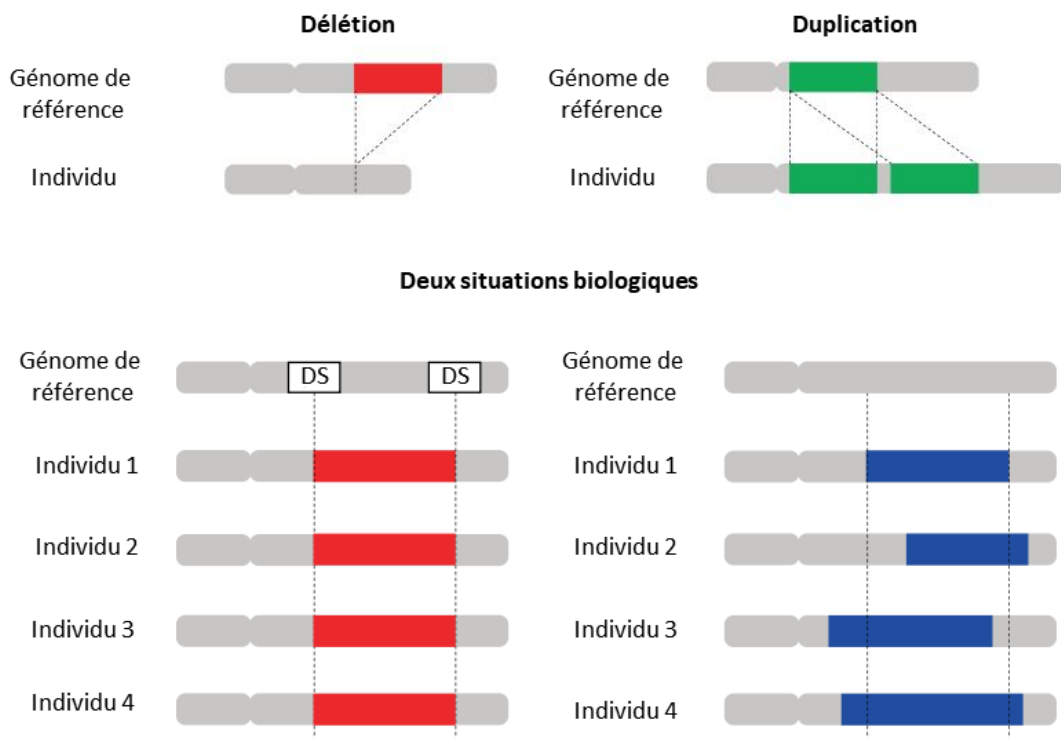


Figure 13. Illustration du concept de variations du nombre de copies. Les deux principaux CNVs impliqués en pathologie humaine sont les délétions et les duplications. Elles peuvent être médiées par des régions répétées appelées Duplications Segmentaires (DS) qui bornent le remaniement. Dans d'autres cas, minoritaires, les bornes peuvent être situées dans n'importe quelle position du génome. Il existe également un certain nombre de remaniements équilibrés – ne faisant pas varier la ploïdie – dont la détection est plus complexe, comme les inversions, que nous nous contentons de citer. Adapté de Nakatochi *et al.*, 2021.

Le développement de puces de détection de CNVs à l'échelle pangénomique a constitué un changement de paradigme. Le développement de la CGH-array (*Array Comparative Genomic Hybridization*, puce d'hybridation génomique comparative) (Pinkel *et al.*, 1998) permet la détection de l'ensemble des CNVs d'un individu. La résolution de détection de ces remaniements a augmenté à mesure des perfectionnements technologiques et subséquemment la taille de ce qui est détectable. On atteint maintenant la limite d'une cinquantaine de pb grâce à l'utilisation de techniques de séquençage de nouvelle génération ou de techniques ciblées (Alkan *et al.*, 2011).

Certains CNVs sont des facteurs de risque de TSA

Les CNVs constituent l'une des causes génétiques de syndromes neurodéveloppementaux pouvant s'exprimer par un TSA que nous avons vu au chapitre précédent. Dans ce cas, les mêmes difficultés se posent pour les syndromes non cliniquement discernables lorsque l'on considère l'utilisation de techniques ciblées comme l'hybridation *in situ* en fluorescence (Bauman *et al.*, 1980).

A l'origine, seuls les CNVs de très grandes tailles, repérés visuellement sur un caryotype, étaient détectables. C'est ainsi qu'un premier CNV, la délétion de la partie terminale du bras court du chromosome 5, a été associé à un syndrome neurodéveloppemental : le syndrome du Cri-du-Chat (Lejeune *et al.*, 1964). Les progrès technologiques ont permis la détection de CNVs de taille toujours plus restreinte. L'observation de corrélations clinico-biologiques entre le phénotype de patients et la position de petits CNVs a permis de délimiter les régions minimales critiques correspondantes et de préciser parfois le ou les gènes causaux (Figure 14).

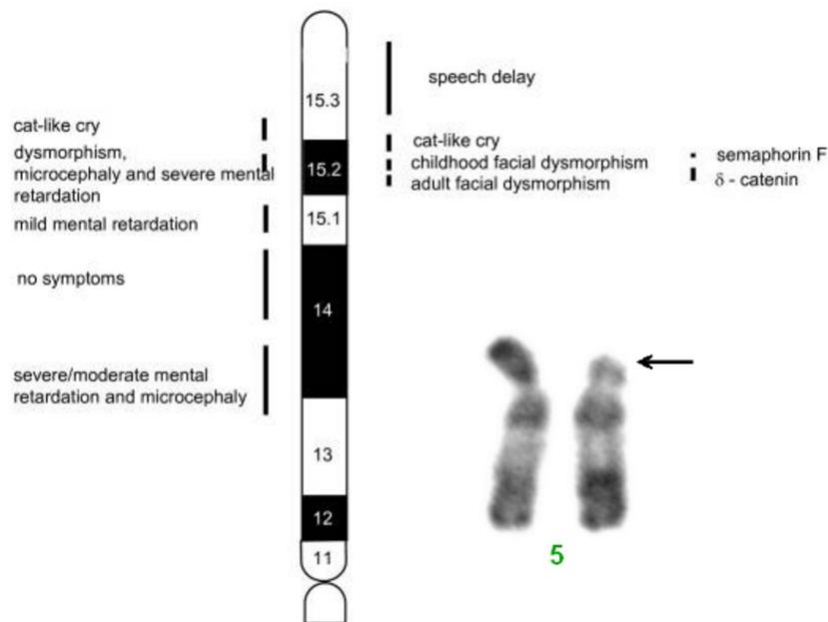


Figure 14. Corrélation entre la localisation de la délétion et les symptômes cliniques des patients présentant une délétion de l'extrémité du bras court du chromosome 5. Représentation schématique extrait de Cerruti Mainardi, 2006, et visualisation d'un exemple de celle-ci sur un caryotype extrait de Ouldin *et al.*, 2012. Ce syndrome n'est pas médié par l'existence de duplications segmentaires et la localisation des délétions pathogènes est donc variable au sein du locus. La corrélation entre la position exacte du remaniement (et donc les gènes qu'il implique) et le phénotype du patient a permis de relier gènes et symptômes.

Le développement des puces pangénomiques ont permis de développer des études cas-témoins entre des patients présentant un TSA, sans phénotype clinique particulier d'orientation, et des témoins issus de la population générale. Dans un premier temps, elles ont montré l'augmentation du taux de CNVs chez les patients TSA par rapport à des contrôles (Girirajan *et al.*, 2013; Pinto *et al.*, 2010) ou par rapport à leurs apparentés sains (Leppa *et al.*, 2016). Puis, avec l'augmentation des tailles d'échantillons, il a été mis en évidence plusieurs loci qui présentent un excès significatif de CNVs récurrents dans le TSA. Tout particulièrement, l'étude

de Girajian et collaborateurs en 2013 a comparé 2588 patients présentant un TSA à 8329 sujets contrôles (Figure 15).

Il est important de noter que tous ces loci n'impliquent pas une maladie génétique mendélienne. Au chapitre précédent nous avons évoqué des syndromes neurodéveloppementaux dont la transmission était issue du dogme classique : une variation pathogène dans un gène implique une maladie correspondante. Ici, et pour un certain nombre de loci, la pénétrance est incomplète. Il existe des porteurs sains dans la population générale. Il s'agit de facteurs de risque importants mais non de facteurs causaux : ils ne sont pas nécessaires et suffisants pour le déclenchement du trouble. Une équipe a estimé la pénétrance d'un certain nombre de ceux-ci (Rosenfeld *et al.*, 2013) à l'aide d'un modèle bayésien, sur l'hypothèse d'une prévalence des troubles neurodéveloppementaux de 5% dans la population. Les résultats doivent être considérés comme une appréciation très large, mais qui vont bien dans le sens d'une pénétrance très incomplète (Tableau 2).

La CGH-*array*, et les autres technologies de détection pangénomique de CNVs, sont donc devenues des outils diagnostiques importants en génétique pour les patients souffrant de TSA. Elles permettent la détection de CNVs pathogènes que ce soit au niveau de loci connus pour être le siège de CNVs récurrents portant un surrisque de TSA ou intersectant les gènes impliqués dans les syndromes neurodéveloppementaux présentant le TSA dans leur spectre phénotypique. Ainsi, le rendement diagnostique chez les patients TSA a été estimé entre **7** % et **9** % (C Yuen *et al.*, 2017a; Y. Shen *et al.*, 2010; Tammimies *et al.*, 2015). Il s'agit d'une amélioration notable par rapport à une stratégie basée uniquement sur la réalisation d'un caryotype

et de la recherche ciblée du syndrome de l’X-fragile (Roesser, 2011; Y. Shen et al., 2010).

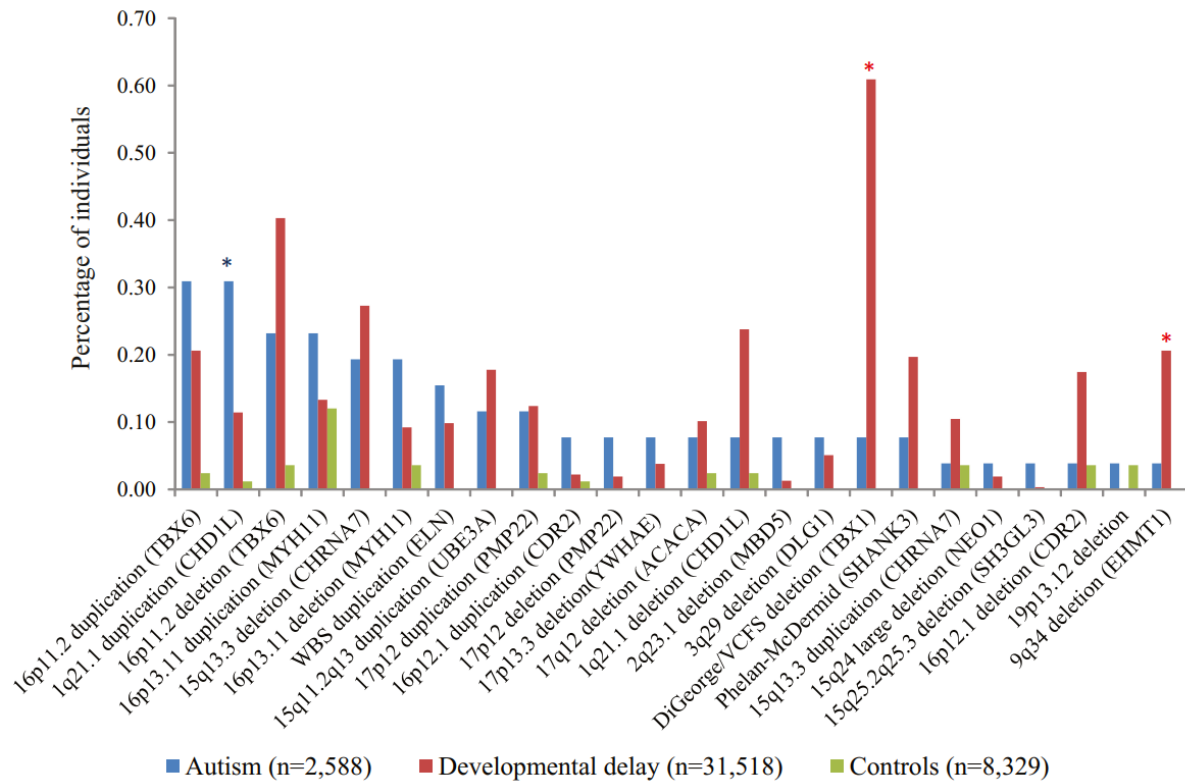


Figure 15. CNVs larges (> 400 kbp) et récurrents chez les sujets présentant un TSA comparés à des contrôles. Les données ont été méta-analysés avec celles issues de deux études de CGH-array chez des patients présentant un retard de développement (groupe « Developmental delay »). L’astérisque indique un excès de CNVs chez les patients présentant un TSA par rapport aux patients présentant un retard de développement en 1q21.1 et vice-versa sur les deux autres loci. La plupart des syndromes associés à ces loci ne sont pas cliniquement reconnaissables. Extrait de Girirajan, Dennis, et al., 2013.

Locus	Type	n cas (%)	n contrôles (%)	p-valeur	n hérités/cas*	Pénétrance (IC _{95%})
1q21.1p	Duplication	85/48 637 (0,17 %)	10/22 246 (0,04 %)	< 0,0001	13/13	17,3% (10,8%-27,4%)
1q21.1d	Délétion	97/33 226 (0,29 %)	6/22 246 (0,03 %)	< 0,0001	32/39 (82 %)	36,9 % (23%-55%)
1q21.1d	Duplication	68/33 226 (0,2 %)	6/22 246 (0,03 %)	< 0,0001	25/30 (83,3 %)	29,1 % (16,9%-49,8%)
15q11.2	Délétion	203/25 113 (0,81 %)	84/22 246 (0,38 %)	< 0,0001	27/27	10,4 % (8,45%-12,7%)
16p13.11	Délétion	50/33 226 (0,15 %)	12/22 246 (0,05 %)	< 0,0005	18/23 (78,5 %)	13,1 % (7,91%-18,8%)
16p12.1	Délétion	62/33 226 (0,11 %)	16/22 246 (0,07 %)	< 0,0002	27/28 (96,4 %)	12,3 % (7,91%-18,8%)
16p11.2d	Délétion	46/33 226 (0,11 %)	1/22 246 (0,005 %)	< 0,0001	7/21 (66,7 %)	62,4 % (26,8%-64,2%)
16p11.2d	Duplication	35/33 226 (0,28 %)	10/22 246 (0,04 %)	< 0,01	7/8 (87,5%)	11,2 % (6,26%-19,8%)
16p11.2p	Délétion	146/33 226 (0,44 %)	6/22 246 (0,03 %)	< 0,0001	14/47 (29,8 %)	46,8 % (31,5%-64,2%)
16p11.2p	Duplication	93/33 226 (0,28 %)	9/22 246 (0,04 %)	< 0,0001	23/30 (76,7 %)	27,2 % (17,4%-40,7%)
17q12	Délétion	29/33 226 (0,09 %)	2/22 246 (0,01 %)	< 0,0001	4/9 (44,4 %)	34,4 % (13,7%-70%)
17q12	Duplication	37/33 226 (0,11 %)	5/22 246 (0,02 %)	< 0,0001	7/9 (77,8 %)	21,1 % (10,6%-39,5%)
22q11.21	Duplication	136/48 637 (0,28 %)	12/22 246 (0,05 %)	< 0,0001	35/47 (74,5 %)	21,9 % (14,7%-31,8%)

Tableau 2. Estimation de la pénétrance de CNVs récurrents, pour tout trouble du neurodéveloppement. *Dans l'étude, la ségrégation familiale n'a pu être effectuée que dans un petit nombre de cas, mais pour ceux-ci on observe une origine parentale majoritaire, ce qui va également dans le sens d'une pénétrance incomplète, à supposer que les parents n'étaient pas symptomatiques. Adapté de Rosenfeld *et al.*, 2013.

ii. Un changement de paradigme avec le séquençage à haut débit

Principe et intérêt du séquençage à haut débit

Bien qu'il s'agisse d'une avancée importante, la CGH-array ne permet la détection que d'une seule classe de remaniement génétique : les CNVs, et ne permet pas la détection des plus petits d'entre eux. Pour les variations nucléotidiques, le séquençage Sanger n'est adapté que pour le séquençage de quelques centaines de pb. Il s'agit toujours de la méthode de référence à cette échelle (Beck *et al.*, 2016), mais elle devient laborieuse au-delà. Le séquençage de plusieurs gènes ou même d'un seul gène de grande taille par ce moyen est beaucoup trop dispendieux en temps et en ressources.

La mise au point du séquençage à haut débit (ou séquençage massif en parallèle), au début des années 2010, a donné accès à la détection de l'ensemble des variations génétiques d'un individu en une seule opération. Ce séquençage de nouvelle génération (NGS, *Next-Generation Sequencing*) a mis la lecture du génome humain à la portée des unités de recherche et des laboratoires de génétique médicale, et son coût est en diminution constante (Figure 16).

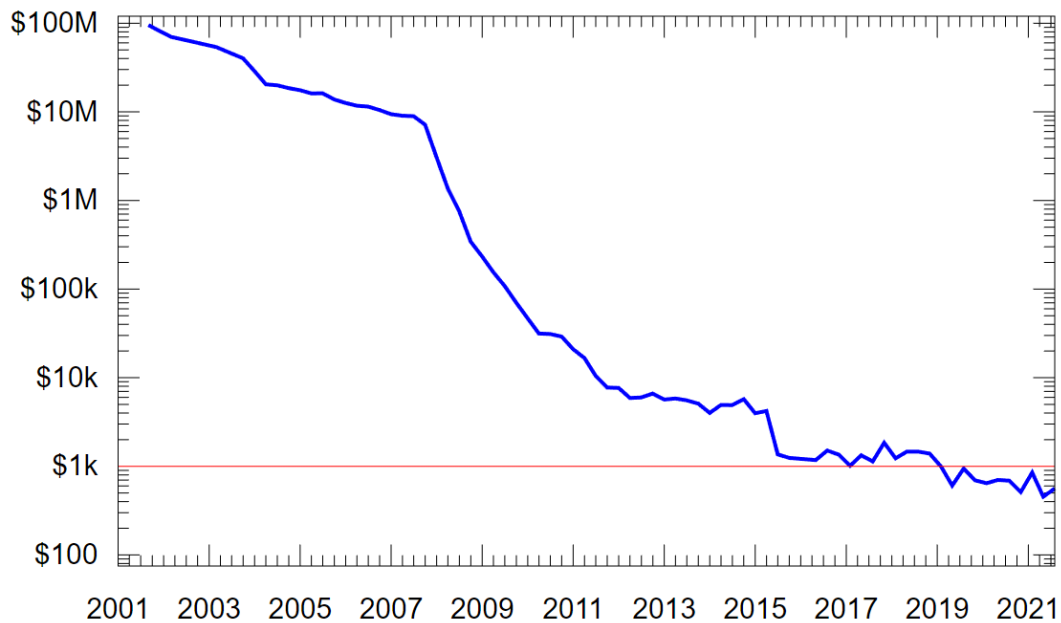


Figure 16. Evolution du coût du séquençage du génome humain entre 2001 et 2021. Figure originale de Ben Moore, réécrite en gnuplot par grendel\khan., libre de droit, via Wikimedia Commons. A noter qu’une annonce très récente de la société Illumina, principal fournisseur dans le domaine du NGS, place le coût du génome humain à ~260\$ grâce au nouveaux séquenceurs Illumina NovaseqX qui seront disponibles dès 2023, et un objectif fixé à 100\$ à l’avenir.

A l’origine, plusieurs technologies concurrentes ont été développées (Goodwin *et al.*, 2016). La technologie Illumina est actuellement en situation de quasi-monopole¹² et est celle qui a été utilisée dans le cadre de ce travail (Figure 17).

L’une des applications du NGS est le séquençage de l’ensemble des séquences codantes (des exons) d’un individu. Ce séquençage d’exome

¹² <https://frontlinegenomics.com/how-did-illumina-monopolize-the-sequencing-market>

(WES, *Whole-Exome Sequencing*) permet la détection de toutes les variations génétiques des séquences exprimées sous la forme de protéines ou d'ARNs non codants, quelques soient leurs fréquences alléliques. Pour un coût et des difficultés bioinformatiques moindres, le WES restreint l'analyse aux séquences pour lesquelles l'impact des variations sur la séquence protéique, et donc la pathogénicité, est plus simple à définir en l'absence de tests fonctionnels (Richards et al., 2015a). Les travaux basés sur le WES ne devraient être considérés que comme une étape intermédiaire de l'étude des facteurs de risque génétique de TSA. Le *whole-genome sequencing* permet d'élargir la palette de facteurs de risque détectés. Bien que les interprétations soient plus difficiles à l'échelle individuelle, le séquençage *whole-genome* montre que certaines variations génétiques non détectées en WES seraient également facteurs de risque de TSA. Ces observations concernent notamment les éléments transposables (Borges-Monroy et al., 2021) et les séquences régulatrices intergéniques (An et al., 2018; Xu et al., 2020; J. Zhou et al., 2019). *In-fine*, même si à l'échelle populationnelle la majorité du risque génétique serait situé dans les régions codantes (Werling et al., 2018), la génération du *whole-genome sequencing* devrait améliorer substantiellement le taux de détection de facteurs de risque de TSA.

Pour chaque position de l'ADN, l'amplification en pont d'Illumina produit un nombre important de fragments séquencés (*reads*) chevauchants. Le nombre de lecture de chaque base, appelé profondeur de lecture, est garant de la fiabilité de détection des variations génétiques et de la définition du statut homo ou hétérozygote. De plus, pour une région donnée, la comparaison de la profondeur de lecture par rapport à la moyenne des individus issus du même lot de données (ou *batch*) informe sur la présence ou non de CNVs. Le logiciel bioinformatique *Canoës* (Backenroth et al.,

2014) utilise notamment ce principe. Des travaux de notre équipe ont montré une sensibilité supérieure à 85% et une valeur prédictive positive de 85% pour la détection de CNVs par rapport à la CGH-*array*, avec ce logiciel (Quenez *et al.*, 2021). De plus, certains petits CNVs n'impliquant qu'un seul exon n'étaient détectés qu'en WES car ils étaient de taille inférieure à la résolution de la CGH-*array*.

Le WES est donc une technologie efficace pour la détection de l'ensemble des variations génétiques (variations nucléotidiques et CNVs) des séquences codantes du génome dont l'interprétation est plus aisée. Elle nécessite néanmoins une architecture bioinformatique conséquente pour le stockage et le traitement des données. Elle a permis une augmentation importante de la taille des cohortes séquencées. C'est par cette technologie que la très grande variabilité du génome humain, dont nous avons parlé en début de chapitre, a été mise à jour. L'effort de recherche principal a été mené par le *broad institute of MIT and Harvard* au travers des projets *Exome Aggregation Consortium* (ExAC) (Lek *et al.*, 2016) puis *Genome Aggregation Database*(GnomAD) (Karczewski *et al.*, 2020).

En plus de ces données de génétique des populations, elle a permis de poursuivre l'identification des gènes responsables de syndromes neurodéveloppementaux d'origine génétiques (Hansen *et al.*, 2019; Zhang, 2014). Le premier exemple date de 2010 avec l'identification du gène *KMT2D* comme l'une des causes du syndrome de Kabuki (Ng *et al.*, 2010).

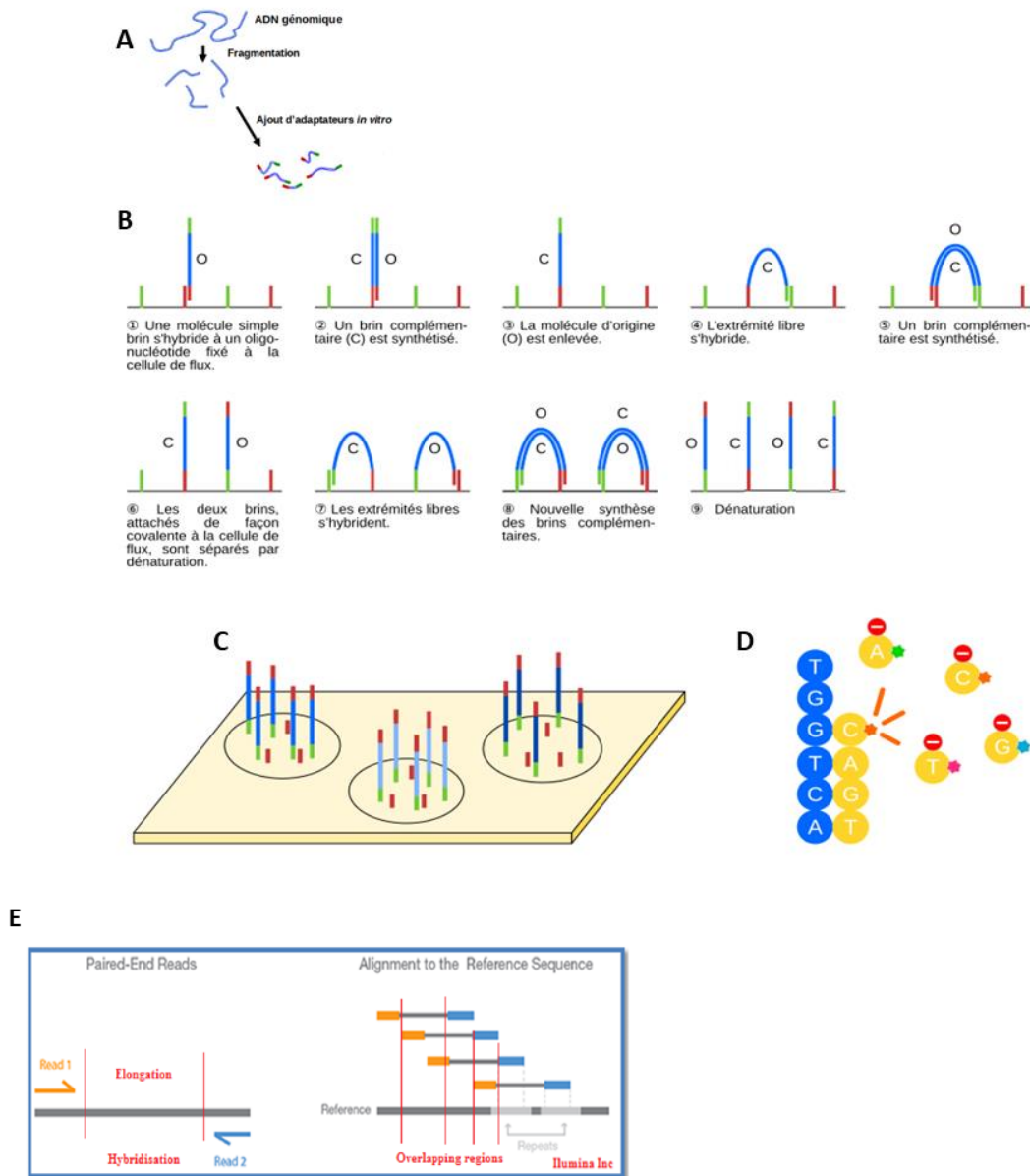


Figure 17. Le principe de la technologie de séquençage à haut débit Illumina. (A) Fragmentation de l'ADN et ajouts d'adaptateurs pour l'amplification en pont (B) représentation schématique du processus d'amplification en pont au sein d'un cluster de la cellule de flux ; (C) de nombreuses molécules d'ADN sont amplifiées puis séquencées simultanément sur les différents clusters d'une cellule de flux ; (D) réaction de séquençage : une polymérase synthétise un brin complémentaire au brin situé sur la cellule de flux en incorporant des nucléotides associés à des fluorochromes et des groupements bloquant de manière réversible la polymérisation, ainsi la réaction de séquence peut se poursuivre après lecture du nucléotide incorporé ; (E) alignement des fragments séquencés (read) sur le génome de référence, dans cet exemple, le séquençage Paired-End reads sequencing est représenté où seules les extrémités des reads sont séquencés.

Les variations génétiques détectées en WES

A ce stade du propos, il est nécessaire de détailler les différents types de variations génétiques détectés en séquençage d'exome. En propos liminaires, nous exposerons en Figure 18, la répartition des différents types de variations génétiques retrouvés en population générale dans l'étude GnomAD (Karczewski et al., 2020). Il faut garder à l'esprit que ces chiffres sont variables d'une étude à l'autre en fonction de la technologie utilisée, de la stringence des critères de qualité utilisés pour différencier artefacts et variations authentiques, de la profondeur de lecture et des populations étudiées.

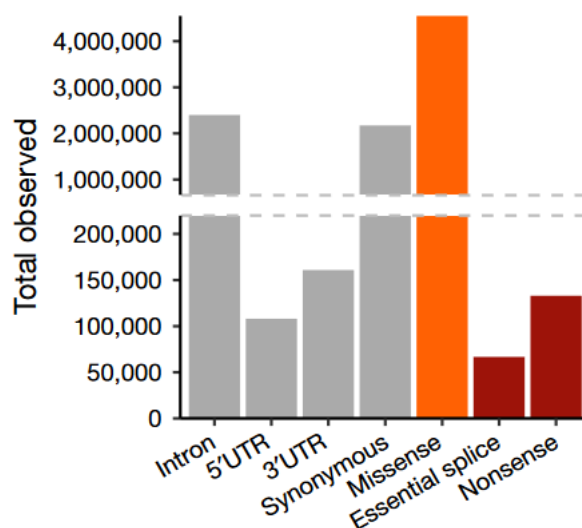


Figure 18. Répartition des variations détectées dans l'étude de Karczewski et collaborateurs (2020). Ces variations ont été détectées chez plus de 125 000 individus issus de la population générale et ayant bénéficié de WES. Les variations faux-sens sont les plus fréquemment détectées dans l'exome.

En effet, la probabilité *a priori* de pathogénicité d'une variation génétique n'est pas homogène. Elle varie en fonction de l'impact théorique

sur la structure protéique. Il convient d'insister sur le fait que, pour une variation génétique donnée, l'impact protéique prédit in-silico est très probable mais jamais certain. Seuls des tests fonctionnels bien menés permettent de le confirmer (Richards et al., 2015a). Les grandes catégories de variations génétiques, et leurs caractéristiques, sont les suivantes, sauf précisions ces informations sont issues de l'article exposant les critères de classification des variations génétiques pathogènes de l'American College of Medical Genetics and Genomics (ACMG), (Richards et al., 2015a) :

Les variations tronquantes ou PTVs (*protein truncating variant*).

Ces variations induisent un raccourcissement de la séquence protéique du fait de l'apparition prématurée d'un codon stop (TAA, TAG, TGA) soit par i) le remplacement direct du codon (variation non-sens) ou, ii) le décalage du cadre de lecture par une indel non multiple de trois ou enfin, iii) la modification des sites canoniques d'épissage entraînant un épissage aberrant et conduisant également à un décalage du cadre de lecture si l'exon n'est pas en phase. Dans ces deux derniers cas, la modification de l'ensemble du code génétique en 3' aboutit tôt ou tard à l'apparition d'un codon stop. Les transcrits ainsi tronqués sont potentiellement détruits avant leur traduction par la voie *nonsense-mediated RNA decay* (Lykke-Andersen & Jensen, 2015). Cette voie de dégradation des transcrits anormalement raccourcis peut cependant être contournée si la variation est située à l'extrémité 3' du gène et n'est pas forcément totale, si bien qu'il peut exister des protéines tronquées traduites (Lindeboom et al., 2019; Supek et al., 2021).

En pratique, les PTVs ont le même impact que les délétions de tout ou partie d'une des deux copies du gène en diminuant de moitié l'expression protéique dans le cas d'une variation hétérozygote, en la supprimant dans

le cas d'une variation homozygote. Ce mécanisme de perte de fonction par haploinsuffisance est fréquemment pathogène dans les syndromes neurodéveloppementaux que la transmission soit récessive, liée à l'X ou dominante. Dans ce dernier cas, il s'agit de gènes dont la régulation est critique pour le développement cérébral au point que la perte de fonction d'une des copies n'est pas tolérable. Ces gènes dits sensibles à l'haploinsuffisance ont donc la particularité d'être extrêmement déplétés en PTVs en population générale (Lek *et al.*, 2016). Le consortium gnomAD¹³ a mis au point une mesure de la sensibilité à l'haploinsuffisance de chaque gène en comparant le nombre attendu de PTVs (vu la taille et la séquence nucléotidique) au nombre observé (chez plus de 140 000 individus (Karczewski *et al.*, 2020). Pour définir les gènes qui subissent une pression de sélection importante à l'encontre des PTVs, l'équipe de gnomAD propose de considérer une borne supérieure de l'intervalle de confiance du ratio PTVs observées/attendues (LOEUF, *loss-of-function observed/expected upper bound fraction*) ≤ 0.35 . Les gènes causaux de maladies mendéliennes sont surreprésentés parmi ceux-ci (Figure 19).

¹³ <https://gnomad.broadinstitute.org/>

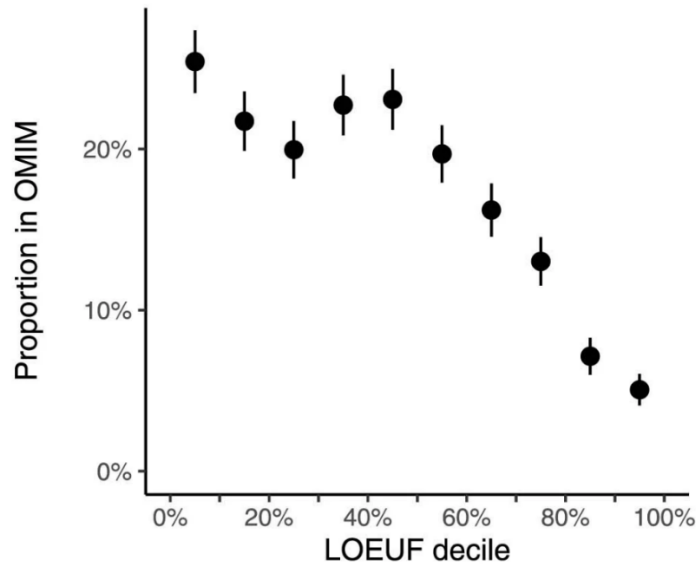


Figure 19. Proportion de gènes de la base de données OMIM (Online Mendelian Inheritance in Man) pour chaque décile de LOEUF. La base de données OMIM référence les gènes impliqués dans les syndromes génétiques mendéliens. L'implication fréquente des PTVs dans différents syndromes génétiques (neurodéveloppementaux ou non) avec la pression de sélection qui en résulte, explique leur rareté en population générale.

- **Les variations entraînant une modification de la taille de la protéine**

Les indels en phase (ajoutant ou retirant un nombre de nucléotides multiple de trois), la perte du codon stop terminal et les variations impactant l'épissage lorsque l'exon correspondant est en phase, modifient la taille finale du transcrit mature. La conséquence en termes de fonction protéique est variable et difficilement prévisible. Elle est plus souvent délétère lorsque ces variations sont situées au sein de domaines fonctionnels ou de zones particulièrement conservées entre les espèces.

- Les variations faux-sens

Les variations faux-sens entraînent la modification d'un acide aminé de la protéine par un autre. La majorité de ces variations sont bénignes mais certaines peuvent avoir un effet sur l'ARN ou sur la protéine comme sa déstabilisation, l'inactivation d'un domaine fonctionnel critique ou un gain de fonction aberrant. De nombreux logiciels de prédictions bioinformatiques proposent de quantifier l'ampleur du caractère délétère d'une variation faux-sens pour la fonction protéique, généralement en fonction de la proximité physico-chimique du nouvel acide aminé par rapport à l'ancien et du niveau de conservation phylogénétique de la position. En pratique, c'est surtout le niveau de congruence entre les différents logiciels qui est pris en compte pour aider à déterminer le potentiel caractère délétère pour la protéine d'une variation.

Au-delà des caractéristiques intrinsèques de la variation faux-sens, une tolérance à ce type de variation, tout comme pour les PTVs, est également calculée pour chaque gène dans la base de données gnomAD. Enfin, il convient de prendre en compte la fréquence des variations faux-sens pathogènes répertoriées dans le gène ainsi que la présence éventuelle de positions récurrentes ou de *hot-spots* mutationnels.

Enfin, les variations modifiant le site d'initiation de la traduction sont à part. Leur pathogénicité dépend en grande partie de la position de la méthionine suivante dans la séquence codante et de son contexte nucléotidique. Celle-ci pourra ainsi faire fonction ou non de site d'initiation alternatif.

- **Les variations synonymes**

Le code génétique est dégénéré, c'est-à-dire que plusieurs codons codent pour le même acide aminé. Cet état de fait, dont les causes et conséquences physiologiques ne sont que partiellement dévoilées (Chaney & Clark, 2015) entraîne la présence de variations génétiques qui ne modifient pas la séquence protéique. Bien que généralement silencieuses certaines d'entre elles peuvent être délétères, par exemple en impactant l'épissage. En dehors de tests fonctionnels il est rarement possible de conclure sur leur pathogénicité.

- **Les variations des séquences non codantes**

Enfin, le séquençage d'exome peut détecter les variations dans les régions UTRs (*UnTranslated Regions*, régions transcrites mais non traduites), en fonction du kit de capture utilisé, et dans les extrémités des introns. Ces variations peuvent être pathogènes, notamment lorsqu'elles affectent une séquence régulatrice ou encore une fois l'épissage. Certains logiciels de prédiction bioinformatiques proposent également un score pour ces variations non codantes (Kircher *et al.*, 2014) mais leur analyse *in-silico* est très difficile.

iii. Variations *de novo* et études de trios dans le TSA

En plus des avancées que nous avons citées, le séquençage de l'ensemble de l'exome d'un individu et de ses deux parents (formant un trio) a ouvert la voie à un nouveau paradigme pour l'identification de gènes impliqués dans le TSA. Ce paradigme est basé sur une première prémisse : bien que la quasi-totalité de la séquence génétique d'un individu soit issue de celle de ses parents, certaines positions ne sont pas héritées mais apparaissent *de novo*. A l'échelle de l'exome, un individu est en moyenne porteur d'environ 1,7 variations nucléotidiques *de novo* (Karczewski et al., 2020) en plus d'environ une variation structurale pour dix individus, essentiellement des CNVs (Belyeu *et al.*, 2021).

Ces variations génétiques sont typiquement issues d'erreurs de la machinerie de réplication de l'ADN au sein des gamètes. Il s'agit probablement d'une explication au risque plus élevé de TSA en fonction de l'âge des parents, que nous avons cité au chapitre précédent. En effet, on observe une augmentation du taux de variations génétiques *de novo* chez les enfants selon l'âge parental (Goldmann *et al.*, 2016). L'âge paternel présente un effet d'autant plus marqué que le taux de variations *de novo* dans les spermatoocytes augmente 4 fois plus vite avec l'âge que dans les ovocytes (Jónsson *et al.*, 2017), essentiellement du fait du nombre de mitoses des spermatogonies, se poursuivant tout au long de la vie chez les hommes. En revanche, une augmentation du risque d'aberration chromosomique *de novo* chez l'enfant en fonction de l'âge maternel (mais non paternel) (Wadhawan *et al.*, 2020) a été montré chez des patients présentant un retard de développement mais non répliqués chez les patients TSA (Belyeu *et al.*, 2021). En réalité, des données récentes évoquent une augmentation du risque sur l'allèle maternel ou paternel en

fonction du mécanisme qui entraîne le CNV. Ceux causés par recombinaison homologue non allélique seraient plus fréquents sur l'allèle maternel alors que les autres mécanismes seraient plus fréquents sur l'autre allèle (Fu et al., 2022).

De surcroît, ces variations génétiques non héritées peuvent également apparaître *in-utero* au stade de zygote (Derijck *et al.*, 2008) ou au stade post-zygotique (D'Gama, 2021). Dans ce second cas, elles ne seront détectables que dans certaines cellules en fonction de la position de la cellule présentant la mutation initiale au sein des voies de différenciations tissulaires. En l'état des connaissances, il n'est pas possible de quantifier la fréquence des variations *de novo* zygotiques dans le TSA. En revanche, il est établi que les variations post-zygotiques sont parfois impliquées dans le TSA (Freed & Pevsner, 2016; Lim *et al.*, 2017). Nous ne détaillerons pas la problématique des variations post-zygotiques, par définition présente en mosaïque chez l'individu, c'est-à-dire uniquement dans une certaine proportion de ses cellules ou tissus. Nous nous contenterons de signaler la nécessité d'une profondeur de lecture plus importante (D'Gama, 2021), pour les différencier d'artefacts, lorsqu'elles sont présentes dans une faible fraction des éléments figurés du sang. L'alternative pour les détecter est le séquençage d'ADN issu d'un autre tissu, comme un prélèvement salivaire (D'Gama, 2021). L'origine des variations *de novo* est résumé dans la Figure 20. Ces mutations peuvent apparaître à n'importe quelle position du génome selon un phénomène stochastique (mais statistiquement biaisé en faveur de *hot-spots* mutationnels (Michaelson et al., 2012)). Lorsqu'elles surviennent sur une position critique, elles peuvent donc naturellement représenter un facteur de risque génétique non hérité de TSA.

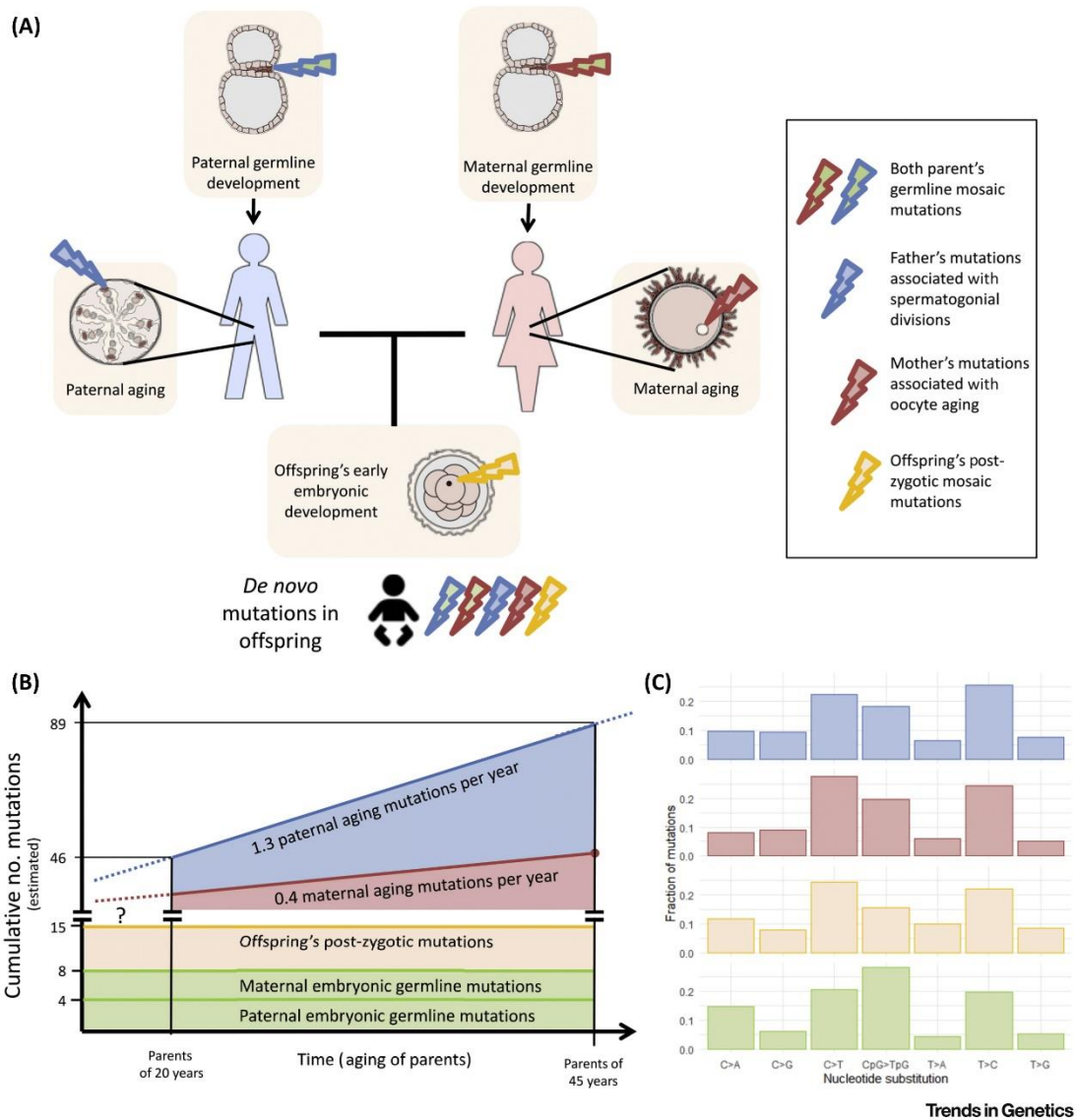


Figure 20. Source des variations de novo. (A) Une variation de novo peut survenir au sein des gamètes parentales, très précocement lors du développement embryologique ou plus tardivement lors du vieillissement. Elles peuvent également survenir au stade zygotique ou post-zygotique et affecter une proportion d'autant plus importante des cellules que l'apparition est précoce ; (B) la probabilité d'apparition d'une variation de novo dépend de l'âge des parents. L'augmentation est particulièrement marquée chez l'homme du fait de la division mitotique des spermatoocytes ; (C) le taux d'apparition de ces variations n'est pas homogène selon le type de substitution nucléotidique, il faut donc en tenir compte lorsque l'on estime le taux attendu de variation de novo d'un gène. D'après Goldmann et al., 2019.

La deuxième prémisse à ce paradigme, est l'observation d'une augmentation de la charge de certains types de variations génétiques *de novo* chez les patients présentant un TSA par rapport aux témoins et à leurs apparentés normotypiques. Cela concerne les CNVs (D. Levy *et al.*, 2011; Sebat *et al.*, 2007; T. N. Turner *et al.*, 2017; Yoon *et al.*, 2021) avec un ratio de 3.5-10:1, les PTVs (Iossifov *et al.*, 2014; Sanders *et al.*, 2012; T. N. Turner *et al.*, 2017; Yoon *et al.*, 2021) avec un ratio de 1.5-2:1, dans une moindre mesure les variations faux-sens (Iossifov *et al.*, 2014; Sanders *et al.*, 2012; T. N. Turner *et al.*, 2017; Yoon *et al.*, 2021) et plus récemment les variations non-codantes situées dans les domaines d'interaction tridimensionnelle de la chromatine, impliqués dans la régulation de l'expression (Kim *et al.*, 2022). Les auteurs ne retrouvent pas de différence significative concernant les variations synonymes ce qui est cohérent avec leur caractère plus souvent silencieux (Figure 21.A). Il est logique que ces variations *de novo* soient plus souvent détectées dans les formes sporadiques de TSA que dans les familles *multiplex* (Yoon *et al.*, 2021). En effet, dans ce deuxième cas, on s'attend à ce que les déterminants majeurs du TSA soient liés à des facteurs génétiques hérités (ou à la rigueur à l'environnement partagé). Au contraire, dans les cas d'autisme sporadique, une partie du déterminisme du trouble est portée par des variations *de novo* qui représentent, selon leur localisation, des facteurs de risque de TSA.

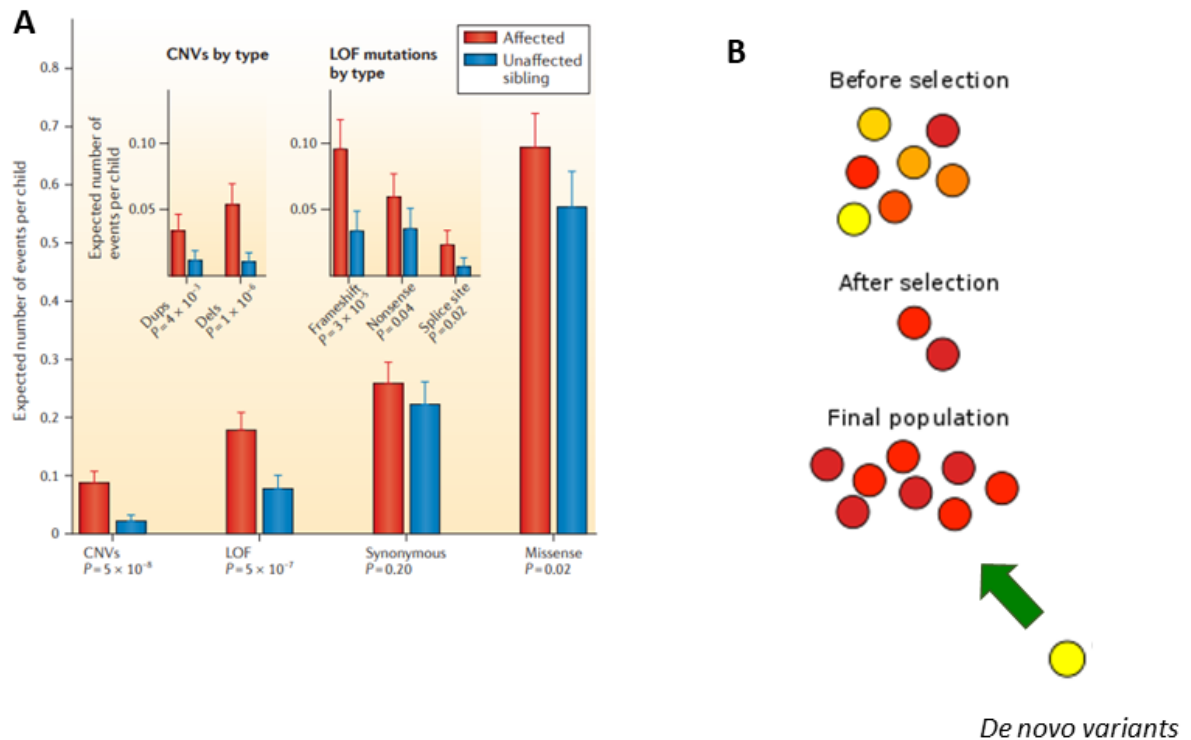


Figure 21. Variations de novo dans le TSA. (A) Détection différentielle de variations génétiques de novo entre les patients TSA et les contrôles. Concernant les variations nucléotidiques, l'étude de O'Roak et al., 2012 – qui ne retrouvait pas de différence significative – est incluse en plus des études de Sanders et al., 2012, et Iossifov et al., 2014. D'après Ronemus et al., 2014 ; (B) : schéma illustrant l'équilibre entre pression de sélection et apparition de variations de novo dans les TSA.

L'implication de variations génétiques *de novo* dans le déterminisme du TSA permet de résoudre, au moins partiellement, un paradoxe. Nous avons vu le rôle des facteurs de risque-génétiques dans le TSA, or ces patients présentent une fécondité effondrée en raison du handicap social qu'ils subissent, avec un ratio de 0,48, IC_{95%} : [0,44-0,51] ; $p < 10^{-10}$ par rapport

à la population générale, pour les femmes (Power *et al.*, 2013)¹⁴. Ainsi, la pression de sélection devrait entraîner la disparition rapide des facteurs de risque génétiques, que nous avons vu fortement impliqués dans le TSA, et donc entraîner une diminution de la prévalence de ce phénotype. Comme nous l'avons vu, ce n'est pas le cas, et c'est même le contraire qui est observé. Certains auteurs posent l'hypothèse qu'un flux régulier de ces variations *de novo* pathogènes viendrait compenser la diminution du taux de fécondité des patients TSA (Pugsley *et al.*, 2022) (Figure 21.B).

L'implication de ces variations non héritées est substantielle car les études les plus récentes estiment qu'elles contribuent au risque génétique de TSA chez 30 à 40% des individus et jusqu'à 50-65% dans les formes sporadiques (Yoon *et al.*, 2021). Comme l'étude de Howrigan et collaborateurs le montre (Figure 23), après la déficience intellectuelle, le TSA est la pathologie neuropsychiatrique la plus impactée par ces variations *de novo*. Le taux de PTVs et variations faux-sens *de novo* est nettement plus importante dans ce trouble que dans la schizophrénie, par exemple.

¹⁴ La fertilité déclarée est encore plus faible chez l'homme, mais la détermination pourrait être soumise à plus de biais de déclaration.

Disease	trios	DNM rate	enrichment	pval
Control	2182	0.09	1.06	0.378
Schizophrenia	2541	0.1	1.18	0.01
Autism Spectrum Disorder	3982	0.14	1.67	1.4e-29
Developmental Delay	4293	0.22	2.61	2.5e-144
Intellectual Disability	971	0.23	2.72	7.9e-38

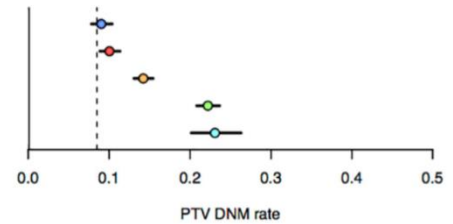


Figure 22. Comparatif du taux de variations de novo chez (de haut en bas sur le graphique) des sujets normotypiques, souffrants de schizophrénie, de TSA, d'un retard de développement ou d'une déficience intellectuelle. Bien qu'un déterminisme génétique soit également fort dans la schizophrénie (Hilker et al., 2018), l'impact des variations de novo est nettement plus faible que pour le TSA. D'après Howrigan et al., 2020.

De la sorte, puisqu'il est détecté un surcroît de variations génétiques *de novo* chez les patients TSA, dont une partie est pathogène, on peut en déduire que les gènes impliqués dans le risque génétique de TSA devraient être **plus souvent impactés par des variations de novo que ce qui est attendu par hasard.**

Les travaux de Samocha et collaborateurs ont permis l'élaboration d'un cadre statistique pour définir les gènes avec un excès de variations *de novo* (Samocha et al., 2014). Le principe est expliqué dans la Figure 23. Brièvement, la première étape a été de définir la probabilité d'apparition de chacun des grands types de variations nucléotidiques (variations synonymes, faux-sens, apparition de codons stops et variations entraînant un décalage du cadre de lecture et sites canoniques d'épissage). Le taux de mutation dépend du contexte nucléotidique, c'est-à-dire qu'il peut être calculé pour chacune des 64 combinaisons de trinuécléotides (Krawczak et al., 1998; Kryukov et al., 2007). *In fine*, celui-ci est déduit de la divergence entre le chimpanzé (allèles primordiaux) et l'homme (allèles mutés) sur des séquences orthologues entre les deux espèces.

A ce stade, le taux d'apparition de chaque type de mutation par génération est calculé pour chaque gène. Il dépend de sa taille et des trinuécléotides qui le composent. Enfin, chez un grand nombre de trio, le taux de variations *de novo* attendu est comparé à celui observé pour chaque gène. Les PTVs se prêtent plus particulièrement à une telle analyse. Elles ont un impact moins équivoque que les autres car elles entraînent théoriquement une dégradation du transcrit et donc une perte de fonction. De fait, toutes les pertes de fonction sur un même gène ont un effet globalement comparable. On en déduit que l'enrichissement en PTVs *de novo* d'un gène chez les patients implique que toutes les PTVs de ce gène sont facteurs de risque du trouble. Qu'elles soient héritées ou non. Comme attendu ces gènes sont généralement intolérants aux PTVs en population générale.

Ce paradigme a été utilisé par plusieurs équipes pour définir des listes de gènes pour lesquels la présence de PTVs est facteur de risque de TSA (C Yuen et al., 2017a; De Rubeis et al., 2014). Certains étaient déjà connus pour leur implication dans un trouble neurodéveloppemental avec surrisque de TSA, comme *SHANK3* (Durand *et al.*, 2007). D'autres étaient connus dans un autre trouble neurodéveloppemental et ces études ont permis de faire le lien avec le TSA. C'est le cas de *SCN2A* qui était à l'origine connu pour son implication dans plusieurs formes d'épilepsie syndromique (Wolff *et al.*, 2017). Mais une forte proportion de ces gènes n'était pas connue pour être impliqués en pathologie humaine avant ces études.

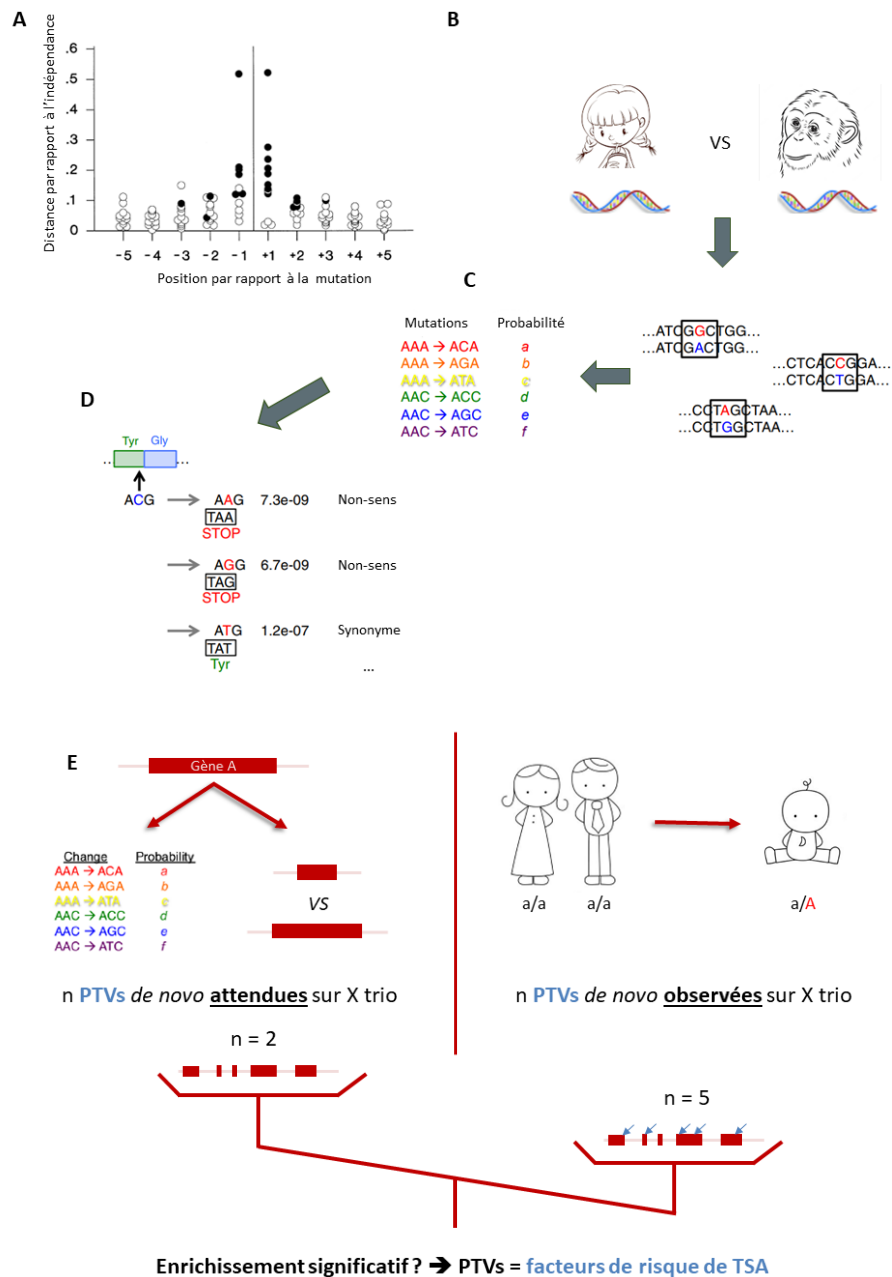


Figure 23. Modèle pour la détermination de gènes avec excès de variations de novo dans le TSA. (A) Sur une localisation donnée, seuls les nucléotides flanquants ont un impact significatif sur le taux de mutation. Traduit de Krawczak et al., 1998 ; (B-C) la probabilité de mutation de chaque trinucleotide est calculée à partir des différences de séquences apparues depuis la divergence entre le chimpanzé et l'homme ; (D) A partir de là, la probabilité d'apparition des différentes mutations dans chaque gène est calculée à partir de la taille de ceux-ci et de sa composition nucléotidique. (C-D) adapté de Samocho et al., 2014 (E) le taux de PTVs attendu et observé est comparé gène à gène en tenant compte des comparaisons multiples.

Le WES est donc une stratégie efficace pour le diagnostic moléculaire de patients présentant un TSA. Il permet la détection *a priori* de l'ensemble des variations nucléotidiques et des CNVs des séquences codantes des gènes. L'ensemble des facteurs de risque forts de TSA que nous avons cités¹⁵ sont donc accessibles au séquençage par cette technologie : syndromes mendéliens avec TSA dans le spectre syndromique, loci avec CNVs récurrents portant un risque de TSA et gènes avec excès de PTVs *de novo*.

Il est important d'insister encore une fois sur cette notion de facteur de risque, et tout particulièrement lors des rendus des résultats aux familles. Aucun de ces facteurs génétiques ne peut être considéré comme étant la cause pleine et entière de l'autisme chez un individu qui en est porteur. Des facteurs modificateurs influent de manière notable sur l'expression phénotypique.

Interaction entre variations rares et variations fréquentes

Ces dernières années, plusieurs équipes se sont attachées à résoudre la question du poids respectif et de l'interaction entre les variations génétiques rares et les polymorphismes que nous avons décrit au chapitre précédent. Notons que, quelle que soit la nature de leur interaction, la variance phénotypique du trouble à l'échelle de la population est affecté de

¹⁵Comme expliqué dans l'entrée *FMR1* de l'encyclopédie *GeneReviews* (<https://www.ncbi.nlm.nih.gov/books/NBK1384/>), les expansions de triplets impliqués dans le syndrome de l'X-fragile ne peuvent pas être détectés en WES. Une méthode ciblée est donc toujours nécessaire en systématique vue la prévalence relativement élevée de ce syndrome.

manière similaire par ces deux types de variations génétiques (Antaki *et al.*, 2022).

Selon certaines équipes, ils semblent agir selon un modèle additif et non multiplicatif (Klei *et al.*, 2021; Weiner *et al.*, 2017). Ces observations se fondent sur les observations suivantes. Dans l'étude de Weiner et collaborateurs, les auteurs étudient la sur-transmission de polymorphismes associés au TSA par leurs parents chez les patients par rapport au reste de la fratrie. Les patients porteurs de variations génétiques rares facteur de risque de TSA présentent également cette transmission déséquilibrée. Klei et collaborateurs calculent que les porteurs de variations génétiques rares pathogènes présentent un PRS en moyenne équidistant entre les autres patients TSA et les contrôles. Ces deux équipes concluent qu'un modèle additif explique mieux ces données. Le postulat implicite est, encore une fois, l'existence de cette variable *liability*, sur laquelle tous les facteurs de risque génétique TSA influeraient, avec un effet seuil pour le déclenchement de la maladie. L'observation que les variations rares et les polymorphismes impactent de manière similaire le « trait » autistique pourrait être en faveur de ce postulat (Robinson *et al.*, 2016).

Quoi qu'il en soit, ces deux classes de variations génétiques présenteraient un impact clinique dissemblable. En effet, le risque génétique médié par les polymorphismes agit spécifiquement sur la dyade autistique (Antaki *et al.*, 2022; Warrier *et al.*, 2019, 2022). Incidemment, la co-héritabilité des symptômes liés à la communication d'une part et au caractère restreint des activités d'autre part semble limitée (Warrier *et al.*, 2019), ce qui est cohérent avec le découpage de la clinique TSA en deux facteurs distincts (correspondant à la dyade autistique que nous avons déjà évoqué) (Frazier *et al.*, 2008; Guthrie *et al.*, 2013; Mandy & Skuse, 2008;

Shuster *et al.*, 2014). En revanche, le PRS de TSA ne semble pas impacter les fonctions cognitives des patients. Au contraire, chez ces derniers (Grove *et al.*, 2019) et dans la population générale (Clarke *et al.*, 2016), il est positivement corrélé au quotient intellectuel et au niveau scolaire atteint. (Figure 25.B.)

Concernant les variations rares, les études d'Antaki et collaborateurs et de Warriar et collaborateurs, toutes deux publiées cette année, ont calculé une charge en variations génétiques rares, que nous appellerons « critiques »¹⁶, chez les patients TSA et leurs apparentés (Antaki *et al.*, 2022; Warriar *et al.*, 2022). Ce calcul prend en compte la présence de variations *de novo* délétères sur la fonction protéique (PTVs et faux-sens prédits délétères par plusieurs logiciels bioinformatiques) dans les gènes sensibles à l'haploinsuffisance ou intolérants aux variations faux-sens et les PTVs rares hérités dans les gènes sensibles à l'haploinsuffisance. Par contraste avec l'effet des polymorphismes génétiques, cette charge en variations rares critiques ne semble pas être directement et indépendamment corrélée aux symptômes cardinaux de TSA. En revanche, elle est significativement négativement corrélée avec la coordination motrice, le quotient intellectuel (Fig 25.A.), les comportements adaptatifs et le développement des acquisitions précoces (âge des premiers mots et

¹⁶ Nous avons choisi ce terme pour éviter une confusion qui n'est pas réellement levée dans ces articles. Pour chaque variation génétique rares probablement délétère sur la fonction protéique, la probabilité d'un effet délétère cliniquement significatif est relativement forte mais pas certaine. De plus, l'effet spécifique sur le TSA (ou même n'importe quel phénotype neurocognitif) est inféré vu la surreprésentation de telles variations chez les patients mais se révélerait probablement fausse sur un certain nombre de ces variations.

de la marche). Toutes choses, par elles-mêmes, associées au risque de TSA. Des travaux précédents appuient indirectement cette démonstration d'un risque induit par les variations rares sur le TSA *via* leur impact sur les fonctions cognitives (Bishop *et al.*, 2017; Buja *et al.*, 2018; Iossifov *et al.*, 2014; O'Roak *et al.*, 2012).

Des gènes spécifiques de TSA ?

A ce jour, il n'existe pas de gène connu pour impacter uniquement le risque de TSA (Myers *et al.*, 2020), puisque ces gènes associent toujours un effet négatif sur le développement des fonctions cognitives. Mais, tout de même, il faut noter que le ratio TSA avec déficience sur TSA sans déficience varie selon ces gènes de *developmental disorder*. Par exemple, dans le cas des gènes *CHD8*, *SNC2A* ou *NRXN1* la moitié des patients TSA répertoriés dans la littérature ne présentent pas de déficience intellectuelle (Myers *et al.*, 2020). De la sorte, il est possible que certaines variations génétiques rares impactent directement le risque de TSA en plus d'un effet *via* les fonctions cognitives. Pour corroborer cette hypothèse, une étude récente montre que dans une population de patients TSA et de patients présentant divers autres troubles neurodéveloppementaux, certains gènes sont plus spécifiquement impactés chez les patients TSA (Figure 24).

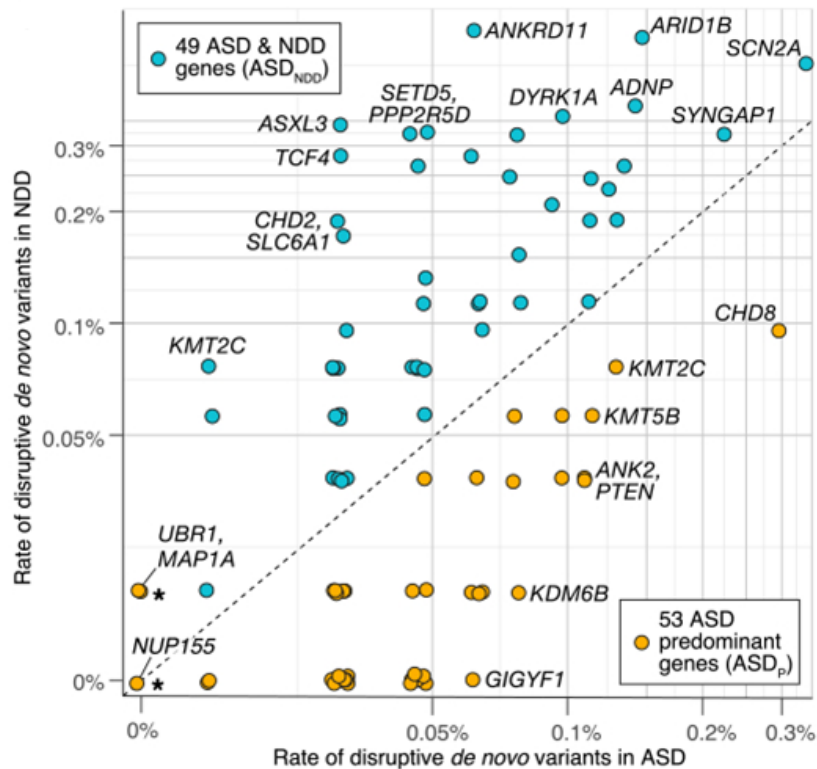


Figure 24. Taux comparé de variations de novo délétères chez les patients présentant un trouble neurodéveloppemental en ordonné et les patients présentant un TSA en abscisse. Certains gènes impactés par ces variations délétères sur la fonction protéique semblent plus spécifiques du TSA que d'autres. Les trois gènes marqués d'une étoile sont à prédominance TSA du fait de données cas-contrôles non représentées ici. D'après Satterstrom et al., 2020.

Conclusion

Pour conclure, une image encore très fragmentaire se dessine quant aux facteurs de risque génétiques de TSA. Comme nous l'avons vu, ils associent :

- Des polymorphismes génétiques avec un effet spécifique sur la dyade autistique et avec un effet plus neutre sur les fonctions cognitives.

Leur effet est trop faible pour que les considérer à l'échelle individuelle.

- Des variations génétiques rares avec un impact direct sur les fonctions cognitives dont la dégradation est elle-même facteur de risque de TSA.

Le modèle le plus courant considère que leur addition explique la distribution des individus sur le trait autistique. Passé un certain poids génétique, le seuil diagnostique serait atteint et le patient présenterait des symptômes suffisamment importants pour qu'un diagnostic de TSA soit posé. Des facteurs environnementaux, essentiellement toxiques au stade *in-utero*, jouent également sur le risque de TSA. La nature de l'interaction gène-environnement mériterait d'être précisée. Nous avons relevé les failles de ce modèle qu'il convient de considérer avec précaution.

Concernant les variations génétiques rares, objet de ce travail, certaines précisions doivent être apportées. Dans les travaux cités dans ce chapitre, elles ont été étudiées à l'échelle de la population permettant de définir leur impact moyen. En réalité cette charge en variations génétiques rares « critiques » agrège d'authentiques variations pathogènes, portant un risque très variable, et des variations qui ne le sont pas (ou en tous cas, pas pour le TSA).

Cette modélisation n'est pas pertinente à l'échelle individuelle où des critères explicites définissant la pathogénicité doivent être utilisés. L'expression singulière de ces facteurs génétiques d'effet fort dépend du gène en question et des facteurs modificateurs génétiques et environnementaux. Un exemple théorique est illustré en fin de ce chapitre (Figure 26).

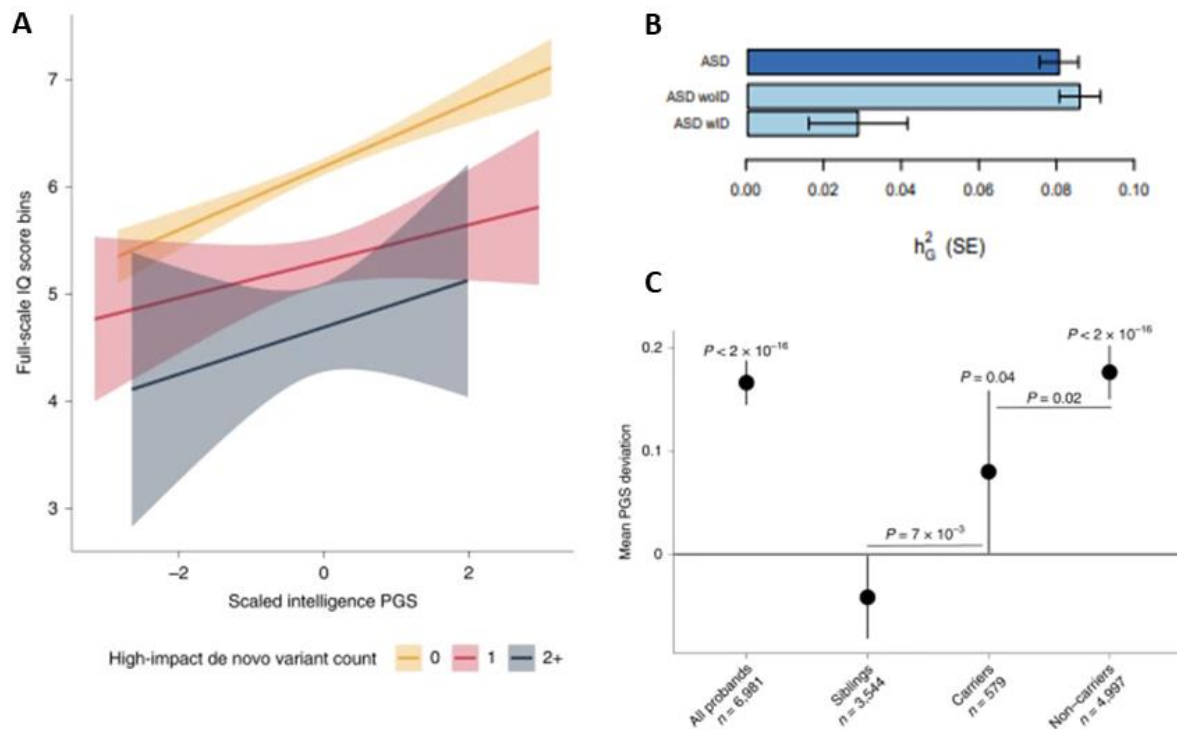


Figure 25. Intégration des différents facteurs de risque génétiques. (A) QI en fonction du PRS pour l'intelligence chez les patients présentant 0, 1 ou 2+ variations génétiques rares critiques, d'après Antaki *et al.*, 2022 ; (B) Héritabilité du TSA expliquée par les polymorphismes génétiques chez l'ensemble des patients et ceux sans ou avec déficience intellectuelle, d'après Grove *et al.*, 2019 ; (C) biais de transmission des polymorphismes génétiques facteur de risque de TSA chez l'ensemble des patients, leur fratrie normotypique, les patients porteurs de variations génétiques rares critiques et les non-porteurs, d'après Antaki *et al.*, 2022.

Aujourd'hui plusieurs défis se révèlent en pratique clinique pour l'utilisation de ces variants rares chez patients TSA dans le cadre d'une consultation génétique. Premièrement, contrairement à la situation conceptuellement plus simple des syndromes neurodéveloppementaux d'origine génétique, il n'y a pas de recommandation claire pour la classification des variations génétiques. Par exemple, les *guidelines* de

l'ACMG-AMP (Richards et al., 2015a) n'ont pas été conçues pour des facteurs de risque. Deuxièmement, il n'y a pas eu de travail d'unification des gènes impliqués dans le TSA dans une démarche diagnostique. La base de données du SFARI¹⁷, que nous avons déjà citée et qui regroupe 1295 gènes, est très hétérogène dans sa conception et difficile à utiliser en pratique. Elle intègre les données issues des études GWAS, de transcriptomique ou des modèles animaux. Nous avons déjà expliqué pourquoi les études GWAS ne peuvent être utilisées dans une démarche diagnostique. Concernant les deux derniers champs d'étude, ils sont pertinents pour comprendre la physiopathologie du TSA, mais ne se traduisent pas en des données utiles en diagnostic pour le moment.

¹⁷ <https://gene.sfari.org/>

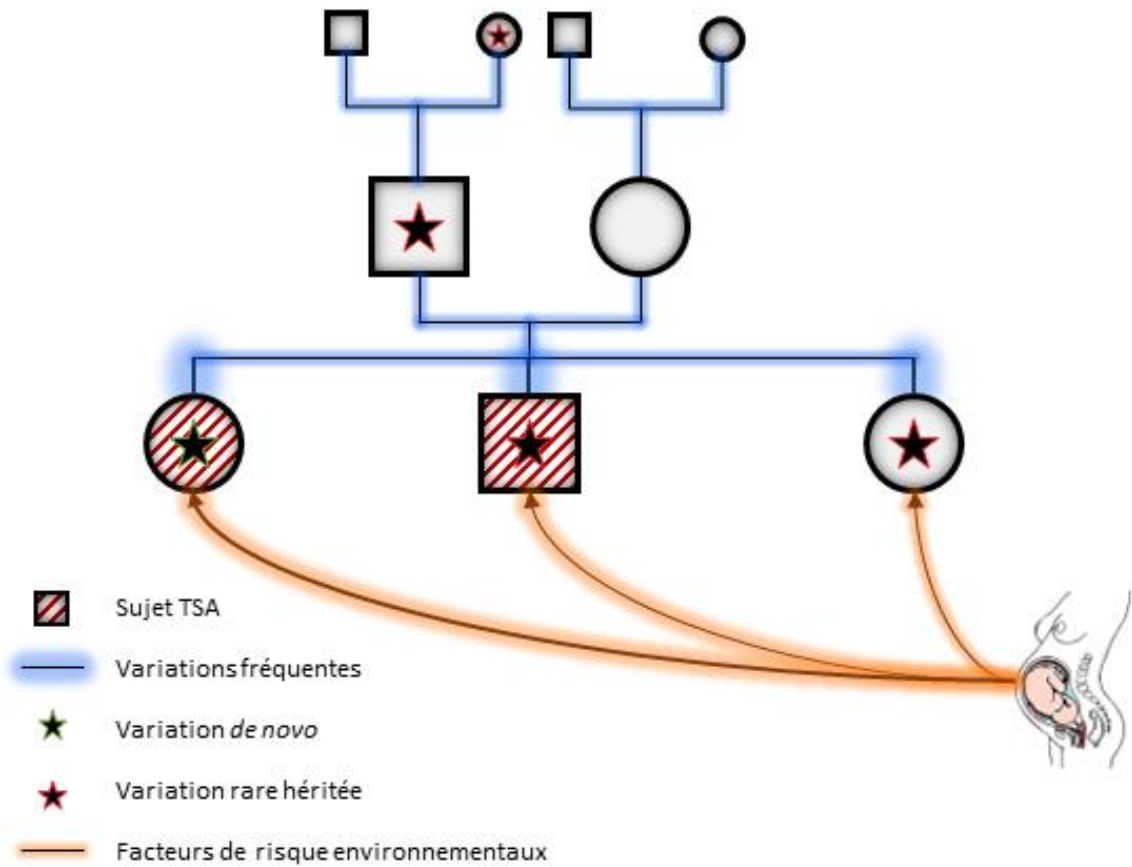


Figure 26. Intégration des différents facteurs de risque de TSA au sein d'un pédigré. La survenue d'un TSA chez un sujet dépend de la transmission d'une charge en variation fréquente, d'une ou plusieurs variations rares pathogènes, de la survenue de variation de novo et des facteurs environnementaux essentiellement in-utero. Lorsque la combinaison de ces facteurs atteint un seuil suffisant (qui semble différent entre les sujets féminins et masculins), le syndrome survient.

Chapitre II. Taux de détection de facteurs de risque de TSA

i. Objectifs et résumé du travail publié

Comme nous l'avons vu, les variations génétiques rares font parties des facteurs causaux majeurs de TSA. De plus, il s'agit des seuls qui portent un risque individuel suffisant pour être rendus aux patients dans une démarche diagnostique.

La littérature manque de critères opérants pour systématiser l'étude génétique dans le TSA. De ce fait, le rendement rapporté du séquençage d'exome était difficile à évaluer chez ces patients. Cet écueil est d'autant plus criant que les populations incluses varient dans leur sévérité alors que les listes de gènes étudiées n'étaient pas fondées sur des critères clairs.

Nous avons donc constitué en 2019 une liste de gènes la plus complète possible eu égard à l'état de la littérature de l'époque, proposé des recommandations concernant l'évaluation de la pathogénicité des variations génétiques de ces gènes et appliqué celles-ci sur un échantillon de patients recrutés dans un service de génétique clinique.

Nous avons constitué la liste de gènes comme suit :

- i) Gènes impliqués dans les syndromes monogéniques – où le TSA fait partie du spectre phénotypique. Cette liste a été conçue à partir de la base de données SFARI (Abrahams *et al.*, 2013). Les travaux de cette équipe visent à répertorier tous les gènes associés aux TSA dans la littérature scientifique. Nous avons revu individuellement chaque gène et conservé ceux pour lesquels le

niveau de preuve et la taille d'effet permettait un rendu diagnostique. En pratique, les études *GWAS*, de transcriptomique, de modèles cellulaires ou animaux n'ont pas été considérées car non probantes pour un rendu diagnostique. Pour évaluer la pathogénicité des variations génétiques de ces gènes nous avons considéré les recommandations de l'ACMG-AMP (Richards et al., 2015a) afin de ne conserver que les variations considérées comme pathogènes ou probablement pathogènes (respectivement classe 4 ou 5).

- ii) Les loci où des CNV récurrents associés aux TSA sont décrits ont été également analysés. Nous nous sommes basés majoritairement sur l'étude de Girirajan et collaborateurs que nous avons déjà citée (Girirajan *et al.*, 2013), complété par les travaux de Yuen et collaborateurs (C Yuen et al., 2017a) et par la base de données du SFARI.
- iii) Enfin, nous avons défini une liste originale de gènes sensibles à l'haploinsuffisance, présentant un excès statistiquement significatif de PTVs *de novo* dans les TSA. Nous avons réanalysé les données publiées de cinq cohortes (C Yuen et al., 2017a; De Rubeis et al., 2014; Dong et al., 2014; Iossifov et al., 2014; Krumm et al., 2015; O'Roak et al., 2012; Stessman et al., 2017; T. Wang et al., 2016) totalisant plus de 5000 trios, en prenant garde à éliminer les doublons. Au total, 79 gènes ont été classés en deux groupes en fonction de leur degré de significativité : « impliqués » dans le TSA (FDR (*False Discovery Rate*, taux de fausses découvertes) < 5%) et « probablement impliqués » dans le TSA (5% < FDR < 10%).

L'équipe de génétique clinique du CHU de Rouen a recruté plus de 600 patients diagnostiqués TSA entre 2009 et 2017, dans le cadre d'une étude promue par le Centre Hospitalier du Rouvray. Ceux-ci y étaient adressés par les psychiatres ayant effectué le diagnostic, comme recommandé par la Haute Autorité de Santé. Une évaluation clinique a été effectuée par les cliniciens de l'unité avant inclusion dans l'étude. Nous avons sélectionné 253 de ces patients pour un séquençage d'exome.

L'ADN des patients a été séquencé par le Centre National de Recherche en Génomique Humaine (CNRGH, Evry) puis la détection et l'interprétation des variations ont été effectuées au sein de l'unité Inserm 1245. Enfin, les données biologiques ont été confrontés à l'évaluation du clinicien. Pour les résultats « positifs », c'est-à-dire permettant d'identifier un variant ayant probablement contribué au TSA au minimum, un compte-rendu des analyses effectuées au titre de la recherche, insistant sur la notion de facteur de risque, a été envoyé aux cliniciens investigateurs. Ceux-ci ont effectué le rendu des résultats auprès du patient et/ou de sa famille. Le diagramme de flux de l'étude est représenté à la suite.

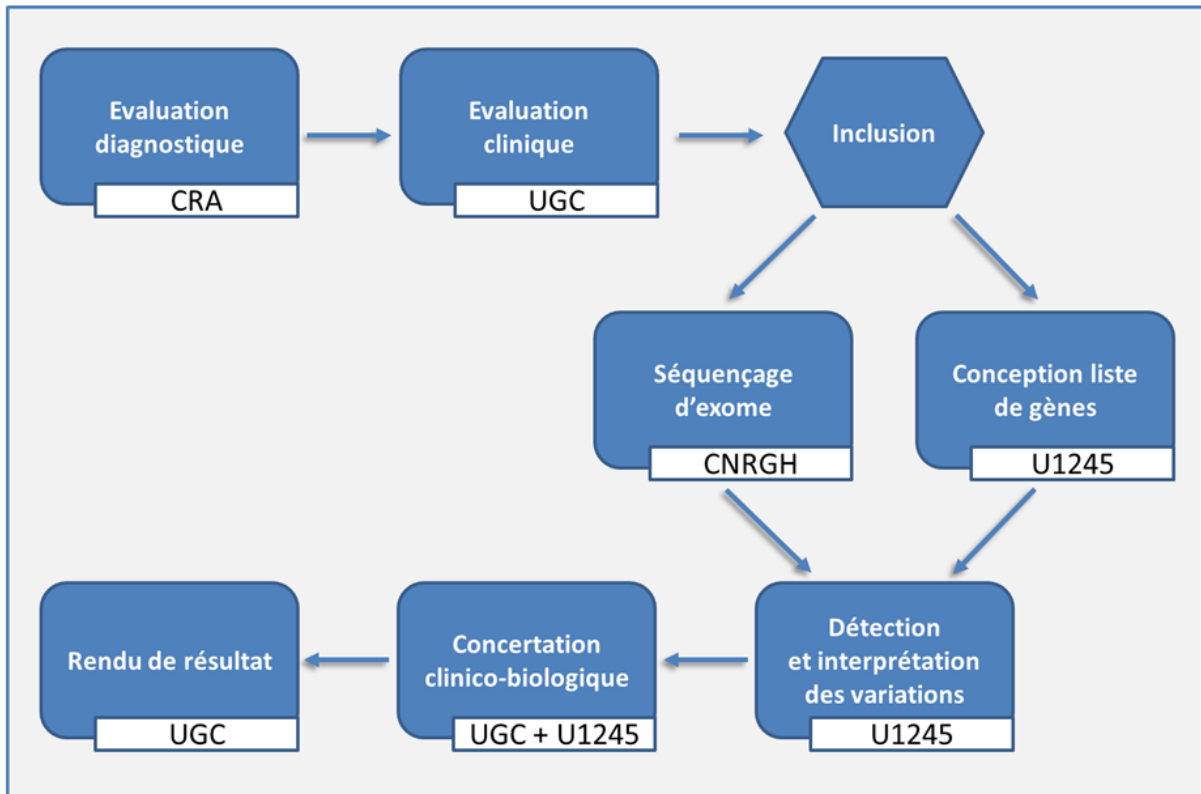


Figure 27. Diagramme de flux de l'étude. UGC : Unité de Génétique Clinique. Seuls les résultats positifs ont été rendus.

Sur le plan clinique, 35% des patients présentaient un diagnostic de syndrome d'Asperger (les critères du DSM-IV-TR étaient encore en pratique) et, à l'autre extrême du spectre, 27% une déficience intellectuelle d'intensité variable.

Brièvement, 19,7%, IC95% : [15%-25,2%] des patients présentaient une variation génétique facteur de risque majeur de TSA. 7,1%, IC95% : [4,3%-11%] présentaient un CNV impliqué dans les TSA alors que pour 12,6%, IC95% : [8,8%-17,4%] une variation nucléotidique avait été détectée. Comme attendu, les patients avec déficience intellectuelle présentaient un taux de détection particulièrement élevé de 30,1%, IC95%

: [20,2%-43,2%]. En plus de ces résultats, nous avons participé à la description d'un nouveau syndrome neurodéveloppemental lié à une altération du gène *CDC42BPB*. Cette collaboration a eu lieu par l'intermédiaire du site *genematcher*¹⁸ (Chilton *et al.*, 2020) mais n'était pas comptée dans le rendement diagnostique de l'article princeps.

Ce travail a été valorisé par une publication dans le journal *Translational Psychiatry*, jointe au manuscrit. Mes contributions à ce travail ont été les suivantes :

- Mise à jour, extension et maintien de la base de données clinique des patients présentant un TSA reçus dans le service de génétique.
- Recrutement de patients lors de mon stage d'internat dans le service de génétique clinique du CHU de Rouen.
- Participation à la revue des gènes répertoriés dans la base de données SFARI et au répertoriage des cohortes de trio – notamment recherche de doublons pour ces cohortes utilisées dans plusieurs articles.
- Analyse des résultats de séquençage d'exome et coordination des réunions de concertation clinico-biologique.
- Ecriture et soumission de l'article.

¹⁸ <https://genematcher.org/statistics/>

ii. Rare genetic susceptibility variants assessment in autism spectrum disorder: detection rate and practical use¹⁹

¹⁹ <https://doi.org/10.1038/s41398-020-0760-7>

ARTICLE

Open Access

Rare genetic susceptibility variants assessment in autism spectrum disorder: detection rate and practical use

Thomas Husson^{1,2}, François Lecoquierre¹, Kevin Cassinari¹, Camille Charbonnier¹, Olivier Quenez¹, Alice Goldenberg¹, Anne-Marie Guerrot¹, Anne-Claire Richard¹, Valérie Drouin-Garraud¹, Anne-Claire Brehin¹, Maryam Soleimani¹, Romain Taton³, Maud Rotharmel², Antoine Rosier³, Pascal Chambon¹, Nathalie Le Meur¹, Géraldine Joly-Helas¹, Pascale Saugier-Veber¹, Anne Boland⁴, Jean-François Deleuze⁴, Robert Olaso⁴, Thierry Frebourg¹, Gael Nicolas¹, Olivier Guillin^{1,2} and Dominique Champion^{1,2}

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder with a strong genetic component whose knowledge evolves quickly. Next-generation sequencing is the only effective technology to deal with the high genetic heterogeneity of ASD in a clinical setting. However, rigorous criteria to classify rare genetic variants conferring ASD susceptibility are currently lacking. We have performed whole-exome sequencing to identify both nucleotide variants and copy number variants (CNVs) in 253 ASD patients, including 68 patients with intellectual disability (ID) and 90 diagnosed as Asperger syndrome. Using explicit criteria to classify both susceptibility genes and susceptibility variants we prioritized 217 genes belonging to the following categories: syndromic genes, genes with an excess of de novo protein truncating variants and genes targeted by rare CNVs. We obtained a susceptibility variant detection rate of 19.7% (95% CI: [15–25.2%]). The rate for CNVs was 7.1% (95% CI: [4.3–11%]) and 12.6% (95% CI: [8.8–17.4%]) for nucleotide variants. The highest rate (30.1%, 95% CI: [20.2–43.2%]) was obtained in the ASD + ID subgroup. A strong contributor for at risk nucleotide variants was the recently identified set of genes ($n = 81$) harboring an excess of de novo protein truncating variants. Since there is currently no evidence that the genes targeted here are *necessary and sufficient* to cause ASD, we recommend to avoid the term “causative of ASD” when delivering the information about a variant to a family and to use instead the term “genetic susceptibility factor contributing to ASD”.

Introduction

Autism spectrum disorder (ASD), whose prevalence is estimated to be around 1 per 100 children, encompass a wide range of phenotypic manifestations ranging from severe behavioral impairment, often associated with

intellectual disability (ID), to mild difficulties in social interaction. The fact that, clinically, autism is not a single disease entity is mirrored by the extreme genetic heterogeneity underlying this condition. Chromosomal microarray analysis is usually used as the first-tier genetic test for individuals with autism with a diagnostic yield of 7–9%^{1,2}. However, such analysis cannot detect single nucleotide variations and small Indels but only loss or gain of genomic DNA material. In light of recent advances in ASD genetics, especially regarding the importance of de novo truncating variants, it is estimated that rare variants within several hundreds of genes may contribute to ASD

Correspondence: Dominique Champion (Dominique.champion@univ-rouen.fr)

¹Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics and Reference Center for Developmental Disorders, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France

²Department of Research, Centre hospitalier du Rouvray, Sotteville-Lès-Rouen, France

Full list of author information is available at the end of the article.

© The Author(s) 2020



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

susceptibility, each of them being present only in a very small proportion of cases³. In this context, next-generation sequencing is the only reasonable and cost-effective approach to search for variants in these genes in a diagnostic perspective.

The aims of the present study were threefold: (i) to establish an accurate list of susceptibility genes from literature data and to provide guidelines to categorize rare susceptibility variants (ii) based on these criteria, to estimate the whole exome sequencing (WES) detection rate of rare susceptibility variants in a sample of ASD patients typical of those attending genetic consultations and (iii) to propose some recommendations for susceptibility factors assessment and interpretation in clinical practice.

Methods

Patient recruitment

During the 2009–2017 period, 679 unrelated subjects received a diagnosis of ASD or Asperger syndrome by clinicians of our local expert center. This study was carried out on a subset of 253 cases, including all patients from whom parental DNA was available ($n = 159$) and 94 randomly selected patients.

Clinical evaluation was mainly based on ADOS-2. In addition, ADI-R and CARS, were also used for 76 and 145 patients, respectively. All diagnoses were made according to DSM-IV-TR criteria. Accordingly, the diagnosis of Asperger syndrome included the absence of language impairment, as documented during clinical examination. IQ was assessed in subjects with age >5 years with the Wechsler Intelligence Scale for Children (WISC) or the Raven Progressive matrix. PsychoEducational Profil third edition (PEP-3) and Vineland Adaptive Behavior Scales (VABS-2) assessment, when available, were used to refine the classification of cases with heterogeneous cognitive functions or mild cognitive impairment. Subjects with age <5 years were classified in the ASD subgroup, without further specification.

Following national recommendations, diagnostic assessment included a genetic consultation. Every patient was received by a senior clinical geneticist for thorough questioning about development, previous medical records and family history, followed by a complete physical examination.

This genetic study was approved by our legal ethics committee. Parents or legal representative of the children signed an informed consent for genetic analyses.

Molecular genetics

All included patients had previously been negatively screened for Fragile X expansion and 54 had been tested by array Comparative Genomic Hybridization (aCGH) Agilent 180K (Santa Clara, United-States), yielding to the prior identification (subsequently confirmed by exome

sequencing) of four CNVs conferring susceptibility for ASD. For each proband, DNA was extracted from fresh blood samples. DNA was also obtained from affected siblings and from parents when available.

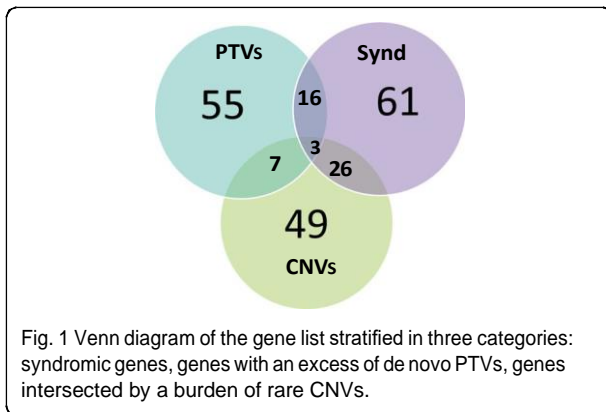
Exome sequencing was performed on every proband. Exomes were captured using the Agilent SureSelect All Exons V5-UTR or V6 Kit (Santa Clara, United-States). Sequencing was performed on an Illumina HiSeq4000 (Illumina, San Diego, CA, USA) at the CNRGH (*Centre National de Recherche en Génomique Humaine, Evry, France*) with paired end mode, 150 base pairs (bp) reads. Nucleotide variants and copy number variants (CNVs) were analyzed through a bioinformatics pipeline to perform variant calling, quality check, annotation and CNVs detection (see Sup Info). When possible, the segregation of retained variants was examined by Sanger sequencing for nucleotide variants and by ddPCR or QMPSF for CNVs. Parenthood was checked by poly-morphic microsatellites for patients carrying de novo mutations.

List of genes and interpretation of variants

To establish a panel of genes firmly contributing to ASD susceptibility, we retained three sets of genes:

(i) genes showing a statistically significant excess of de novo protein truncating variants (PTVs), (*i.e.* nonsense, canonical splice site variants and frameshift indels) in large published cohorts. Since these genes have been identified in several, sometimes redundant recent trio studies^{1,4–10}, we reanalyzed (see Sup Info) all de novo events reported in each of these studies (Sup Table 1). Depending on the level of statistical evidence for the enrichment in de novo PTVs, we ranked these genes into two categories: class A (false discovery rate (FDR) <5%) meaning definitely involved and class B ($5\% \leq \text{FDR} \leq 10\%$), meaning probably involved in ASD susceptibility. This led to a list of 81 genes (Sup Table 2). For these genes, we retained as susceptibility factor all rare PTVs found in patients, whatever the inheritance (*i.e.* de novo occurrence or not).

(ii) syndromic ASD-related genes, *i.e.* genes causing monogenic diseases for which ASD can be part of the phenotype ($n = 107$). This set of genes was carefully curated from the Simon Foundation Autism Research Initiative (Sfari) database and from literature data. For each syndromic gene, the mode of inheritance was specified and rare non-synonymous variants were prioritized. For these genes, pathogenicity of rare variants ($\text{MAF} < 1\%$ in GnomAD) was evaluated using the guidelines from the American College of Medical Genetics-Association for Molecular Pathology (ACMG-AMP). Only variants considered as likely pathogenic or pathogenic (class 4 or 5) were retained as advised by the guidelines. For our analysis, we considered these variants as probable (class 4) or



definite (class 5) ASD-susceptibility variants and then evaluated whether patients carrying them harbored or not the full syndromic presentation typical of each gene.

(i) genes intersected by a burden of rare CNVs, listed in the Sfari database ($n = 85$).

Combining these three sets yielded a high confidence list of 217 ASD susceptibility genes (Sup Table 3). Note that some of these genes were simultaneously present in several sets. (Fig. 1).

Besides this main panel and despite the fact that the interpretation was more problematic, we also examined rare missense variants in genes carrying a burden of de novo and transmitted mutations identified by TADA analyses^{4,11} as well as in genes recently reported as carrying an excess of de novo missense mutations in ASD and developmental disorders (DD)^{1,12}. This analysis was conducted on 89 genes (61 already included in the main list and 28 additional genes). For this analysis, only de novo missense variants predicted as damaging by three in silico prediction tools were considered.

Finally, for patients with ID or additional sensory impairment, we examined all OMIM referenced genes.

While CNV detection was performed on all genes of the main list, variant interpretation depended on the gene subsets. We retained all previously associated CNVs (i.e. duplications or deletions) within the set of large recurrent CNVs associated with various DDs including ASD¹³. In contrast, we used a conservative approach to interpret the remaining CNVs, and only retained rare deletions targeting Loss of Function (LoF) intolerant genes as probable ASD susceptibility variants.

Results

Among the 253 ASD probands (81.4% males), 68 were categorized in the ASD+ID subgroup while 90 were classified as having Asperger syndrome. The average age at inclusion was 8.5 ± 5.6 years (range 3–41). Positive family history (among first, second or third-degree relatives) for ASD, DD or ID was reported for 27.7% of

Table 1 Demographic and phenotypic characteristics of the sample.

	% (N/253)
Males	81.4% (206)
Family history	27.7% (70)
Consanguinity	1.6 % (4)
Asperger syndrome	35.6% (90)
Intellectual disability	26.9% (68)
Dysmorphic features	26.5% (67)
Premature birth	3.95% (10)
Microcephaly	3.7% (9)
Macrocephaly	7.9% (20)
Epilepsy	7.1% (18)

probands. Demographic and phenotypic characteristics of the sample are displayed in Table 1.

As shown in Table 2, ASD-contributing CNVs were present in 18/253 cases (7.1%, 95% CI: [4.3–11%]) while ASD-contributing nucleotide variants (Table 3) were detected in 32/253 subjects (12.6%, 95% CI: [8.8–17.4%]), yielding to a global detection rate of 50/253 (19.7%, 95% CI: [15–25.2%]). Of note, no patient carried multiple susceptibility variants. A slightly higher detection rate (29.8%, 95% CI: [17.3–44.9%]) was found in females. When the sample was split into three clinical subgroups, the rate of CNVs and nucleotide susceptibility-variants was particularly high in the ASD+ID subgroup (30.1%, 95% CI: [20.2–43.2%]) (Table 4). In the whole sample, two CNVs (*NRXN1* E1–3 del and 15q11 BP 1–2 del) were recurrent and four genes were found to carry nucleotide variants more than once: *CHD8*, *SHANK3*, *NRXN1*, and *KMT2A*. Regarding nucleotide susceptibility variants, the contribution of the set of genes previously shown as enriched in de novo PTVs in large ASD cohorts should be highlighted. Ten genes carrying ASD-susceptibility variants belonged to this category vs seven syndromic genes and nine genes belonging simultaneously to the two categories. Overall, the high frequency of de novo events is noteworthy: we identified 21/28 de novo nucleotide variants and 6/14 de novo CNVs among 42 patients from whom parental DNA was available. Regarding the mode of inheritance of ASD contributing variants, no autosomal recessive inheritance was encountered in this mainly outbred population.

Finally, variants in two genes not included in our list, but which cause nevertheless part of the phenotype, were found in two patients presenting ASD, ID, and sensory deficits. A patient with hearing impairment carried compound heterozygous *c.2485C > T, p.Gln829 / c.5665T > C*,

Table 2 CNVs conferring susceptibility for autism identified among the 253 ASD subjects.

ID	Chr	Start (hg19)	End (hg19)	Size (bp)	Type	Syndrome/gene	Inheritance	Diag	Age	Sex	Family history
<i>Pathogenic CNVs</i>											
266	15	22833359	23330473	497114	Del	15q11 del syndrome (BP1-BP2)	Maternal	Asperger	10	M	S
175	15	22833359	23330473	497114	Del	15q11 del syndrome (BP1-BP2)	Maternal	ASD	6	M	S
298	15	22833359	23330473	497114	Del	15q11 del syndrome (BP1-BP2)	NA	Asperger	13	F	S
277	15	22833359	23330473	497114	Del	15q11 del syndrome (BP1-BP2)	NA	Asperger	16	M	Fa
207	15	23686022	28567298	4881276	Dup	15q11.2q13 dup syndrome (BP2-BP3)	DN	ASD	9	M	Fa
405	15	23686022	30114707	6428685	Dup	15q11.2q13 dup syndrome (BP2-BP4) ^c	DN	ASD	5	F	S
225	16	14968893	16267261	1298368	Del	16p13 del syndrome	Maternal	ASD+ID	5	F	S
106	16	14968893	16267261	1298368	Dup	16p13 dup syndrome	Paternal	ASD+ID	11	F	S
40	16	29674271	30198544	524273	Dup	16p11.2 dup syndrome	DN	Asperger	11	M	Fa
392	16	29674271	30198544	524273	Del	16p11.2 del syndrome	DN	ASD+ID	7	M	S
288	22	51123001	51220846	97845	Del	SHANK3 (E9–22)+2 other genes	DN	ASD+ID	32	M	S
394	X	144899347	148678216	3778869	Del	Del Xq27.3-q28 (15 genes including FMR1 and AFF2) ^c	NA	ASD+ID	24	F	Fa
52	1	146500897	147806796	1305899	Dup	1q21.1 dup syndrome	Not maternal	ASD	8	M	S
304	2	51253508	51259674	6166	Del	NRXN1 (E1–3) ^c	Paternal	ASD+ID	6	M	Fa
396	2	51253508	51259674	6166	Del	NRXN1 (E1–3) ^c	DN	ASD+ID	7	M	S
318	2	51253508	51259674	6166	Del	NRXN1 (E1–3)	Paternal	Asperger	12	M	S
<i>Likely pathogenic CNVs</i>											
55	18	44624417	44641759	17342	Del	KATNAL2 (E 16–18)	Paternal	ASD	6	M	Fa ^a
365	9	119373136	119495896	122760	Del	ASTN2 (E1–4)	Maternal	Asperger	19	F	Fa ^b

NA not assessed, E exon, BP breakpoint, DN de novo, Del deletion, Dup duplication, F female, M male, S sporadic, Fa familial.

^a1stSib (ASD): wild type and 2nd Sib (language delay): variant carrier.

^bSib (Asperger): variant carrier.

^cAlready identified by aCGH.

p.Trp1889Arg (NM_194248.2) *OTOF* (MIM: 601071) pathogenic variants, while a patient with severe visual impairment, born from consanguineous parents, carried a homozygous c.389del, p.Pro130Leu fs*36 (NM_000180.3) *GUCY2D* (MIM: 204000) pathogenic variant. These mutations explain the sensory deficits of these patients but there is currently no evidence that they are related to the ASD phenotype.

Genotype/phenotype correlations

Consistency with the previously described phenotypes was checked for syndromic genes and was fair to good for most genes (*DNMT3A*, *ASXL3*, *EBF3*, *NR2F1*, *DYRK1A*, *DEAF1*, *ANKRD11*, *SHANK3*). (See Sup Info) but in some cases unusual phenotypes were encountered:

Patients carrying *MECP2* or *KMT2A* mutations did not harbor the classical features of Rett or Wiedemann-Steiner syndrome, respectively. A *MECP2* mutation (not present in the International Rett database) was identified

in a girl who presented with mild ASD. She had no developmental delay (walk at 15 months, no language delay) and was enrolled in an ordinary school at the age of 14. She demonstrated no ID and no dysmorphic features. *KMT2A* mutations were identified in two patients. One of them carrying the p.Gln3192Pro de novo mutation was the first child of a couple with epilepsy and mild cognitive impairment. He had a global developmental delay with a sitting age at 1 year but a normal walking age at 17 months, a language delay and a motor instability. Clinical evaluation showed a normal growth, an hyper-laxity and a diffuse hypertrichosis but no hairy elbow or morphological features typical of Wiedemann–Steiner Syndrome. The second one was the first child of a couple without medical history. He carried a c.10835 + 1G > A splice mutation already described in 3 related ID patients, not diagnosed as Wiedemann-Steiner. This variant leads to the in frame deletion of exon 28 and likely disrupts the stabilizing interaction between the C terminal *KMT2A*

Table 3 Nucleotide variants conferring susceptibility for autism (stratified by category) identified among the 253 ASD subjects. Nucleotide variants in genes harboring a statistically significant excess of de novo PTVs were classified either as susceptibility variants in class A genes (false discovery rate (FDR) <5%) or as probable susceptibility variants in class B genes (10% < FDR < 5%). Nucleotide variants in syndromic ASD-related genes were classified using the guidelines from the American College of Medical Genetics-Association for Molecular Pathology.

ID	Gene	Variatio	Transcript	inheritance	o/e	GnAD	Diag	Age	Sex	Family history	
(a) Nucleotide variants in genes harboring a statistically significant excess of de novo PTVs											
<i>Susceptibility variants</i>											
213	CHD8	p.Gln1171* c.3511C>T	NM_001170629.1	Paternal	3.94%	_	ASD+ID	7	F	S	
223	CHD8	p.Lys528Glyfs*14 c.1582_1583delAA	NM_001170629.1	DN	3.94%	_	Asperger	19	M	S	
220	CHD8	p.Leu980Argfs*5 c.2937_2939delinsTC	NM_001170629.1	DN	3.94%	_	Asperger	30	M	S	
98	NRXN1	p.Leu158Alafs*29 c.471dupG	NM_001135659.1	DN	12.30%	1/242	430	Asperger	12	M	S
307	NRXN1	p.Glu389Serfs*3 c.1165delG	NM_001135659.1	Maternal	12.30%	_	ASD	3	M	S	
253	ANK2	p.Glu3062* c.9184G>T	NM_001148.4	DN	6.20%	_	Asperger	11	M	S	
325	GIGYF1	p.Ter1036Glyext*72 c.3106T>G	NM_022574.4	DN	35.64%	_	ASD	12	M	Fa ^a	
341	KDM5B	p.Arg1397* c.4189C>T	NM_006618.4	NA	44.95%	2/247	406	ASD	4	M	S
<i>Probable susceptibility variants</i>											
122	RIMS1	p. ? c.100+1G>A	NM_001168407.1	Maternal	18.65%	1/31	404	ASD+ID	14	F	S
188	SETD2	p.Arg1708* c.5122C>T	NM_014159.6	DN	13%	_	Asperger	8	F	Fa ^b	
301	DIP2A	p. ? c.1429+2T>G	NM_015151.3	Maternal	23.89%	_	Asperger	6	F	S	
285	HECTD4	p.Ala3108Glyfs*5 c.9322dupG	NM_001109662.3	Maternal	7.56%	_	Asperger	12	M	S	
168	SPEN	p.Ser2411* c.7232C>A	NM_015001.2	DN	2.31%	_	Asperger	6	M	S	
(a) Nucleotide variants in syndromic ASD-related genes											
<i>Susceptibility variants</i>											
370	EBF3	p.Gln78* c.232C>T	NM_001005463.2	DN	3%	_	ASD	4	M	S	
79	TSC2	p.Val1062Glyfs*7 c.3185_3188del	NM_000548.3	DN	2%	_	ASD+ID	8	F	S	
<i>Probable susceptibility variants</i>											
30	DEAF1	p.Arg224Gln c.671G>A ^c	NM_021008.3	Maternal	41%	1/246	022	ASD+ID	14	M	S
328	DNMT3A	p.Gln249* c.745C>T	NM_022552.4	DN	126%	1/244	420	ASD+ID	10	M	S
215	AUTS2	p.Thr328Argfs*19 c.983_984del	NM_015570.3	DN	13%	_	ASD+ID	4	M	S	
222	DYNC1H1	p. ? c.4396-1G>C	NM_001376.4	DN	5%	_	ASD	27	F	Fa ^a	
121	NR2F1	p.Gly395Ala c.1184G>C	NM_005654.5	DN	0.00%	_	ASD+ID	20	M	S	
(b) Nucleotide variants in genes belonging to both categories											
<i>Susceptibility variants</i>											
296	SHANK3	p.Pro922Argfs*34 c.2765del	NM_033517.1	DN	3.98%	_	ASD+ID	8	F	S	
377	SHANK3	p.Leu1030Cysfs*48 c.3088delC	NM_033517.1	DN	3.98%	_	ASD+ID	6	M	S	
353	SHANK3	p.Ala1227Glyfs*69 c.3679dupG ^d	NM_033517.1	NA	3.98%	_	ASD+ID	31	F	S	
136	GRIN2B	p.Ser9Phefs*50 c.23_24insC	NM_000834.3	DN	0.00%	_	ASD	4	M	S	
85	ASXL3	p.Ser1246* c.3737C>A	NM_030632.1	DN	9.86%	_	ASD+ID	11	M	Fa	
78	SCN2A	p.Phe601Leufs*40 c.1800delC	NM_001040142.1	DN	6.33%	_	ASD+ID	12	F	S	
382	CHD2	p.Val332Glyfs*25 c.995_999delTGAAG	NM_001271.3	NA	2.75%	_	ASD+ID	7	M	S	
47	DYRK1A	p.Arg255* c.763C>T ^e	NM_101395.2	Not maternal	8.45%	_	ASD+ID	18	M	S	

Table 3 continued

395	<i>ANKRD11</i>	p.Glu1154Glyfs*16c.3460dup	NM_001256182.1	DN	4.72%	_	Asperger	11	M	S
<i>Probable susceptibility variants</i>										
291	<i>KMT2A</i>	p. ? c.10835+1G>A ^f	NM_001197104	Paternal	3.12%	_	Asperger	6	M	S
324	<i>KMT2A</i>	p.Gln3192Pro c.9575A>C	NM_001197104	DN	3.12%	_	ASD	9	M	S
157	<i>MECP2</i>	p.Glu282Gly c.845A>G	NM_004992.3	DN	10.72%	_	ASD	13	F	S

Ffemale, Mmale, Ssporadic, Fa familial, DNde novo, NA not assessed, o/e observed/expected ratio of PTVs in GnomAD, GnAD allelic frequency in the GnomAD database.

^aSib (ASD): wild type.

^bSib (Asperger): wild type.

^cp.Arg224Trp found as a de novo event in an ASD case²⁷.

^dVariation found as de novo event in ASD+ID cases²⁸.

^eVariation found as a de novo event in a syndromic case of DYRK1A-haploinsufficiency²⁹.

^fVariation found in three related ID patients¹⁴.

Table 4 Detection rate for CNVs and nucleotide variants among the 253 ASD subjects stratified by clinical subgroups.

	ASD+ID <i>n</i> = 68	ASD <i>n</i> = 95	Asperger <i>n</i> = 90	Total <i>n</i> = 253
CNVs	7	5	6	18–7.1% [4.3–11%]
Nucleotide variants	14	8	10	32–12.6% [8.8–17.4%]
Total	21–30.1% [20.2–43.2%]	13–13.7% [7.5–22.3%]	16–17.7% [10.5–27.3%]	50–19.7% [15–25.2%]

fragments¹⁴. Our patient had no motor delay but a language delay. Electroencephalogram showed left occipital spikes, but he never had seizures. MRI showed an isolated agenesis of the corpus callosum. The diagnosis of Wiedemann-Steiner syndrome was not clinically evoked. Likewise, although *DYNC1H1* missense or truncating variations are usually associated with a severe phenotype including ID, neuronal migration defects and epileptic encephalopathy¹⁵, our patient had an unremarkable ASD phenotype without neurological features. *CHD2* truncating mutations are mostly associated with pediatric refractory epilepsy¹⁶, but our patient had only experienced one tonic-clonic seizure. In the same vein and in line with recent reports underlining the phenotypic heterogeneity of *SCN2A*¹⁷ and *GRIN2B*¹⁸ mutation carriers, none of the patient carrying de novo truncating mutations on these genes had epilepsy, which is a feature generally associated with these mutations.

It should also be noted that besides these well-characterized syndromic genes, another patient carrying the *RIMS1* splice variant received a diagnosis of Dravet syndrome but bore no mutation on genes causing this phenotype. Interestingly, the RIM1a protein encoded by the mouse ortholog of *RIMS1* has been characterized as a presynaptic active zone protein mainly expressed in neurons from the cerebellum¹⁹ and controlling epileptogenesis following pharmacologically induced status epilepticus²⁰. This suggests that the *RIMS1* truncating variant identified here, which affects a minor transcript

expressed in the cerebellum (<https://gtexportal.org>), could be associated with a peculiar neurological phenotype.

Discussion

In this report, the rate of identified CNVs, largely based on well-known recurrent CNVs, was similar to that obtained in several recent studies^{1,2}. It is notable that four of our patients carry the 15q11.2 recurrent deletion which is known to have a mild effect on developmental disorder. Hence, it is very probable that additional factors are involved in this subset. Regarding nucleotide variants, we used rigorous criteria to classify both ASD susceptibility genes and variants. As a result, we retained only 217 genes as compared for example with the 880 genes present in the Sfar database (october 2018 release), often with minimal evidence of involvement in ASD susceptibility. We nevertheless obtained a detection rate of 12.6% for nucleotide variants. Combining CNVs and nucleotide variants, we reached a higher detection rate (19.7%) than the one obtained in 105 ASD patients previously analyzed by WES and aCGH (15.8%)². This is partly due to the contribution of the set of genes with an excess of de novo PTVs in ASD patients, which were identified in the recent years thanks to the study of more than 6000 ASD trios^{1,4–10}. Clearly, this huge effort has obvious consequences for clinical practice. The finding of a particularly high rate of de novo PTVs in this set of genes (14/20 PTVs = 70% of the total number of PTVs

in patients for whom parental DNA was available) reinforces the conclusion that this set is truly enriched in de novo PTVs in ASD subjects. As seen in Sup Table 2, the gene which is the most frequently hit by de novo PTVs in ASD patients from the literature is *CHD8*²¹. Interestingly, among the three *CHD8* PTVs identified in this study, two were de novo events while the third was inherited from a parent whose sole phenotypic feature was macrocephaly. More generally, the presence of 30% of PTVs inherited from parents with no ASD diagnosis underlines the incomplete penetrance and/or variable expressivity of these variants.

Due to the stringent criteria used to classify variants, it should be stressed that our detection rate is likely to be underestimated. First, concerning CNVs, partial duplications intersecting genes were conservatively considered of uncertain significance (Sup Table 4). However, it is well known that tandem intragenic duplications often disrupt reading frames and can lead to haploinsufficiency. In a recent analysis of all structural variation detected in the gnomAD dataset, the authors showed a significant correlation between constraint metrics reflecting intolerance to LoF mutations and the depletion of partial genic duplications at the gene level²². However, despite these findings, the individual interpretation of partial duplications remains challenging and will require additional investigations. Second, since all missense variants impacting genes exhibiting a statistically significant excess of de novo missense mutations in ASD^{1,12} were inherited, we decided to consider them as variants of unknown significance (VUS) (Sup Table 5) even when they were absent from gnomAD.

The detection rate was particularly high in the ASD+ID subgroup. It was in the same range as those recently reported in ID or DD patients (25–39%)^{23–25}, or in a cohort of 163 ID/DD patients “with reported ASD or autistic features” (25.8%)²⁶. Of note, all these studies used larger panels of genes than ours. In our study, although ASD+ID patients were more likely hit, susceptibility variants were also found with a high frequency in the Asperger group. Several genes carried susceptibility variants in both ASD+ID and Asperger subjects. For example, in accordance with a previous report outlining that cognitive ability of *CHD8* patients ranges from profoundly disabled to age-appropriate²¹, of the three patients carrying *CHD8* PTVs, one had ID associated to ASD while two others received a diagnosis of Asperger syndrome. However, macrocephaly, a shared feature among *CHD8* patients, was present in all three patients. Likewise, *NRXN1* exon 1–3 deletions were present in patients with each clinical presentation (ASD, ASD+ID, Asperger).

The genotype/phenotype correlation for syndromic genes was generally good, but in some cases was not typical. As

already outlined^{23,25}, it is noteworthy that the phenotypic spectrum of mutations in several “syndromic genes” for which a specific phenotype has previously been described is broader than initially thought. Clearly, for a large set of genes, prioritizing by double entries (genes with both an excess of de novo PTVs and syndromic genes) allows the identification of ASD susceptibility factors in more patients than prioritizing upon a single criteria based on strict phenotypic concordance with described syndromic features. In summary, the strategy we advocate here is an efficient and cost-effective manner to deal with the detection of both rare nucleotide variants and CNVs in a clinical setting. Of course, exome data should be periodically re-analyzed and the list of genes updated. Due to the extreme rarity of variants studied here, it is currently difficult to establish for each of them if they are fully causative of an ASD phenotype. Most likely, only a subset of them are *necessary and sufficient* to cause ASD. For this reason, we recommend to avoid the term “causative of ASD” when delivering the information about these variants to a family and to use instead the term “genetic susceptibility factor contributing to ASD phenotype”. At the present stage, with the exception of pathogenic variants in syndromic genes causing well-defined monogenic diseases, using these results for prenatal diagnosis is premature.

Acknowledgements

Supported by Fondation Deniker and Inserm grants. We are indebted to families B. and H. for their generous gifts. This study was cosupported by European Union and Région Normandie in the context of Recherche Innovation Normandie (RIN2018). Europe gets involved in Normandie with the European Regional Development Fund (ERDF). Funding sources did not have any involvement in the study design.

Author details

¹Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics and Reference Center for Developmental Disorders, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France. ²Department of Research, Centre hospitalier du Rouvray, Sotteville-Lès-Rouen, France. ³Centre Ressources Autisme Normandie Seine Eure, Centre hospitalier du Rouvray, Sotteville-Lès-Rouen, France. ⁴Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, CEA, Université Paris-Saclay, F-91057 Evry, France

Conflict of interest

The authors have full access to all the data in the study and had final responsibility for the decision to submit the article. The authors declare no conflict of interest regarding the publication of this article.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41398-020-0760-7>).

Received: 2 December 2019 Revised: 10 December 2019 Accepted: 10 January 2020

Published online: 24 February 2020

References

1. Yuen, C. et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* 20, 602–611 (2017).
2. Tammimies, K. et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *JAMA* 314, 895–903 (2015).
3. Ronemus, M., Iossifov, I., Levy, D. & Wigler, M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat. Rev. Genet.* 15, 133–141 (2014).
4. De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209–215 (2014).
5. Dong, S. et al. De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. *Cell Rep.* 9, 16–23 (2014).
6. Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216–221 (2014).
7. Krumm, N. et al. Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* 47, 582–588 (2015).
8. O’Roak, B. J. et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nat. Commun.* 5, 5595 (2014).
9. Stessman, H. A. F. et al. Targeted sequencing identifies 91 neurodevelopmental disorder risk genes with autism and developmental disability biases. *Nat. Genet.* 49, 515–526 (2017).
10. Wang, T. et al. De novo genic mutations among a Chinese autism spectrum disorder cohort. *Nat. Commun.* 7, 13316 (2016).
11. Sanders, S. J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* 87, 1215–1233 (2015).
12. Geisheker, M. R. et al. Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains. *Nat. Neurosci.* 20, 1043–1051 (2017).
13. Rosenfeld, J. A., Coe, B. P., Eichler, E. E., Cuckle, H. & Shaffer, L. G. Estimates of penetrance for recurrent pathogenic copy-number variations. *Genet. Med.* 15, 478–481 (2013).
14. Bogaert, D. J. et al. Early-onset primary antibody deficiency resembling common variable immunodeficiency challenges the diagnosis of Wiedeman-Steiner and Roifman syndromes. *Sci. Rep.* 7, 3702 (2017).
15. Willemsen, M. H. et al. Mutations in DYNC1H1 cause severe intellectual disability with neuronal migration defects. *J. Med. Genet.* 49, 179–183 (2012).
16. Lamar, K.-M. J. & Carvill, G. L. Chromatin remodeling proteins in epilepsy: lessons from CHD2-associated epilepsy. *Front Mol. Neurosci.* 11, 208 (2018).
17. Wolff, M. et al. Genetic and phenotypic heterogeneity suggest therapeutic implications in SCN2A-related disorders. *Brain* 140, 1316–1336 (2017).
18. Platzer, K. et al. GRIN2B encephalopathy: novel findings on phenotype, variant clustering, functional consequences and treatment aspects. *J. Med. Genet.* 54, 460–470 (2017).
19. Kintscher, M., Wozny, C., Jochenning, F. W., Schmitz, D. & Breustedt, J. Role of RIM1 α in short- and long-term synaptic plasticity at cerebellar parallel fibres. *Nat. Commun.* 4, 2392 (2013).
20. Pitsch, J. et al. The presynaptic active zone protein RIM1 α controls epileptogenesis following status epilepticus. *J. Neurosci.* 32, 12384–12395 (2012).
21. Bernier, R. et al. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* 158, 263–276 (2014).
22. Collins, R. L. et al. An open resource of structural variation for medical and population genetics. *bioRxiv* 578674 <https://doi.org/10.1101/578674> (2019).
23. Chérot, E. et al. Using medical exome sequencing to identify the causes of neurodevelopmental disorders: experience of 2 clinical units and 216 patients. *Clin. Genet.* 93, 567–576 (2018).
24. Martínez, F. et al. High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. *J. Med. Genet.* 54, 87–92 (2017).
25. Redin, C. et al. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *J. Med. Genet.* 51, 724–736 (2014).
26. Rossi, M. et al. Outcomes of diagnostic exome sequencing in patients with diagnosed or suspected autism spectrum disorders. *Pediatr. Neurol.* 70, 34–43. e2 (2017).
27. Vulto-van Silfhout, A. T. et al. Mutations affecting the SAND domain of DEAF1 cause intellectual disability with severe speech impairment and behavioral problems. *Am. J. Hum. Genet.* 94, 649–661 (2014).
28. Durand, C. M. et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* 39, 25–27 (2007).
29. Earl, R. K. et al. Clinical phenotype of ASD-associated DYRK1A haploinsufficiency. *Mol. Autism* 8, 54 (2017).

Chapitre III. Evaluation du bénéfice patient du séquençage d'exome dans le TSA

Le travail précédent montrait quelques limites. Tout d'abord, il s'agissait d'un travail rétrospectif. Les patients séquencés avaient été sélectionnés parmi ceux reçus en consultation de génétique les années précédentes. De plus, les critères diagnostiques majoritairement utilisés étaient ceux du DSM-IV-TR et non les derniers publiés. Il est concevable que le taux de détection diffère en fonction de la modification des limites du regroupement syndromique.

Différents éléments se combinent pour poser l'hypothèse d'une évolution de la population de patients. Tout d'abord, plusieurs études ont montré l'impact de l'évolution des critères diagnostiques sur les caractéristiques des patients considérés comme présentant un TSA. La nature de cet impact reste controversée mais il semble que les populations de patients diagnostiqués comme présentant un TSA ne soient pas tout à fait similaires selon la version du DSM utilisée (Kulage *et al.*, 2020; Peters & Matson, 2020; Smith *et al.*, 2015; Wiggins *et al.*, 2019). Ensuite, dans une perspective nationale, les différents « plans autisme » ont amélioré le dépistage et le diagnostic des patients présentant un TSA, notamment concernant les formes plus légères²⁰. Enfin, la Haute Autorité de Santé

²⁰<https://solidarites-sante.gouv.fr/archives/archives-handicap/les-propositions-du-cih-2016/article/preparation-du-4eme-plan-autisme>

<https://handicap.gouv.fr/la-strategie-nationale-autisme-et-troubles-du-neuro-developpement>

recommande qu'une consultation de génétique soit proposée pour chaque patient diagnostiqué, quelle que soit la sévérité du trouble ou les comorbidités, ce qui a pu élargir l'accès à la consultation de génétique, y compris aux patients sans comorbidités neurodéveloppementales²¹.

Pour tenter de dépasser ces difficultés, nous avons conçu l'étude REDIA (Rendement Diagnostic Autisme). Cette étude prévoyait le séquençage d'exome de 300 patients TSA inclus consécutivement dans les unités de génétique clinique des CHU de Rouen et Caen à partir de 2019. L'étude REDIA, promue par le CHU de Rouen, et dont l'investigateur coordonnateur est le Pr Gaël Nicolas, a obtenu un financement au titre du Programme Hospitalier de Recherche Clinique Interrégional 2018 pour la réalisation de 300 WES.

Les objectifs sont d'évaluer le taux de détection sur une cohorte prospective, d'évaluer l'évolution de celui-ci par rapport aux patients recrutés dans l'étude précédente et d'estimer les variables cliniques qui corrélerent avec la détection de variations contribuant au TSA. Enfin, nous avons conçu des questionnaires afin d'évaluer l'utilité perçue de la démarche génétique par les familles, ainsi que leurs appréhensions.

Actuellement, 297 patients ont bénéficié de séquençage d'exome. De plus, des questionnaires ont été prévus également au moment du rendu de résultats. Ceux-ci n'ont pas tous été passés. Ils seront bien sûr analysés par la suite. Nous présentons ici une version préliminaire d'un article qui

²¹https://www.has-sante.fr/upload/docs/application/pdf/2018-02/trouble_du_spectre_de_lautisme_de_lenfant_et_ladolescent__recommandations.pdf

sera soumis lorsque nous disposerons des données finales. Elle devrait être fidèle au manuscrit final.

Ma contribution à ce travail est la suivante :

- Conception des questionnaires
- Informatisation et codage des données cliniques
- Participation à l'interprétation des données de séquençage d'exome
- Participation à l'analyse statistique et écriture du manuscrit

Title

Assessing the clinical utility of whole exome sequencing in Autism Spectrum Disorder: a prospective study

Authors

Thomas Husson, François Lecoquierre, Kévin Cassinari, Alice Goldenberg, Anne-Marie Guerrot, Gabriella Vera, Juliette Coursimault, Maryam Soleimani, Anne-Claire Bréhin, Florian Perrin de Brichambaut, Antoine Bonnevalle, Elise Petit, Margaux Clément Le Choismier, Anaïs Henry, Camille Desjardins, Aline Vincent, Varoona Bizaoui, Marion Gérard, Pascal Chambon, Géraldine Joly-Hélas, Pascale Saugier-veber, Olivier Guillin, Olivier Quenez, Sophie Coutant, Anne Boland, Jean-François Deleuze, Robert Olasso, Vincent Meyer, Camille Charbonnier, Dominique Champion, Gaël Nicolas

Introduction

Autism Spectrum Disorders (ASD) affects from 1 to 3% of individuals (Shenouda et al., 2022; Sun et al., 2019). Population wise, a part of the genetic risk is spread through thousands of common risk alleles with pocket-sized effects (Grove et al., 2019). Besides, a subset of probands carries a rare genetic variant that confers a high risk for ASD (Antaki et al., 2022; Fu et al., 2022). Only these can be regarded as risk factors with a large enough effect to be reported individually in a clinical setting. Nowadays, following clinical guidelines, ASD probands are a significant proportion of patients who consult in clinical genetics departments (Harris et al., 2020). Providing an assessment of the clinical utility of the molecular diagnosis process in the context of ASD is not straightforward. Very few actionable genes have been reported for this population, beyond obvious impact on genetic counselling in case of a Mendelian variant identification. However, clinical utility cannot be only focused on health-related outcomes only as the delivery of genetic tests results can have positive psychological effects on probands and caregivers (Hayeems et al., 2016; Johannessen et al., 2017). It could also have detrimental effects such as perceived stigmatization (Johannessen et al., 2017).

Hundreds of genes have been associated with ASD, notably by large-scale next-generation sequencing studies leveraging frameworks to interpret *de novo* mutation (Samocha et al., 2014). Globally, they partially overlap with those associated with Developmental Delay (DD), intellectual disability (ID), epilepsy and macrocephaly as well as all comorbid conditions with ASD. Some of them are even

involved in rare neurogenetic disorders which encompass ASD in their phenotypic spectrum. They add up to known loci with recurrent Copy Number Variants (CNV) bearing a risk for ASD and other neurodevelopmental disorders. Thus, genetic risk factors for ASD comprise a broad range of situations making it arduous to define explicit criteria for pathogenicity.

To obtain enough statistical power to identify ASD-associated variants, huge cohorts have been set up by international *consortia*. Unfortunately, they lack precise phenotype assessment most of the time, and the recruitment process is sometimes unclear. It is then difficult to know whether those samples are representative of patients referred to clinical genetics departments. From these large cohorts, it is also hard to determine whether pathogenic variant carriers come from a distinct ASD subpopulation with more severe or more syndromic phenotypes or not, especially as some evidences are suggestive of a causative effect of the cognitive impairment carried by these rare variants on ASD susceptibility (Antaki et al., 2022; Bishop et al., 2017; Buja et al., 2018).

At the biological level, two gene sets have been described (Satterstrom et al., 2020). A first group of genes is involved in Neuronal Communication (NC). It consists in genes involved in synapse building and neurotransmission (Courchesne et al., 2020). They are mainly expressed in brain tissue. A second, and bigger set of genes is implicated in Gene Expression Regulation (GER). These broadly expressed genes are involved in the epigenetic machinery and in chromatin remodelling (Courchesne et al., 2020). The GER gene set has an earlier expression peak during prenatal development and could be associated with more severe phenotypes (Satterstrom et al., 2020).

These scientific discoveries bring however some unresolved interrogations at the clinical level. Firstly, no detection rate evaluation has been performed in a prospective cohort. Secondly, whether variant detection yields have evolved since ASD diagnostic criteria were modified (through DSM-IV to DSM-5 switch) is yet to be determined. Thirdly, what are the perceived benefits and concerns of going through the molecular diagnostic process for probands and their family. Fourthly, no guidelines have been reported for variant interpretation in a clinical setting for ASD, which is often considered as a difficult task as compared to intellectual disability for example. Fifthly and most importantly given the clinical, biological and genetical heterogeneity, what clinical features could predict the highest chances of receiving a positive/conclusive molecular diagnostic.

To bridge these gaps, we conducted a Whole-Exome Sequencing (WES) study in a prospective cohort of 300 probands with ASD. Clinical and genetic features were then combined with that of a retrospective series from our group, adding up to a total of 656 patients. Clinical predictor of pathogenic variant detection was evaluated and a clinical score predictive of conclusive molecular diagnosis was built. Eventually, probands and caregivers were asked to fill in a questionnaire about their expectations and worries about the molecular diagnosis process.

Methods

Recruitment

Probands referred to the clinical genetics departments of the Rouen and Caen University Hospitals, Normandy, France, were offered to participate to this study until 300 inclusions were reached, between June 2019 and June 2021. All ASD diagnoses were made according to the DSM-5 criteria by a senior psychiatrist or a pediatric neurologist prior to the clinical genetics consultation. Inclusion criteria were: diagnosis of ASD as defined above, absence of known genetic cause or strong genetic risk factor for ASD, absence of prior next-generation sequencing explorations, without any age limit or any family history criterion and availability of both parents' DNA.

Record of developmental milestones, previous medical records and family history, followed by a complete physical examination was undertaken by the clinical geneticist. Information on current knowledge about ASD genetics was explained to all patients and both parents, including what to expect and not to expect from WES in this context, in the same face-to-face consultation. Patients or legal representatives as well as both parents provided informed written consent for genetic analyses in a medical setting as well as informed written consent to participate to the study. This study was approved by the CPP OUEST 1 institutional review board. Blood samples were taken from the proband and both parents (fresh blood sample on EDTA tube).

Following patient inclusion, parents and probands (when able to understand) were asked to fill in a 5-point Likert scale self-assessment questionnaire about their putative expectations and concerns regarding genetic testing in the context of ASD (see below).

Assessment and classification of clinical phenotypes

A physical anomaly score was encoded from the medical record as inspired from Miles et al. and Tammimies et al. work (Miles et al., 2005; Tammimies et al., 2015) : 1 point for facial dysmorphism or each organ congenital anomalies, 0.2 points for each embryologically unrelated minor physical anomaly and for each measurement (size and growth) abnormality beyond 2 SD from the mean.

A cognitive deficit score was encoded as following: 0 if IQ > 130 or advanced learning, 1 for normal IQ without any cognitive deficit, 2 for specific cognitive disorder such as ADHD, dyspraxia or learning disorder with normal IQ, 3 for mild to moderate intellectual disability, 4 for severe to profound intellectual disability. Notably, not all patients underwent a full neuropsychological assessment, as some patients were too young or neuropsychological assessment was not possible to organize for some other patients. Hence, investigators were asked to rate intellectual disability, when possible, based on clinical examination and personal history as well as psychiatrists' reports.

Other clinical items (language or walking delay, neurological symptoms such as epilepsy or neurological syndromes, abnormal head circumference and neonatal complications) were independently binarily rated. Genetic and clinical data of the current prospective study were then combined with those of 356 ASD samples retrospectively analyzed (i.e. probands from Husson et al. (Husson et al., 2020) and 100 additional patients following the exact same methods and inclusion criteria as in Husson et al., but with WES performed after publication and hence not included in Husson et al.), and diagnosed using DSM-IV criteria.

Molecular genetics

Exomes were captured using Agilent SureSelect All Exons V6 UTR Kit (Santa Clara, United-States) and sequencing was performed on an Illumina Novaseq6000 (Illumina, San Diego, CA, USA) at the CNRGH (Centre National de Recherche en Génomique Humaine, Evry, France) with paired end mode, 150 base pairs (bp) reads. Nucleotide variants were detected through a bioinformatics pipeline to perform variant calling, quality check and annotation as previously described (Husson et al., 2020). CNVs detection from WES data was performed using the CANOES software (Backenroth et al., 2014). Diagnostic performances of our CANOES-based workflow were previously assessed with high sensitivity (87%) and high predictive positive value (86%) (Quenez et al., 2021). Following interpretation by an expert molecular geneticist, rare variants were prioritized based on molecular and

phenotypic features, and segregation of retained variants was examined by Sanger sequencing for nucleotide variants and by ddPCR or QMPSF for CNVs. Parenthood was checked by polymorphic informative microsatellites for patients carrying *de novo* mutations. Final variant classification was performed following segregation analysis based on the ACMG-AMP recommendations (Richards et al., 2015a) for genes belonging to a list of Mendelian disorders as provided by the OMIM database. Variants rated as likely pathogenic (class 4) or class 5 (pathogenic) were reported to the referring clinical geneticist, as well as class 3 variants (variants of unknown significance) when phenotype information was consistent with the corresponding gene, and when further investigations may help reclassify the variants later (such as further segregation analysis, RNA assessment, X-chromosome inactivation in a female, etc.). Thus, in some patients, a final diagnosis of a rare neurodevelopmental disorder including ASD phenotype was made. In addition, for genes or CNVs not fulfilling ACMG-AMP rules because falling outside lists of Mendelian genes as provided by OMIM, but still considered as significantly enriched in a certain type of nonsynonymous rare variants in ASD probands (see lists of genes below), we added a category of “risk factor” or “probable risk factor” and these results were also reported to the clinical geneticists. An exome results report was sent to clinical geneticist for every patient. Finally, clinical geneticists explained the results of this genetic investigation (all cases, including negative reports) in a face-to-face consultation or by teleconsultation. Patients (when able to understand) and parents were asked to fill in a second post-test self-assessment questionnaire following the results consultation.

List of ASD susceptibility genes

Several partially overlapping gene (n = 448) sets were considered:

- (i) Genes (n = 80) showing statistically significant excess of *de novo* protein truncating variants (PTVs), (i.e., nonsense, canonical splice site variants and frameshift indels) in meta-analysed (Husson et al., 2020) published cohorts (C Yuen et al., 2017b; De Rubeis et al., 2014; Dong et al., 2014; Iossifov et al., 2014; Krumm et al., 2015; Sanders et al., 2012; Stessman et al., 2017; T. Wang et al., 2016) with a False Discovery Rate (FDR < 10%). In those, all PTVs, inherited or *de novo*, and all *de novo* missense variants were considered.
- (ii) Genes (n = 166) involved in monogenic neurodevelopmental disorders for which ASD is described as belonging to the possible phenotypic spectrum. This set of genes was

carefully curated from the Simons Foundation Autism Research Initiative (SFARI) database and from literature data. Pathogenicity was evaluated using the guidelines from the American College of Medical Genetics-Association for Molecular Pathology (ACMG-AMP) (Richards et al., 2015b).

- (iii) Gene (n = 353) from more recently published studies with Transmission and De novo Association (TADA) paradigm (Kaplanis et al., 2020; Satterstrom et al., 2020). For every gene, we have incorporated the relative burden of PTVs and damaging missense variants in our pathogenicity interpretation.
- (iv) Loci with known recurrent CNVs that lead to an increased ASD risk from the SFARI database and Girirajan et al. (Girirajan et al., 2013).
- (v) In patients with extra-features than ASD itself (physical features and/or ID), Mendelian genes were interpreted based on OMIM lists (version at the time of clinical genetics interpretation, between years 2020 and 2022, last update: may 2020), allowing diagnosing rare disorders explaining part of the phenotype outside ASD (thus not considered in the ASD diagnostic yield) but also possibly rare disorders actually associated with ASD but not listed in above-mentioned lists because ASD clinical assessments are not always precise enough in the literature, most of the time because patients are not examined by a psychiatrist not allowing a formal ASD phenotype but still incorporating “autistic features” information on the OMIM phenotype details)

Overall, we considered here as “ASD-contributing variants”, likely pathogenic or pathogenic variants in Mendelian genes as well as probable risk factors or risk factors.

Self-assessment questionnaires

We underwent a careful review of qualitative research regarding psychological effects of genetic testing for probands with neurodevelopmental disorders and their parents. Building on it, we designed a self-assessment 5-point Likert-type questionnaire about expectations of probands and parents concerning genetic testing. Cronbach's alpha was calculated for measuring internal consistency, that is, how closely related a set of items are as a group.

Questions for parents (english translation herein) were:

Please indicate how much you agree with following proposals regarding this genetic procedure for your child. From 1 (not at all) to 5 (fully agree):

- Q1. This genetic test can help me understand what the cause of my child's autism is
- Q2. This genetic test can be useful for my family
- Q3. This genetic test can help me anticipate what my child's development will be
- Q4. This genetic test can help me improving care for my child (treatment, rehabilitation, administrative procedures)
- Q5. This genetic test can improve support for my child (from relatives, patient associations, or healthcare institutions)
- Q6. This genetic test can give me an estimate about recurrence of autism for a possible new pregnancy for me or other relatives
- Q7. This genetic test can be harmful in any way for my child
- Q8. This genetic test can be harmful in any way for my family
- Q9. This genetic test is susceptible to generate anxiety to me
- Q10. This genetic test is susceptible to generate some anxiety in my child
- Q11. It is important for me to understand why my child has autism spectrum disorders

The same questionnaires were proposed to patients when they were old enough and cognitively able enough to understand and fill in (with formulation adapted like "me" instead of "my child").

Following the second genetics consultation aiming at delivering the personal results of WES, the same self-assessment questionnaire was proposed. In addition, four questions were added:

- Q12. I do consider that my child's ASD is genetic.
- Q13. This genetic test allowed me to understand why my child has ASD.
- Q14. Would you recommend consulting in a genetic department for a relative, if he/she would be in the same situation as you?
- Q15. I wish that the blood sample of my child go through other investigations or reanalysis in the future, in order to make me understand better why my child has ASD.

Results

Recruitment, exome sequencing, and ASD-contributing variant detection rates

Between June 2019 and June 2021, 300 ASD probands attending medical genetics departments were included in this study and 297 underwent WES in the proband, with DNA of both parents available in case of followed-up targeted segregation analyses. Clinical characteristics of the whole cohort are presented in Table 1. Overall, 65/297 (21.6%) probands had a diagnosis of ASD with intellectual disability (ID) while 136/297 (46%) showed ASD without ID. For the remaining patients, cognitive functioning could not be assessed, mainly because they were too young. For detailed clinical data, see supplementary table 1.

Among these 297 probands, 42 (14.1%) carried at least one ASD-contributing variant, including 8% with a (likely) pathogenic variant in a Mendelian gene, and 6% a (probable) risk factor. Detailed information on variants identified in this series is available in supplementary Table 1. Nineteen patients with a rare Mendelian disease diagnosis presented a *de novo* mutation, while 2 had an autosomal recessive inheritance or inherited X-linked variants from the mother, thus with putatively important consequences for genetic counselling. Of note, the rate of CNV detection was rather low (3.3%) as compared to previous studies including ours (C Yuen et al., 2017b; Husson et al., 2020; Tammimies et al., 2015). Interestingly, 3.3% of the ASD-contributing variants were inherited from an asymptomatic or paucisymptomatic parent, so that this kind of analysis may be difficult to interpret in a clinical setting without clear guidelines for variant interpretation in this context. Molecular genetic results are displayed in table 2.

We then gathered data from a retrospective analysis (Husson et al., 2020), and its follow up (unpublished) to that of the current prospective study, so that we analysed phenotype and genetic data of a total of 653 patients here (molecular genetic results in supplementary table 1). These data were analysed through the same bioinformatic pipeline and along the same criteria.

In the current prospective study, the variant detection yield is significantly below that of the retrospective series: 14.1% vs 20.3% ($p = 0.049$, chi-square test). This is not mirrored by a significant difference in ID rate between cohorts: 21.6 % vs 23.3% ($p = 0.25$, chi-square test).

Association between clinical features and detection of an ASD-contributing variant

To explain the probability of ASD-susceptibility variant detection from clinical features, we built a logistic regression model by grouping together both cohorts.

Missingness is often an issue regarding clinical data. Overall, the dataset presented an exhaustive picture of clinical features since only 65 (9.9%) probands suffered from any missing values. These were mostly about developmental milestones with 5.9% for language delay (3.7% for this prospective cohort alone) and 6.9% for walking delay (3.7% for this prospective cohort alone). On the other hand, the cognitive score could not be calculated for 189 probands because they were under 5 years of age, hence too young for meaningful intellectual assessment. Thereby, we computed two iterations of our model. The main one excludes the cognitive deficit score variable and is built on the subset of 588 probands with complete remaining clinical information. The second one is built on the subset of 415 patients over the age of 5 with the cognitive score variable and other clinical information available.

In the main model, clinical features associated with variant detection rate were walking delay (OR = 3.6, CI_{95%}: [2.2-5.9]) $p = 2.11 \cdot 10^{-7}$), an abnormal head circumference (OR = 3, CI_{95%}: [1.7-5.3], $p = 0.0001$), neurological symptoms (OR = 2.5, CI_{95%}: [1.3-4.6], $p = 0.004$) and birth complications (OR = 0.36, CI_{95%}: [0.1-0.9], $p = 0.045$).

Results from the second model were coherent to the main model, as displayed in table 3. The birth complication feature was not significantly associated anymore, however the OR remained in the close range (OR = 0.5, CI_{95%}: [0.2-1.3], $p = 0.21$). Of note, this feature is relatively rare (only 7% of probands), it is, then, probable that the second model lacks statistical power to detect the association. In addition, the cognitive deficit score was significantly associated with the detection rate (OR = 1.3, CI_{95%}: [1.02-1.8], $p = 0.04$).

Expectation questionnaires

A proportion of 245/300 (82%) mothers, 243/300 (81%) fathers and 36/300 (12%) probands filled the questionnaire. Of note, correlation between spouse responses was large ($r = 0.65$, CI_{95%}: [0.58-0.71], $p < 2.2 \cdot 10^{-16}$), which is expected because they often fill in the questionnaires at the same time, following the genetics consultation. In contrast, correlation between proband and parent responses was weaker

($r = 0.36$, $CI_{95\%}: [0.03-0.6]$, $p = 0.03$). No questions had more than 5% missing responses. Summaries of responses by questions are displayed in fig.1.

By design, Q1-Q6 and Q11 corresponded to an “interest for genetic testing” variable while Q7-Q10 corresponded to a “concerns about genetic testing” variable. When considered independently for each cluster, Chronbach's Alpha was respectively 0.82 ($IC_{95\%}: [0.79-0.85]$) and 0.85 ($IC_{95\%}: [0.82-0.88]$).

In that way, each responder had two composite scores: an “interest” one and a “concerns” one. Results for those two scores are displayed in fig.2b-c. Finally, each family could be placed on a dual axis chart, made up from the mean “interest” and “concerns” of the family members as show on fig.3a. Post-test questionnaires will be examined after the completion of the study as well as correlations with exome sequencing results.

Discussion

We have shown that in a cohort of ASD probands prospectively recruited from our clinical genetics departments, the rate of ASD-contributing variant detection was 14.1%, which the previously observed rate. Rate of ID was slightly lower albeit with a non-significant difference. We can still speculate that part of the reduced variant detection rate could result from less severe phenotypes.

The difference in the ASD-contributing variant detection rate between both series was mainly related to the CNV detection rate, which was lower when comparing both series each other. As WES CNV detection tools were not different, we hypothesize that there could be a difference in the study population between both series. CNV screening has been made available in a clinical setting before sequencing, and because inclusion criteria required patients not to carry a known ASD-contributing variant here, patients with an ASD-contributing CNV detected by array CGH before were not included and may not even have re-consulted in a genetics department, contrary to patients having underwent negative or unconvulsive array CGH. This might thus reduce the detection rate of CNV, knowing that the geographic population of recruitment was almost the same (15% of probands from the current series were recruited from an independent center (Caen) from the main one, used in the previous study (Rouen)). An alternative and putatively complementary hypothesis is that moving from DSM-IV to DSM-5 criteria may have led to including less severe patients with a genetic component somewhat

different from the one identified in trio-based ASD studies. However, this did not seem to translate into SNV/indels variant detection rates.

Several clinical features were associated with variant detection rates. It is rather expected that neurological symptoms and abnormal head circumference were good predictors of a variant detection, as they both can be part of the phenotypic spectrum of rare genetic neurodevelopmental disorders. Of more interest is that the strongest association is with a delay in walking without assistance as rare variants in ASD susceptibility genes are known to be associated with poorer motor skills development (Antaki et al., 2022; Buja et al., 2018). Finally, probands with history of neonatal complications have a decreased rate of detection. It is likely because neonatal complications have already been described as a strong ASD risk factor (Bölte et al., 2019), and it would be less likely to find a second one.

The main limit of our study is that the sample size remains modest. This could also contribute to explain, by sampling fluctuation, our lower rate of recurrent CNV detection. Additionally, increasing our sample size may improve the accuracy of our clinical score. In addition, WES was performed in the proband only, sequencing the parents at the same time in a trio-based study could have allowed us to increase the detection of *de novo* mutations or bi-allelic variants of interest, especially among missense variants. However, it is noteworthy that some of the ASD-contributing variants were inherited from an asymptomatic parent here. Trio-based studies may be associated with a risk for geneticists to focus on classical models for sporadic rare disease, i.e. *de novo* mutations, autosomal recessive inheritance, X-linked inheritance, and inherited heterozygous variants on autosomes might be overlooked in such a setting.

Finally, our questionnaire regarding parents and probands expectation showed good internal consistency. Interest was rather high while concerns were somewhat low in average which is consistent with the literature (Hayeems et al., 2016; Johannessen et al., 2017). It seems to be particularly important for parents to understand the causation of the ASD of their child, with a high perceived usefulness of genetic testing. In contrast, while some anxiety may be induced, few parents fear that the genetic test can be harmful. Obviously, a bias may exist from the fact that patients who care to come to the genetics consultation may be interested by genetics questions. However, psychiatrists were asked to systematically suggest to all patients with an ASD diagnosis to attend a genetics consultation, and some of the patients indeed attended without having previously requested

any kind of genetic counselling, thus following their psychiatrist's advice. However, this was not systematically noted at the time of inclusion in the questionnaire, so that this bias cannot be accurately measured.

Overall, our study provides insights into clinical utility of exome sequencing in ASD probands. In a prospective series of patients with ASD attending genetics consultations, the detection rate of ASD-contributing variants was 14.1%. Although our study may suffer from a bias based on the larger availability of diagnostic genome-wide CNV screening in France since previous reports, the rate of indels and single nucleotide variant-contributing to ASD was 11.6 % in exome data, which is consistent with previous retrospective series, despite some degree of extension of the ASD gene lists. Patients with abnormal growth parameters, neurological features, and/or developmental delay (especially motor skill developmental delays) as well as with intellectual disability, and patient without birth complications showed the highest likelihood of rare ASD-contributing variant detection and should be prioritized for exome/genome sequencing in a clinical setting. In addition, it is probably among these patients that the consequences for genetic counselling of detecting an ASD-contributing variant may be higher. Finally, clear information should be provided to patients to what to expect and not to expect from whole exome sequencing and different types of variants that can be identified in diverse clinical presentations, to help patients and their parents decide, depending on their personal clinical situation.

Acknowledgements.

We would like to thank all patients and their families as well as all psychiatrists and child neurologists having referred patients to the genetics consultation of both university hospitals.

This study was funded by a grant from the French ministry of health (PHRC-I 18-38, REDIA study).

Collaboration CEA-DRF-Jacob-CNRGH—CHU de Rouen. This work did benefit from support of the France Génomique National infrastructure, funded as part of the «Investissements d'Avenir» program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09).

References

Antaki, D., Guevara, J., Maihofer, A. X., Klein, M., Gujral, M., Grove, J., Carey, C. E., Hong, O., Arranz, M. J., Hervas, A., Corsello, C., Vaux, K. K., Muotri, A. R., Iakoucheva, L. M., Courchesne, E., Pierce, K., Gleeson, J. G., Robinson, E. B., Nievergelt,

C. M., & Sebat, J. (2022). A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. *Nature Genetics*, 1-9. <https://doi.org/10.1038/s41588-022-01064-5>

Backenroth, D., Homsy, J., Murillo, L. R., Glessner, J., Lin, E., Brueckner, M., Lifton, R., Goldmuntz, E., Chung, W. K., & Shen, Y. (2014). CANOES : Detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Research*, 42(12), e97. <https://doi.org/10.1093/nar/gku345>

Bishop, S. L., Farmer, C., Bal, V., Robinson, E. B., Willsey, A. J., Werling, D. M., Havdahl, K. A., Sanders, S. J., & Thurm, A. (2017). Identification of Developmental and Behavioral Markers Associated With Genetic Abnormalities in Autism Spectrum Disorder. *The American Journal of Psychiatry*, 174(6), 576-585. <https://doi.org/10.1176/appi.ajp.2017.16101115>

Buja, A., Volfovsky, N., Krieger, A. M., Lord, C., Lash, A. E., Wigler, M., & Iossifov, I. (2018). Damaging de novo mutations diminish motor skills in children on the autism spectrum. *Proceedings of the National Academy of Sciences of the United States of America*, 115(8), E1859-E1866. <https://doi.org/10.1073/pnas.1715427115>

C Yuen, R. K., Merico, D., Bookman, M., L Howe, J., Thiruvahindrapuram, B., Patel, R. V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., Pellecchia, G., Buchanan, J. A., Walker, S., Marshall, C. R., Uddin, M., Zarrei, M., Deneault, E., D'Abate, L., Chan, A. J. S., ... Scherer, S. W. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature Neuroscience*, 20(4), 602-611. <https://doi.org/10.1038/nn.4524>

Courchesne, E., Gazestani, V. H., & Lewis, N. E. (2020). PRENATAL ORIGINS OF ASD : THE WHEN, WHAT AND HOW OF ASD DEVELOPMENT. *Trends in neurosciences*, 43(5), 326-342. <https://doi.org/10.1016/j.tins.2020.03.005>

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., Kou, Y., Liu, L., Fromer, M., Walker, S., Singh, T., Klei, L., Kosmicki, J., Shih-Chen, F., Aleksic, B., Biscaldi, M., Bolton, P. F., Brownfeld, J. M., Cai, J., ... Buxbaum, J. D. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515(7526), 209-215. <https://doi.org/10.1038/nature13772>

Dong, S., Walker, M. F., Carriero, N. J., DiCola, M., Willsey, A. J., Ye, A. Y., Waqar, Z., Gonzalez, L. E., Overton, J. D., Frahm, S., Keaney, J. F., Teran, N. A., Dea, J., Mandell, J. D., Hus Bal, V., Sullivan, C. A., DiLullo, N. M., Khalil, R. O., Gockley, J., ... Sanders, S. J. (2014). De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. *Cell Reports*, 9(1), 16-23. <https://doi.org/10.1016/j.celrep.2014.08.068>

Fu, J. M., Satterstrom, F. K., Peng, M., Brand, H., Collins, R. L., Dong, S., Wamsley, B., Klei, L., Wang, L., Hao, S. P., Stevens, C. R., Cusick, C., Babadi, M., Banks, E., Collins, B., Dodge, S., Gabriel, S. B., Gauthier, L., Lee, S. K., ... Talkowski, M. E. (2022). Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nature Genetics*, 54(9), Art. 9. <https://doi.org/10.1038/s41588-022-01104-0>

Girirajan, S., Johnson, R. L., Tassone, F., Balciuniene, J., Katiyar, N., Fox, K., Baker, C., Srikanth, A., Yeoh, K. H., Khoo, S. J., Nauth, T. B., Hansen, R., Ritchie, M., Hertz-Picciotto, I., Eichler, E. E., Pessah, I. N., & Selleck, S. B. (2013). Global increases in both common and rare copy number load associated with autism. *Human Molecular Genetics*, 22(14), 2870-2880. <https://doi.org/10.1093/hmg/ddt136>

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awasthi, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., ... Børghlum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, 51(3), 431-444. <https://doi.org/10.1038/s41588-019-0344-8>

Harris, H. K., Sideridis, G. D., Barbaresi, W. J., & Harstad, E. (2020). Pathogenic Yield of Genetic Testing in Autism Spectrum Disorder. *Pediatrics*, 146(4), e20193211. <https://doi.org/10.1542/peds.2019-3211>

Hayeems, R. Z., Babul-Hirji, R., Hoang, N., Weksberg, R., & Shuman, C. (2016). Parents' Experience with Pediatric Microarray : Transferrable Lessons in the Era of Genomic Counseling. *Journal of Genetic Counseling*, 25(2), 298-304. <https://doi.org/10.1007/s10897-015-9871-3>

Husson, T., Lecoquierre, F., Cassinari, K., Charbonnier, C., Quenez, O., Goldenberg, A., Guerrot, A.-M., Richard, A.-C., Drouin-Garraud, V., Brehin, A.-C., Soleimani, M., Taton, R., Rotharmel, M., Rosier, A., Chambon, P., Le Meur, N., Joly-Helas, G., Saugier-veber, P., Boland, A., ... Champion, D. (2020). Rare genetic susceptibility variants assessment in autism spectrum disorder : Detection rate and practical use. *Translational Psychiatry*, 10(1), 77. <https://doi.org/10.1038/s41398-020-0760-7>

Iossifov, I., O’Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., Stessman, H. A., Witherspoon, K. T., Vives, L., Patterson, K. E., Smith, J. D., Paepier, B., Nickerson, D. A., Dea, J., Dong, S., Gonzalez, L. E., Mandell, J. D., Mane, S. M., Murtha, M. T., ... Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515(7526), Art. 7526. <https://doi.org/10.1038/nature13908>

Johannessen, J., Nærland, T., Hope, S., Torske, T., Høyland, A. L., Strohmaier, J., Heiberg, A., Rietschel, M., Djurovic, S., & Andreassen, O. A. (2017). Parents’ Attitudes toward Clinical Genetic Testing for Autism Spectrum Disorder—Data from a Norwegian Sample. *International Journal of Molecular Sciences*, 18(5), 1078. <https://doi.org/10.3390/ijms18051078>

Kaplanis, J., Samocha, K. E., Wiel, L., Zhang, Z., Arvai, K. J., Eberhardt, R. Y., Gallone, G., Lelieveld, S. H., Martin, H. C., McRae, J. F., Short, P. J., Torene, R. I., de Boer, E., Danecek, P., Gardner, E. J., Huang, N., Lord, J., Martincorena, I., Pfundt, R., ... Retterer, K. (2020). Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature*, 586(7831), 757-762. <https://doi.org/10.1038/s41586-020-2832-5>

Krumm, N., Turner, T. N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B. P., Stessman, H. A., He, Z.-X., Leal, S. M., Bernier, R., & Eichler, E. E. (2015). Excess of rare, inherited truncating mutations in autism. *Nature Genetics*, 47(6), 582-588. <https://doi.org/10.1038/ng.3303>

Llera, A., Brammer, M., Oakley, B., Tillmann, J., Zabihi, M., Amelink, J. S., Mei, T., Charman, T., Ecker, C., Dell’Acqua, F., Banaschewski, T., Moessnang, C., Baron-Cohen, S., Holt, R., Durston, S., Murphy, D., Loth, E., Buitelaar, J. K., Floris, D. L., & Beckmann, C. F. (2022). Evaluation of data imputation strategies in complex, deeply-phenotyped data sets : The case of the EU-AIMS Longitudinal European Autism Project. *BMC Medical Research Methodology*, 22(1), 229. <https://doi.org/10.1186/s12874-022-01656-z>

Miles, J. H., Takahashi, T. N., Bagby, S., Sahota, P. K., Vaslow, D. F., Wang, C. H., Hillman, R. E., & Farmer, J. E. (2005). Essential versus complex autism : Definition of fundamental prognostic subtypes. *American Journal of Medical Genetics. Part A*, 135(2), 171-180. <https://doi.org/10.1002/ajmg.a.30590>

Quenez, O., Cassinari, K., Coutant, S., Lecoquierre, F., Le Guennec, K., Rousseau, S., Richard, A.-C., Vasseur, S., Bouvignies, E., Bou, J., Lienard, G., Manase, S., Fourneaux, S., Drouot, N., Nguyen-Viet, V., Vezain, M., Chambon, P., Joly-Helas, G., Le Meur, N., ... Nicolas, G. (2021). Detection of copy-number variations from NGS data using read depth information : A diagnostic performance evaluation. *European Journal of Human Genetics: EJHG*, 29(1), 99-109. <https://doi.org/10.1038/s41431-020-0672-2>

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and Guidelines for the Interpretation of Sequence Variants : A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics*, 17(5), 405-424. <https://doi.org/10.1038/gim.2015.30>

Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., Kosmicki, J. A., Rehnström, K., Mallick, S., Kirby, A., Wall, D. P., MacArthur, D. G., Gabriel, S. B., DePristo, M., Purcell, S. M., Palotie, A., Boerwinkle, E., Buxbaum, J. D., Cook, E. H., ... Daly, M. J. (2014). A framework for the interpretation of de novo mutation in human disease. *Nature Genetics*, 46(9), 944-950. <https://doi.org/10.1038/ng.3050>

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., Walker, M. F., Ober, G. T., Teran, N. A., Song, Y., El-Fishawy, P., Murtha, R. C., Choi, M., Overton, J. D., Bjornson, R. D., ... State, M. W. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485(7397), 237-241. <https://doi.org/10.1038/nature10945>

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., ... Buxbaum, J. D. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, 180(3), 568-584.e23. <https://doi.org/10.1016/j.cell.2019.12.036>

Shenouda, J., Barrett, E., Davidow, A. L., Halperin, W., Silenzio, V. M. B., & Zahorodny, W. (2022). Prevalence of autism spectrum disorder in a large, diverse metropolitan area : Variation by sociodemographic factors. *Autism Research: Official Journal of the International Society for Autism Research*, 15(1), 146-155. <https://doi.org/10.1002/aur.2628>

Stessman, H. A. F., Xiong, B., Coe, B. P., Wang, T., Hoekzema, K., Fenckova, M., Kvarnung, M., Gerds, J., Trinh, S., Cosemans, N., Vives, L., Lin, J., Turner, T. N., Santen, G., Ruivenkamp, C., Kriek, M., van Haeringen, A., Aten, E., Friend, K., ... Eichler, E. E. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nature Genetics*, 49(4), 515-526. <https://doi.org/10.1038/ng.3792>

Sun, X., Allison, C., Wei, L., Matthews, F. E., Auyeung, B., Wu, Y. Y., Griffiths, S., Zhang, J., Baron-Cohen, S., & Brayne, C. (2019). Autism prevalence in China is comparable to Western prevalence. *Molecular Autism*, 10, 7. <https://doi.org/10.1186/s13229-018-0246-0>

Tammimies, K., Marshall, C. R., Walker, S., Kaur, G., Thiruvahindrapuram, B., Lionel, A. C., Yuen, R. K. C., Uddin, M., Roberts, W., Weksberg, R., Woodbury-Smith, M., Zwaigenbaum, L., Anagnostou, E., Wang, Z., Wei, J., Howe, J. L., Gazzellone, M. J., Lau, L., Sung, W. W. L., ... Fernandez, B. A. (2015). Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. *JAMA*, 314(9), 895-903. <https://doi.org/10.1001/jama.2015.10078>

Wang, T., Guo, H., Xiong, B., Stessman, H. A. F., Wu, H., Coe, B. P., Turner, T. N., Liu, Y., Zhao, W., Hoekzema, K., Vives, L., Xia, L., Tang, M., Ou, J., Chen, B., Shen, Y., Xun, G., Long, M., Lin, J., ... Eichler, E. E. (2016). De novo genic mutations among a Chinese autism spectrum disorder cohort. *Nature Communications*, 7, 13316. <https://doi.org/10.1038/ncomms13316>

Figure 1. Barplot representing responses distribution by questions.

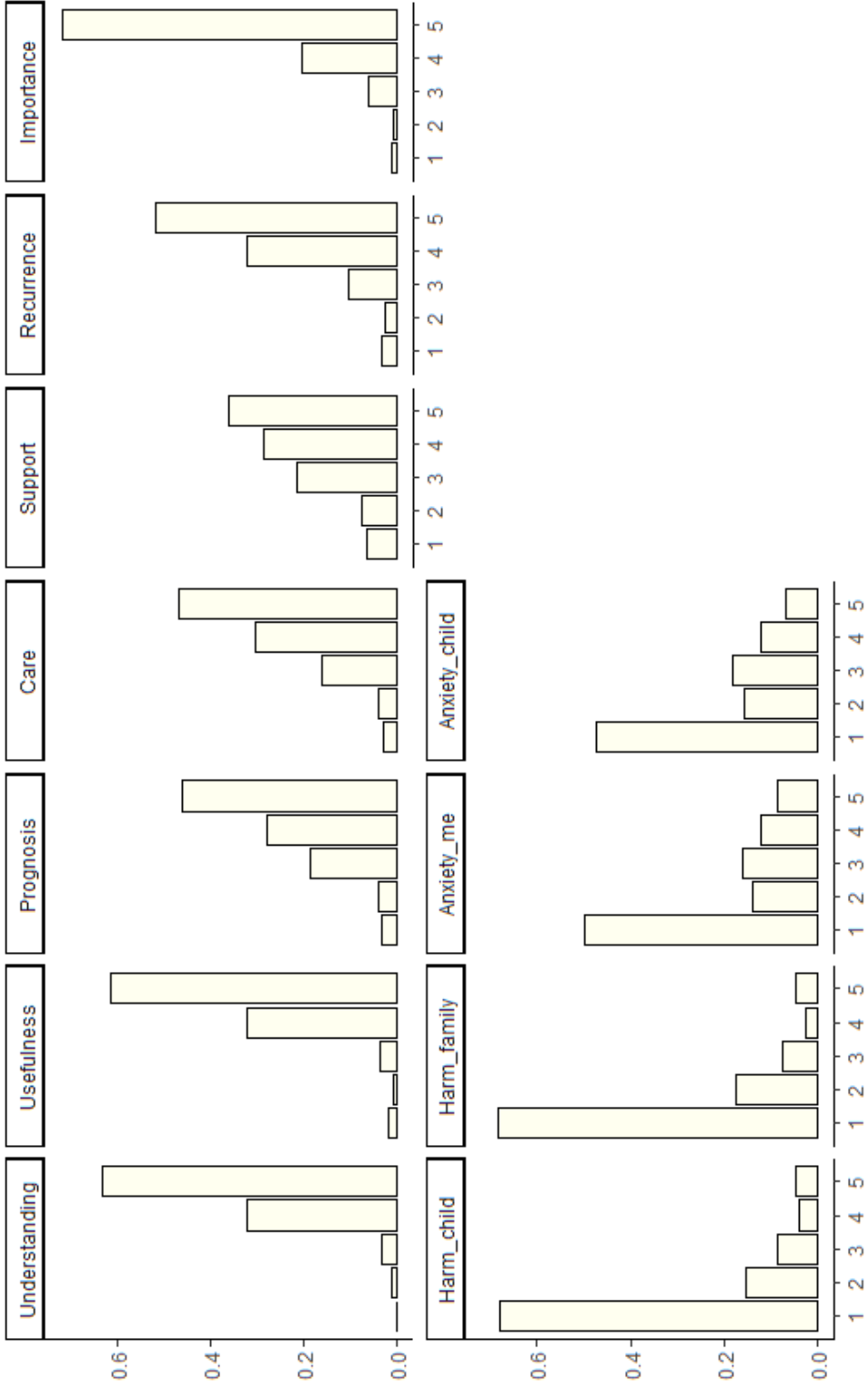


Figure 2. Expectation for the genetic test’s questionnaire results. a. scatterplot representing each family depending on these two dimensions. Each dot is a composite score from the mean of each family members who fill in a questionnaire. Bulk of the families show a relatively high interest and few concerns about the genetic test. c-d. Score distributions of the “interests” (Q1-6 and Q11) and “concerns” (Q7-10) clusters.

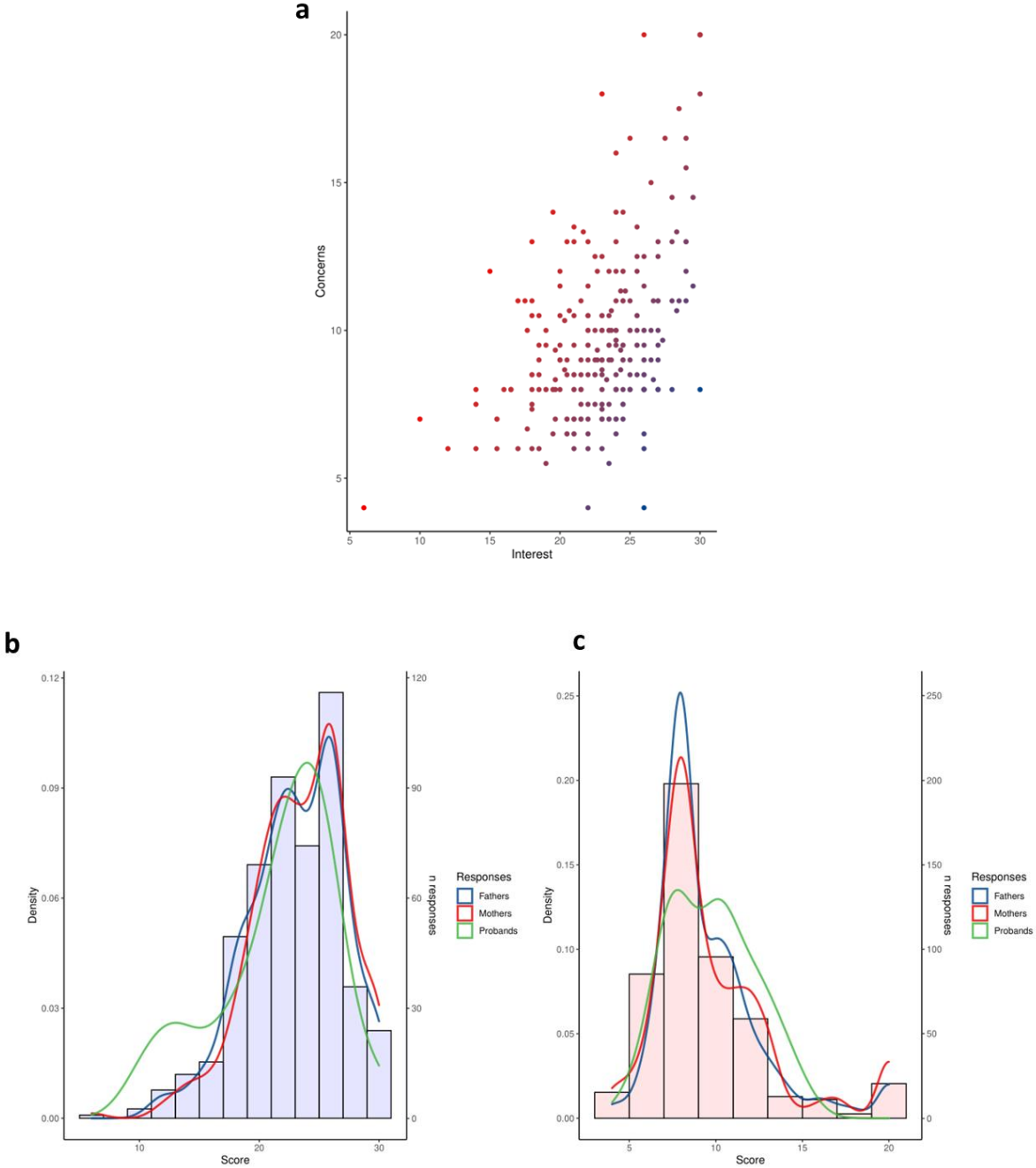


Table 1. Samples clinical characteristics

N (%)	Retrospective cohort (n = 356)	Prospective cohort (n = 297)
Male	295 (82.9%)	234 (78.8%)
Cognitive score 0	13 (3.6%)	2 (0.7%)
Cognitive score 1	115 (32.3%)	97 (32.6%)
Cognitive score 2	34 (9.5%)	56 (18.8%)
Cognitive score 3	58 (16.3%)	49 (16,5%)
Cognitive score 4	27 (5.9%)	13 (4.4%)
Language delay	241 (67.8%)	206 (69.4%)
Walking delay	84 (23.6%)	63 (21.2%)
Physical anomaly < 1	219 (61.5%)	153 (51.5%)
Neurological syndrome	40 (11.2%)	29 (9.8%)
Abnormal head circumference	51 (14.3%)	39 (13.1%)
Neonatal complication	18 (5%)	31 (10.4%)
Family history	105 (29.5%)	97 (32.6%)

Table2a. Nucleotide variants in ASD-susceptibility genes detected among the 300 ASD probands

ID	Gene	Variant	Protein_consequence	Inheritance
ACA_5	POGZ	NM_001194937.1:c.1391A>G	p.(Tyr464Cys)	DN
ACA_13	ASXL2	NM_018263.4:c.2326A>C	p.(Thr776Pro)	DN
ACA_14	ASXL3	NM_030632.2:c.4409del	p.(Pro1470Argfs*12)	DN
ACA_18	PQBP1	NM_001167992.1:c.992C>T	p.(Arg143Trp)	M
ARO_623	KCNMA1	NM_002247.3:c.1845C>G	p.(Phe615Leu)	DN
ARO_629	RORB	NM_006914.3:c.971A>G	p.(Tyr324Cys)	DN
ARO_635	CHD2	NM_001271.3:c.3427T>G	p.(Tyr1143Asp)	DN
ARO_643	KDM6B	NM_001080424.2:c.4750A>C	p.(Asn1584His)	DN
ARO_646	PTEN	NM_000314.6:c.1027-2A>C	p.(?)	DN
ARO_650	RFX3	NM_134428.2:c.1358-8A>G	p.(?)	DN
ARO_664	SETD1A	NM_014712.2:c.2981_2982del	p.(Val994Glyfs*28)	M
ARO_669	SCN2A	NM_001040143.1:c.2548C>T	p.(Arg850*)	DN
ARO_670	CSNK2B	NM_001282385.1:c.72+1G>T	p.(?)	DN
ARO_684	KAT6A	NM_006766.4:c.4674C>A	p.(Ser1558Arg)	DN
ARO_710	TCF20	NM_181492.2:c.2803C>T	p.(Gln395*)	DN
ARO_712	KDM6A	NM_001291418.1:c.2797dup	p.(Trp933Leufs*13)	DN
ARO_732	KDM5B	NM_001314042.1:c.883C>T	p.(Arg295*)	DN
ARO_739	PBX1	NM_002585.4:c.550C>T	p.(Arg184*)	DN
ARO_744	CIC	NM_001304815.1:c.4980_4981insCTTCC	p.(Thr1661fs)	M
ARO_749	MEF2C	NM_001308002.1:c.301C>A	p.(Pro101Thr)	DN
ARO_762	GIGYF1	NM_022574.5:c.332del	p.(LeuArgfs*234)	DN
ARO_772	SMARCB1	NM_001007468.2:c.541C>T	p.(Arg181Trp)	DN
ARO_783	NBEA	NM_015678.4:c.5206dupA	p.(Ser176fs)	DN
ARO_791	NAA10	NM_001256120.1:c.217C>T	p.(Arg76Cys)	DN
ARO_804	EED	NM_003797.4:c.767T>C	p.(Met256Thr)	DN
ARO_814	DEPDC5	NM_014662.4:c.2485C>T	p.(Arg2899*)	P
ARO_827	ANK2	NM_001148.4:c.9154_9161del	p.(Asp3052fs)	P
ARO_830	TRAPPC9	NM_031466.6:c.3214C>T/NM_031466.6:c.227C>T	p.(Arg1072*)/p.(Pro76Leu)	AR
ARO_831	SCN2A	NM_021007.2:c.3296_3317del	p.(Pro1099fs)	DN
ARO_844	SKI	NM_003036.4:c.2147_2156dup	p.(Ala720Profs*114)	DN
ARO_869	KMT2C	NM_170606.2:c.5716C>T	p.(Arg1906*)	DN
ARO_870	CASZ1	NM_001079843.3:c.4089delC	p.(Met1364fs)	M

Table2b. CNVs detected in ASD-susceptibility genes or loci among the 300 ASD probands

ID	Locus/gene	Inheritance
ACA_1	22q11.23 dup	M
ACA_34	Large chrX del	M
ARO_651	16p11.2 dup	P
ARO_833	STXBP1 del (E27)	DN
ARO_871	4q21.3 del (WDFY3)	M
ARO_57	15q13.3dup	P
ARO_653	4q35 del/3p26dup	M
ARO_619	17q12 del (HNF1B)	DN
ARO_745	15q11.2 BP1-BP2 del	P
ARO_759	16p13.11 del	P

Table 3. Logistic regression model of the probability of ASD-contributing variant detection based on clinical features.

Clinical feature	Main model (n = 588)			Model with probands > 5 years (n = 415)		
	OR	95% CI	p.value	OR	95% CI	p.value
Sex	1.4	0.8-2.4	0.18	1.6	0.8-2.9	0.14
Physical anomaly score	1.1	0.8-1.5	0.59	1.2	0.8-1.7	0.29
Abnormal head circumference	3	1.7-5.3	0.0001	2.4	1.2-4.7	0.01
Neurological symptoms	2.4	1.3-4.6	0.004	2.4	1.2-4.7	0.01
Birth complication	0.36	0.1-0.9	0.04	0.5	0.2-1.3	0.21
Language delay	0.8	0.5-1.4	0.46	0.9	0.5-1.8	0.8
Walking delay	3.32	2.2-5.9	2.11×10^{-7}	2.2	1.3-4	0.004
Cognitive deficit score	NA	NA	NA	1.3	1.02-1.8	0.04

INTERPRETATION DE VSI A L'AIDE D'EPISIGNATURES

Nous débuterons cette partie par une introduction sur la pertinence de l'étude de la méthylation de l'ADN pour classer des VSI dans les syndromes neurodéveloppementaux. Nous détaillerons par la suite le *pipeline* bioinformatique nécessaire pour l'analyse des données de méthylation.

Par la suite, nous décrirons les travaux réalisés dans le cadre de ma thèse concernant la recherche d'une épisignature originale chez les patients portant une variation pathogène du gène *MBD5*. Nous verrons les difficultés méthodologiques rencontrées. Ce travail sera difficilement valorisable par une publication scientifique. Enfin, nous présenterons le manuscrit d'une étude s'attachant à l'évaluer plusieurs épisignatures déjà publiées. Celui-ci sera soumis en vue d'une publication scientifique d'ici la fin d'année après relecture finale par les différents auteurs.

Chapitre I. Analyse de méthylation dans les syndromes neurodéveloppementaux

i. Quel biomarqueur pour aider à classifier des VSI ?

Du TSA à la machinerie épigénétique

Par essence, les données de séquençage d'exome produisent un grand nombre de VSI (Petersen et al., 2017). Pour un certain nombre d'entre eux, un faisceau d'arguments pointe vers la pathogénicité sans atteindre un niveau de certitude suffisant. Par exemple, au cours de l'étude REDIA, nous avons détecté 20 variations, pour lesquels nous avons de fortes présomptions quant à leur implication dans le phénotype. Sans prétention d'exhaustivité, les situations prototypiques que nous avons rencontrées sont la détection :

- D'une variation faux-sens ou d'une indel en phase, prédite délétère par les logiciels bioinformatiques, inconnue des bases de données de témoins mais héritée de parents normotypiques.
- D'une variation faux-sens ou d'une indel en phase apparue *de novo*, dans un gène avec excès de PTVs *de novo* dans le TSA
- D'une variation prédite comme impactant fortement l'épissage
- D'une variation impactant de façon délétère un exon alternatif ou d'une PTVs située dans le dernier exon ou dans les 50 dernières pb du gène et qui n'entraînerait pas de NMD.
- D'une PTV dans un gène à pLI élevée non recensé en pathologie humaine

Il est ici fréquent qu'un argument supplémentaire soit nécessaire pour conclure à la pathogénicité (Richards et al., 2015b). Théoriquement, les

tests fonctionnels représentent une alternative idéale pour aider à la classification de ces variations. Ils permettent de mesurer les conséquences fonctionnelles d'une variation sur l'ARN messager ou sur la protéine dans un modèle cellulaire ou animal (Marwaha *et al.*, 2022). Outre une amélioration des connaissances fondamentales sur la physiopathologie, ces essais lorsqu'ils sont bien menés et reproductibles représentent un argument de pathogénicité important et souvent décisif (Richards *et al.*, 2015a). Dans les situations cliniques où quelques gènes majeurs sont impliqués et très bien décrits, la mise au point de tels outils est inestimable. C'est le cas en oncogénétique, par exemple avec le gène *BRCA1* impliqué dans la prédisposition génétique aux cancers du sein et de l'ovaire (Woods *et al.*, 2016).

En génétique du neurodéveloppement en général, et du TSA en particulier, la situation ne se prête pas à l'utilisation de ce type de techniques. Plusieurs centaines, voire milliers, de gènes sont impliqués, dont l'activité et les domaines fonctionnels sont souvent très mal caractérisés. La mise au point de tout autant de tests fonctionnels est actuellement trop complexe et dispendieuse, particulièrement à l'échelle d'un laboratoire unique.

Plusieurs solutions alternatives existent pour pallier ce problème. Selon les situations, les études transcriptomiques ou métabolomiques (Marwaha *et al.*, 2022) peuvent aider à reclassifier les variations. Cependant, une approche prometteuse dans le cadre du TSA est l'étude des particularités épigénétiques. En effet, les gènes impliqués dans le TSA peuvent être schématiquement classés en deux catégories, représentées en Figure 28 (Courchesne *et al.*, 2020; Satterstrom, Kosmicki, Wang, Breen, De Rubeis, An, Peng, Collins, Grove, Klei, Stevens, Reichert, Mulhern, Artomov,

Gerges, Sheppard, Xu, Bhaduri, Norman, Brand, Schwartz, Nguyen, Guerrero, Dias, Autism Sequencing Consortium, et al., 2020). La première – intuitive pour un trouble neurodéveloppemental – est un groupe de gènes impliqués dans la communication neuronale. Mais il ne représente qu’environ un tiers des gènes. L’autre groupe représente des gènes associés plus globalement à la régulation de l’expression génique. Ces gènes sont généralement d’expression ubiquitaire, avec un pic lors des deux premiers trimestres de la période fœtale et leur atteinte pourrait être associée à des formes plus sévères en moyenne (Satterstrom, Kosmicki, Wang, Breen, De Rubeis, An, Peng, Collins, Grove, Klei, Stevens, Reichert, Mulhern, Artomov, Gerges, Sheppard, Xu, Bhaduri, Norman, Brand, Schwartz, Nguyen, Guerrero, Dias, Betancur, et al., 2020).

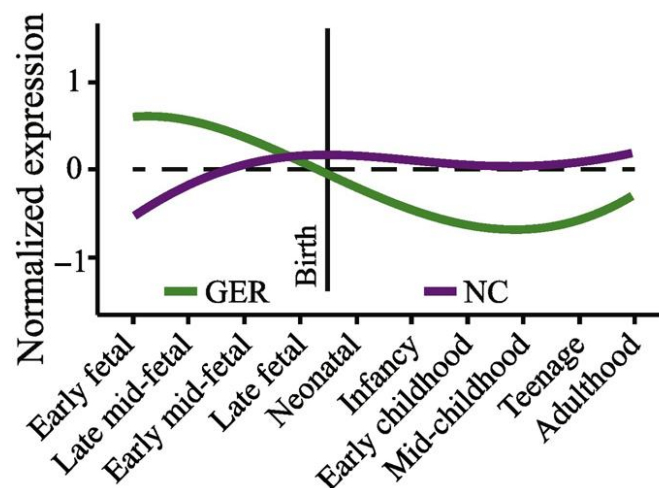


Figure 28. Expression normalisée des gènes associés au TSA. Environ 2/3 des gènes associés au TSA sont impliqués dans la régulation de l’expression. Ils sont préférentiellement exprimés en période in-utero précoce. Le 1/3 restant est composé de gènes impliqués dans la communication neuronale, d’expression en moyenne plus tardive. GER : Gene Expression Regulator – gènes impliqués dans la régulation de l’expression ; NC : Neuronal Communication – gènes impliqués dans la communication neuronale et la construction synaptique. Extrait de Courchesne et al., 2020.

Directement ou non, ces gènes interagissent avec la machinerie épigénétique pour contrôler l'expression d'autres gènes (Berdasco & Esteller, 2013; Gibney & Nolan, 2010; Peixoto *et al.*, 2020). Le terme machinerie épigénétique est une désignation générique pour tout un ensemble de systèmes protéiques qui modifient l'ADN en préservant sa séquence codante (*i.e* la succession singulière des bases azotées d'un individu). Elle permet de faire le lien entre un patrimoine génétique stable et la diversité des types cellulaires et des conditions environnementales auxquelles ceux-ci doivent s'adapter. Elle comprend les complexes protéiques initiant, maintenant, modifiant et interagissant avec les quatre composants principaux de l'épigénétique que sont les ARN non-codants, les variations de structure tridimensionnelle de la chromatine, les modifications post-traductionnelles des histones et l'apposition de marques sur l'ADN (Cavalli & Heard, 2019). A l'échelle du TSA, l'étude des particularités épigénétiques présentent un champ d'étude d'autant plus fertile, que celles-ci pourraient représenter des cibles thérapeutiques (Mbadiwe & Millis, 2013; Miyake *et al.*, 2012; Siu & Weksberg, 2017).

L'intérêt particulier de la méthylation de l'ADN

La première mention de la méthylation de l'ADN comme une modification épigénétique potentielle date de 1975 (Holliday & Pugh, 1975; Riggs, 1975). Depuis, elle représente la marque épigénétique la plus étudiée. L'essentiel des travaux chez le mammifère concernent l'ajout d'un groupement méthyl sur le 5^{ème} carbone d'une cytosine précédé d'une guanine (abrégé usuellement en CpG) sur l'ADN. La méthylation des autres nucléotides est rare dans le monde animal et sa fonction n'est pas connue (Jones, 2012). Nous nous concentrerons donc sur la 5-MethylCytosine (5-mc) dans un contexte CpG. Ces doublés cytosine-guanine sont globalement

déplétés dans le génome car la 5-mc a tendance à être spontanément désaminée en thymine. Il a donc été supposé que les CpG conservés sont enrichis en éléments fonctionnellement actifs (Jones, 2012). *A contrario*, il existe au niveau des régions promotrices d'environ 75% des gènes des zones de quelques centaines de pb enrichies en CpG (Saxonov *et al.*, 2006). Ces régions appelées îlots CpG, très conservées dans l'évolution (Long *et al.*, 2016), sont globalement appauvries en 5-mc mais lorsqu'elles sont méthylées, elles participent à l'inhibition de la transcription du gène en regard. Ce phénomène est particulièrement marqué lorsque le site d'initiation de la transcription est à proximité directe d'un îlot (Jones, 2012). La lyonisation du deuxième chromosome X chez les femelles mammifères est l'exemple le plus connu de l'inactivation de gènes par le biais de la méthylation des promoteurs (Deaton & Bird, 2011).

Les recherches sur la méthylation de l'ADN se sont longtemps concentrées sur la répression de la transcription. Toutefois, des études plus récentes ont révélé une action plus large et complexe de cette modification épigénétique. Les fonctions et les différentes caractéristiques de la méthylation de l'ADN sont représentées dans les Figures 29 et 30.

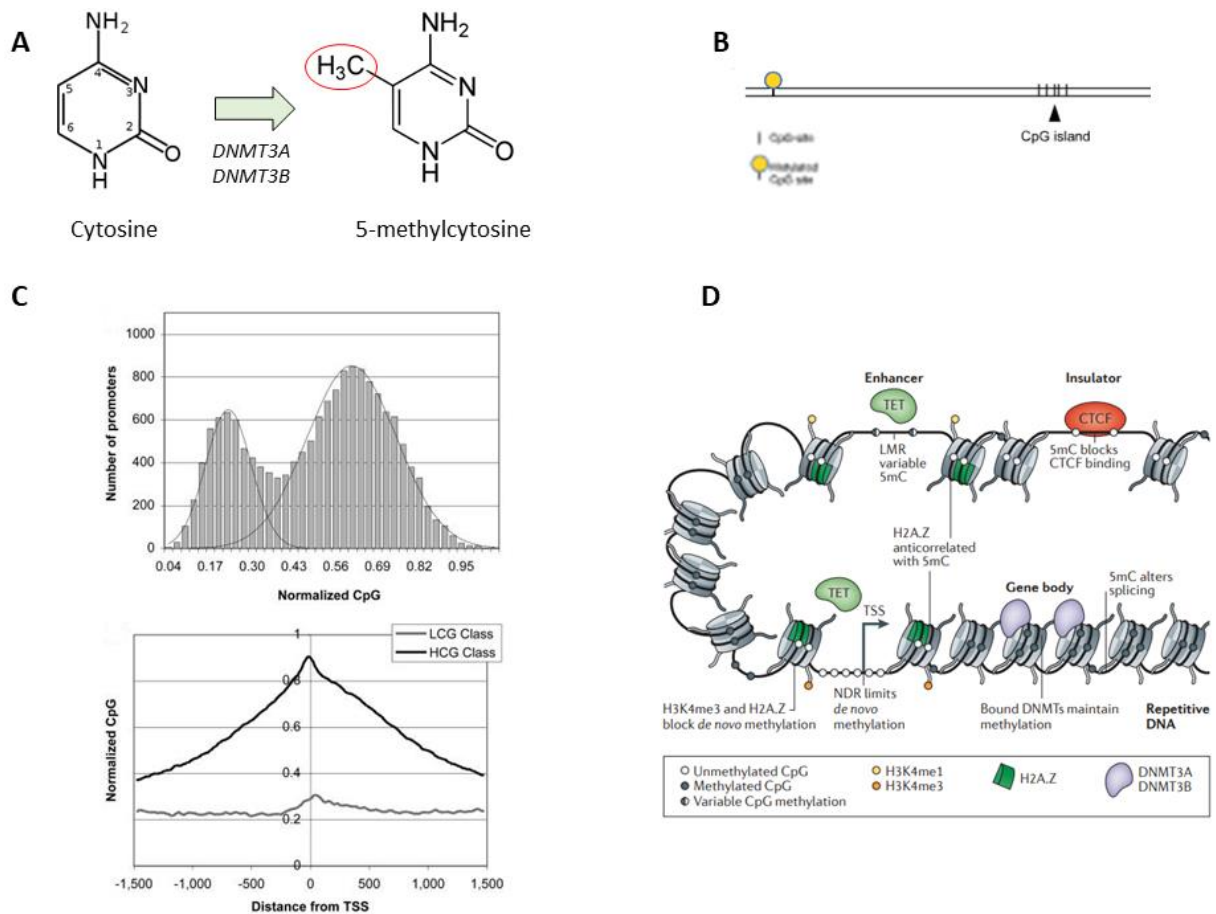


Figure 29. Caractéristiques de la méthylation de l'ADN. (A) La méthylation de novo de l'ADN est catalysée par les protéines DNMT3A et B ; (B) il existe une raréfaction des CpG dans l'ADN, sauf en regard d'environ 75% des promoteurs des gènes où des zones denses en CpG existent (les îlots). Ils n'y sont généralement pas méthylés contrairement au reste de l'ADN ; (A-B) tiré de Wikipedia sous licence creative commons CC BY-SA 3.0 ; (C) distribution des promoteurs par rapport à leur taux de CpG. Le taux de CpG observé est normalisé en fonction du taux attendu vu la composition en CG de la séquence. On observe un taux de CpG plus important pour $\frac{3}{4}$ des gènes, ceux présentant des îlots CpG au niveau des promoteurs (en haut). Pour les promoteurs enrichis en CpG, cet enrichissement est particulièrement important à proximité des sites d'initiation de la transcription (TSS). Extrait de Saxonov et al., 2006 ; (D) répartition de la 5-mc sur la chromatine et rôle dans la régulation de l'expression. Les régions déméthylées permettent le recrutement des insulator et enhancer et l'initiation de la transcription. D'après Jones, 2012.

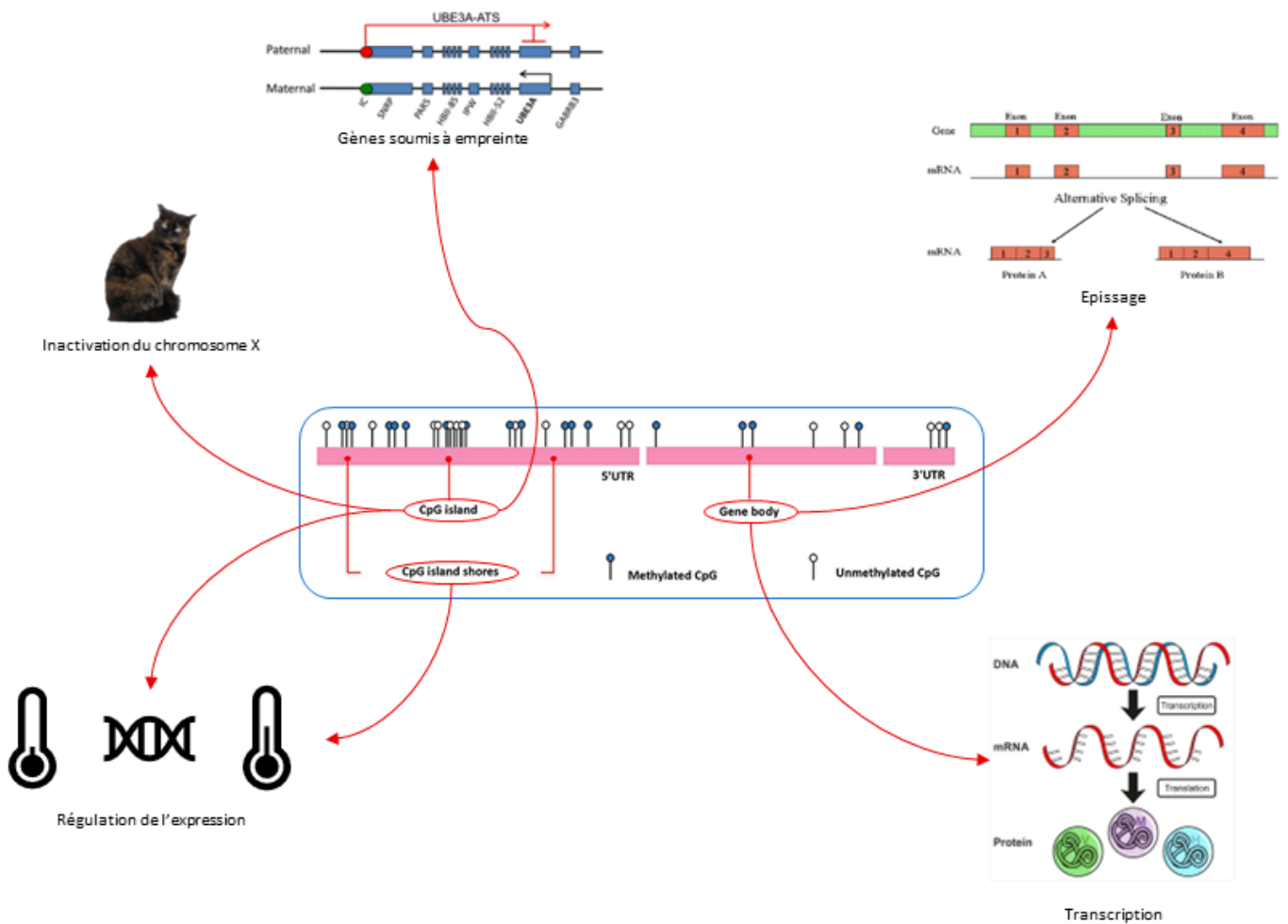


Figure 30. Différentes fonctions de la méthylation de l'ADN. La méthylation des îlots CpG est impliquée, sans en être l'initiatrice, dans l'inactivation stable et à long termes des gènes soumis à empreinte, du second chromosome X des femelles mammifères – lorsqu'un gène de couleur est situé sur le chromosome X, le motif du pelage peut-être une illustration macroscopique de l'inactivation aléatoire de ce chromosome dans les cellules – et plus globalement de gènes actifs uniquement lors de l'embryogénèse. Par ailleurs la méthylation d'un promoteur au dépend d'un autre permet également l'activation préférentielle d'un transcript alternatif. La méthylation des CpG en dehors des îlots, et tout particulièrement au niveau des « shores » de part et d'autre est impliquée dans la régulation dynamique et tissu spécifique des cellules. Enfin, le degré de méthylation du corps des gènes est proportionnel au niveau de transcription : la méthylation empêche l'initiation mais pas l'élongation de la transcription. La méthylation du corps du gène exprimé permet d'éviter l'activation d'éventuels éléments transposables en son sein et serait impliqué dans l'épissage. Informations issues de Jones, 2012. D'après Fu et al., 2017, Godwill, 2015, Khatri & Man, 2019.

La méthylation de l'ADN comme biomarqueur de pathogénicité

Comme nous l'avons vu, la méthylation de l'ADN est un acteur majeur et ubiquitaire de la régulation de l'expression. Plusieurs gènes associés au TSA sont directement impliqués dans la régulation de la localisation de 5-mc. Par exemple, le gène DNMT3A (DNA Methyltransferase 3 Alpha) présente un excès de variations génétiques délétères de novo dans le TSA. Ce gène catalyse la méthylation de novo d'une cytosine en 5-mc.

D'autres codent pour des protéines qui interagissent directement avec l'ADN méthylé pour effectuer leurs actions comme MBD5 (Methyl-CpG Binding Domain 5), un gène impliqué dans un syndrome neurodéveloppemental aspécifique pour lequel le TSA fait partie du spectre syndromique (Ichino et al., 2021).

Plus généralement, il est possible de poser l'hypothèse que la dysfonction d'une forte proportion des gènes impliqués dans la régulation de l'expression impliquera des modifications extensives du pattern de méthylation de l'ADN. Ce postulat repose sur l'existence i) de nombreuses couches d'intégration de l'information au niveau cellulaire aboutissant à une adaptation de l'expression génique, ii) de l'interrelation et de l'existence de voies de suppléances et de rétroactions sur les différents mécanismes de régulation de celle-ci, iii) et de leur aboutissement en fin de compte sur des modifications épigénétiques ou l'interaction avec celles-ci (Law & Holland, 2019; Marchal & Miotto, 2015; Murr, 2010). De plus, des modifications dynamiques et de particulièrement grande ampleur de l'épigénomes sont observées dans les phases précoces de l'embryogénèse (Gopinathan & Diekwisch, 2022) puis dans les premières années de vie (Urdinguio et al., 2016; D. Wang et al., 2012), tout en étant relativement stable par la suite (Pérez et al., 2019). Le développement harmonieux du fœtus puis du

nourrisson dépend donc de la régulation de la machinerie épigénétique, lors des deux pics d'activité de la majorité des gènes associés au TSA. On peut également supposer que plus une dysrégulation de la méthylation est précoce au cours de l'embryogénèse plus elle sera ubiquitaire au niveau des différents tissus.

L'étude pangénomique de la méthylation de l'ADN a donc été proposée comme un biomarqueur d'intérêt pour participer à la classification des VSI de gènes impliqués dans la régulation de l'expression (Marwaha et al., 2022; Sadikovic et al., 2020, 2021). L'idée n'est pas de corrélérer le niveau de méthylation d'une unique position du génome à une condition mais plutôt un pattern reconnaissable d'altérations plus ou moins modérées de la méthylation à l'échelle pangénomique à la présence d'un facteur de risque génétique dans un gène donné. Pour le dire autrement, l'idée n'est pas d'effectuer une epigenome-wide association study (ou EWAS par référence aux GWAS) (Wei et al., 2021) mais bien de détecter différentes signatures épigénétiques (ou épisignatures) comme biomarqueurs de l'altération d'un gène donné.

Le concept d'épisignature

La création d'une épisignature peut être résumée par les deux étapes suivantes : i) définir un panel de positions différentiellement méthylées entre cas et contrôles d'une cohorte d'apprentissage et ii) utiliser ce panel pour prédire le phénotype des cohortes suivantes. Si les « cas » sont des patients diagnostiqués avec des altérations pathogènes d'un gène spécifique, alors cette signature pourra être utilisée pour aider à la classification des VSI de ce gène (Haghshenas et al., 2020; Sadikovic et al., 2019, 2021). Ces positions différentiellement méthylées sont donc des marqueurs de pathogénicité. Certains auteurs vont jusqu'à considérer les

épisignatures comme des annotations fonctionnelles de l'impact transcriptionnel des variations pathogènes renseignant sur la pathogénicité de la maladie (Sadikovic et al., 2020). Un exemple théorique est illustré dans la Figure 31.

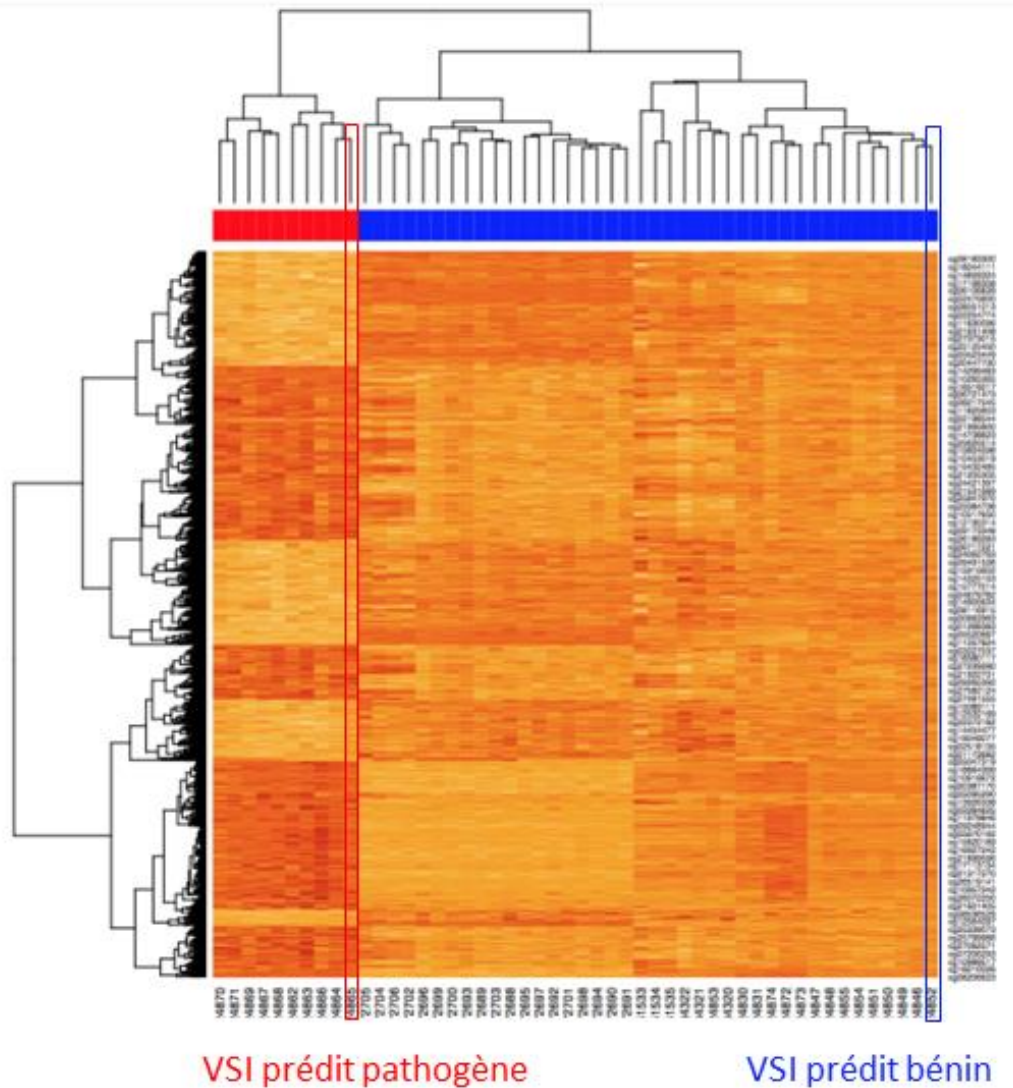


Figure 31. Illustration théorique du concept d'épissignature pour aider à la classification de deux VSI. Données analysées par classification hiérarchique ascendante et visualisées sous la forme d'une heatmap. Chaque ligne représente une position différenciellement méthylée entre cas et contrôles, chaque colonne représente un individu. Les porteurs de variations pathogènes d'un même gène sont représentés en rouge, alors que les témoins sont en bleu. Dans cet exemple théorique, l'un des VSI ségrége avec les cas, à gauche, et est donc prédit pathogène. Ce n'est pas le cas du deuxième qui ségrége avec les témoins, à droite.

Il s'agit d'une notion à différencier de l'impact de l'exposition au tabac (Bollepalli et al., 2019) ou de l'âge (Horvath, 2013) sur la méthylation. D'une part car il s'agit de facteurs dimensionnels. D'autre part car dans le cas des syndromes neurodéveloppementaux, le tissu incriminé (le cerveau) n'est pas accessible.

Un certain nombre de technologies ont été développées pour l'analyse de méthylation pangénomique. Elles ont été revues en détail par Li & Tollefsbol (S. Li & Tollefsbol, 2021). Brièvement, les premières technologies du début des années 2010 comme la *liquid chromatography-mass spectroscopy* permettaient la mesure du taux global de méthylation d'un tissu. Le développement, par la suite, des techniques immunohistochimiques n'a pas été probant. De fait, leurs performances, tant d'un point de vue de la résolution spatiale que du caractère quantitatif des mesures étaient considérées comme décevantes. En parallèle, des équipes ont conçu des mesures par enzyme de restriction. Brièvement, le principe repose sur l'utilisation conjointe de deux isoschizomères, l'une sensible à la méthylation et l'autre non. Celles-ci peuvent détecter le niveau de méthylation à la résolution d'une base avec une sensibilité très importante. Néanmoins, seules les CpG à proximité des sites de restriction peuvent être évaluées ce qui limite le choix des sites accessibles.

C'est l'application du NGS à l'analyse de méthylation qui a donné accès à une méthode pangénomique de référence pour mesurer le niveau de méthylation de chacune des positions. Le principe repose sur le traitement au bisulfite de l'ADN qui vient transcrire les cytosines méthylées en thymine. En comparant les résultats de séquençage de molécules d'ADN traités ou non au bisulfite, il est possible d'inférer le caractère méthylé ou non de chaque position.

En revanche, contrairement à son application au séquençage de l'ADN génomique, le NGS appliqué à la méthylation est resté très dispendieux : par essence, cette méthode nécessite la réalisation de deux WGS par échantillon. Son accessibilité est donc limitée avec peu d'études – rarement reproduites – et des tailles d'effectifs faibles.

L'utilisation des technologies de génotypage sur puce forme une alternative efficace pour la conception d'épisingatures. Ces puces ont la capacité de génotyper de manière quantitative plusieurs centaines de milliers de CpG réparties sur le génome, avec une couverture plus dense dans les régions géniques et péri-géniques en fournissant une information sur leur niveau de méthylation. Dans la mesure où il existe une corrélation spatiale de méthylation des CpG (Luo et al., 2020), le génotypage d'une portion de ceux-ci constitue un compromis efficace. Cette corrélation est particulièrement forte à l'échelle de la kilobase ce qui pourrait représenter des unités fonctionnelles (Bell et al., 2011; Lehne et al., 2015). Comme représenté dans la Figure 32, c'est la technologie la plus adaptée pour l'étude de grands volumes d'échantillons (S. Li & Tollefsbol, 2021; Pidsley et al., 2016).

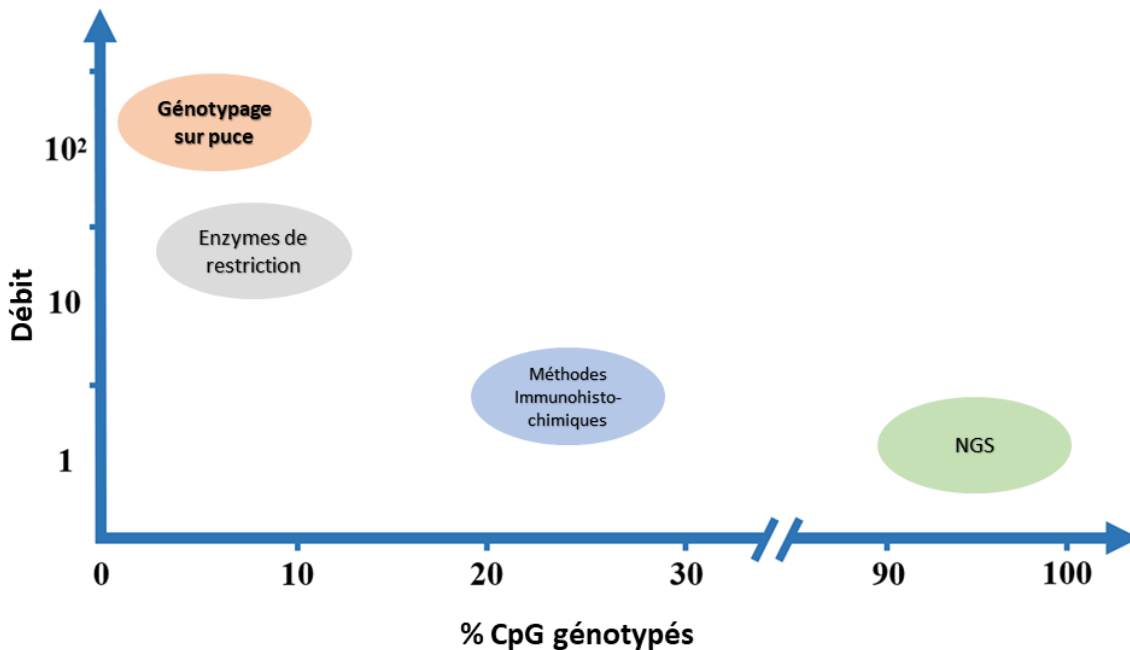


Figure 32. Comparaison de l'efficacité de différentes technologies d'analyse de méthylation pangénomique. Le génotypage sur puce est la méthode présentant le meilleur débit (en nombre d'échantillons par analyse) tout en conservant une couverture satisfaisante. De plus, contrairement à la méthode par enzymes de restriction, les positions génotypées peuvent être sélectionnées lors de la conception de la puce et les régions d'intérêts être, ainsi, priorisées. Adapté de Li et Tollefsbol, 2021.

En pratique, à ce jour, une seule technologie est commercialisée, par la société Illumina (Bibikova *et al.*, 2009). Les différentes itérations de cette technologie permettent le génotypage d'environ 25 000 (*HumanMethylation27K BeadChip*), 480 000 (*HumanMethylation450K BeadChip*) puis 850 000 (*Infinium MethylationEPIC Beadchip*) CpG réparties sur le génome, que nous abrègerons respectivement en 27K, 450K et EPIC (Pidsley *et al.*, 2016). Les concepteurs des puces ont prêté attention à couvrir préférentiellement les régions géniques et péri-géniques, les îlots CpG et les régions régulatrices intergéniques connues (Pidsley *et al.*, 2016).

Les puces illumina ont montré leur capacité à détecter des épisignatures concernant le statut fumeur ou non des individus (Dogan *et al.*, 2014) ou à prédire l'âge d'un sujet (Horvath, 2013). L'âge épigénétique pourrait d'ailleurs être un bon marqueur du risque de mortalité d'un individu, nonobstant l'âge chronologique de celui-ci (B. H. Chen *et al.*, 2016; Fransquet *et al.*, 2019) (Figure 33). Il n'y a en revanche, et c'est attendu vu l'hétérogénéité du syndrome, pas d'épisignature séparant les patients TSA des contrôles (S. V. Andrews *et al.*, 2018). Bien que cette dernière étude ait présenté une puissance de 80% pour montrer une différence de méthylation de 4% ou plus, aucune position n'était significativement différenciellement méthylée en tenant compte du nombre de tests statistiques. La recherche d'épisignature n'est donc concevable que gène par gène.

Concernant les gènes impliqués dans les syndromes neurodéveloppementaux, deux équipes canadiennes ont posé les bases de la recherche d'épisignatures en 2015 (Choufani *et al.*, 2015; Strong *et al.*, 2015) pour l'équipe de Toronto et 2016 pour celle de London et Ottawa (Hood *et al.*, 2016; Kernohan *et al.*, 2016). Les deux équipes ont par la suite démontré l'efficacité de cette stratégie pour reclasser des VSI pour de nombreux syndromes (Aref-Eshghi *et al.*, 2020; M. A. Levy, McConkey, *et al.*, 2022) et facteurs de susceptibilité au TSA (Siu *et al.*, 2019).

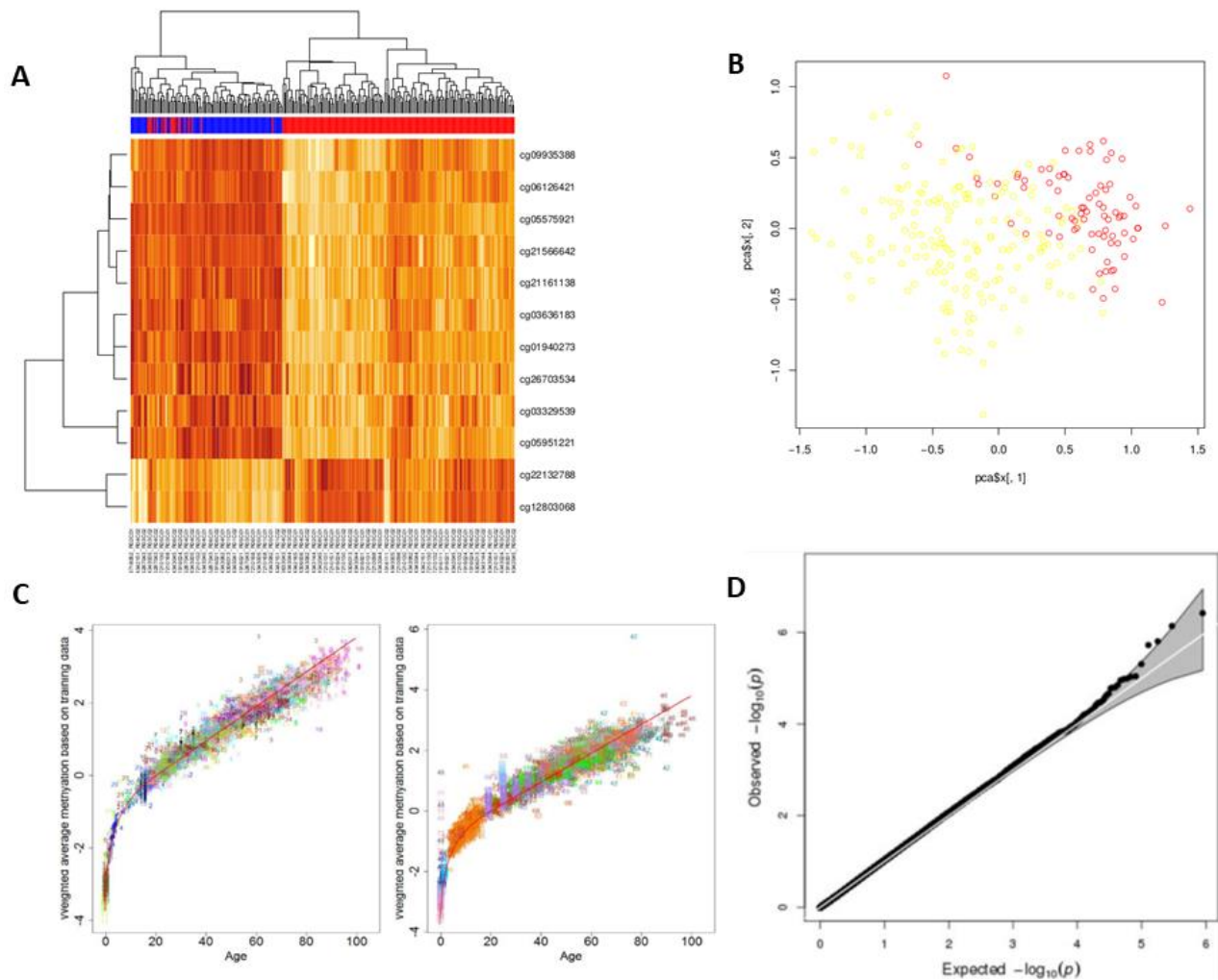


Figure 33. Présence ou non d'épissignatures dans différentes conditions. (A) Classification hiérarchique ascendante et heatmap et (B) analyse en composante principale du jeu de données GSE53045²² disponible sur GeneExpressionOmnibus entre fumeurs et non-fumeurs permettant de visualiser la séparation entre ces deux populations à l'aide de données de méthylation ; (C) cohorte d'apprentissage et de test pour détermination de l'âge épigénétique, d'après Horvath, 2013 ; (D) QQ-plots entre p -valeurs observées et attendues sous l'hypothèse nulle entre 453 patients présentant un TSA et 515 témoins : aucune position n'est différenciellement méthylée : la distribution des p -valeurs est compatible avec l'hypothèse nulle, d'après Andrews et al., 2018.

²² <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53045>

ii. Caractéristiques techniques des puces Illumina

Description de la technologie Infinium

Dans le chapitre suivant, nous allons considérer les caractéristiques techniques des puces 450K et EPIC. Les puces 27K²³ ne sont plus utilisées et il n'existe pas de jeu de données disponible pertinent dans le cadre de notre travail.

Cette technologie tire parti du traitement de l'ADN au bisulfite de sodium, dont l'effet est illustré dans la Figure 34. Ce sel entraîne la transversion des cytosines non méthylées en uracile (puis thymine après amplification par PCR). A l'inverse, la réaction est ralentie pour les cytosines méthylées au point qu'elle ne soit plus apparente en pratique (Laird, 2010). De ce fait, le génotypage d'une position CpG sur un échantillon de molécules d'ADN aboutira à un ratio cytosine/thymine proportionnel au ratio cytosine méthylée/non-méthylée de cette position chez le sujet (Laird, 2010).

²³ 94% des CpG de la puce 27k sont présents sur la puce 450K.

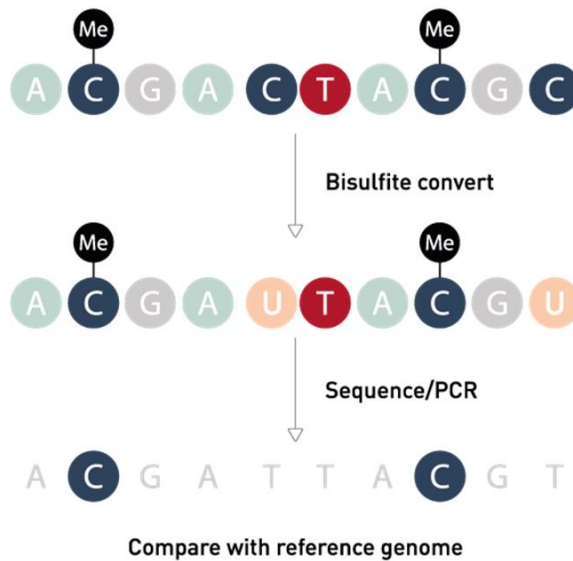


Figure 34. Schéma simplifié de l'effet du traitement de l'ADN au bisulfite de sodium. Les cytosines méthylées sont en quelque sorte protégées de l'effet du bisulphite de sodium.

Concernant les puces Illumina, après traitement au bisulfite, chaque position est génotypée à l'aide d'une amorce de 50 pb complémentaire de la séquence en 5' de la CpG concernée. Après hybridation de l'amorce, des didésoxyribonucléotides fluorescents sont incorporés à l'extrémité 3' de celle-ci. La lecture de la fluorescence permet d'en déduire le ratio cytosine méthylée/cytosine non-méthylée (Pidsley *et al.*, 2016). Le processus est exposé figure 35. En pratique, il existe deux types de sondes : les sondes de type I mesurent de manière séparée la méthylation et la non-méthylation alors que les sondes type II mesurent l'un et l'autre simultanément. La première conception est nécessaire pour les régions denses en CpG (Pidsley *et al.*, 2016).

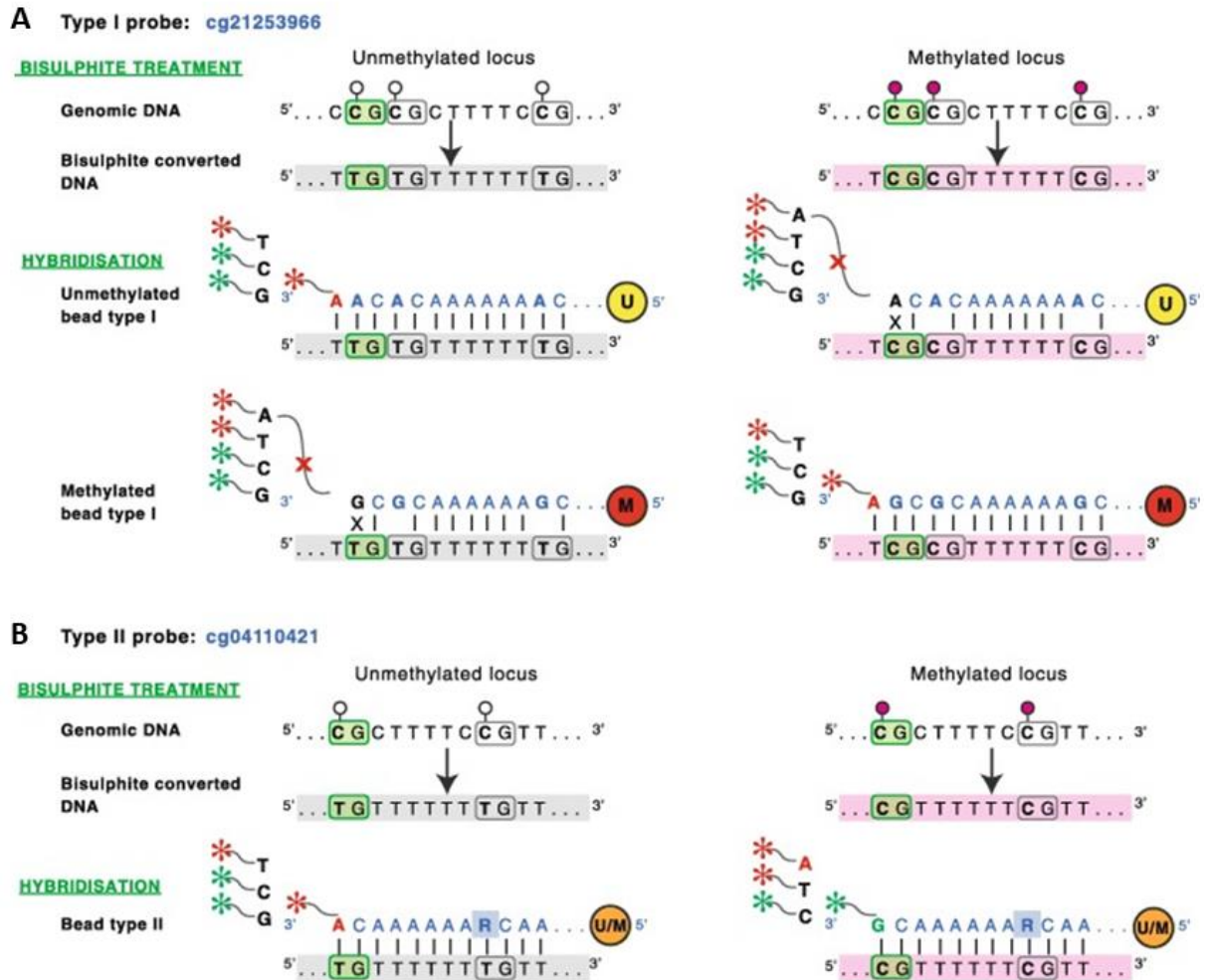


Figure 35. Fonctionnement des deux types de sondes de la technologie Infinium d'Illumina. (A) les sondes de type I mesurent la méthylation à l'aide de deux amorces différentes (appelées beads, en référence à la perle de silice sur laquelle l'amorce est fixée à la surface de la puce). Dans ce cas l'extrémité 3' de l'amorce correspond à la cytosine (ou à la thymine si la position n'est pas méthylée) à génotyper. L'hybridation et l'extension par une base fluorescente sont spécifiques du caractère méthylé ou non de la position ; (B) au contraire, pour les sondes de type II, la position finale de l'amorce correspond à la position précédent la position à génotyper, la nature méthylée ou non-méthylée dépendra de la couleur de la fluorescence détectée. D'après Pidsley et al., 2016.

La puce 450K cible 482 421 CpG. Ceux-ci couvrent l'ensemble des promoteurs, UTRs et corps des gènes ainsi qu'une partie des séquences

régulatrices (Bibikova *et al.*, 2011). La conception initiale de cette puce a été réalisée sur les conseils d'un consortium regroupant plusieurs spécialistes de la méthylation de l'ADN (Pidsley *et al.*, 2016). La puce EPIC, plus récente, cible 863 904 positions comprenant 93,3% des CpG de la 450K – certaines sondes ont été retirées car identifiées pour être particulièrement peu fiables (Naeem *et al.*, 2014). Les positions ajoutées visent un catalogue extensif des séquences régulatrices cellules-spécifiques et permettent une couverture densifiée du corps des gènes (Pidsley *et al.*, 2016). En pratique, l'analyse de jeux de données qui comprennent des puces 450K et EPIC se concentre sur l'intersection entre les deux listes de sondes (Figure 36). La concordance entre les deux générations de puces semble bonne pour la majorité des positions (Logue *et al.*, 2017; Solomon *et al.*, 2018). Il existe également des sondes de contrôles qualité permettant d'estimer l'efficacité des différentes étapes que nous avons citées et la détection du bruit de fond par le lecteur de fluorescence (Min *et al.*, 2018).

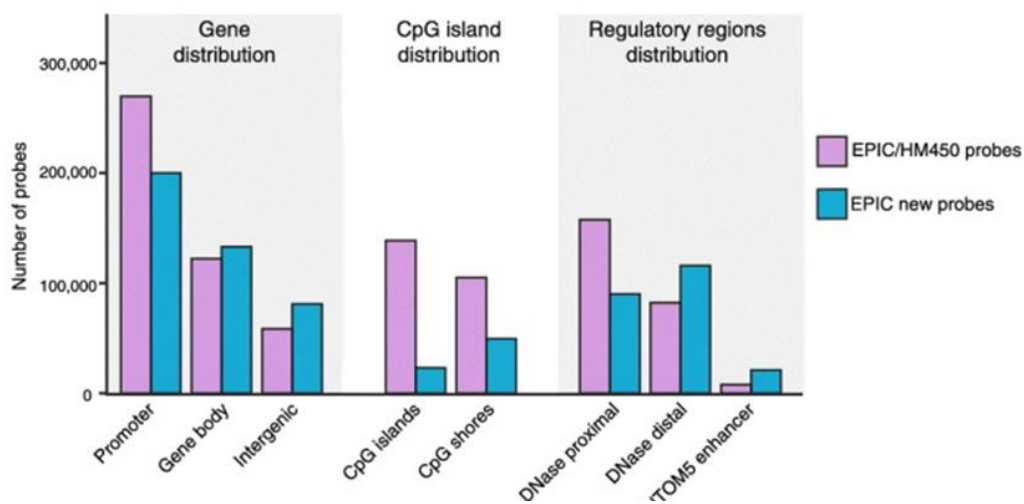


Figure 36. Distribution de la position des sondes des puces 450K et EPIC. Les sondes ajoutées sur la puce EPIC, améliorent la couverture globale mais sont distribuées préférentiellement dans les régions intergéniques où elles ciblent les régions régulatrices connues. D'après Pidsley *et al.*, 2016.

Données analysables en sortie des puces Illumina

L'analyse de méthylation sur puce Infinium aboutit à un fichier informatique du format spécifique des puces Illumina (extension « .idat ») qui contient les données brutes. Ce fichier peut être traité par différents packages du logiciel de statistique R²⁴ comme *minfi* (Aryee *et al.*, 2014) ou *wateRmelon* (Pidsley *et al.*, 2013) pour les plus utilisés. Ils donnent accès aux données de séquençage sous la forme de plusieurs matrices comprenant une sonde par ligne et un échantillon par colonne. De la sorte, pour chaque CpG de chaque échantillon les valeurs de méthylation, non-méthylation, les p-valeurs de fiabilité de détection, le nombre de beads hybridés sont accessibles. Pour l'analyse finale, c'est une valeur composite appelée « β » qui est utilisée (Pidsley *et al.*, 2016). Elle se calcule comme suit :

$$\beta = \frac{M}{(M + U + 100)}$$

Avec M = mesure de méthylation ; U = mesure de non-méthylation. La constante au dénominateur évite les valeurs aberrantes pour un très petit nombre de beads.

La valeur β représente, pour chaque position, le pourcentage de CpG méthylées. Une valeur de 0 représente une position qui n'est jamais lue méthylée sur les molécules d'ADN du patient, une valeur (approchant) de 1 représente une position (quasi) systématiquement méthylée. Certains

²⁴ R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

auteurs ont exprimé des réserves quant aux qualités mathématiques des β de méthylation et ont proposé de travailler sur une valeur logarithmique (Du *et al.*, 2010). Néanmoins, la distribution des β est bimodale avec deux pics aux extrêmes (Luo *et al.*, 2020; Varley *et al.*, 2013), c'est à dire dans l'intervalle où la transformation en logarithme aura tendance à surestimer de petits changements de valeurs. En pratique, il n'y a pas de différence démontrée entre valeurs β et logarithmes pour la détermination de sondes différenciellement méthylées (Zhuang *et al.*, 2012).

Bien qu'une matrice de β soit un fichier informatique relativement léger et maniable, une partie des étapes de contrôle qualité et de normalisation des données qui président à son obtention nécessitent un traitement des données brutes. De ce fait plusieurs matrices de β issues de population, de laboratoire, de centres de séquençages pipelines bio-informatiques différents ne peuvent être raisonnablement fusionnées. Il est nécessaire de fusionner les données brutes. Plusieurs équipes ont mis à disposition, en accès libres, leurs fichiers « .idat » sur la base de données du NCBI Gene Expression Omnibus (GEO). La plus grande étude sur le sujet a tiré parti de ce développement de la science open access pour compléter son propre recueil international d'échantillons (M. A. Levy, McConkey, et al., 2022). Malheureusement, le développement d'un outil commercial à partir de ces données a fait obstacle à la mise à disposition complète des fichiers « .idats » générés par de nombreux laboratoires pour la communauté scientifique.

iii. Traitement bioinformatique des données de méthylation

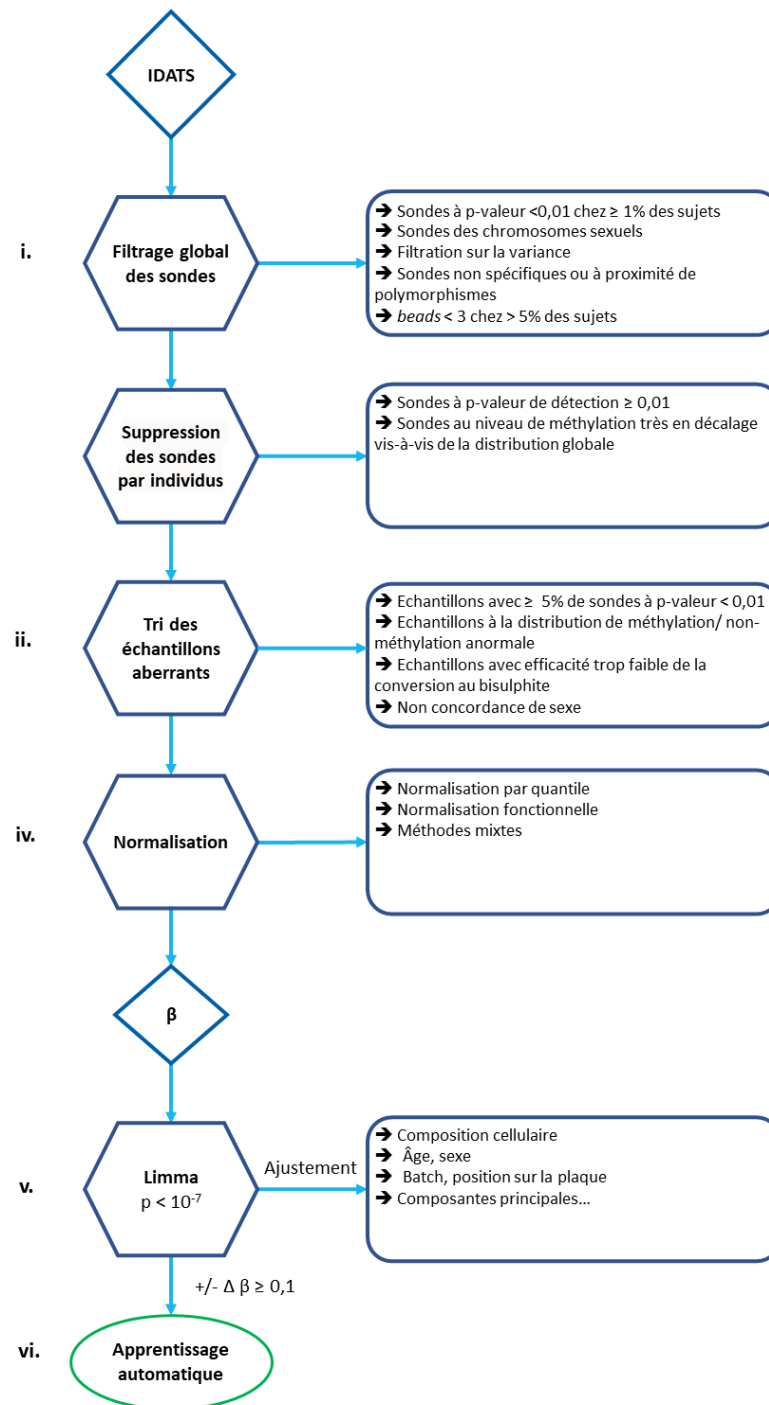


Figure 37. Proposition d'un algorithme de traitement des données de méthylation.

Les données brutes issues des puces Illumina ne peuvent être analysées directement. Plusieurs étapes préliminaires sont nécessaires pour limiter les effets artéfactuels qui peuvent biaiser les conclusions statistiques. Une fois obtenue une matrice de β satisfaisante, il existe également plusieurs stratégies d'analyse des données possibles. Le paramétrage des différents outils disponibles dépend de l'analyse que l'on souhaite effectuer et de la taille de la cohorte. Les grandes étapes sont présentées dans la figure 37 puis détaillées dans ce chapitre.

Filtrage des sondes

La qualité de la mesure de chaque sonde est définie par sa p-valeur de détection. Si celle-ci est supérieure ou égale à 0,05, le signal détecté n'est pas suffisant pour être discriminant vis-à-vis du bruit de fond et la sonde n'est pas considérée²⁵. Malgré cette limite fournie par Illumina, la plupart des auteurs travaillent à partir d'un seuil de 0,01 pour fiabiliser au maximum leurs données. Ils suivent, ce faisant, les recommandations de Lehne et collaborateurs (Lehne *et al.*, 2015). Lorsqu'une sonde dépasse le seuil de détection sur un trop grand nombre d'échantillons, il est d'usage de ne plus la considérer dans l'analyse (Maksimovic *et al.*, 2017). Le seuil pour définir « trop » n'est pas équivoque et dépend du nombre d'échantillons inclus. Pour la conception d'épisingatures le seuil d'1% des échantillons est habituellement retenu²⁶.

²⁵ Documentation Illumina

²⁶ *Illumina 450 and EPIC methylation array normalization and metrics*, <https://bioc.ism.ac.jp/packages/devel/bioc/manuals/wateRmelon/man/wateRmelon.pdf>

Une autre mesure de la qualité de la mesure d'une position est le nombre de *beads* pour lequel une amorce s'est hybridée. L'idée est que lorsqu'une sonde capte un signal par un tout petit nombre d'amorces (Pidsley *et al.*, 2016), le risque que celui-ci soit un artefact est plus important. La pratique habituelle est de supprimer toute sonde avec moins de 3 *beads* chez plus de 5% des sujets²⁷.

Plus globalement, les sondes des chromosomes sexuels sont rarement conservées. Bien qu'elles puissent porter un signal biologiquement intéressant, des difficultés techniques justifient ce choix. L'analyse conjointe d'échantillons des deux sexes rend le problème du nombre de chromosomes X insoluble. En outre, chez les sujets féminin, les sondes portées par les chromosomes Y ont tendance à capter le bruit de fond (Lehne *et al.*, 2015) (la Figure 38).

De plus, un grand nombre de sondes doivent être retirées *a priori* de l'analyse (Y. Chen *et al.*, 2013; Pidsley *et al.*, 2016; W. Zhou *et al.*, 2017), soit :

- Qu'elles manquent de spécificité et s'hybrident à distance de la zone pour laquelle elles ont été conçues, typiquement sur les régions pseudo-autosomales des chromosomes sexuels.
- Qu'elles soient à proximité immédiate d'un polymorphisme génétique avec un impact important sur la méthylation détectée. Le signal de

²⁷ *Ibidem*

méthylation est biaisé par la présence ou non de cette variation génétique chez l'individu.

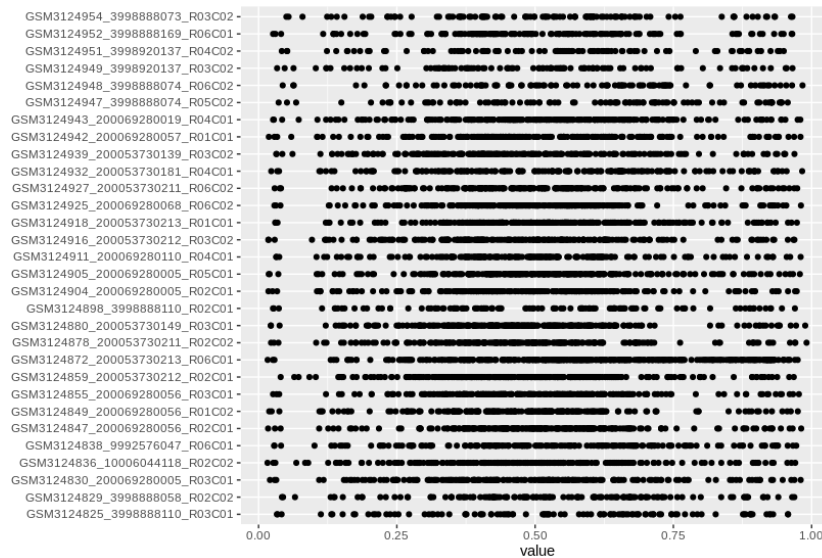


Figure 38. Distribution des valeurs β des sondes situées sur le chromosome Y sur un échantillon d'individus de sexe féminin. Jeu de données téléchargé sur GEO avec analyse restreinte aux sujets féminins. Le lecteur détecte un taux de méthylation significatif pour 50% des sondes situées sur le chromosome Y.

Un certain nombre de sondes ont également montré des problèmes de fiabilité sur des réplicats techniques et nous proposons de ne pas les considérer (Bose et al., 2014).

L'effet des polymorphismes génétiques sur la méthylation en regard ne peut être totalement atténué en supprimant une liste de sondes. Celle-ci ne pourra jamais être exhaustive pour toutes les populations au vu de la

diversité du génome humain. Plusieurs outils²⁸ filtrent les sondes dont le taux de méthylation est anormal chez un individu par rapport à sa distribution dans le reste de la cohorte. Des bornes très conservatrices sont conseillées pour ne retirer que les sondes très fortement en décalage avec le reste des données et ne pas masquer un signal statistique authentique.

Enfin, certains auteurs proposent de filtrer les sondes qui présentent une variance quasi-nulle afin d'augmenter la puissance statistique (Bourgon *et al.*, 2010; Siu *et al.*, 2019) ou d'améliorer la concordance entre les données des deux types de puces (Logue *et al.*, 2017; Solomon *et al.*, 2018). L'idée est que supprimer *a priori* les positions où le niveau de méthylation ne varie, pour ainsi dire, jamais, autorise à diminuer le nombre de tests statistiques et augmenter d'autant la puissance statistique. Par exemple, Siu et collaborateurs filtrent toutes les sondes avec une déviation standard < 2% de la déviation standard maximale observée pour l'ensemble d'entre elles. Dans le contexte d'analyse de méthylation, ce choix risque cependant de biaiser l'analyse. En effet, la valeur β n'a pas, par construction mathématique, une variance stable : celle-ci est plus faible au niveau des pics (Du *et al.*, 2010). De plus, le comportement des deux types de sondes est différent, avec une distribution de valeur β des types I formant deux pics plus proches des extrêmes (Pidsley *et al.*, 2016). Pour la raison que nous venons d'évoquer, la filtration sur la variance entraînerait la suppression préférentielle des sondes aux valeurs de méthylation extrême

²⁸ *Pwod par exemple, voir Illumina 450 and EPIC methylation array normalization and metrics,*
<https://bioc.ism.ac.jp/packages/devel/bioc/manuals/wateRmelon/man/wateRmelon.pdf>

et donc tout particulièrement des sondes de type I. Lorsque les tailles d'effets attendues sont faibles, ce qui est le cas pour une épisignature, cette étape pourrait même détériorer la puissance statistique (Zhuang *et al.*, 2012), contrairement à l'effet recherché.

Tri des échantillons aberrants

Une autre source de biais dans les études est l'existence d'individus présentant une distribution de leurs valeurs β , c'est-à-dire de la méthylation de l'ADN, trop différente par rapport au reste de la cohorte. Il s'agit généralement d'un artefact technique. Ces *outliers* doivent être retirés avant analyse pour éviter de créer ou fausser des associations statistiques. Différentes solutions existent pour déterminer les échantillons aberrants. Il peut s'agir d'une vérification visuelle d'un diagramme ou d'une technique de réduction dimensionnelle (Aryee *et al.*, 2014) comme dans la figure 39.

Des critères plus objectifs sont également proposés. Par exemple, la fonction *outlyx*²⁹ implémentée dans plusieurs *packages* définit les *outliers* en combinant une analyse en composante principale et la distance de chaque individu par rapport au gros des données. Les auteurs du package *meffil* (Min *et al.*, 2018) proposent quant à eux de considérer les individus présentant un ratio méthylation sur non-méthylation aberrant par rapport au reste des données.

²⁹*Illumina 450 and EPIC methylation array normalization and metrics*,
<https://bioc.ism.ac.jp/packages/devel/bioc/manuals/watermelon/man/watermelon.pdf>

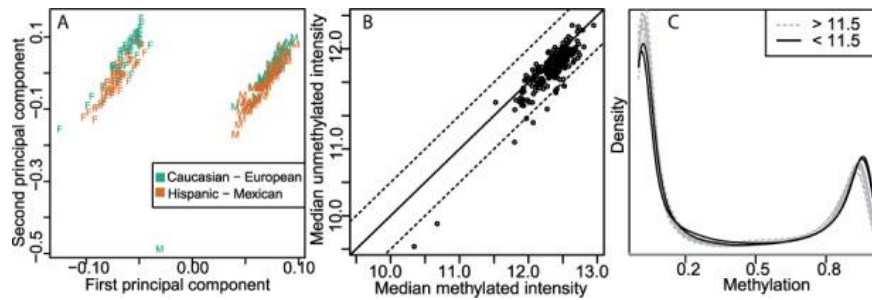


Figure 39. Trois méthodes de visualisation des données de méthylation pour la détection d'échantillons aberrants. (A) Analyse en composante principale des données de méthylation. La première composante dépend du sexe des individus. On observe que deux échantillons présentent une structure de méthylation très différente des autres et ne devraient pas être inclus selon les auteurs ; (B) nuage de point représentant pour chaque individu l'intensité médiane de méthylation en fonction de l'intensité médiane de non-méthylation. Une demi-douzaine d'individus présente un ratio $< 0,5$ qui semble excessif aux auteurs ; (C) diagramme de densité de la distribution des mesures de méthylation par individus. Plusieurs d'entre eux présentent une déviation par rapport à la distribution bimodale de la cohorte et leur retrait est discuté. Tiré de Aryee *et al.*, 2014.

D'un point de vue plus technique, il est fréquent dans la littérature de ne pas considérer les échantillons pour lesquels une efficacité trop faible de la conversion au bisulfite est constatée ou présentant plus de 5% de sondes défailtantes. Enfin, il est possible de prédire le sexe des individus à partir des données de méthylation, ce que la plupart des *packages* proposent. Il est utile de vérifier que le sexe prédit correspond aux individus afin de se prémunir contre une inversion de ceux-ci sur la plaque (Maksimovic *et al.*, 2017).

Normalisation

Plus les données comparées par un modèle statistique sont hétérogènes, moins les résultats de celui-ci seront robustes. Les données de méthylation varient en fonction des caractéristiques des sujets et d'effets techniques divers, imparfaitement compris. La normalisation statistique des données vise à harmoniser la distribution des données dans la cohorte afin

de rendre comparables les différentes populations qui la compose. La normalisation par quantile est une stratégie qui a été introduite spécifiquement pour l'analyse des puces pangénomiques (Bolstad *et al.*, 2003), comme les puces Illumina. L'implémentation de cette méthode sous le nom de *dasen* a été montrée comme la normalisation statistique la plus efficace pour l'analyse de méthylation sur puces par rapport à une normalisation, dite fonctionnelle, basée sur les sondes de contrôle qualité (Pidsley *et al.*, 2013; Shiah *et al.*, 2017). Bien que très peu utilisée en pratique, une normalisation intégrant les deux approches serait plus efficace pour l'étude de données issues de jeux hétérogènes (Min *et al.*, 2018) et atténuerait l'effet *batch*.

Détection des positions différenciellement méthylées

La normalisation utilise les signaux de méthylation et de non-méthylation de l'ensemble des échantillons pour aboutir à la matrice de β utilisable – en tenant parfois compte d'une correction en fonction du bruit de fond. L'étape suivante consiste à déterminer les positions différenciellement méthylées entre les cas et les contrôles. Deux approches co-existent. La méthode principale consiste en l'adaptation d'un modèle de régression linéaire multiple pour les données issues de puces pangénomiques. Ce modèle, appelé limma (pour *Linear Model for MicroArray*) permet de définir des positions différenciellement méthylées entre les cas et les contrôles en tenant compte des facteurs confondants (Campagna *et al.*, 2021; Ritchie *et al.*, 2015). Cette approche à l'échelle de la position est celle qui permet d'aboutir à une épisignature, c'est-à-dire un ensemble de marqueur de possibles dysrégulations transcriptionnelles.

L'autre approche consiste en la détermination de régions différenciellement méthylées. Le niveau de méthylation de chaque CpG est

localement fortement corrélé, ce qui représente des plages entières méthylées ou non au sein du génome (Lehne *et al.*, 2015), reflets des unités fonctionnelles régulatrices. La procédure de *bump hunting* permet la mise à jour de ces régions différentiellement méthylées (Jaffe *et al.*, 2012). D'autres stratégies existent mais sont peu employées en pratique et semblent moins robustes sur le plan statistique (D. Li *et al.*, 2015). La détermination des régions différentiellement méthylées donne accès à un début de description des effets d'une épisignature, et donc de son impact fonctionnel – en gardant à l'esprit qu'il s'agit de l'effet mesuré sur le tissu sanguin. C'est particulièrement le cas lorsqu'elles sont situées au niveau de régions régulatrices de gènes cibles

Ces modèles statistiques tiennent compte de variables d'ajustement qui ont souvent un impact plus fort sur la stratification des données que la variable d'intérêt (Lehne *et al.*, 2015). Certains de ces facteurs confondants sont bien connus. Un effet « jeu de données » (ou *batch effect*) est bien décrit sur les données à hautes dimensions comme celles de méthylation (Leek *et al.*, 2010). Brièvement, de multiples petites différences sont introduites selon les conditions expérimentales (laboratoire, machines, procédures, techniciens différents). Leur sommation influe sur la structure des données en dehors de toute causalité scientifique (Leek *et al.*, 2010). Plusieurs stratégies peuvent être appliquées, parfois conjointement, pour mitiger l'effet *batch*, mais leur efficacité reste partielle et peut parfois masquer le signal d'intérêt (Fortin *et al.*, 2014; Leek *et al.*, 2010; Zindler *et al.*, 2020). L'utilisation de témoins pour chaque *batch*, ou à tout le moins par laboratoire et lecteur de puce paraît la solution minimale pour atténuer cet effet en autorisant le modèle à ajuster sur cette variable (Zindler *et al.*, 2020).

Intra-*batch*, les variables qui influencent le plus sur les données de méthylation sont l'âge, (Lehne *et al.*, 2015) la composition cellulaire du prélèvement (Lehne *et al.*, 2015), le genre (Grant *et al.*, 2022) et les effets positionnels sur la plaque (Jiao *et al.*, 2018). L'idéal est donc de tenir compte de l'âge et du genre de l'individu dans le modèle et de randomiser les échantillons sur la plaque (Verdugo *et al.*, 2009). Lorsque ceux-ci ne sont pas connus, ils peuvent-être estimés à partir des données de méthylation (Horvath, 2013; Yousefi *et al.*, 2015) – en prenant garde à ne pas mêler données cliniques et prédictions afin de ne pas ajouter un biais potentiel. Lorsque les données cliniques sont connues, ces prédictions peuvent également constituer un outil d'identitovigilance supplémentaire.

Pour la composition cellulaire, une étape supplémentaire est nécessaire. En effet, la méthylation est mesurée sur de l'ADN issu de sang total, dont les éléments figurés varient d'un prélèvement à l'autre. Pour pallier cette difficulté, les marqueurs épigénétiques spécifiques des différents types leucocytaires sont utilisés pour inférer la formule leucocytaire de chaque individu (Houseman *et al.*, 2012). Cette fonctionnalité est présente dans la plupart des *packages* disponibles. La composition cellulaire peut ensuite être utilisée comme variable d'ajustement du modèle de régression.

Tous les facteurs confondants ne sont pas connus ou mesurés. Des méthodes d'apprentissage non supervisées sont également souvent utilisées pour atténuer leur effet (Leek *et al.*, 2010; Lehne *et al.*, 2015).

Apprentissage automatique

Cette dernière étape n'est pas indispensable pour la détermination *per se* d'une épisignature. Certaines équipes se contentent donc de rendre un jeu de positions différenciellement méthylées, en tentant compte des

comparaisons multiples. Le modèle de prédiction permettant de définir si un individu tiers est plus proche des cas ou des contrôles est donc construit en donnant un poids égal à chaque sonde (Siu *et al.*, 2019). D'autres équipes utilisent des logiciels d'apprentissage automatique (Aref-Eshghi *et al.*, 2020). Toutes les techniques d'apprentissage automatique supervisé (*machine learning*) adaptées aux données de très grandes dimension (car le nombre de sonde est supérieur au nombre de sujets) peuvent-être utilisés. Elles consistent, très basiquement, en la détermination d'un système de classification en pondérant les sondes différemment méthylées. L'équipe de London (Aref-Eshghi *et al.*, 2020; M. A. Levy, McConkey, *et al.*, 2022), qui a publié le plus grand nombre d'épisingatures, utilise le *Support Vector Machine (SVM)*. Sur les puces 27K, cette méthode de classification a été montrée comme la plus efficace, avec l'*elastic net* (Zhuang *et al.*, 2012).

Pour conclure, l'analyse sur puces de méthylation Illumina aboutit à des données à haute dimension qui doivent bénéficier d'un traitement bioinformatique complexe. Celui-ci a pour but de limiter au maximum les erreurs systémiques qui pourraient, par artefact, créer (ou masquer) des sondes différemment méthylées. Ce chapitre n'est en aucun cas un panorama exhaustif des stratégies applicables pour la génération d'épisingatures *via* puces Illumina. Il s'agit plutôt d'une revue sélective des méthodes applicables en fonction de leur fréquence constatée dans la littérature et des éléments de preuve, souvent limités, en leur faveur. Le *pipeline* que nous en avons déduit doit donc se lire comme une proposition, discutable, parmi d'autres. Il est regrettable que la plupart des études ne commentent pas, ou peu, le rationnel pour le choix de tel ou tel critère de traitement des données. Une comparaison systématique et bien menée de ces stratégies reste à faire pour définir un *gold-standard* dans la génération

d'épissignatures fiables pour l'utilisation en diagnostic clinique (Chater-Diehl *et al.*, 2021).

Il s'agit toutefois d'une approche innovante et prometteuse pour l'aide au reclassement de VSI dans les syndromes neurodéveloppementaux d'origine génétique, dont beaucoup comprennent le TSA dans leur spectre phénotypique. Nous avons inclus en annexe un tableau issu de la dernière étude de l'équipe de London, qui reprend la quasi-totalité des gènes pour lesquels une épissignature est rapportée (M. A. Levy, McConkey, *et al.*, 2022). Une découverte surprenante est que certains gènes sont associés à plusieurs épissignatures en fonction des domaines fonctionnels atteints par les variations génétiques pathogènes. C'est le cas par exemple du gène *ADNP*. A l'inverse dans d'autres cas, tous les gènes causaux d'un même syndrome présentent une même épissignature, comme le syndrome Kabuki (gènes *KMT2D* et *KDM6A*).

Plus généralement, la comparaison des différentes épissignatures laisse entrevoir une taxinomie des patterns de méthylation (Figure 40). Il existe un rapprochement voire un chevauchement entre certains syndromes concernant les anomalies de méthylation mais il persiste des positions permettant de les différencier. L'ampleur de l'épissignature est également variable avec certains syndromes plus ou moins éloignés de la population générale.

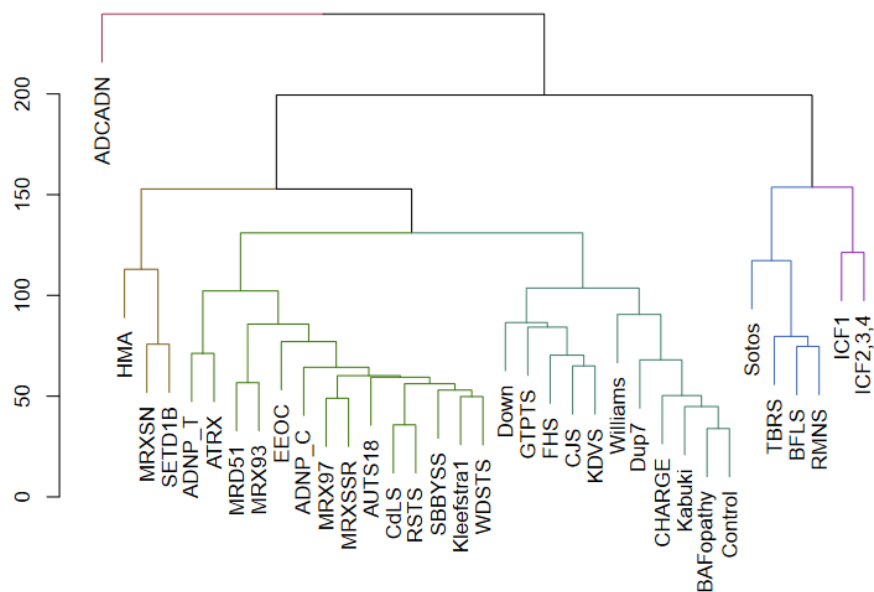


Figure 40. Distance et classification hiérarchique entre 34 épisignatures rapportées par l'étude d'Aref-Eshghi et collaborateurs de 2020. La distance horizontale représente la dissemblance entre les différents clusters. La distance verticale précédant la séparation représente la dissemblance entre deux syndromes. On constate l'existence d'un groupe de syndrome avec hypométhylation à la droite de la classification particulièrement différencié des syndromes avec une hyperméthylation prédominante. Certaines épisignatures comme celle regroupant les différentes bafopathies entraîne un pattern de méthylation très discrètement différencié de celui des sujets issus de population générale.

Dans un premier temps, nous avons cherché à détecter une épisignature pour le gène *MBD5*, responsable d'un syndrome neurodéveloppemental non reconnaissable (MIM : 156200). Les difficultés méthodologiques rencontrées dans ce travail nous ont mené à une réflexion quant à la fiabilité des épisignatures rapportées.

De ce fait, et afin de généraliser leur emploi pour classifier des VSI en diagnostic, nous avons considéré qu'une réplication sur des données indépendantes était nécessaire. C'est d'autant plus vrai pour les données publiées par l'équipe de London. En effet, les données brutes et l'ensemble

de la méthodologie ne sont pas accessibles. De plus, pour guider le clinicien, une évaluation de la sensibilité et spécificité des signatures publiées devrait être réalisée en parallèle. C'est l'objet du premier travail.

Chapitre II. Conception d'une épisignature pour le gène *MBD5*

Avant l'année 2019, nous ne disposions pas, au sein du laboratoire, d'expérience concernant l'analyse de données de méthylation sur puces illumina. Nous avons donc dû concevoir la méthodologie *ex-nihilo* à partir des données de la littérature et de nos connaissances préalables sur la méthylation de l'ADN. Après coup, force est de constater que les choses auraient pu être mieux réalisées. Peut-être avec plus de prévoyance ou de prévenance. Mais je pense, surtout, que certaines erreurs étaient nécessaires afin de nous construire une réelle expertise concernant ce paradigme expérimental, ses possibilités et ses limites. Ce travail sera malheureusement difficile à valoriser par une publication scientifique.

Contexte scientifique concernant MBD5

Les microdélétions récurrentes du locus 2q23.1 sont connues comme responsable de troubles neurodéveloppementaux (van Bon et al., 2010). La région critique du locus ne contient que le gène *MBD5* (Talkowski et al., 2011, p. 5; Walz & Young, 2014). Ainsi, les variants pathogènes du gène *MBD5* sont considérés comme responsables d'un syndrome neurodéveloppemental autosomique dominant répertorié dans la base de données OMIM sous le nom de *intellectual developmental disorder, autosomal dominant 1* (MIM#156200).

La pathogénicité est médiée par un mécanisme d'haploinsuffisance, y compris pour les délétions des exons non codants en situés 5' du gène (Talkowski et al., 2011). De manière intéressante, *MBD5* semble être un gène soumis à un effet de dosage car les duplications sont également

pathogènes, avec une moindre sévérité (Mullegama et al., 2021). Les CNV récurrents associés au locus de *MBD5* ne présentent pas de point de cassures fixes (van Bon et al., 2010). De ce fait, ce syndrome n'est pas médié par la recombinaison homologue non-allélique, mais plutôt par un mécanisme décrit plus récemment, appelé *microhomology-mediated repair* (van Bon et al., 2010; Vissers et al., 2009). Ces points de cassure variables entraînent de fréquentes délétions et duplications partielles ou introniques pour lequel il est parfois difficile de conclure quant à la pathogénicité.

Le spectre phénotypique comprend une très fréquente déficience intellectuelle, un risque de TSA (Mullegama et al., 2014), d'épilepsie et diverses anomalies crânio-faciales (en particulier une microcéphalie) et squelettiques sans *gestalt* reconnaissable (Mullegama & Elsea, 2016; van Bon et al., 2010). Certains cas cliniques suggèrent un phénotype extrêmement variable, y compris au sein d'un même pedigree (Orrico et al., 2021), alors qu'un risque de démence précoce est évoquée par d'autres (González-Ortega et al., 2021; Verhoeven et al., 2019). Compte tenu de cette hétérogénéité clinique, sans corrélation claire entre le génotype et le phénotype, l'interprétation et la classification des variations de *MBD5* peuvent être difficiles, en particulier lorsqu'elles sont héritées de parents non (ou peu) symptomatiques.

La famille de gènes présentant un domaine de liaison méthyl-CpG (MBD, *Methyl-cytosine Binding Domain*) comprend notamment les gènes *MBD1* à *6* (Hendrich & Bird, 1998). Les protéines qu'ils encodent ont un rôle dans l'organisation de la chromatine et le maintien de l'épigénome (Fournier et al., 2012) *via* leur domaine de liaison aux CpG méthylées. Contrairement aux autres protéines à domaine MBD, *MBD5* ne se lie pas aux CpG méthylées, même s'il colocalise avec l'hétérochromatine fortement

méthylée (Laget et al., 2010). En effet, dans le cas de MBD5, le domaine MBD est un domaine d'interaction protéine et non de liaison aux CpG (Baymaz et al., 2014). Une étude menée sur un modèle végétal concluait, pour cette protéine, à une fonction d'interprétation de la méthylation des CpG pour le recrutement de facteurs de répression de la transcription (Ichino et al., 2021).

Concernant *MBD5*, des données de transcriptomique issues de progéniteurs neuronaux provenant de sujets porteurs de variations génétiques pathogènes de *MBD5* ont révélé que les variations d'expression concernent particulièrement un réseau de gènes impliqués dans la modification de la chromatine et la régulation de la machinerie épigénétique (Mullegama et al., 2021).

Le caractère fréquent des VSI de ce gène et son implication dans la machinerie épigénétique en général et dans l'interaction avec le méthylome en particulier font de *MBD5* un bon candidat pour la recherche d'une épisignature.

Nous avons donc entrepris un recueil national (au sein du réseau AnDDI-Rares) d'échantillons d'ADN de patients porteurs d'une variation pathogène ou d'une VSI dans le gène *MBD5* afin détecter la présence d'une épisignature et d'évaluer ses éventuelles performances diagnostiques en vue de l'utiliser comme aide à la classification des VSI.

Précisions méthodologiques

L'augmentation de la taille de la cohorte de découverte d'une épisignature permet d'augmenter la puissance statistique et donc la probabilité de la détecter (Chater-Diehl et al., 2021). Néanmoins la dimension optimale de celle-ci ne peut pas être prédite *a priori*. Comme

nous l'avons vu, l'impact sur le profil de méthylation est variable en fonction du gène (Aref-Eshghi et al., 2020) et la puissance statistique pour détecter l'épissignature varie donc en fonction de l'ampleur de la différence entre le pattern de méthylation des cas et des témoins, que l'on ne peut estimer par avance. On observe que les signatures rapportées dans la littérature ont été détectées sur un nombre minimum de cinq à dix individus (M. A. Levy, McConkey, et al., 2022). Nous verrons dans le chapitre suivant qu'une épissignature fiable et précise nécessite généralement un plus grand nombre de sujets.

Dans le cas des syndromes neurodéveloppementaux rares, comme ceux liés à l'haploinsuffisance de *MBD5*, le nombre de patients accessibles est le facteur limitant. Nous avons cherché à maximiser ce nombre pour optimiser notre puissance statistique sans fixer de limite *a priori*.

Pour l'ensemble des résultats détaillés par la suite, la méthodologie suivante a été employée :

- i- Analyse de méthylation sur puce EPIC puis téléchargement des fichiers bruts (.idat)
- ii- Chargement des données de méthylation *via* l'un des packages dédiés du logiciel R (watermelon (Pidsley et al., 2013) dans un premier temps, puis meffil (Min et al., 2018) par la suite) et normalisation des données de méthylation avec la méthode recommandée par les auteurs.
- iii- Suppression des sondes avec échec de détection dans plus de 5% des échantillons, des sondes connues pour être à proximité d'un polymorphisme génétique (Y. Chen et al., 2013), pour présenter une liaison non spécifique (Y. Chen et al., 2013), comme étant peu fiable sur des réplicats techniques (coefficient de corrélation

interclasse $< 0,2$) (Bose et al., 2014), ou ciblant les chromosomes sexuels.

- iv- Suppression des échantillons avec un profil aberrant de méthylation par rapport à l'ensemble des échantillons, avec une anomalie détectée par les sondes de contrôle (efficacité trop faible du traitement par bisulfite par exemple) ou avec plus d'1% de sondes avec échec de détection de méthylation.
- v- Recherche de sondes différentiellement méthylées entre porteurs de variations pathogènes et 10 témoins par cas appariés sur l'âge et le sexe.

Plusieurs jeux de contrôles disponibles sur les plateformes du NCBI GEO ou dbGaP³⁰ ont été utilisés. Lorsque cela était nécessaire, nous avons restreint notre analyse aux sondes communes entre les puces 450K et EPIC.

Données préliminaires

Une première série de 18 échantillons d'ADN de patients porteurs de variations pathogènes du gène *MBD5* – ainsi que 6 échantillons d'ADN de porteurs de VSI du gène – ont bénéficié d'une analyse de méthylation sur puce EPIC effectuée par le CNRGH. Le *pipeline* décrit précédemment a été appliqué *via* le package *wateRmelon*. Nous les avons comparés à 180 témoins appariés sur l'âge, le sexe et les dix premières composantes principales des données de méthylation, parmi 557 témoins issus de la base

³⁰ <https://www.ncbi.nlm.nih.gov/gap/>

de données GEO, ce qui est une pratique courante dans la littérature (Aref-Eshghi et al., 2020; Butcher et al., 2017; Siu et al., 2019).

Le modèle de régression linéaire (limma) permettait la détection de 31 sondes différentiellement méthylées ($p < 5.10^{-7}$), après ajustement sur les dix premières composantes principales. Ces données étaient encourageantes, néanmoins la projection des deux premières composantes principales des β des 31 sondes montrait principalement une ségrégation des données en fonction du *batch* sur le premier axe, avant la séparation en fonction du statut cas/témoin sur le deuxième axe, comme nous pouvons le voir dans la figure 41.

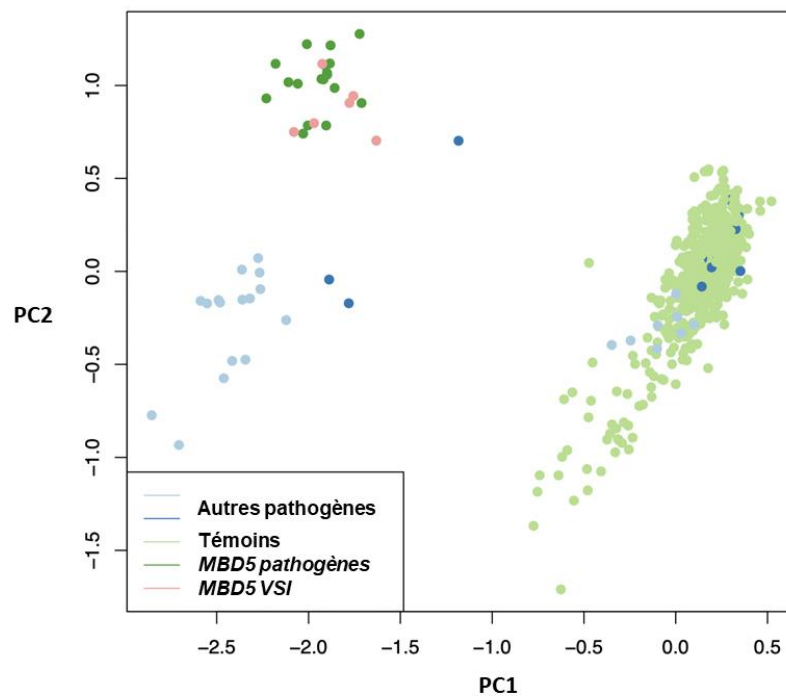


Figure 41. Projection des deux premières composantes principales des valeurs β des sondes différentiellement méthylées des données issues du CNRGH et des contrôles appariés. On observe que l'axe de la première composante principale sépare l'ensemble des échantillons issus du CNRGH (porteurs de variations du gène MBD5 en vert et rose et autres variations inclus sur la plaque, en bleu à gauche du graphique) des témoins et d'autres porteurs de variations pathogènes issus de GEO et dbGaP.

Nous n'avions pas fait le choix d'inclure des témoins sur la plaque, nous inspirant en cela de la méthodologie de l'équipe de London (Aref-Eshghi et al., 2020). Ces données nous ont fait suspecter un effet artéfactuel lié au jeu de données (effet *batch*). Malheureusement, les stratégies d'ajustement sur les différents *batch* nécessitent des témoins sur l'ensemble d'entre eux (Ross et al., 2022).

Pour confirmer ces données préliminaires nous avons pris la décision d'analyser trois échantillons supplémentaires. Les puces de méthylations supplémentaires ont été effectués par l'entreprise Diagenode dans une série comprenant – entre autres – des témoins. Ainsi il nous a été possible de vérifier si ces nouveaux échantillons pathogènes présentaient, ou non, un profil de méthylation plus proche des témoins ou des cas précédemment inclus. Comme nous pouvons le voir dans la figure 42, la structure de batch est toujours visible et les nouveaux échantillons ségrégaient avec les témoins.

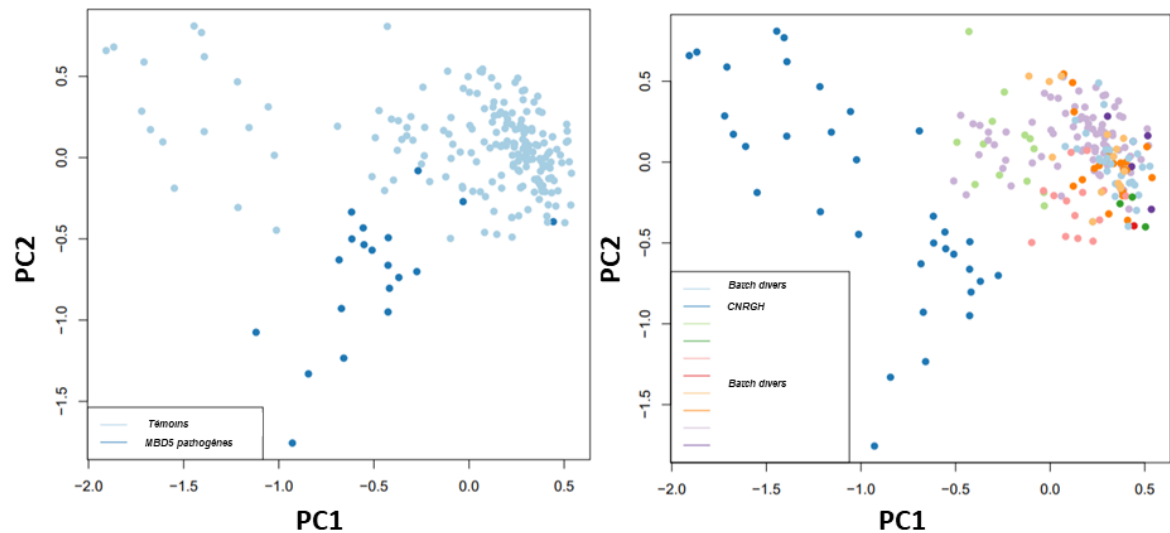


Figure 42. Projection des deux premières composantes principales des valeurs β des sondes différenciellement méthylées des données issues du CNRGH, des contrôles appariés et des nouveaux cas pathogènes. Comme on le voit, l'effet du jeu de données est toujours majeur et les nouveaux échantillons porteurs de variations pathogènes de *MBD5* (en vert foncé) ne ségrègent pas avec les précédents.

Résultats finaux

Afin de d'essayer de conclure quant à l'existence ou non d'une épisignature pour les pertes de fonction du gène *MBD5*, nous avons pris la décision de réanalyser une partie des échantillons de la première série. Notre objectif était de répéter la recherche de sondes différenciellement méthylées, avec suffisamment de témoins pour rajouter le *batch* comme co-variable du modèle de régression linéaire. Pour cette analyse finale nous avons utilisé le *package* meffil dans le *pipeline* bio-informatique. La technologie utilisée par les auteurs permet une accélération substantielle des analyses par rapport aux *packages* précédents. En plus de l'ajustement sur le *batch* (dorénavant possible avec l'inclusion de témoins), nous espérons une diminution des effets technique avec l'utilisation de meffil pour deux raisons. Premièrement, les auteurs ont montré une diminution

des effets batchs par leur stratégie de normalisation fonctionnelle. Deuxième, ils intègrent la possibilité d'ajouter des *surrogate variables* au modèle de régression linéaire. Cette méthode d'apprentissage non supervisé (*surrogate variable analysis, SVA*) définit des co-variables d'ajustement diminuant les effets techniques et est fréquemment utilisée pour l'étude des données de puces à ADN (Leek et al., 2010).

Finalement, 14 échantillons de patients porteurs d'une variation pathogène de *MBD5* ont bénéficié d'une analyse de méthylation sur puce EPIC auprès de la société Diagenode, dont 8 échantillons préalablement analysés par le CNRGH³¹. Tous ces échantillons ont été analysés sur le même lecteur de puce. Nous avons également inclus 24 témoins. L'ensemble des variations pathogènes et des VSI que nous avons incluses sont visualisables dans la figure 43 et répertoriées en annexe.

³¹ Pour ces 8 échantillons que nous avons en double, seules les données de la seconde analyse de méthylation ont été considérées dans l'analyse finale.

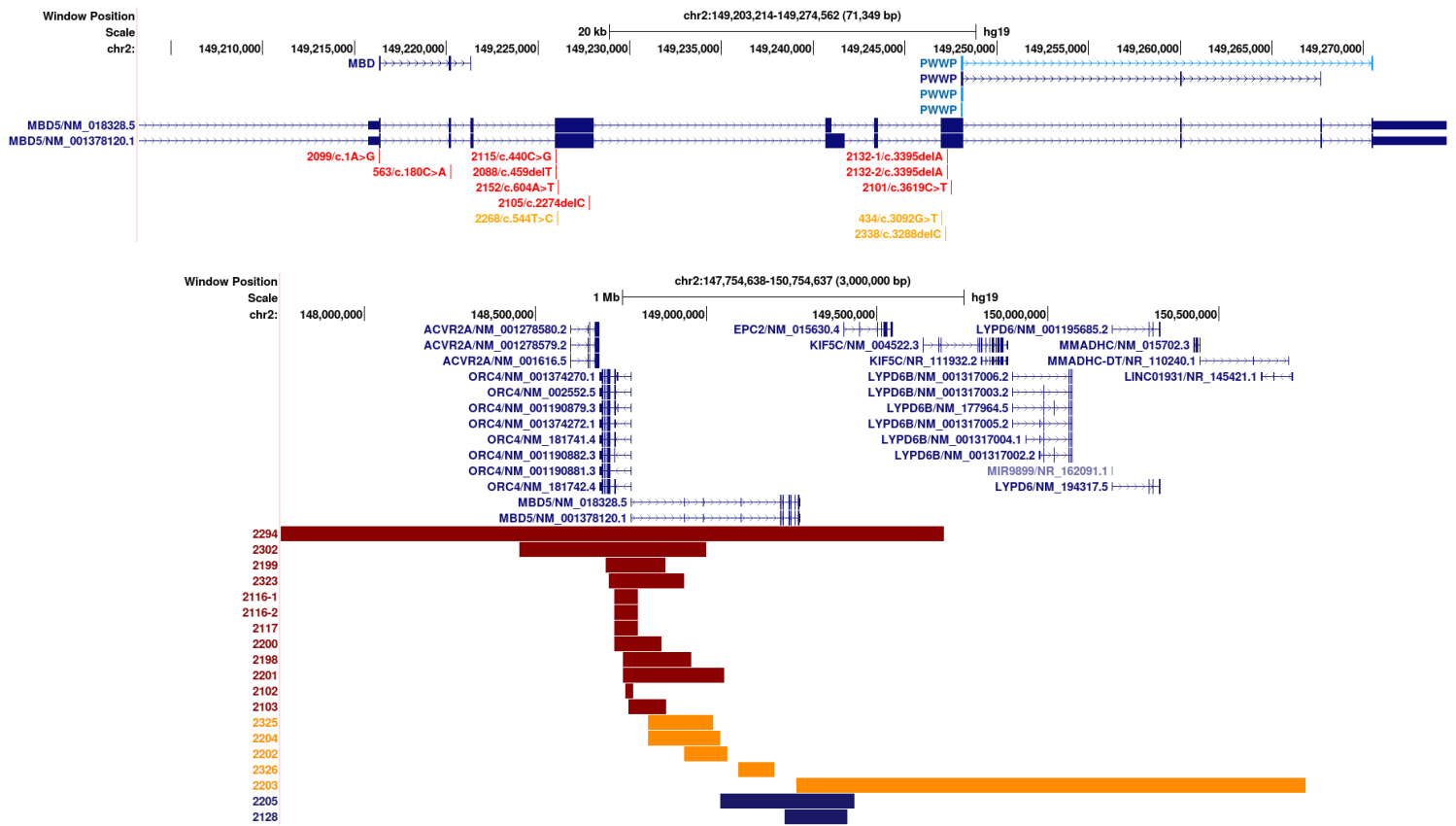


Figure 43. Localisation des variations nucléotidiques et des CNVs de MBD5 incluent dans l'étude. Les variations nucléotidiques (partie haute de l'image) sont indiquées en rouge lorsque classées pathogènes et en orange pour les VSI. Concernant les CNVs (partie basse de l'image), les délétions pathogènes sont représentées en rouge alors que les VSI sont représentées en orange et en bleu pour les délétions et les duplications, respectivement.

Nous avons répété l'analyse précédente dans ce contexte. Comme nous le voyons dans la figure 44.A, un effet technique majeur est toujours présent. L'importance du facteur d'inflation génomique λ ($\lambda = 3,5$) montre un excès de p-valeurs significatives par rapport à ce qui est attendu sous l'hypothèse nulle. Toutefois, la figure 44.B montre une efficacité du SVA pour diminuer cet effet. Dans cette configuration le modèle ne détecte plus que deux sondes différenciellement méthylées. La projection des deux

premières composantes principales des résidus des données de méthylation – c'est-à-dire les valeurs β indépendantes des co-variables d'ajustement du modèle, soit les conditions ayant menées au diagramme quantile-quantile (*Q-Q plot*) de la figure 44.A – sont représentés en figure 44.C. Ces données nous ont fait suspecter un effet technique entre les différentes puces de méthylation utilisées (450K et EPIC) qui pourrait masquer l'épissage. En effet, la corrélation des niveaux de méthylation n'est pas parfaite entre les deux technologies (Logue et al., 2017).

Pour nous affranchir de cette contrainte technique, nous avons ensuite restreint notre analyse aux puces EPIC. L'analyse, dont les résultats sont représentés sur la figure 45 a permis la détection de 20 sondes différenciellement méthylées. La majorité de ces sondes présentent un différentiel de valeur β moyenne $>$ à 10%, ce qui est considéré comme biologiquement significatif (Pidsley et al., 2016). L'analyse en composante principale montre un *cluster* de 11 échantillons (issus des trois séries) séparés des contrôles, alors que les autres ségrégent avec les contrôles.

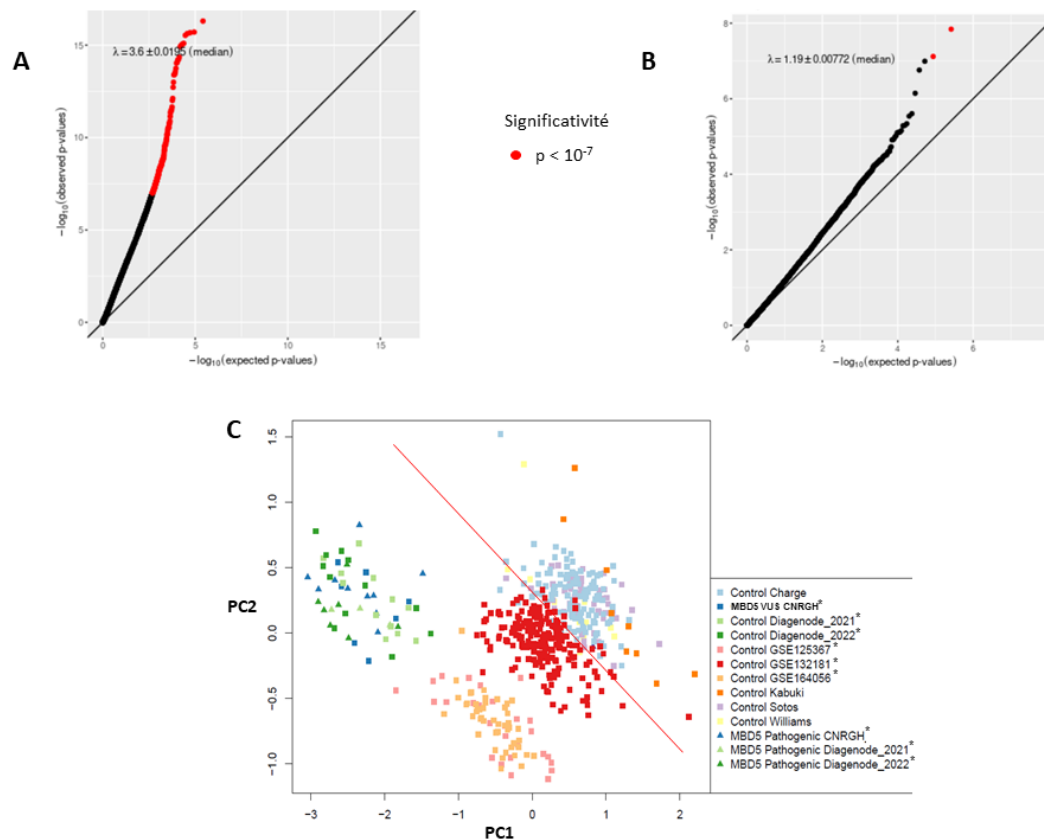


Figure 44. Recherche d'une épigraphie pour le gène *MBD5* sur les sondes communes entre puces EPIC et 450K. A. Q-Q plot du modèle de régression linéaire entre porteurs de variations pathogènes de *MBD5* et témoins, sans ajustement sur les surrogate variables ; B. Analyse avec ajustement sur les surrogate variables. Cette stratégie diminue l'inflation des p-valeurs significatives mais nous ne détectons plus que 2 sondes statistiquement significatives en tenant compte des comparaisons multiples ; C. Projection des deux premières composantes principales des résidus des données de méthylation : les données issues de puces EPIC ségrègent à gauche du trait rouge, ajouté à la figure par soucis de lisibilité. *Donnes issues de puces EPIC.

En revanche, l'analyse en composante principale restreinte aux résidus de méthylation de ces sondes ne montre pas une différenciation parfaite des cas par rapport aux témoins.

A l'issue de ce travail, il persiste une ambiguïté quant à l'existence et la nature d'une épisignature pour les pertes de fonction du gène *MBD5*. D'une part, l'existence d'un *hot-spot* d'hypométhylation dans la partie proximale du gène *MBD5* est un indice fort en réalité de sa nature biologique. D'autre part, nous n'avons pas trouvé d'explication satisfaisante quant à l'hétérogénéité des données de méthylation sur l'analyse en composante principale. De plus, une différenciation semble persister entre nos données et celles téléchargées sur GEO sur le premier axe de l'analyse en composante principale. Les puces à ADN en général (Leek et al., 2010) et les puces *infinium* en particulier (Ross et al., 2022) sont soumis à de nombreuses variations du fait de facteurs techniques qui ne peuvent être complètement contrôlés. De plus, les variations génétiques situées à proximité des sondes peuvent influencer le niveau de méthylation mesuré. Les données sont donc soumises à un risque de biais de stratification de population (Daca-Roszak et al., 2015; W. Zhou et al., 2017).

Les 11 patients pour lesquels un *cluster* est visible sur la projection des deux premières composantes principales ne présentent pas de particularité clinique. Il n'y a pas non plus de corrélation parfaite avec la localisation de la variation génétique. En effet, 9/11 des patients qui présentent une délétion sont présents dans ce *cluster* contre 2/9 des patients qui présentent une variation nucléotidique. Pour ces deux derniers, l'un présente une perte du codon d'initiation et l'autre une délétion d'une paire de bases située dans l'exon 8, soit en dehors de la localisation du *hot-spot* d'hypométhylation.

La localisation de cette hypométhylation, englobe un îlot CpG situé dans le promoteur des gènes *MBD5* et *ORC4*. Dans la mesure où l'hypométhylation d'un îlot CpG est corrélé à une augmentation de la

transcription du gène correspondant (Jones, 2012), cette observation pourrait signifier une boucle de régulation compensatoire de la dysfonction d'une des deux copies du gène – du moins chez certains patients. La comparaison de l'expression de *MBD5* en fonction de la méthylation des sondes de ce *hot-spot* permettrait de corroborer cette hypothèse. Il faut tout de même noter que cette éventuelle compensation à l'échelle biologique ne semble pas se traduire par un phénotype atténué chez les patients : deux d'entre eux présentent par exemple une encéphalopathie épileptique.

En tout état de cause, et malgré leur intérêt sur le plan biologique, ces données ne nous permettent pas d'utiliser l'épissignature *MBD5* pour classer des VSI.

Le processus de conception d'un *pipeline* pour la détection d'un épissignature a été très instructif sur les qualités et limites de l'analyse des puces de méthylation *Infinium* d'Illumina. Nous en avons tiré plusieurs préoccupations concernant l'utilisation des épissignatures en pratique clinique.

Premièrement, de nombreux biais viennent modifier le niveau de méthylation, et ceux-ci ne peuvent être totalement atténués par les contrôles qualité et le traitement statistique des données. La nature du *pipeline* bioinformatique utilisé est en lui-même un élément modificateur. De ce fait, il semble à tout le moins nécessaire de disposer de témoins pour chaque lecteur de puce utilisé (avec une standardisation du protocole en amont pour limiter les variations techniques) et de valider toute épissignature sur une cohorte indépendante.

Deuxièmement, il est possible qu'une partie des individus porteurs d'une variation pathogène du gène ne présente pas l'épissignature ou bien que des profils intermédiaires existent. La détection de sondes différenciellement méthylées ne permet pas d'obtenir cette information. L'utilisation d'une technique d'apprentissage non supervisée – par exemple la visualisation des premières composantes principales des données de méthylation – nous semble donc nécessaire.

Ces préoccupations nous ont mené à conduire un travail d'évaluation des épissignatures publiées, que nous présenterons dans la partie suivante.

Chapitre III. Evaluation des épisignatures publiées

i. Objectifs et résumé du travail publié

Comme nous l'avons vu, un nombre important – et en extension régulière – d'épisignatures ont été rapportées et plusieurs équipes proposent d'utiliser ce paradigme pour aider à la classification de VSI.

Toutefois, nous considérons qu'il existe actuellement plusieurs obstacles à l'utilisation des épisignatures en pratique clinique. A savoir l'absence : i) de méthodologie standardisée pour la conception d'épisignature, ii) de critères explicites permettant de définir si une épisignature est cliniquement significative, iii) de la réplication de leur existence par des équipes indépendantes, iv) d'évaluation de leurs performances diagnostiques et même dans certains cas, v) de la validation de leur existence sur un échantillon indépendant de patients.

Bien que certaines épisignatures ont été publiées par deux équipes différentes, cela ne peut pas être considéré comme une réplication car les méthodologies utilisées ne sont pas comparables. La publication de plusieurs épisignatures informe quant à un effet réel des variations pathogènes du gène sur la méthylation des leucocytes mais pas sur la liste de sondes à considérer pour séparer de manière optimale les cas des témoins. Il n'y a pas d'exemple dans la littérature de comparaison des performances de deux épisignatures rapportées pour un même syndrome.

Enfin, l'une des deux équipes principales impliquée dans la description d'épisignatures commercialise une solution diagnostique à l'aide de ses données. Ses travaux sont donc à risque de conflit d'intérêt et une validation indépendante de ces épisignatures est impérative. De plus, les données

brutes et la méthodologie exacte ne sont pas disponibles, car protégées par un brevet. L'équipe de London ne publie qu'une liste de sondes et non la pondération attribuée à chacune d'entre elle par leur algorithme d'apprentissage automatique.

Pour avancer sur cette problématique, nous avons mené un effort collaboratif à l'échelle nationale pour obtenir des échantillons d'ADN de patients porteurs de variations pathogènes dans des gènes associés aux épisignatures rapportées. Vu la rareté des différents syndromes neurodéveloppementaux, ce travail n'a été rendu possible qu'avec l'apport des différents laboratoires de biologie moléculaire français contactés par l'intermédiaire du réseau AnDDI-Rares. Cet effort visait à répliquer l'existence d'épisignatures, d'évaluer leurs performances diagnostiques et de démontrer la possibilité d'utiliser cette liste de sonde pour l'aide à la classification de VSI.

Nous avons recueilli 105 échantillons d'ADN de contrôles positifs pour 6 syndromes neurodéveloppementaux pour lesquels deux épisignatures étaient répertoriés dans la littérature et 3 pour lesquels une épisignature était publiée. Nous avons également inclus 58 porteurs de VSI dans ces gènes et 6 individus présentant un phénotype clinique classique de ces syndromes sans variation pathogène retrouvée dans le ou les gènes correspondants. 24 témoins sans pathologie neurodéveloppementale connue ont également été inclus. Tous les sujets ont bénéficié d'une analyse de méthylation sur puce EPIC.

Pour évaluer les épisignatures nous avons utilisé la méthodologie suivante pour chaque épisignature :

- I- La liste de sondes était extraite de la publication correspondante. L'évaluation portait uniquement sur ces sondes.
- II- Un modèle de régression linéaire était ajusté sur les valeurs β de ces sondes avec l'âge, le sexe et la composition cellulaire comme co-variables. Les résidus (valeurs β indépendante de l'effet des co-variables) étaient considérés pour les étapes suivantes. Cette étape est nécessaire car la méthodologie pour définir les sondes différentiellement méthylées passe invariablement par une étape de régression linéaire.
- III- Le statut de chaque individu était prédit à l'aide d'un algorithme de classification appelé méthode des plus proches voisins (Cover & Hart, 1967; J. H. Friedman et al., 1975). Cet algorithme transforme des données quantitatives à haute dimension (ici les résidus des valeurs β) en une matrice de distance à deux dimensions. Le statut d'un individu est déduit du statut de ses plus proches voisins sur la matrice de distance. Cet algorithme permet donc d'estimer la spécificité et la sensibilité de la signature à partir des contrôles positifs et des témoins, ainsi que sa capacité à classer des VSI. Il a intérêt de se passer de la pondération attribuée à chaque sonde par les modèles d'apprentissage automatique utilisés dans les différentes publications : celle-ci n'est généralement pas publiée.
- IV- La bonne séparation des cas et des contrôles était confirmée par visualisation de la projection deux premières composantes principales et par partitionnement hiérarchique.

La spécificité était invariablement parfaite. En revanche, la sensibilité était très variable selon les épisignatures. En particulier, deux épisignatures (l'une des deux épisignatures publiées pour le gène *CHD8* et celle publiée

pour le gène *CREBBP*) présentait une sensibilité insuffisante pour l'utilisation en pratique clinique.

Le modèle permettait effectivement d'aider à la classification de VSI. De plus, un patient présentant un phénotype typique de syndrome de Sotos (gène *NSD1*) sans variation pathogène connue dans le gène présentant effectivement l'épissignature.

Ces données nous ont permis de faire plusieurs recommandations. Premièrement, certaines épissignatures n'ont pas une fiabilité suffisante pour être utilisée en pratique clinique. Deuxième, dans plusieurs cas, des patients porteurs de VSI ou même des contrôles positifs présentaient un profil de méthylation intermédiaire entre les contrôles positifs et les témoins. Ce cas de figure était connu pour le gène *ATRX* (Schenkel et al., 2017) mais nos travaux indiquent qu'il pourrait être plus répandu. Il nous semble impératif d'associer une méthode d'apprentissage non supervisée (comme une analyse en composante principale) à un algorithme de *machine learning* pour détecter ces situations limites pour lesquels il n'est pas possible de conclure avec un niveau de certitude suffisant. Plus généralement, une expertise gène à gène en fonction des singularité biologiques est nécessaire et la simple utilisation d'un algorithme de classification unique n'est pas suffisante dans une démarche diagnostique. Troisièmement, la réalisation d'une puce de méthylation pour un nombre limité de contrôles positifs et de témoins permet l'utilisation des listes de sonde publiées pour la classification de VSI. Enfin, concernant les syndromes pour lesquels plusieurs épissignatures sont publiés, nous observons deux cas (gènes *CHD8* et *KMT2D*) où les performances diagnostiques de l'une sont nettement supérieures à l'autre.

Ce travail mènera à une publication scientifique, je joins à la suite le manuscrit de l'article qui sera prochainement soumis. Ma contribution à ce travail a été la suivante :

- Recueil des échantillons et des informations cliniques
- Participation à la conception de la méthodologie et à l'analyse des résultats
- Ecriture du manuscrit

Title

Independent evaluation of ten published neurodevelopmental disorder epesignatures in a diagnostic perspective

Authors

Thomas Husson, François Lecoquierre, Anne-Claire Richard, Anne Boland, Jean-François Deleuze, Alice Goldenberg, Anne-Marie Guerrot, Gabriella Vera, Pascale Saugier-Veber, AnDDI-Rares collaborators, Olivier Guillin, Gaël Nicolas, Dominique Campion, Camille Charbonnier.

Abstract

Variants of Unknown Significance (VUS) are a major challenge upon genetic analysis interpretation in the context of rare disease diagnostics. The publication of epesignatures as effective biomarkers of certain neurodevelopmental disorders has recently raised a lot of hopes. In case of a VUS detection, they may help interpreting the results and reach a final diagnosis. More than 50 epesignatures are now available in the literature to sort out the pathogenicity of more than 60 monogenic syndromes. Yet, prediction abilities of most of these epesignatures have not been independently investigated so far, although this is a prerequisite for informed and rigorous use in a diagnostic setting. In that aim, we generated DNA methylation data from 106 independent carriers of (likely) pathogenic variants in ten different genes with available epesignature and 24 age and sex-matched controls. Combining published lists of epesignature probes with a k-nearest-neighbour algorithm, we provide specificity and sensitivity estimates for each of the signatures. We show that although specificity satisfies diagnostic use requirements, the sensitivities of the ten signatures unexpectedly span a very large spectrum. While *ATRX*, *DNMT3A*, *KMT2D*, and *NSD1* signatures displayed a 100% sensitivity, *CREBBP* (the only signature) and *CHD8* (one of the two signatures) reached less than 40% sensitivity on our dataset. Remaining Cornelia de Lange syndrome, *KDM5C*, *CHD7*, and *KMT2A* signatures reached fair but unstable sensitivity, suffering from heterogeneous methylation profiles among cases and rare discordant samples. This study clearly advocates for the analysis of larger validation sample sizes, including expert genotype/methylation and phenotype/methylation correlations to better characterize the validity perimeter and interpretation of each epesignature and precisely understand the clinical implications of specific methylation profiles.

Introduction

The efficient etiological diagnosis of neurodevelopmental disorders represents an important matter of public health. However, behind this single denomination, NDDs encompass a wide spectrum of clinical

manifestations, arising from a large and heterogeneous set of rare disorders, from monogenic, fully penetrant Mendelian disorders to non-syndromic presentations with variable expression of traits. Providing a timely and accurate molecular diagnosis of monogenic disorders is of utmost importance for patients and their families. A precise and definite molecular diagnosis helps define personalized care for each patient, paves the way to genetic counseling and offers the possibility of prenatal diagnosis. In some cases, the final diagnosis is obtained only after years of diagnostic odyssey, while some cases remain unsolved.

Large trio-based exome (ES) and genome sequencing (GS) studies have shown that, beyond environmental factors, genetic factors considerably contribute to the determinism of NDDs. However, genetic causes are extremely heterogeneous. A thousand genes are now considered to contribute to the pathophysiology of NDDs. In the recent years, implementation of ES and GS in a diagnostic setting has largely contributed to an increase in the rate of patients with a diagnosis. An important source of improvement is that diagnostic yields results from the rigorous interpretation of (likely) pathogenic (class 4-5) variants according to the ACMG-AMP (American College of Medical Genetics and Genomics – American for Molecular Pathology) –classification (Richards et al., 2015b). However, the interpretation of ES/GS is plagued by a large number of variants that remain of uncertain significance (VUS, ACMG-AMP class 3 variants), in up to 10% of analyzes. When reported, these variants add some level of uncertainty to undiagnosed patients. Indeed, further analysis would be required to reclassify them as pathogenic variants or benign. Thus, sorting benign from pathogenic variants from VUS is one of the most important tasks in medical genomics in the post-sequencing era.

In the last decade, a new and promising strategy has emerged as an alternative to hard-to-design functional studies to classify such variants. The rationale is that a large number of NDD genes are involved in the regulation of gene expression, from transcription factors/regulators to critical players of the dynamic 3D-chromatin organization and the epigenetic machinery, including histone and DNA methylation regulation (Kaplanis et al., 2020; Satterstrom, Kosmicki, Wang, Breen, De Rubeis, An, Peng, Collins, Grove, Klei, Stevens, Reichert, Mulhern, Artomov, Gerges, Sheppard, Xu, Bhaduri, Norman, Brand, Schwartz, Nguyen, Guerrero, Dias, Betancur, et al., 2020). Therefore, the main idea is to investigate the pathogenicity of genetic variants through the identification of epesignatures, namely vast but specific disruptions in the DNA methylation patterns, that are homogeneously shared among affected samples, but absent in unaffected samples. Over the last ten years, several research groups have published such epesignatures, using Illumina 450K and more recently 850K EPIC Infinium Beadchips (Aref-Eshghi et al., 2017, 2018, 2020; Cherik et al., 2022; Foroutan et al., 2022; Haghshenas et al., 2021; M. A. Levy, McConkey, et al., 2022; Mirza-Schreiber et al., 2022; Radio et al., 2021; Rooney et al., 2021; Rouxel et al., 2022; Schenkel et al., 2017, 2021; Siu et al., 2019). Thereby, DNA methylation analysis, mostly issued from blood samples, can be used as an additional piece of evidence in the characterization of pathogenic variants. In practice, an epesignature consists of an optimal subset of differentially methylated positions (ranging from hundreds to thousands) between patients affected by a given monogenic condition due to a pathogenic variant in a specific set of genes, unique gene or

even a specific functional domain, and controls. In most cases, only uncorrelated probes which best discriminate carriers from controls are retained in an episignature and are weighted by a machine learning algorithm. Ideally, these genome-wide disruptions of the methylation pattern have to be specific and homogeneously found in individuals carrying pathogenic genetic variations in the gene of interest. Ultimately, such positions are used to train a machine learning algorithm that can be used in combination with other arguments from the ACMG-AMP variants interpretation recommendations to discriminate whether a VUS is eventually pathogenic or not. More than 50 episignatures have been reported in the literature so far, in the context of more than 60 syndromes (M. A. Levy, McConkey, et al., 2022; M. A. Levy, Relator, et al., 2022).

This strategy raises hopes to end the plague of variants of uncertain significance (VUS), offering an efficient alternative to hard-to-design and time-consuming in vitro studies, not easily suitable to a diagnostic setting (Marwaha et al., 2022). Several research groups have repeatedly provided evidence of the effectiveness of this concept from a general point of view (Kerkhof et al., 2022; Sadikovic et al., 2020, 2021; Siu et al., 2019), but the robustness, repeatability, and actual sensitivity of episignatures still need to be assessed in independent datasets, for the following reasons. Firstly, because the field remains too young to benefit from gold standard quality controls and statistical analyses, the methodology used to define episignatures suffers from a lack of consensus. Several genes have been associated with episignatures resulting from diverse methodologies by several groups. In the absence of an independent validation set, it is impossible to decide objectively which episignature performs best. Secondly, DNA methylation datasets suffer from notorious batch effects. Added to the fact that a substantial number of episignatures have been identified on less than ten samples (M. A. Levy, McConkey, et al., 2022), there is a need to validate the existence of each episignature on independent data. Last but not least, no independent and accurate evaluation of diagnostic performances (sensitivity and specificity) has been provided.

Our objective was threefold. Primarily, we aimed to validate the ability of ten episignatures to discriminate cases from controls on an independent validation dataset. Secondly, we aimed at providing an unbiased assessment of the diagnostic accuracy of corresponding episignatures, both in terms of specificity, inter-syndrome specificity and sensitivity. Eventually, we applied our prediction strategy to the classification of VUS and describe the practical challenges encountered. For these purposes, we generated an independent validation set of a target of approximately 10 new carrier samples of each tested episignature along with aged and sex-matched controls and carriers of VUS. Because the actual classification algorithms as well as most of the raw data that could be used to replicate the training steps are not openly-shared, we obtained accurate sensitivity and specificity estimates from our validation dataset by embedding a multiple class k-Nearest Neighbor classification algorithm (kNN) into a leave-one out scheme to check the predictive abilities of each published list of probes and applied it to our VUS samples.

Methods

Sample collection

We leveraged a national collaborative effort to gather DNA samples isolated from fresh blood of probands harboring (likely) pathogenic variants in a list of twelve genes spanning ten neurodevelopmental disorders: *ATRX* (Alpha-thalassemia/mental retardation syndrome, MIM#301040), *CHD7* (syndrome CHARGE, MIM#214800), *CHD8* (Autism, Susceptibility to, 18 (AUTS18), MIM#615032), *CREBBP* (Rubinstein-Taybi syndrome 1 (RSTS), MIM#180849), *DNMT3A* (Tatton-Brown-Rahman syndrome (*TBRS*), MIM#615879), *KDM5C* (Intellectual developmental disorder, X-linked syndromic, Claes-Jensen type (MRXSCJ), MIM#300534), *KMT2D* (Kabuki syndrome 1 MIM# 147920), *KMT2A* (Wiedemann-Steiner syndrome, MIM#605130), *NIPBL/SMC1A/SMC3* (Cornelia de Lange syndrome 1-3 (CdL), MIM#122470, MIM#300590, and MIM#610759), *NSD1* (Sotos syndrome, MIM# 117550), as described in Table 1. All these variants were classified as class 4 or 5 variants according to ACMG-AMP interpretation guidelines by experienced geneticists. We included germline variations detectable from DNA extracted leucocytes. Along with 24 controls free of any neurodevelopmental condition, these samples whose status is perfectly and a priori known constitute the *validation set* and are described in Supp Table 1.

As a proof of concept, we also constituted a *testing set*, including 58 samples with variants of unknown significance (VUS), as well as 9 “clinical hypothesis” cases, namely probands without definite molecular diagnosis despite a suggestive clinical presentation of a syndrome within one of the ten syndromes under investigation, as described in Table 2. All but one variation were germline. We also included a sample harboring somatic mosaicism in *CREBBP*, that could be identified on saliva but undetected from blood.

All patients or legal representative provided informed written consent for genetic analyses in a medical setting that contains a query on the use of residual samples for research.

DNA methylation analysis

Genomic DNA was extracted from blood leukocytes and bisulphite converted. DNA methylation profile was then derived using Illumina’s Infinium EPIC array, in accordance with the manufacturer’s protocol. Carriers were randomly assigned a chip position while controls were homogeneously distributed along the chip. DNA methylation arrays were generated by Diagenode SA and the CNRGH, Evry. These data were newly generated without overlap with original epigenome training sets.

Raw IDAT data were processed and normalized using the Meffil R package (Min et al., 2018). This package efficiently handles large DNA methylation datasets. Briefly, probes which failed methylation detection (detection p-value > 0.01) in more than 5% of samples were removed. Samples with > 1% of failed probes or an outlier methylation distribution (methylation/unmethylation ≥ 3 s.d. from mean) were flagged. Three samples (MRXSCJ_17, Sotos_12 and Sotos_18) failed these quality controls. At

last, all remaining samples were functionally normalized as advocated in the Meffil documentation, width adjustment on array, sentrix column and row.

Several DNA methylation predictions were added to the sample information table. Sex predictions were extracted from the standard meffil normalized object. No inconsistencies between reported and predicted sex were noted. Blood cell counts were estimated with the `meffil.cell.count.estimate` function. DNA methylation age was predicted with the `DNAMAge` function (`methyclock` R package, Pelegri-Siso et al, 2021 (Pelegí-Sisó et al., 2021)). Among all available epigenetic clocks, the skinHorvath clock (Horvath et al., 2018), which was trained on skin and blood 450K samples, displayed the strongest correlation with actual age at inclusion on our dataset, and was thereby used as age predictor. For statistical analyses, we adjusted on age and sex predictions for consistency across samples with known or unknown age and sex.

Statistical analysis

For each considered episignature we performed the following steps:

- i) The probe list was retrieved from supplemental information of the corresponding publication (Table 1). All analyses regarding this episignature were then restricted to this list of probes.
- ii) A linear regression model was fitted to adjust probe β -values on predicted age, sex and cell composition as covariables. The following steps were performed on model residuals, in other words methylation variations that remain independent from main known confounders.
- iii) Following a leave-one-out scheme, the kNN R implementation from the “class” package was used to predict the status (case or control) of every validation sample, case or control, using all remaining validation samples as training set. To control inter-syndrome specificity, a multiclass kNN was fitted to all syndromes simultaneously. Eventually, we let the number k of nearest neighbours and required level of consensus vary to challenge the robustness of our results. Parameters ranged from “2/2”, perfect consensus between the two nearest neighbors, to “5/5”, perfect consensus between the five nearest neighbours. For 4 and 5 nearest neighbour predictions, because some syndromes display close signature profiles, we also allowed for the possibility of one discordant nearest neighbour (respectively “3/4” and “4/5”). Sensitivity and specificity were computed for each setting using this leave-one-out strategy from cases and controls. *KDM5C* and *ATRX* genes being located on the X chromosome, it is expected that female carriers should not fully present the same episignature as male carriers. We therefore restricted sensitivity computations for these two genes to male samples. Inter-syndrome specificity was computed from other variant carriers. Sensitivity and specificity 95% confidence intervals were computed from an exact binomial distribution.

- iv) The reliability of case/control discrimination by the episcapature was visually inspected on the first two principal components as well as on a heatmap of step (ii) DNA methylation residuals.
- v) Finally, VUS and clinical hypothesis samples from the testing set were classified using by the “3/4” kNN algorithm.

Results

Recruitment

A total of 106 samples from ten neuro-developmental disorders described in Table 1 along with 24 age and sex-matched controls free from neurodevelopmental disorders passed basic meffil quality checks and were included in our evaluation. Sample characteristics are displayed in Supplementary Table 1.

Independent evaluation of episcapature prediction accuracy

For every episcapature, we combined a quantitative evaluation of prediction performances as obtained from kNN clustering with a visual inspection of DNA methylation profiles on the validation dataset. Sensitivity and specificity were estimated using a leave-one-out strategy on our independent dataset, guaranteeing the absence of overfitting and thereby unbiased estimates. Because most differential methylation analysis workflows rely on age, sex and blood cell count adjustments, all analyses and graphical representations were made from residual methylation levels, after adjustment for such major confounders.

In all but one setting, our multiclass kNN prediction algorithm reached 100% specificity (95%CI [85.7%-100%]), whatever the kNN parametrization. The one control misclassification occurred in the AUTS18/*CHD8_2* episcapature with a 2/2 NN configuration, which must be two permissive. As displayed on Figure 1, we observed 100% sensitivity on four gene specific episcapatures: *ATRX*, *TBR1/DNMT3A*, *Kabuki/KMT2D*, *Sotos/NSD1*), with concordance between both available episcapatures in the last two settings and robustness to kNN parametrization. A subset of episcapatures showed good but highly variable sensitivities in the 40%-90% range according to the kNN configuration used, namely *CdL*, *CHARGE/CHD7*, *AUTS18/CHD8_1*, *MRXCSJ/KDM5C_1* and *2*, *WDST1/KMT2A*. The sensitivity of these episcapatures decreased as we increased the number of concordant nearest neighbours required for prediction. Two episcapatures, *AUTS18/CHD8_2* and *RSTS/CREBBP*, showed less than 40% sensitivities, whatever the configuration, with close to 0 sensitivity for increasing numbers of concordant nearest neighbours.

Visual inspection of PCA plots and heatmaps correlated well with sensitivity estimates. Three PCA plots and corresponding heatmaps, namely *Sotos/NSD1*, *MRXCSJ/KDM5C* and *RSTS/CREBBP*, chosen to best illustrate the wide spectrum of episcapature performances, are presented in Figure 1. Graphical representations for other episcapatures can be viewed in Supplementary Figures 1 to 13. Notably,

syndromes with 100% sensitivity displayed a separation of cases and controls by a large margin, while syndromes with intermediate and unstable sensitivity suffered from a large dispersion of variant carriers, with a juxtaposition of what could be called extreme and intermediate DNA methylation profiles. This scenario explains the increasing difficulty of the kNN algorithm to find large numbers of concordant nearest neighbours to accurately predict these intermediate profiles, which are located halfway between controls and more extreme cases. This is expected for female carriers of variants in the *KDM5C* and *ATRX* genes, which are located on the X chromosome, but it is also observed among male carriers of *ATRX* mutations and carriers of *CHD7* mutations. The intermediate status of a subset of profiles is clearly visible on the heatmaps, with intermediate shade of yellow or red. *AUTS18/CHD8_2* and *RSTS/CREBBP* episignatures showed incomplete separation between cases and controls.

The strength of episignatures was also reflected by the percentage of methylation variance within the signature that is explained by the first axis. Indeed, the separation between cases and controls was observed on the first and major axis of variation. On an episignature like *Sotos/NSD1*, this separation accounts for 86,7% of the variance. For *MRXCSJ/KMD5C*, this number falls down to 66,5%. On the *RSTS/CREBBP* episignature, the separation between cases and controls accounts for no more than 26,8% of the variance of methylation level residuals, leaving more than two thirds of the variance to unexplained factors or noise.

Surprisingly, two carriers of *KDM5C* class 4 and 5 variants appeared as outliers on both the heatmap and first principal components. Sample identity was double checked by sanger sequencing and a global status prediction of all validation samples in search for obvious sample swaps. The first outlier is *MRXSCJ_8*, carrying a *de novo* missense variant, NM_004187.3:c.593G>A, p.(Arg198Gln), classified as likely pathogenic. The second outlier is *MRXSCJ_10*, carrying a small deletion NM_004187.3:c.645_657+5del, p.? (inheritance unknown), classified as pathogenic. ACMG-AMP classification of the variants was based on a collegial decision taking *in silico* pathogenicity predictions, mode of inheritance, clinvar co-occurrences as well as phenotype concordance into account and is not questioned.

Among the ten syndromes under investigation, six had two distinct published episignatures. Except for *AUTS18/CHD8*, performances were somewhat similar between the two available signatures. In contrast, our data shows that the episignatures that we called *AUTS18/CHD8_1* (Siu et al., 2019) clearly outperformed the other one (*AUTS18/CHD8_2*) (Aref-Eshghi et al., 2020). Of note, the *AUTS18/CHD8_2* episignature was trained on only n=5 samples. In Levy et al 2022, the same team reported an increased sample size of 28 *CHD8* cases but no update was published regarding the probe list for this episignature.

Episignatures as a biomarker for VUS classification

Overall, we included 58 VUS carriers and 9 probands with a suggestive phenotype but negative exome sequencing. Sample characteristics and classification results are displayed in Table 1. Several details deserve special attention.

Twelve VUS carriers showed methylation profiles compatible with the corresponding epesignatures. From these, two samples had a discordant prediction between the two available epesignatures. CHARGE_18 presented a positive epesignature with the CHARGE/CHD7_1 epesignature but not by the CHARGE/CHD7_2 epesignature. The opposite was true for CHARGE_16 sample. Both segregated properly with cases in the PCA plot made from the CHARGE/CHD7_2 probes but separation between cases and controls was milder for this epesignature, which can cause classification errors.

In general, a lot of epesignatures find themselves challenged by VUS with such intermediate profiles. A similar discrepancy was observed for sample Kabuki_11, which presented a positive Kabuki/KMT2D_2 epesignature but a negative Kabuki/KMT2D_1 signature, because of an intermediate methylation profile. Such profiles were only expected from female carriers of X-linked syndromes, but oddly enough, even in *ATRX* and *KDM5C* genes, a few males VUS carriers presented intermediate profiles which remain impossible to interpret without further knowledge.

We also included one proband carrying a pathogenic *CHD7* complete duplication among VUS. This duplication presented a negative epesignature, but PCA plot inspection revealed that it was projected on the opposite end of the first axis. This observation is consistent with the haploinsufficiency mediated pathogenicity in CHARGE syndrome (Lalani et al., 2006).

A somatic mosaicism in *CREBBP* was also included as a VUS. This variant was identified on saliva but undetected from blood. This sample was classified as control by the kNN but remains as the intersection between cases and controls. Given the weak sensibility of this epesignature, it is not possible to draw a firm conclusion about the detection of this somatic variant.

Among clinical hypotheses, only one proband with a Sotos syndrome clinical phenotype harbored positive Sotos/NSD1 epesignatures, suggesting that the patient actually presents this disorder and that further genetic analyses should be proposed to identify the causal variant. A second sample, RSTS_12, would be close to class 4 and 5 variants carriers both on PCA and heatmap, but the separation was too mild and the dispersion of cases too strong for the kNN algorithm to reach enough consensus.

Discussion

We sought to provide an independent evaluation of ten epesignatures in terms of accuracy and robustness. For every signature, the multiclass and stringent kNN strategy guaranteed perfect specificity, with regard to both negative controls and other potentially close syndromes. On the other hand, sensitivity was highly heterogeneous among syndromes, ranging from perfectly separable syndromes (*ATRX*, *TBR1/DNMT3A*, Kabuki/*KMT2D*, Sotos/*NSD1*) to very mild signatures with less than

40% sensitivity (RSTS/*CREBBP* and AUTS18/*CHD8_2* signatures). Visual inspection of PCA plots and heatmaps revealed that, essentially, epesignatures can be split up in three groups.

A first group with robust epesignatures, in which pathogenic variants lead to a strong enough methylation disruption to distinguish cases from controls without ambiguity. Herein, VUS can be reliably classified as pathogenic or benign. This group includes both Sotos/*NSD1* signatures and TBRS/*DNMT3A*. Both syndromes are known to cause a drastic hypomethylated signature among carriers (Aref-Eshghi et al., 2020; Choufani et al., 2015 ; Levy, McConkey, et al., 2022; Levy, Relator, et al., 2022).

In the other hand, a second group showed very weak epesignatures, namely RSTS/*CREBBP* and AUTS18/*CHD8_2*. Reasons may be either biological or technical. Perhaps the methylation impact is not strong enough to allow for a consistent distinction between cases and controls or the methylation profiles among cases suffer from a diversity which has not been understood yet. It is also possible that the published epesignatures were contaminated by batch effects or suffered from overfitting to a too small training set, which prevents generalization to other independent datasets. Our data suggest caution against any use of these epesignatures in a diagnostic setting. Progress could be made by increasing the sample size and exploring methylation/genotype or methylation/phenotype correlations with the help of gene-specific expert centers.

In between, remaining epesignatures have proven their reasonable sensitivity, which makes them effective biomarkers for VUS classification but display clearly intermediate profiles. Such intermediate profiles are not usually reported. One noteworthy exception concerns the *ATRX* associated epesignature. Authors who have reported this epesignature have found that the methylation profile change was variable between *ATRX* pathogenic variant carriers: when viewed individually, several probands showed only a small proportion of loci involved in the epesignature actually differentially methylated with respect to controls (Schenkel et al., 2017). A critical insight of our work is that this heterogenous pattern of methylation was far from uncommon in published epesignatures. We also identified two samples with completely discordant methylation patterns in *KDM5C*. Close attention to those two variants revealed that they affected the same exon, which is expressed in two close isoforms with a one-codon difference. One of these variants affects a canonical splice site with predictions in favor of either in-frame exon skipping or in-frame intronic sequence retention of 36 base pairs although with a stop codon in this sequence, based on SpliceAI predictions (Jaganathan et al., 2019). The other one is a missense variant without any arguments as a splicing modulator. Of note, two truncating variants are reported as pathogenic in the Clinvar database in the same exon (one nonsense, one frameshift indel). It remains possible, if the two variants found to be discordant regarding the published epesignature do not result into haploinsufficiency, that altering this region of the *KDM5C* protein leads to a distinct or absent epesignature. More samples are required to confirm this hypothesis.

In both instances, be it intermediate profiles or the existence of local exceptions, it is essential that more samples should be analysed to (i) confirm and understand the reason and implications of intermediate methylation profiles (ii) investigate and validate the existence of regions that may escape the signature. A larger sample size would allow for genotype/methylation and phenotype/methylation correlation analyses. A better characterization of these profiles is required to provide help reclassify variants and hence provide accurate genetic counselling to carriers. Interestingly, one *CHD8* pathogenic variant carrier (AUTS18_2) was the father of the proband (AUTS18_1), also carrying the same *CHD8* pathogenic variant. While the proband showed typical features of AUTS18/*CHD8* (intellectual disability, autism spectrum disorders and macrocephaly), her father only displayed macrocephaly (which belong to the phenotypic spectrum of AUTS18/*CHD8*) without any known neurodevelopmental features. Nevertheless, his methylation profile was without any doubt within the bulk of other AUTS18/*CHD8* variant carriers. That single observation does not support the hypothesis that milder phenotype coincides inevitably with intermediate profiles, at least regarding AUTS18/*CHD8*.

A limitation of our study is that all probes in the signature contributed equally to the prediction, contrary to other machine learning models that assign weights to each differentially methylated probe. However, with only about ten samples per syndrome, kNN provided a more reasonable and robust approach. We can assume that most of the biological significance lies in the overall list of differentially methylated probes.

So far, our work is the first independent evaluation of multiple episignatures. Moreover, to our knowledge, it is the only existing attempt to compare competing episignatures for the same genes. It also provides a proof of concept that, by investing a limited set of positive controls, a kNN classifier is able to leverage a differentially methylated probes list from summary statistic's publications to help classify VUS. The perfect inter-syndrome specificity allows for the interpretation of probands with a suggestive phenotype but with negative exome diagnostic. The fact that our algorithm detected a positive episignature in such a proband reinforces the incentive to carry on the search for deep intronic or complex structural variants in this proband.

Last but not least, the extreme disparity between episignature performances, and the complexity of some scenarios emphasize the fact that the use of episignature as biomarkers should not in any case be used blindly as an automatized prediction tool, at least not yet and not for all available signatures in the literature. It is important that automatic predictions should be challenged by careful inspection of heatmaps and principal components. Methylation profiles will have to be accurately discussed and interpreted by gene-specific experts in the light of essential genotype/methylation and phenotype/methylation correlation analyses. As always in the molecular biology diagnostic process, biological and clinical expertise about the neurodevelopmental syndrome is mandatory to make an analysis reliable. Unless robust information is provided, episignature predictions should be reluctant to keep "inconclusive" results until knowledge improves.

List of AnDDI-Rares collaborators who contributed by sending samples

Alexandra Afenjar, Sélerine Audebert-Bellanger Catherine Badens, Frédéric Bilan, Varoona Bizaoui, Roseline Borie, Elise Boucher, Tiffany Busa, Céline Bonnet, Marie Bournez, Odile Boute, Roseline Caumes, Perrine Charles, Nicolas Chassaing, Nicolas Chatron, Benjamin Cogne, Estelle Colin, Cindy Colson, Valérie Cormier, Rodolphe Dard, Benjamin Dauriat, Audrey Dayez, Julian Delanne, Florence Demurger, Anne-Sophie Denommé-Pichon, Christel Depienne, Anne Dieux, Christele Dubourg, Salima El Chehadeh, Laurence Faivre, Claire Francastel, Mélanie Fradin, Aurore Garde, David Geneviève, Marion Gerard, Brigitte Gilbert-Dussardier, Godeliève Morel, Cyril Goizet, Anne Guimmier, Inès Harzallah, Delphine Héron, Bertrand Isidore, Boris Keren, Elodie Lacaze, Alinoë Lavillaureix, Gwenael Leguyader, Daphné Lehalle, James Lespinasse, Jonathan Levy, Stanislas Lyonnet, Nolwenn Jean-Marçais, Sandrine Marlin, Luisa Marsili, Cyril Mignot, Sophie Nambot, Mathilde Nizon, Laurent Pasquier, Laurine Perrin, Veronique Pingault, Amélie Piton, Fabienne Prieur, Marc Planes, Sylvie Odent, Chloé Quélin, Sylvia Quemener, Melanie Rama, Marlène Rio, Elise Schaefer, Sophie Rondeau, Thomas Smol, Sabine Sigaudy, Renaud Touraine, Frederic Tran Mau-Them, Aurélien Trimouille, Clémence Vanlerberghe, Valérie Vantalon, Alain Verloes, Marie Vincent, Alban Ziegler

Acknowledgement

Grants from La Fondation de l’Avenir and from la Fondation Deniker

We are grateful to all members of the AnDDI-Rares network for their participation

Collaboration CEA-DRF-Jacob-CNRGH—CHU de Rouen. This work did benefit from support of the France Génomique National infrastructure, funded as part of the «Investissements d’Avenir» program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09).

References

Aref-Eshghi, E., Bend, E. G., Hood, R. L., Schenkel, L. C., Carere, D. A., Chakrabarti, R., Nagamani, S. C. S., Cheung, S. W., Campeau, P. M., Prasad, C., Siu, V. M., Brady, L., Tarnopolsky, M. A., Callen, D. J., Innes, A. M., White, S. M., Meschino, W. S., Shuen, A. Y., Paré, G., ... Sadikovic, B. (2018). BAFopathies’ DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaidis-Baraitser syndromes. *Nature Communications*, 9(1), 4885. <https://doi.org/10.1038/s41467-018-07193-y>

Aref-Eshghi, E., Kerkhof, J., Pedro, V. P., Groupe DI France, Barat-Houari, M., Ruiz-Pallares, N., Andrau, J.-C., Lacombe, D., Van-Gils, J., Fergelot, P., Dubourg, C., Cormier-Daire, V., Rondeau, S., Lecoquierre, F., Saugier-Veber, P., Nicolas, G., Lesca, G., Chatron, N., Sanlaville, D., ... Sadikovic, B. (2020). Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *American Journal of Human Genetics*, 106(3), 356-370. <https://doi.org/10.1016/j.ajhg.2020.01.019>

Aref-Eshghi, E., Schenkel, L. C., Lin, H., Skinner, C., Ainsworth, P., Paré, G., Rodenhiser, D., Schwartz, C., & Sadikovic, B. (2017). The defining DNA methylation signature of Kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance. *Epigenetics*, 12(11), 923-933. <https://doi.org/10.1080/15592294.2017.1381807>

- Cherik, F., Reilly, J., Kerkhof, J., Levy, M., McConkey, H., Barat-Houari, M., Butler, K. M., Coubes, C., Lee, J. A., Le Guyader, G., Louie, R. J., Patterson, W. G., Tedder, M. L., Bak, M., Hammer, T. B., Craigen, W., Démurger, F., Dubourg, C., Fradin, M., ... Genevieve, D. (2022). DNA methylation episcapature in Gabriele-de Vries syndrome. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 24(4), 905-914. <https://doi.org/10.1016/j.gim.2021.12.003>
- Choufani, S., Cytrynbaum, C., Chung, B. H. Y., Turinsky, A. L., Grafodatskaya, D., Chen, Y. A., Cohen, A. S. A., Dupuis, L., Butcher, D. T., Siu, M. T., Luk, H. M., Lo, I. F. M., Lam, S. T. S., Caluseriu, O., Stavropoulos, D. J., Reardon, W., Mendoza-Londono, R., Brudno, M., Gibson, W. T., ... Weksberg, R. (2015). NSD1 mutations generate a genome-wide DNA methylation signature. *Nature Communications*, 6, 10207. <https://doi.org/10.1038/ncomms10207>
- Foroutan, A., Haghshenas, S., Bhai, P., Levy, M. A., Kerkhof, J., McConkey, H., Niceta, M., Ciolfi, A., Pedace, L., Miele, E., Genevieve, D., Heide, S., Alders, M., Zampino, G., Merla, G., Fradin, M., Bieth, E., Bonneau, D., Dieterich, K., ... Lebre, A.-S. (2022). Clinical Utility of a Unique Genome-Wide DNA Methylation Signature for KMT2A-Related Syndrome. *International Journal of Molecular Sciences*, 23(3), 1815. <https://doi.org/10.3390/ijms23031815>
- Haghshenas, S., Levy, M. A., Kerkhof, J., Aref-Eshghi, E., McConkey, H., Balci, T., Siu, V. M., Skinner, C. D., Stevenson, R. E., Sadikovic, B., & Schwartz, C. (2021). Detection of a DNA Methylation Signature for the Intellectual Developmental Disorder, X-Linked, Syndromic, Armfield Type. *International Journal of Molecular Sciences*, 22(3), 1111. <https://doi.org/10.3390/ijms22031111>
- Horvath, S., Oshima, J., Martin, G. M., Lu, A. T., Quach, A., Cohen, H., Felton, S., Matsuyama, M., Lowe, D., Kabacik, S., Wilson, J. G., Reiner, A. P., Maierhofer, A., Flunkert, J., Aviv, A., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., ... Raj, K. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging*, 10(7), 1758-1775. <https://doi.org/10.18632/aging.101508>
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S. J., & Farh, K. K.-H. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, 176(3), 535-548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>
- Kaplanis, J., Samocha, K. E., Wiel, L., Zhang, Z., Arvai, K. J., Eberhardt, R. Y., Gallone, G., Lelieveld, S. H., Martin, H. C., McRae, J. F., Short, P. J., Torene, R. I., de Boer, E., Danecek, P., Gardner, E. J., Huang, N., Lord, J., Martincorena, I., Pfundt, R., ... Retterer, K. (2020). Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature*, 586(7831), 757-762. <https://doi.org/10.1038/s41586-020-2832-5>
- Kerkhof, J., Squeo, G. M., McConkey, H., Levy, M. A., Piemontese, M. R., Castori, M., Accadia, M., Biamino, E., Della Monica, M., Di Giacomo, M. C., Gervasini, C., Maitz, S., Melis, D., Milani, D., Piccione, M., Prontera, P., Selicorni, A., Sadikovic, B., & Merla, G. (2022). DNA methylation episcapature testing improves molecular diagnosis of Mendelian chromatinopathies. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 24(1), 51-60. <https://doi.org/10.1016/j.gim.2021.08.007>
- Lalani, S. R., Safiullah, A. M., Fernbach, S. D., Harutyunyan, K. G., Thaller, C., Peterson, L. E., McPherson, J. D., Gibbs, R. A., White, L. D., Hefner, M., Davenport, S. L. H., Graham, J. M., Bacino, C. A., Glass, N. L., Towbin, J. A., Craigen, W. J., Neish, S. R., Lin, A. E., & Belmont, J. W. (2006). Spectrum of CHD7 Mutations in 110 Individuals with CHARGE Syndrome and Genotype-Phenotype Correlation. *American Journal of Human Genetics*, 78(2), 303-314.
- Levy, M. A., McConkey, H., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., Bralo, M. P., Cappuccio, G., Ciolfi, A., Clarke, A., DuPont, B. R., Elting, M. W., Faivre, L., Fee, T., Fletcher, R. S., Cherik, F., Foroutan, A., Friez, M. J., Gervasini, C., ... Sadikovic, B. (2022). Novel diagnostic DNA methylation episcapatures expand and refine the epigenetic landscapes of Mendelian disorders. *HGG Advances*, 3(1), 100075. <https://doi.org/10.1016/j.xhgg.2021.100075>

Levy, M. A., Relator, R., McConkey, H., Pranckeviciene, E., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., Palomares Bralo, M., Cappuccio, G., Ciolfi, A., Clarke, A., DuPont, B. R., Elting, M. W., Faivre, L., Fee, T., Ferilli, M., Fletcher, R. S., Cherick, F., ... Sadikovic, B. (2022). Functional correlation of genome-wide DNA methylation profiles in genetic neurodevelopmental disorders. *Human Mutation*, 43(11), 1609-1628. <https://doi.org/10.1002/humu.24446>

Marwaha, S., Knowles, J. W., & Ashley, E. A. (2022). A guide for the diagnosis of rare and undiagnosed disease : Beyond the exome. *Genome Medicine*, 14, 23. <https://doi.org/10.1186/s13073-022-01026-w>

Min, J. L., Hemani, G., Davey Smith, G., Relton, C., & Suderman, M. (2018). Meffil : Efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics (Oxford, England)*, 34(23), 3983-3989. <https://doi.org/10.1093/bioinformatics/bty476>

Mirza-Schreiber, N., Zech, M., Wilson, R., Brunet, T., Wagner, M., Jech, R., Boesch, S., Škorvánek, M., Necpál, J., Weise, D., Weber, S., Mollenhauer, B., Trenkwalder, C., Maier, E. M., Borggraefe, I., Vill, K., Hackenberg, A., Pilshofer, V., Kotzaeridou, U., ... Oexle, K. (2022). Blood DNA methylation provides an accurate biomarker of KMT2B-related dystonia and predicts onset. *Brain: A Journal of Neurology*, 145(2), 644-654. <https://doi.org/10.1093/brain/awab360>

Pelegí-Sisó, D., de Prado, P., Ronkainen, J., Bustamante, M., & González, J. R. (2021). methylclock : A Bioconductor package to estimate DNA methylation age. *Bioinformatics (Oxford, England)*, 37(12), 1759-1760. <https://doi.org/10.1093/bioinformatics/btaa825>

Radio, F. C., Pang, K., Ciolfi, A., Levy, M. A., Hernández-García, A., Pedace, L., Pantaleoni, F., Liu, Z., de Boer, E., Jackson, A., Bruselles, A., McConkey, H., Stellacci, E., Lo Cicero, S., Motta, M., Carrozzo, R., Dentici, M. L., McWalter, K., Desai, M., ... Tartaglia, M. (2021). SPEN haploinsufficiency causes a neurodevelopmental disorder overlapping proximal 1p36 deletion syndrome with an epismutation of X chromosomes in females. *American Journal of Human Genetics*, 108(3), 502-516. <https://doi.org/10.1016/j.ajhg.2021.01.015>

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and Guidelines for the Interpretation of Sequence Variants : A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics*, 17(5), 405-424. <https://doi.org/10.1038/gim.2015.30>

Rooney, K., Levy, M. A., Haghshenas, S., Kerkhof, J., Rogaia, D., Tedesco, M. G., Imperatore, V., Mencarelli, A., Squeo, G. M., Di Venere, E., Di Cara, G., Verrotti, A., Merla, G., Tedder, M. L., DuPont, B. R., Sadikovic, B., & Prontera, P. (2021). Identification of a DNA Methylation Epismutation in the 22q11.2 Deletion Syndrome. *International Journal of Molecular Sciences*, 22(16), 8611. <https://doi.org/10.3390/ijms22168611>

Rouxel, F., Relator, R., Kerkhof, J., McConkey, H., Levy, M., Dias, P., Barat-Houari, M., Bednarek, N., Boute, O., Chatron, N., Cherik, F., Delahaye-Duriez, A., Doco-Fenzy, M., Faivre, L., Gauthier, L. W., Heron, D., Hildebrand, M. S., Lesca, G., Lespinasse, J., ... Genevieve, D. (2022). CDK13-related disorder : Report of a series of 18 previously unpublished individuals and description of an epigenetic signature. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 24(5), 1096-1107. <https://doi.org/10.1016/j.gim.2021.12.016>

Sadikovic, B., Levy, M. A., & Aref-Eshghi, E. (2020). Functional annotation of genomic variation : DNA methylation epismutations in neurodevelopmental Mendelian disorders. *Human Molecular Genetics*, 29(R1), R27-R32. <https://doi.org/10.1093/hmg/ddaa144>

Sadikovic, B., Levy, M. A., Kerkhof, J., Aref-Eshghi, E., Schenkel, L., Stuart, A., McConkey, H., Henneman, P., Venema, A., Schwartz, C. E., Stevenson, R. E., Skinner, S. A., DuPont, B. R., Fletcher, R. S., Balci, T. B., Siu, V. M., Granadillo, J. L., Masters, J., Kadour, M., ... Alders, M. (2021). Clinical epigenomics : Genome-wide DNA methylation analysis for the diagnosis of

Mendelian disorders. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 23(6), 1065-1074. <https://doi.org/10.1038/s41436-020-01096-4>

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., ... Buxbaum, J. D. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, 180(3), 568-584.e23. <https://doi.org/10.1016/j.cell.2019.12.036>

Schenkel, L. C., Aref-Eshghi, E., Rooney, K., Kerkhof, J., Levy, M. A., McConkey, H., Rogers, R. C., Phelan, K., Sarasua, S. M., Jain, L., Pauly, R., Boccuto, L., DuPont, B., Cappuccio, G., Brunetti-Pierri, N., Schwartz, C. E., & Sadikovic, B. (2021). DNA methylation epi-signature is associated with two molecularly and phenotypically distinct clinical subtypes of Phelan-McDermid syndrome. *Clinical Epigenetics*, 13(1), 2. <https://doi.org/10.1186/s13148-020-00990-7>

Schenkel, L. C., Kernohan, K. D., McBride, A., Reina, D., Hodge, A., Ainsworth, P. J., Rodenhiser, D. I., Pare, G., Bérubé, N. G., Skinner, C., Boycott, K. M., Schwartz, C., & Sadikovic, B. (2017). Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome. *Epigenetics & Chromatin*, 10, 10. <https://doi.org/10.1186/s13072-017-0118-4>

Siu, M. T., Butcher, D. T., Turinsky, A. L., Cytrynbaum, C., Stavropoulos, D. J., Walker, S., Caluseriu, O., Carter, M., Lou, Y., Nicolson, R., Georgiades, S., Szatmari, P., Anagnostou, E., Scherer, S. W., Choufani, S., Brudno, M., & Weksberg, R. (2019). Functional DNA methylation signatures for autism spectrum disorder genomic risk loci : 16p11.2 deletions and CHD8 variants. *Clinical Epigenetics*, 11(1), 103. <https://doi.org/10.1186/s13148-019-0684-3>

Table 1 List of investigated syndromes and episignatures with validation sample size. *Even if a testing cohort is referenced, it contains both VUS and pathogenic variant carriers without specifying the status of each one. Thus, it is not possible to determinate the size of the testing cohort.

Episignature name	N validation samples	n probes (% after QC)	Original sample size: discovery/testing	Original publication
<i>ATRX</i>	8	101 (100%)	13/5	Aref-Eshghi et al, 2020
<i>AUTS18/CHD8_1</i>	16	491 (79.6%)	7/13	Siu et al, 2019
<i>AUTS18/CHD8_2</i>		103 (99%)	5/0	Aref-Eshghi et al, 2020
<i>CdL/NIPBL, SMC1A, SMC3A</i>	7, 3, 2	128 (99.2%)	31/10	Aref-Eshghi et al, 2020
<i>CHARGE/CHD7_1</i>	13	165 (93.3%)	19/20	Butcher et al, 2017
<i>CHARGE/CHD7_2</i>		148 (100%)	45/15	Aref-Eshghi et al, 2020
<i>Kabuki/KMT2D_1</i>	10	221 (92.8%)	11/8	Butcher et al, 2017
<i>Kabuki/KMT2D_2</i>		153 (100%)	66/21	Aref-Eshghi et al, 2020
<i>MRXSCJ/KDM5C_1</i>	10	53 (96.2%)	10/0	Grafodatskaya et al, 2013
<i>MRXSCJ/KDM5C_2</i>		127 (100%)	26/8	Aref-Eshghi et al, 2020
<i>RSTS/CREBBP</i>	7	139 (100%)	30/9	Aref-Eshghi et al, 2020
<i>Sotos/NSD1_1</i>	10	7085 (92.8%)	19/19	Choufani et al, 2015
<i>Sotos/NSD1_2</i>		112 (99.1%)	47/15	Aref-Eshghi et al, 2020
<i>TBRS/DNMT3A</i>	10	139 (99.3%)	10/4	Aref-Eshghi et al, 2020
<i>WDSTS/KMT2A_1</i>	10	104 (100%)	12/4	Aref-Eshghi et al, 2020
<i>WDSTS/KMT2A_2</i>		207 (100%)	41/?*	Foroutan et al, 2022

Table 2 Testing set: variants of unknown significance and clinical hypotheses. Pred_1: results from the first published episignature. Pred_2 (when applicable): results from the second published episignature.

Syndrome/Gene	ID	Sex	Age (years)	Variant	Protein	Inheritance	Pred_1	Pred_2
Variants of unknown significance								
ATRX/ATRX	ATRX_9	M	3.5	NM_000489.6:c.371-10A>G	p.(?)	Maternal	-	NA
	ATRX_10	M	5	NM_000489.6:c.5932_5934del	p.(Ser1978del)	Maternal	-	NA
	ATRX_11	M	17	NM_000489.6:c.4156G>A	p.(Glu1386Lys)	Maternal	-	NA
	ATRX_12	M	9.5	NM_000489.6:c.5646A>G	p.(Ile1882Met)	<i>De novo</i>	+	NA
	ATRX_13	M	17	NM_000489.6:c.6418C>T	p.(Arg2140Trp)	Maternal	+	NA
	ATRX_14	M	3	NM_000489.6:c.1235A>G	p.(Asp412Gly)	Maternal	-	NA
	ATRX_15	M	4	NM_000489.6:c.2798C>A	p.(Ser933Tyr)	Maternal	-	NA
	ATRX_16	M	2	NM_000489.6:c.1217A>G	p.(His406Arg)	Maternal	-	NA
	ATRX_17	M	8	NM_000489.6:c.6064C>T	p.(Leu2022Phe)	Maternal	+	NA
	ATRX_18	M	5	NM_000489.5:c.4292A>G	p.(Gln1431Arg)	Maternal	-	NA
AUTS18/CHD8	AUTS18_17	M	5	NM_001170629.1:c.2T>C	p.(Met1?)	Maternal	-	-
	AUTS18_18	M	5	NM_001170629.1c.7508A>C	p.(His2503Pro)	Unknown	-	-
	AUTS18_19	M	13	NM_001170629.2:c.7645_7647del	p.(Asp2549del)	Paternal	-	-
	AUTS18_20	M	4	NM_001170629.1:c.2150C>T	p.(Pro717Leu)	Unknown	-	-
	AUTS18_21	M	12	NM_001170629:c.7273G>T	p.(Asp2425Tyr)	Paternal	-	-
	AUTS18_22	F	8.5	NM_001170629.1:c.2063T>C	p.(Leu688Pro)	Paternal (mildly affected)	-	-
	AUTS18_23	M	3.4	NM_001170629.1:c.4173G>T	p.(Lys1391Asn)	Not maternal	-	-
	AUTS18_24	M	3	NM_001170629.2:c.2364+1G>C	p.(?)	Maternal	-	-
	AUTS18_25	M	11	NM_001170629.1:c.2552G>A	p.(Arg851His)	Maternal	-	-
	CdL/NIPBL	CdL_8	M	5	NM_001170629:c.-457_-456delinsAT	p.(?)	<i>De novo</i>	+
CdL/NIPBL	CdL_9	M	2	NM_001170629.1:c.7307C>T	p.(Ala2436Val)	<i>De novo</i>	+	NA

CdL/SMC3	CdL_15	M	11	NM_005445:c.1417T>C	p.(Trp473Arg)	De novo	+	NA
CdL/SMC1A	CdL_16	M	17	NM_006306.3:c.2776C>T	p.(Arg926Cys)	De novo	-	NA
CHARGE/CHD7	CHARGE_13	M	3.5	arr8q12(61551446-62298932)*3		De novo	-	-
	CHARGE_14	M	8	NM_017780.3 c.6194G>A	p.(Arg2065His)	Maternal (very mild symptoms but not asymptomatic)	-	-
	CHARGE_15	F	15.6	NM_017780.4:c.5744G>A	p.(Arg1915Gln)	De novo	-	-
	CHARGE_16	M	30	NM_017780.4:c.3398C>T	p.(Thr1133Met)	Unknown	-	+
	CHARGE_17	M	11	NM_017780.3:c.4208T>G	p.(Ile1403Arg)	De novo	-	-
	CHARGE_18	F	15	NM_017780.3:c.6193C>T	p.(Arg2065Cys)	Unknown	+	-
Kabuki/KMT2D	Kabuki_11	F	16	NM_003482.3:c.13840-20C>G	p.(?)	Maternal	-	+
	Kabuki_12	F	2	NM_003482.4:c.16529A>G	p.(Tyr5510Cys)	De novo	-	-
	Kabuki_13	F	12	NM_003482.3:c.1844C>G	p.(Ser615Cys)	Unknown	-	-
MRXSCJ/KDM5C	MRXSCJ_11	M	2	NM_004187:c.3749C>T	p.(Ser1250Leu)	Maternal	-	-
	MRXSCJ_12	M	2	NM_004187.3:c.2533G>C	p.(Gly845Arg)	Maternal	-	-
	MRXSCJ_13	F	20	NM_004187.4:c.2230C>T	p.(Arg744Trp)	Not maternal	-	-
	MRXSCJ_14	M	2.5	NM_004187.3:c.1241A>G	p.(His414Arg)	Maternal	+	+
	MRXSCJ_15	F	5	NM_004187.3:c.3199G>T	p.(Glu1067*)	Maternal (mildly affected)	-	-
	MRXSCJ_16	F	1	NM_004187.5: c.1786A>C	p.(Thr596Pro)	Maternal	-	-
	MRXSCJ_18	F	5	NM_004187.5:c.1625C>T	p.(Ala542Val)	De novo	-	-
	MRXSCJ_19	M	4	NM_004187.3:c.2944C>T	p.(Arg982Cys)	Maternal	-	-
	MRXSCJ_20	M	6	NM_004187.3:c.2944C>T	p.(Arg982Cys)	Maternal	-	-
	MRXSCJ_21	M	22	NM_004187.4 :c.2517-7_251	p.(?)	Unknown	-	-
	MRXSCJ_22	M	6	NM_004187:c.860C>T	p.(Ser287Leu)	Maternal	-	-
RSTS/CREBBP	RSTS_3**	M	6	NM_004380.2:c.3779del+3-8	p.(Tyr1234_Thr1260del)	Somatic mosaicism	-	NA
	RSTS_9	M	NA	arr16p13.3(2821573-4003054)x3	CREBBP and several other genes duplication	Unknown	-	NA

	RSTS_10	M	2.5	NM_004380.2:c.5272G>A	p.(Glu1758Lys)	Maternal (affected)	-	NA
	RSTS_11	F	7	NM_004380.2:c.5357G>A	p.(Arg1786His)	<i>De novo</i>	-	NA
Sotos/NSD1	Sotos_16	M	6	arr5q35.2-35.3(176448783-176546244)*3	NSD1 partial duplication	Paternal	-	-
	Sotos_17	M	4.5	NM_022455.5:c.2117A>T	p.(Asn706Ile)	Not maternal	-	-
TBRS/DNMT3A	TBRS_11	M	12	NM_022552.5:c.1280A>T	p.(Glu427Val)	Maternal	-	NA
	TBRS_12	M	5	NM_022552.5:c.302C>T	p.(Pro101Leu)	Not maternal	-	NA
	TBRS_13	F	3m	NM_022552.5:c.1669T>G	p.(Cys557Gly)	<i>De novo</i>	+	NA
WDSTS/KMT2A	WDSTS_14	M	9	NM_001197104:c.9575A>C	p.(Gln3192Pro)	<i>De novo</i>	-	-
	WDSTS_15	M	11	NM_001197104:c.4391A>C	p.(Glu1464Ala)	Maternal	-	-
	WDSTS_16	M	6	NM_001197104:c.10835+1G>A	p.(?)	Paternal	-	-
	WDSTS_17	M	6	NM_001197104.1:c.3542G>A	p.(Gly1181Asp)	Not maternal	+	+
	WDSTS_18	M	2	NM_001197104.1:c.100C>G	p.(Arg34Gly)	<i>De novo</i>	-	-
	WDSTS_19	M	5.5	NM_001197104.1:c.31G>A	p.(Ala11Thr)	<i>De novo</i>	-	-
Clinical hypotheses: no molecular diagnostic but suggestive presentation								
RSTS/CREBBP	RSTS_12	M	24.4				-	NA
Sotos/NSD1	Sotos_11	F	6				-	-
	Sotos_13	M	4				+	+
	Sotos_14	M	8.3				-	-
	Sotos_15	F	4.5				-	-
WDSTS/KMT2A	WDSTS_11	M	15				-	-
	WDSTS_12	F	6				-	-
	WDSTS_13	M	9.5m				-	-

Figure 1 Sensitivity by episignature

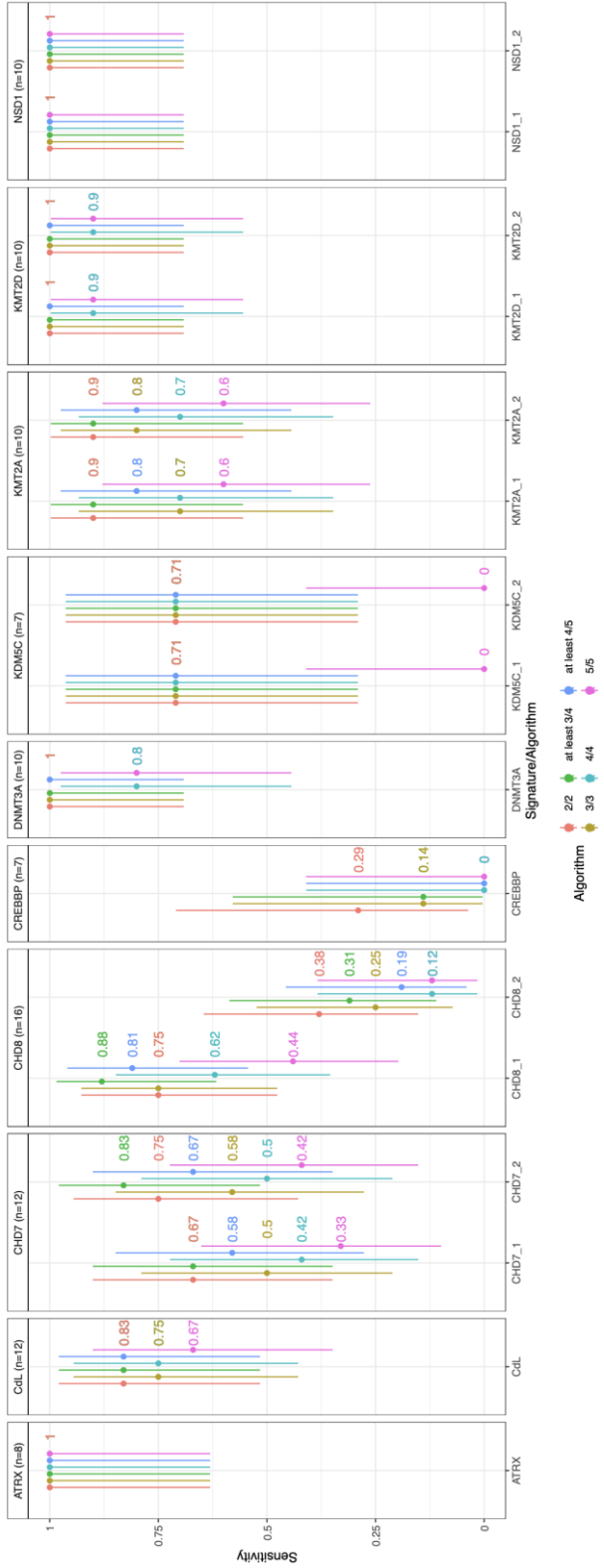
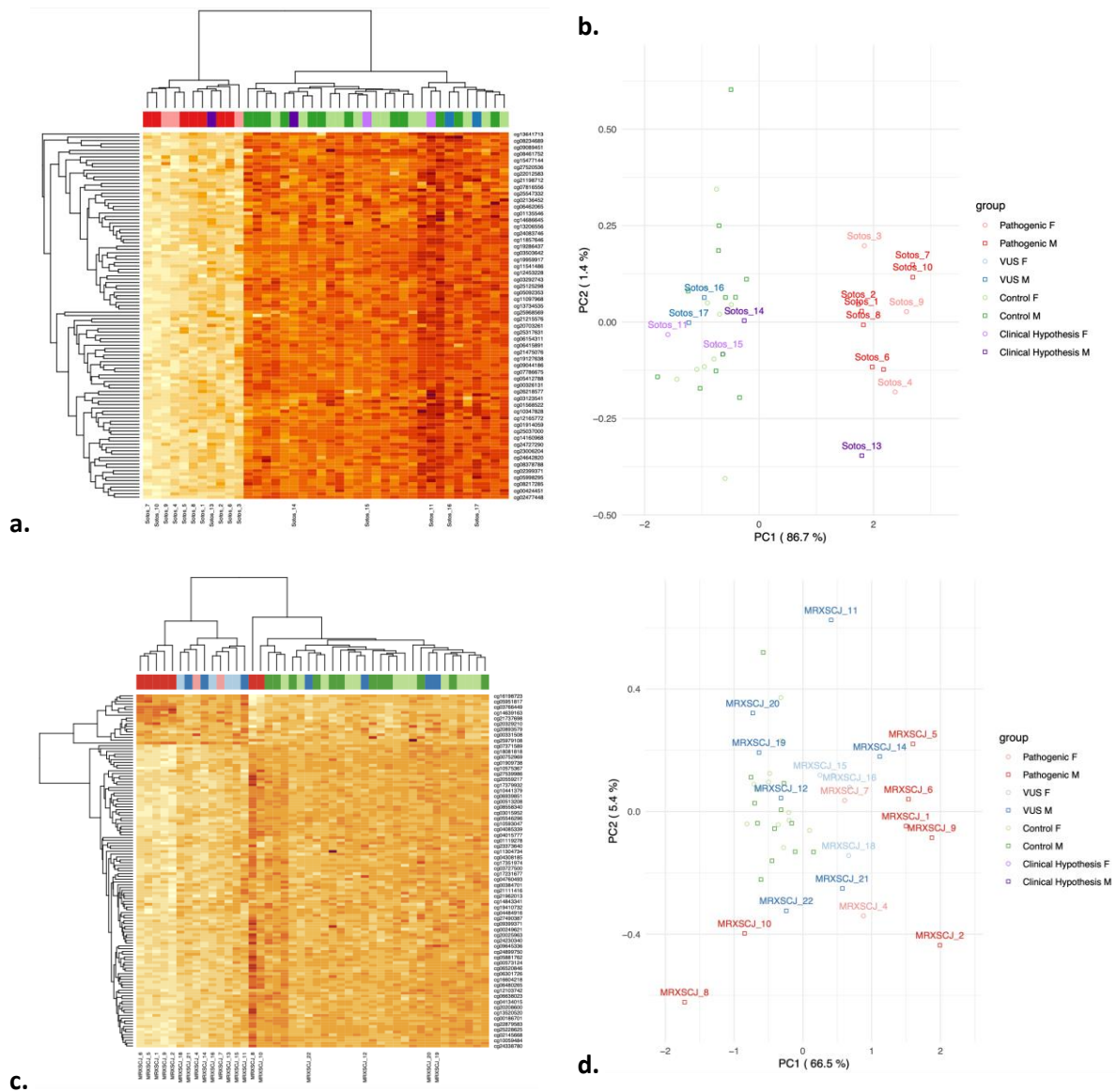
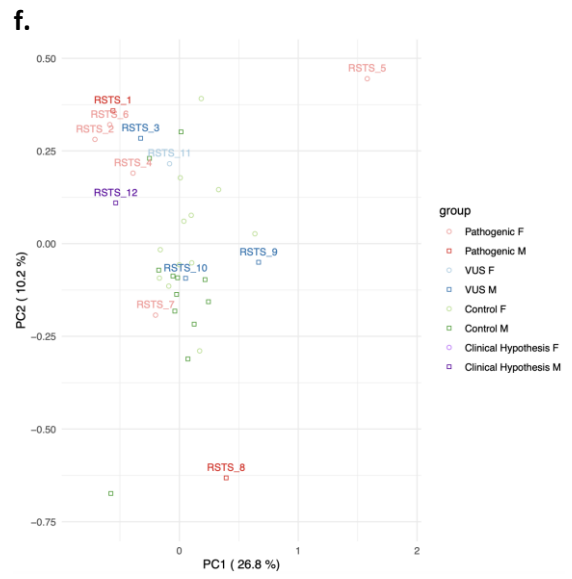
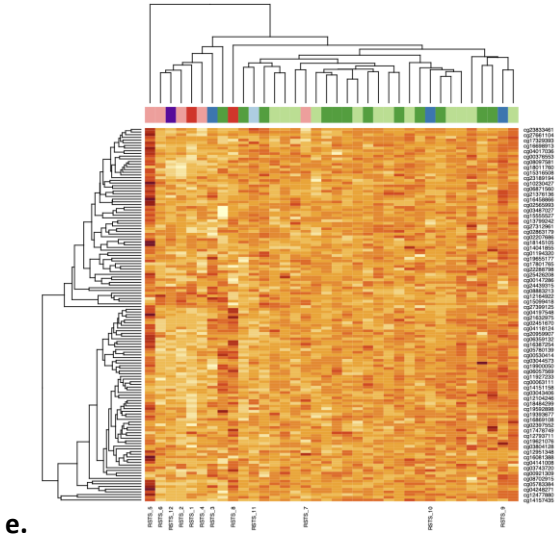


Figure 2. Graphical inspection of epesignature performances: heatmap and first two principal components for a-b. Sotos/*NSD1*, c-d.MRXSCJ/*KDM5C_2*, and e-f.RSTS/*CREBBP* signatures. Sotos/*NSD1*_1 is an example of a clear-cut distinction between cases and controls. For MRXSCJ/*KDM5C_2* a clear structure is visible, but several seems to have an intermediate profile. Whatever the machine learning prediction, as shown herein, reality is not straightforward. At last, RSTS/*CREBBP* doesn't show any noteworthy structure and thereby can't be used to help classify VUS.





DISCUSSION & CONCLUSION

Lors de ce travail nous avons évalué le taux de découverte de facteurs de risque génétiques dans une population de patients présentant un TSA. Le corollaire à ce travail était la constitution d'une liste de gène la plus exhaustive possible – à un instant donné – et la formulation de critères explicites pour définir ces facteurs de risque. Nous avons effectué une première estimation sur un recrutement rétrospectif que nous avons, dans une certaine mesure, confirmée sur une étude prospective. En effet, pour les variations nucléotidiques, le taux de découverte s'est révélé être d'environ 10-12% dans les deux études. Cependant, nous observons une diminution drastique de la détection de CNVs aux loci pour lesquels un surrisque de TSA est connu : plus de 7% dans l'étude rétrospective pour 3.3% dans cette deuxième étude prospective. Un biais de recrutement dans l'une des deux séries est l'explication la plus probables. Historiquement, la *CGH-array* a été accessible bien avant le WES dans les services de génétique. Des patients ayant bénéficié d'une *CGH-array* dans le passé sans détection de facteur de risque ont pu être inclus dans la seconde étude alors que ceux qui pour lesquels un CNV pathogène avait été détecté n'étaient pas incluables. L'évolution de la population de patient est une autre hypothèse, peut-être complémentaire, même si nos données ne confirment pas l'hypothèse d'une sévérité moindre des patients inclus pour la deuxième étude sur le plan du handicap neurocognitif. De plus, ces CNVs sont connus pour présenter une pénétrance modérée et sont fréquemment hérités de parents peu ou pas symptomatiques (Rosenfeld et al., 2013). De plus, cela ne se traduit pas par une diminution du taux de détection de variations nucléotidiques.

Nos travaux se sont révélés les plus proches possibles des conditions réelles de consultation dans un service de génétique. En effet, nous avons travaillé à partir de la population de patients qui consulte effectivement et en étroite collaboration avec les cliniciens. Nos constatations et recommandations sont donc applicables aux conditions réelles. La présence de symptômes neurologiques tel que l'épilepsie ou d'une anomalie du périmètre crânien augmente la probabilité de détection de facteurs de risque génétiques de TSA. Ces données s'expliquent aisément, plusieurs syndromes neurogénétiques présentent dans leur spectre phénotypique à la fois un risque d'épilepsie et TSA comme pour le gène *SCN2A* (Wolff et al., 2017) ou de macrocéphalie et de TSA comme pour le gène *CHD8* (Ostrowski et al., 2019, p. 8). En revanche, l'observation d'une association statistique entre retard à la marche et taux de détection de variations rares pathogènes est plus intéressante. Elle corrobore les données montrant un développement moteur altéré chez les patients diagnostiqués TSA avec détection d'une variation rare pathogène (Antaki et al., 2022; Buja et al., 2018). Un retard à la marche (ou plus généralement aux étapes du développement moteur) pourrait être un meilleur marqueur précoce d'altération du développement cognitif chez ces patients que le retard au langage, car dernier est consubstantiel au diagnostic de TSA. Notre objectif serait d'aboutir à une gradation du taux de découverte de ces facteurs de risque génétiques en fonction de certaines caractéristiques cliniques afin de prioriser les patients qui bénéficient le plus du séquençage d'exome.

Les limites de ces travaux sont inhérentes aux intérêts que nous avons cités. Il s'agit d'une photographie de la situation actuelle, qui sera par définition évolutive d'une part avec la modification des pratiques cliniques – nous avons vu que les critères diagnostiques sont particulièrement changeants en l'absence de marqueurs paracliniques. La population de

patients peut évoluer et n'est peut-être pas représentative de celle rencontrée en d'autres temps ou d'autres lieux.

Mais surtout, tant le taux de détection que son association avec les variables cliniques évolueront en fonction des données de la littérature. La liste de gènes s'est étoffée dans les quelques années qui ont séparé les deux études. Et au moment d'écrire ces lignes elle n'est déjà plus à jour. Les résultats d'une étude de grande ampleur ont été publiés en septembre 2022 (Fu et al., 2022). Les auteurs ont associé plusieurs cohortes de patients diagnostiqués avec un TSA ($n = 20\ 627$). Ils ont utilisé, comme d'autres équipes auparavant un modèle d'étude TADA, ce qui leur permet d'intégrer dans la statistique de test finale des données hétérogènes : détection de variations *de novo* issues de trios et études cas-témoins. Ils décrivent finalement 72 gènes pour lesquels les variations génétiques délétères sont facteurs de risque du trouble (à $FDR \leq 0,001$ ce que les auteurs considèrent comme une correction pour tests multiples à l'échelle de l'exome). Parmi ceux-ci, trois gènes n'étaient pas considérés dans nos analyses. Ils ont également mis en commun cette cohorte avec des données issues de patients présentant un retard de développement ($n = 31\ 058$). Dans ce cas, 304 gènes étaient associés au phénotype – certes très élargi – (à $FDR \leq 0,001$).

Dans cette étude, 13% des gènes associés au TSA ne le sont pas avec le retard de développements. De plus, parmi les gènes associés aux deux phénotypes, 36 le sont plus particulièrement avec les TSA. Cette étude confirme les données précédentes (Satterstrom et al., 2020) : l'hypothèse que les variations génétiques rares n'induisent qu'un risque aspécifique de TSA par l'intermédiaire de leur impact cognitif ne semble pas corroborée pour l'ensemble des gènes.

Un autre syndrome neurodéveloppemental à forte composante génétique est dans une situation similaire. Dans le cas de la schizophrénie, un certain nombre de gènes ont aussi été associés à une augmentation du risque de la maladie lorsqu'ils sont touchés par une variation rare pathogène (Singh et al., 2022). Seule une fraction de ces gènes sont connus pour être associés aux retards de développement.

L'étude de Fu et collaborateurs que nous avons déjà cité conclu de manière suivante : « *Together, these data suggest that one subset of ASD risk genes may overlap DD while a different subset overlaps schizophrenia* ». Cela ne semble pas corroborer l'hypothèse d'une causalité linéaire gènes - impact cognitif - schizophrénie/TSA

Plus généralement des altérations comportementales spécifiques sont bien décrites pour certains syndromes neurodéveloppementaux comme les syndromes de Williams-Beuren (délétion du locus 7q11.23) ou de Smith-Magenis (perte de fonction du gène *RAI1*). A ce stade, il nous semble prématuré de conclure que l'ensemble des gènes impliqués dans le TSA puisse être simplement regroupé dans une catégorie « retard de développement » large et aspécifique.

Le décryptage des similitudes et différences entre ces voies physiopathologiques sera certainement un objectif de recherche majeur des prochaines années.

Une deuxième limitation de nos travaux est inhérente aux critères de pathogénicité que nous avons utilisé, et des paradigmes de recherche qui ont gouverné à la mise à jour des gènes que nous considérons dans le TSA. Dans un cas comme dans l'autre, nous avons insisté sur l'importance du caractère *de novo* des variations. Nous avons plus de difficultés à conclure

lorsque les variations génétiques sont héritées de parents asymptomatiques.

Dans l'introduction, nous avons opposé deux situations extrêmes : les polymorphismes génétiques portant un risque faible et les variations rares amenant un risque fort de développer le phénotype. Comme souvent en biologie, lorsque l'on définit des catégories, c'est que l'on ne considère que les extrêmes d'une dimension sous-jacente. Ici aussi, tout un spectre de variations génétiques existe entre ces extrêmes. Les travaux récemment publiés de Zhou et collaborateurs intègrent comme données la transmission préférentielle de certaines variations génétiques des parents vers leurs enfants atteint de TSA (X. Zhou et al., 2022). Les variations pathogènes de ces gènes impliqueraient une susceptibilité au TSA et une détérioration cognitive moindre que les gènes mis à jour *via* l'excès de variation *de novo*. En l'absence de biomarqueur, il sera très difficile de conclure à la pathogénicité de variations génétique de ces gènes.

L'étude d'un biomarqueur sanguin de la pathogénicité de variations génétiques dans des gènes spécifiques a été l'objet de la deuxième partie de ce travail. La mise à jour et l'utilisation d'épisignatures est un champ d'étude très prometteur pour aider à la classification de VSI. Ces travaux ne sont plus directement concentrés sur la génétique du TSA mais sur l'impact fonctionnel de variations génétiques sur la méthylation de l'ADN. Il s'agit d'un point de vue centré sur le gène et non plus sur le phénotype. Nous avons néanmoins considéré des gènes pour lesquels les variations pathogènes amènent – parmi le spectre phénotypique – une susceptibilité au TSA.

Nous avons détecté une épisignature d'ampleur modérée pour l'un d'entre eux, *MBD5*, sans que sa capacité à discriminer cas et témoins soit suffisante pour une utilisation en clinique.

Nos travaux ont en grande partie porté sur des considérations méthodologiques permettant l'utilisation de ces épisignatures dans un cadre clinique. Nous avons rappelé qu'une nouvelle stratégie diagnostique, aussi élégante et prometteuse soit-elle, ne peut faire l'économie d'une évaluation indépendante de ses performances. Certaines épisignatures semblent matures pour l'utilisation en pratique clinique mais ce n'est pas le cas de toutes. Cette évaluation était d'autant plus nécessaire qu'une solution commerciale (EpiSign développé par l'équipe de London à partir de données collectées auprès du monde académique) existe pour l'utilisation de ces épisignatures. La méthodologie employée y est maintenue à dessein opaque ce qui n'est pas compatible avec une démarche scientifique. Ceci est d'autant plus préjudiciable qu'une évaluation rigoureuse n'a pas été réalisée pour toutes les signatures proposées. De fait, nos travaux montrent que plusieurs d'entre elles ne devraient pas être utilisées en pratique clinique.

Concernant ce domaine d'étude – l'analyse de méthylation sur puce – plusieurs perspectives nous paraissent d'intérêt scientifique. Premièrement, l'évaluation de la performance des différentes épisignatures publiées doit être poursuivie. Deuxièmement, une comparaison des différentes méthodologies utilisées devrait être effectuée pour en déduire un *gold-standard* méthodologique pour la détection d'une épisignature et son utilisation en aide au diagnostic. Troisièmement, l'étude de méthylation sur puce permet l'analyse d'anomalies sévères et localisées de méthylation de l'ADN. Ces hyper ou hypométhylations locales appelées épivariations – par opposition aux épisignatures – sont des événements relativement rares

(Barbosa et al., 2018; Garg et al., 2020) potentiellement impliqués en pathologie humaine. La description de l'hyperméthylation sévère du promoteur d'un gène impliqué en troubles du neurodéveloppement chez un patient au phénotype compatible serait un argument fort pour leur responsabilité dans certains cas de troubles neurodéveloppementaux – surtout si un impact transcriptionnel est démontré en parallèle.

Ces perspectives de travail sur la méthylation de l'ADN sont particulièrement enthousiasmantes. En effet, la machinerie épigénétique est une interface entre le patrimoine génétique de chacun et les contraintes environnementales.

En conclusion, il nous paraît nécessaire de souligner l'importance d'une approche associant un substrat biologique (génétique, épigénétique voir transcriptomique) à une description de phénotypes les plus précis possibles (comportementaux, neurophysiologiques, neuropsychologiques...). Encore une fois, une pensée scientifique « complexe » sera nécessaire pour fonctionner en accord avec les deux prémisses suivantes : d'un côté les grandes catégories syndromiques comme le TSA devraient être considérées comme des regroupements athéoriques de symptômes aux causes multiples, aboutissant sur une voie physiopathologique commune, sans procéder *a priori* d'une cause biologique commune. De l'autre, il convient d'accepter que le fantasme d'une association précise et essentialiste entre l'altération d'un gène donné et un phénotype comportementale précis ne sera pas atteint (dans la plupart des cas). Il existe trop de variables intermédiaires et seules des tendances se dégageront. C'est l'ensemble de ces tendances qui forment l'idiosyncrasie de l'individu.

« *La chose dont nous parlons ne peut jamais être trouvée en cherchant, pourtant seuls les chercheurs la trouvent.* » - Bayazid Bastami, philosophe Soufi du 9^{ème} siècle.

REFERENCES BIBLIOGRAPHIQUES

Abbott, A. E., Nair, A., Keown, C. L., Datko, M., Jahedi, A., Fishman, I., & Müller, R.-A. (2016). Patterns of Atypical Functional Connectivity and Behavioral Links in Autism Differ Between Default, Salience, and Executive Networks. *Cerebral Cortex (New York, N.Y.: 1991)*, 26(10), 4034-4045. <https://doi.org/10.1093/cercor/bhv191>

Abrahams, B. S., Arking, D. E., Campbell, D. B., Mefford, H. C., Morrow, E. M., Weiss, L. A., Menashe, I., Wadkins, T., Banerjee-Basu, S., & Packer, A. (2013). SFARI Gene 2.0 : A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Molecular Autism*, 4(1), 36. <https://doi.org/10.1186/2040-2392-4-36>

Al-Farsi, F. A. H., Al-Alyani, O. B. S., Al-Kumzari, A., & Al-Saadi, T. (2022). Systemic Review and Meta-analysis of the Intellectual Integrity of Children with Neurofibromatosis Type 1. *World Neurosurgery*, 157, 69-74. <https://doi.org/10.1016/j.wneu.2021.10.081>

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, 12(5), Art. 5. <https://doi.org/10.1038/nrg2958>

An, J.-Y., Lin, K., Zhu, L., Werling, D. M., Dong, S., Brand, H., Wang, H. Z., Zhao, X., Schwartz, G. B., Collins, R. L., Currall, B. B., Dastmalchi, C., Dea, J., Duhn, C., Gilson, M. C., Klei, L., Liang, L., Markenscoff-Papadimitriou, E., Pochareddy, S., ... Sanders, S. J. (2018). Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science (New York, N.Y.)*, 362(6420), eaat6576. <https://doi.org/10.1126/science.aat6576>

Anderson, J. S., Druzgal, T. J., Froehlich, A., DuBray, M. B., Lange, N., Alexander, A. L., Abildskov, T., Nielsen, J. A., Cariello, A. N., Cooperrider, J. R., Bigler, E. D., & Lainhart, J. E. (2011). Decreased interhemispheric functional connectivity in autism. *Cerebral Cortex (New York, N.Y.: 1991)*, 21(5), 1134-1146. <https://doi.org/10.1093/cercor/bhq190>

Andrews, D. S., Lee, J. K., Harvey, D. J., Waizbard-Bartov, E., Solomon, M., Rogers, S. J., Nordahl, C. W., & Amaral, D. G. (2021). A Longitudinal Study of White Matter Development in Relation to Changes in Autism Severity Across Early Childhood. *Biological Psychiatry*, 89(5), 424-432. <https://doi.org/10.1016/j.biopsych.2020.10.013>

Andrews, S. V., Sheppard, B., Windham, G. C., Schieve, L. A., Schendel, D. E., Croen, L. A., Chopra, P., Alisch, R. S., Newschaffer, C. J., Warren, S. T., Feinberg, A. P., Fallin, M. D., & Ladd-Acosta, C. (2018). Case-control meta-analysis of blood DNA methylation and autism spectrum disorder. *Molecular Autism*, 9, 40. <https://doi.org/10.1186/s13229-018-0224-6>

Antaki, D., Guevara, J., Maihofer, A. X., Klein, M., Gujral, M., Grove, J., Carey, C. E., Hong, O., Arranz, M. J., Hervas, A., Corsello, C., Vaux, K. K., Muotri, A. R., Iakoucheva, L. M., Courchesne, E., Pierce, K., Gleeson, J. G., Robinson, E. B., Nievergelt, C. M., & Sebat, J. (2022). A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. *Nature Genetics*, 1-9. <https://doi.org/10.1038/s41588-022-01064-5>

Aref-Eshghi, E., Bend, E. G., Hood, R. L., Schenkel, L. C., Carere, D. A., Chakrabarti, R., Nagamani, S. C. S., Cheung, S. W., Campeau, P. M., Prasad, C., Siu, V. M., Brady, L., Tarnopolsky, M. A., Callen, D. J., Innes, A. M., White, S. M., Meschino, W. S., Shuen, A. Y., Paré, G., ... Sadikovic, B. (2018). BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaidis-Baraitser syndromes. *Nature Communications*, 9(1), 4885. <https://doi.org/10.1038/s41467-018-07193-y>

Aref-Eshghi, E., Kerkhof, J., Pedro, V. P., Groupe DI France, Barat-Houari, M., Ruiz-Pallares, N., Andrau, J.-C., Lacombe, D., Van-Gils, J., Fergelot, P., Dubourg, C., Cormier-Daire, V., Rondeau, S., Lecoquierre, F., Saugier-Weber, P., Nicolas, G., Lesca, G., Chatron, N., Sanlaville, D., ... Sadikovic, B. (2020). Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *American Journal of Human Genetics*, 106(3), 356-370. <https://doi.org/10.1016/j.ajhg.2020.01.019>

Aref-Eshghi, E., Schenkel, L. C., Lin, H., Skinner, C., Ainsworth, P., Paré, G., Rodenhiser, D., Schwartz, C., & Sadikovic, B. (2017). The defining DNA methylation signature of Kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance. *Epigenetics*, *12*(11), 923-933. <https://doi.org/10.1080/15592294.2017.1381807>

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* (Oxford, England), *30*(10), 1363-1369. <https://doi.org/10.1093/bioinformatics/btu049>

Atladóttir, H. Ó., Henriksen, T. B., Schendel, D. E., & Parner, E. T. (2012). Autism after infection, febrile episodes, and antibiotic use during pregnancy: An exploratory study. *Pediatrics*, *130*(6), e1447-1454. <https://doi.org/10.1542/peds.2012-1107>

Backenroth, D., Homsy, J., Murillo, L. R., Glessner, J., Lin, E., Brueckner, M., Lifton, R., Goldmuntz, E., Chung, W. K., & Shen, Y. (2014). CANOES: Detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Research*, *42*(12), e97. <https://doi.org/10.1093/nar/gku345>

Barbaro, J., & Dissanayake, C. (2017). Diagnostic stability of autism spectrum disorder in toddlers prospectively identified in a community-based setting: Behavioural characteristics and predictors of change over time. *Autism: The International Journal of Research and Practice*, *21*(7), 830-840. <https://doi.org/10.1177/1362361316654084>

Barbosa, M., Joshi, R. S., Garg, P., Martin-Trujillo, A., Patel, N., Jadhav, B., Watson, C. T., Gibson, W., Chetnik, K., Tessereau, C., Mei, H., De Rubeis, S., Reichert, J., Lopes, F., Vissers, L. E. L. M., Kleefstra, T., Grice, D. E., Edelmann, L., Soares, G., ... Sharp, A. J. (2018). Identification of rare de novo epigenetic variations in congenital disorders. *Nature Communications*, *9*(1), 2064. <https://doi.org/10.1038/s41467-018-04540-x>

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*(6), 248-254. [https://doi.org/10.1016/S1364-6613\(02\)01904-6](https://doi.org/10.1016/S1364-6613(02)01904-6)

Baron-Cohen, S., Auyeung, B., Nørgaard-Pedersen, B., Hougaard, D. M., Abdallah, M. W., Melgaard, L., Cohen, A. S., Chakrabarti, B., Ruta, L., & Lombardo, M. V. (2015). Elevated fetal steroidogenic activity in autism. *Molecular Psychiatry*, *20*(3), 369-376. <https://doi.org/10.1038/mp.2014.48>

Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). Sex differences in the brain: Implications for explaining autism. *Science (New York, N.Y.)*, *310*(5749), 819-823. <https://doi.org/10.1126/science.1115455>

Bauman, J. G., Wiegant, J., Borst, P., & van Duijn, P. (1980). A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. *Experimental Cell Research*, *128*(2), 485-490. [https://doi.org/10.1016/0014-4827\(80\)90087-7](https://doi.org/10.1016/0014-4827(80)90087-7)

Baymaz, H. I., Fournier, A., Laget, S., Ji, Z., Jansen, P. W. T. C., Smits, A. H., Ferry, L., Mensinga, A., Poser, I., Sharrocks, A., Defossez, P.-A., & Vermeulen, M. (2014). MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics*, *14*(19), 2179-2189. <https://doi.org/10.1002/pmic.201400013>

Beck, T. F., Mullikin, J. C., NISC Comparative Sequencing Program, & Biesecker, L. G. (2016). Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clinical Chemistry*, *62*(4), 647-654. <https://doi.org/10.1373/clinchem.2015.249623>

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., Gilad, Y., & Pritchard, J. K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology*, *12*(1), R10. <https://doi.org/10.1186/gb-2011-12-1-r10>

Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and abnormal development of brain connectivity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *24*(42), 9228-9231. <https://doi.org/10.1523/JNEUROSCI.3340-04.2004>

- Belyeu, J. R., Brand, H., Wang, H., Zhao, X., Pedersen, B. S., Feusier, J., Gupta, M., Nicholas, T. J., Brown, J., Baird, L., Devlin, B., Sanders, S. J., Jorde, L. B., Talkowski, M. E., & Quinlan, A. R. (2021). De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. *American Journal of Human Genetics*, *108*(4), 597. <https://doi.org/10.1016/j.ajhg.2021.02.012>
- Benedetto, L., Cucinotta, F., Maggio, R., Germanò, E., De Raco, R., Alquino, A., Impallomeni, C., Siracusano, R., Vetri, L., Roccella, M., Ingrassia, M., & Gagliano, A. (2021). One-Year Follow-Up Diagnostic Stability of Autism Spectrum Disorder Diagnosis in a Clinical Sample of Children and Toddlers. *Brain Sciences*, *11*(1), 37. <https://doi.org/10.3390/brainsci11010037>
- Berdasco, M., & Esteller, M. (2013). Genetic syndromes caused by mutations in epigenetic genes. *Human Genetics*, *132*(4), 359-383. <https://doi.org/10.1007/s00439-013-1271-x>
- Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. *Brain Research*, *1380*, 42-77. <https://doi.org/10.1016/j.brainres.2010.11.078>
- Bey, A. L., & Jiang, Y. (2014). Overview of mouse models of autism spectrum disorders. *Current Protocols in Pharmacology*, *66*, 5.66.1-26. <https://doi.org/10.1002/0471141755.ph0566s66>
- Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., Delano, D., Zhang, L., Schroth, G. P., Gunderson, K. L., Fan, J.-B., & Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*(4), 288-295. <https://doi.org/10.1016/j.ygeno.2011.07.007>
- Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., & Gunderson, K. L. (2009). Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics*, *1*(1), 177-200. <https://doi.org/10.2217/epi.09.14>
- Bieleninik, Ł., Posserud, M.-B., Geretsegger, M., Thompson, G., Elefant, C., & Gold, C. (2017). Tracing the temporal stability of autism spectrum diagnosis and severity as measured by the Autism Diagnostic Observation Schedule: A systematic review and meta-analysis. *PLoS ONE*, *12*(9), e0183160. <https://doi.org/10.1371/journal.pone.0183160>
- Bishop, S. L., Farmer, C., Bal, V., Robinson, E. B., Willsey, A. J., Werling, D. M., Havdahl, K. A., Sanders, S. J., & Thurm, A. (2017). Identification of Developmental and Behavioral Markers Associated With Genetic Abnormalities in Autism Spectrum Disorder. *The American Journal of Psychiatry*, *174*(6), 576-585. <https://doi.org/10.1176/appi.ajp.2017.16101115>
- Bollepalli, S., Korhonen, T., Kaprio, J., Anders, S., & Ollikainen, M. (2019). EpiSmokEr: A robust classifier to determine smoking status from DNA methylation data. *Epigenomics*, *11*(13), 1469-1486. <https://doi.org/10.2217/epi-2019-0206>
- Bolstad, B. M., Irizarry, R. A., Astrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics (Oxford, England)*, *19*(2), 185-193. <https://doi.org/10.1093/bioinformatics/19.2.185>
- Bölte, S., Girdler, S., & Marschik, P. B. (2019). The contribution of environmental exposure to the etiology of autism spectrum disorder. *Cellular and Molecular Life Sciences*, *76*(7), 1275-1297. <https://doi.org/10.1007/s00018-018-2988-4>
- Bonnet-Brilhault, F., Rajerison, T. A., Paillet, C., Guimard-Brunault, M., Saby, A., Ponson, L., Tripi, G., Malvy, J., & Roux, S. (2018). Autism is a prenatal disorder: Evidence from late gestation brain overgrowth. *Autism Research: Official Journal of the International Society for Autism Research*, *11*(12), 1635-1642. <https://doi.org/10.1002/aur.2036>
- Borges-Monroy, R., Chu, C., Dias, C., Choi, J., Lee, S., Gao, Y., Shin, T., Park, P. J., Walsh, C. A., & Lee, E. A. (2021). Whole-genome analysis reveals the contribution of non-coding de novo transposon insertions to autism spectrum disorder. *Mobile DNA*, *12*(1), 28. <https://doi.org/10.1186/s13100-021-00256-w>

Bose, M., Wu, C., Pankow, J. S., Demerath, E. W., Bressler, J., Fornage, M., Grove, M. L., Mosley, T. H., Hicks, C., North, K., Kao, W. H., Zhang, Y., Boerwinkle, E., & Guan, W. (2014). Evaluation of microarray-based DNA methylation measurement using technical replicates: The Atherosclerosis Risk In Communities (ARIC) Study. *BMC Bioinformatics*, *15*, 312. <https://doi.org/10.1186/1471-2105-15-312>

Bougeard, C., Picarel-Blanchot, F., Schmid, R., Campbell, R., & Buitelaar, J. (2021). Prevalence of Autism Spectrum Disorder and Co-morbidities in Children and Adolescents : A Systematic Literature Review. *Frontiers in Psychiatry*, *12*, 744709. <https://doi.org/10.3389/fpsy.2021.744709>

Bourgon, R., Gentleman, R., & Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(21), 9546-9551. <https://doi.org/10.1073/pnas.0914005107>

Buja, A., Volfovsky, N., Krieger, A. M., Lord, C., Lash, A. E., Wigler, M., & Iossifov, I. (2018). Damaging de novo mutations diminish motor skills in children on the autism spectrum. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(8), E1859-E1866. <https://doi.org/10.1073/pnas.1715427115>

Butcher, D. T., Cytrynbaum, C., Turinsky, A. L., Siu, M. T., Inbar-Feigenberg, M., Mendoza-Londono, R., Chitayat, D., Walker, S., Machado, J., Caluseriu, O., Dupuis, L., Grafodatskaya, D., Reardon, W., Gilbert-Dussardier, B., Verloes, A., Bilan, F., Milunsky, J. M., Basran, R., Papsin, B., ... Weksberg, R. (2017). CHARGE and Kabuki Syndromes : Gene-Specific DNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically Overlapping Conditions. *The American Journal of Human Genetics*, *100*(5), 773-788. <https://doi.org/10.1016/j.ajhg.2017.04.004>

C Yuen, R. K., Merico, D., Bookman, M., L Howe, J., Thiruvahindrapuram, B., Patel, R. V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., Pellecchia, G., Buchanan, J. A., Walker, S., Marshall, C. R., Uddin, M., Zarrei, M., Deneault, E., D'Abate, L., Chan, A. J. S., ... Scherer, S. W. (2017a). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature Neuroscience*, *20*(4), 602-611. <https://doi.org/10.1038/nn.4524>

C Yuen, R. K., Merico, D., Bookman, M., L Howe, J., Thiruvahindrapuram, B., Patel, R. V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., Pellecchia, G., Buchanan, J. A., Walker, S., Marshall, C. R., Uddin, M., Zarrei, M., Deneault, E., D'Abate, L., Chan, A. J. S., ... Scherer, S. W. (2017b). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature Neuroscience*, *20*(4), 602-611. <https://doi.org/10.1038/nn.4524>

Campagna, M. P., Xavier, A., Lechner-Scott, J., Maltby, V., Scott, R. J., Butzkueven, H., Jokubaitis, V. G., & Lea, R. A. (2021). Epigenome-wide association studies : Current knowledge, strategies and recommendations. *Clinical Epigenetics*, *13*, 214. <https://doi.org/10.1186/s13148-021-01200-8>

Careaga, M., Murai, T., & Bauman, M. D. (2017). Maternal Immune Activation and Autism Spectrum Disorder: From Rodents to Nonhuman and Human Primates. *Biological Psychiatry*, *81*(5), 391-401. <https://doi.org/10.1016/j.biopsych.2016.10.020>

Cavalli, G., & Heard, E. (2019). Advances in epigenetics link genetics to the environment and disease. *Nature*, *571*(7766), 489-499. <https://doi.org/10.1038/s41586-019-1411-0>

Cerruti Mainardi, P. (2006). Cri du Chat syndrome. *Orphanet Journal of Rare Diseases*, *1*, 33. <https://doi.org/10.1186/1750-1172-1-33>

Cesta, C. E., Öberg, A. S., Ibrahimson, A., Yusuf, I., Larsson, H., Almqvist, C., D'Onofrio, B. M., Bulik, C. M., Fernández de la Cruz, L., Mataix-Cols, D., Landén, M., & Rosenqvist, M. A. (2020). Maternal polycystic ovary syndrome and risk of neuropsychiatric disorders in offspring : Prenatal androgen exposure or genetic confounding? *Psychological Medicine*, *50*(4), 616-624. <https://doi.org/10.1017/S0033291719000424>

Chance, S. A. (2014). The cortical microstructural basis of lateralized cognition : A review. *Frontiers in Psychology*, *5*, 820. <https://doi.org/10.3389/fpsyg.2014.00820>

Chaney, J. L., & Clark, P. L. (2015). Roles for Synonymous Codon Usage in Protein Biogenesis. *Annual Review of Biophysics*, *44*, 143-166. <https://doi.org/10.1146/annurev-biophys-060414-034333>

- Chater-Diehl, E., Goodman, S. J., Cytrynbaum, C., Turinsky, A. L., Choufani, S., & Weksberg, R. (2021). Anatomy of DNA methylation signatures: Emerging insights and applications. *American Journal of Human Genetics*, 108(8), 1359-1366. <https://doi.org/10.1016/j.ajhg.2021.06.015>
- Chawarska, K., Shic, F., Macari, S., Campbell, D. J., Brian, J., Landa, R., Hutman, T., Nelson, C. A., Ozonoff, S., Tager-Flusberg, H., Young, G. S., Zwaigenbaum, L., Cohen, I. L., Charman, T., Messinger, D. S., Klin, A., Johnson, S., & Bryson, S. (2014). 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: A baby siblings research consortium study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 53(12), 1317-1327.e1. <https://doi.org/10.1016/j.jaac.2014.09.015>
- Chen, B. H., Marioni, R. E., Colicino, E., Peters, M. J., Ward-Caviness, C. K., Tsai, P.-C., Roetker, N. S., Just, A. C., Demerath, E. W., Guan, W., Bressler, J., Fornage, M., Studenski, S., Vandiver, A. R., Moore, A. Z., Tanaka, T., Kiel, D. P., Liang, L., Vokonas, P., ... Horvath, S. (2016). DNA methylation-based measures of biological age: Meta-analysis predicting time to death. *Ageing*, 8(9), 1844-1865. <https://doi.org/10.18632/aging.101020>
- Chen, Uljarevic, M., Neal, J., Greening, S., Yim, H., & Lee, T.-H. (2021). Excessive Functional Coupling With Less Variability Between Salience and Default Mode Networks in Autism Spectrum Disorder. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, S2451-9022(21)00344-X. <https://doi.org/10.1016/j.bpsc.2021.11.016>
- Chen, Y., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., Gallinger, S., Hudson, T. J., & Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*, 8(2), 203-209. <https://doi.org/10.4161/epi.23470>
- Cherik, F., Reilly, J., Kerkhof, J., Levy, M., McConkey, H., Barat-Houari, M., Butler, K. M., Coubes, C., Lee, J. A., Le Guyader, G., Louie, R. J., Patterson, W. G., Tedder, M. L., Bak, M., Hammer, T. B., Craigen, W., Démurger, F., Dubourg, C., Fradin, M., ... Genevieve, D. (2022). DNA methylation epigenature in Gabriele-de Vries syndrome. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 24(4), 905-914. <https://doi.org/10.1016/j.gim.2021.12.003>
- Cheslack Postava, K., & Winter, A. S. (2015). Short and long interpregnancy intervals: Correlates and variations by pregnancy timing among U.S. women. *Perspectives on Sexual and Reproductive Health*, 47(1), 19-26. <https://doi.org/10.1363/47e2615>
- Chilton, I., Okur, V., Vitiello, G., Selicorni, A., Mariani, M., Goldenberg, A., Husson, T., Campion, D., Lichtenbelt, K. D., van Gassen, K., Steinraths, M., Rice, J., Roeder, E. R., Littlejohn, R. O., Srouf, M., Sebire, G., Accogli, A., Héron, D., Heide, S., ... Chung, W. K. (2020). De novo heterozygous missense and loss-of-function variants in CDC42BPB are associated with a neurodevelopmental phenotype. *American Journal of Medical Genetics. Part A*, 182(5), 962-973. <https://doi.org/10.1002/ajmg.a.61505>
- Chisholm, A. K., Haebich, K. M., Pride, N. A., Walsh, K. S., Lami, F., Ure, A., Maloof, T., Brignell, A., Rouel, M., Granader, Y., Maier, A., Barton, B., Darke, H., Dabscheck, G., Anderson, V. A., Williams, K., North, K. N., & Payne, J. M. (2022). Delineating the autistic phenotype in children with neurofibromatosis type 1. *Molecular Autism*, 13(1), 3. <https://doi.org/10.1186/s13229-021-00481-3>
- Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols*, 15(9), 2759-2772. <https://doi.org/10.1038/s41596-020-0353-1>
- Choufani, S., Cytrynbaum, C., Chung, B. H. Y., Turinsky, A. L., Grafodatskaya, D., Chen, Y. A., Cohen, A. S. A., Dupuis, L., Butcher, D. T., Siu, M. T., Luk, H. M., Lo, I. F. M., Lam, S. T. S., Caluseriu, O., Stavropoulos, D. J., Reardon, W., Mendoza-Londono, R., Brudno, M., Gibson, W. T., ... Weksberg, R. (2015). NSD1 mutations generate a genome-wide DNA methylation signature. *Nature Communications*, 6, 10207. <https://doi.org/10.1038/ncomms10207>
- Chun, H., Leung, C., Wen, S. W., McDonald, J., & Shin, H. H. (2020). Maternal exposure to air pollution and risk of autism in children: A systematic review and meta-analysis. *Environmental Pollution (Barking, Essex: 1987)*, 256, 113307. <https://doi.org/10.1016/j.envpol.2019.113307>
- Clarke, T.-K., Lupton, M. K., Fernandez-Pujals, A. M., Starr, J., Davies, G., Cox, S., Pattie, A., Liewald, D. C., Hall, L. S., MacIntyre, D. J., Smith, B. H., Hocking, L. J., Padmanabhan, S., Thomson, P. A., Hayward, C.,

- Hansell, N. K., Montgomery, G. W., Medland, S. E., Martin, N. G., ... McIntosh, A. M. (2016). Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Molecular Psychiatry*, *21*(3), 419-425. <https://doi.org/10.1038/mp.2015.12>
- Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C., Khera, A. V., Lowther, C., Gauthier, L. D., Wang, H., Watts, N. A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C. W., Huang, Y., Brookings, T., ... Talkowski, M. E. (2020). A structural variation reference for medical and population genetics. *Nature*, *581*(7809), 444-451. <https://doi.org/10.1038/s41586-020-2287-8>
- Colvert, E., Tick, B., McEwen, F., Stewart, C., Curran, S. R., Woodhouse, E., Gillan, N., Hallett, V., Lietz, S., Garnett, T., Ronald, A., Plomin, R., Rijsdijk, F., Happé, F., & Bolton, P. (2015). Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry*, *72*(5), 415-423. <https://doi.org/10.1001/jamapsychiatry.2014.3028>
- Constantino, J. N., & Todd, R. D. (2003). Autistic traits in the general population : A twin study. *Archives of General Psychiatry*, *60*(5), 524-530. <https://doi.org/10.1001/archpsyc.60.5.524>
- Courchesne, E., Campbell, K., & Solso, S. (2011). Brain growth across the life span in autism : Age-specific changes in anatomical pathology. *Brain Research*, *1380*, 138-145. <https://doi.org/10.1016/j.brainres.2010.09.101>
- Courchesne, E., Gazestani, V. H., & Lewis, N. E. (2020). PRENATAL ORIGINS OF ASD : THE WHEN, WHAT AND HOW OF ASD DEVELOPMENT. *Trends in neurosciences*, *43*(5), 326-342. <https://doi.org/10.1016/j.tins.2020.03.005>
- Courchesne, E., Karns, C. M., Davis, H. R., Ziccardi, R., Carper, R. A., Tigue, Z. D., Chisum, H. J., Moses, P., Pierce, K., Lord, C., Lincoln, A. J., Pizzo, S., Schreibman, L., Haas, R. H., Akshoomoff, N. A., & Courchesne, R. Y. (2001). Unusual brain growth patterns in early life in patients with autistic disorder: An MRI study. *Neurology*, *57*(2), 245-254. <https://doi.org/10.1212/wnl.57.2.245>
- Courchesne, E., Mouton, P. R., Calhoun, M. E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M. J., Barnes, C. C., & Pierce, K. (2011). Neuron number and size in prefrontal cortex of children with autism. *JAMA*, *306*(18), 2001-2010. <https://doi.org/10.1001/jama.2011.1638>
- Courchesne, E., Pramparo, T., Gazestani, V. H., Lombardo, M. V., Pierce, K., & Lewis, N. E. (2019). The ASD Living Biology: From cell proliferation to clinical phenotype. *Molecular Psychiatry*, *24*(1), 88-107. <https://doi.org/10.1038/s41380-018-0056-y>
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21-27. <https://doi.org/10.1109/TIT.1967.1053964>
- Creeth, H. D. J., Rees, E., Legge, S. E., Dennison, C. A., Holmans, P., Walters, J. T. R., O'Donovan, M. C., & Owen, M. J. (2022). Ultrarare Coding Variants and Cognitive Function in Schizophrenia. *JAMA Psychiatry*, *79*(10), 963-970. <https://doi.org/10.1001/jamapsychiatry.2022.2289>
- Daca-Roszak, P., Pfeifer, A., Żebracka-Gala, J., Rusinek, D., Szybińska, A., Jarząb, B., Witt, M., & Ziętkiewicz, E. (2015). Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip Array: Implications for comparative population studies. *BMC Genomics*, *16*, 1003. <https://doi.org/10.1186/s12864-015-2202-0>
- De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., Kou, Y., Liu, L., Fromer, M., Walker, S., Singh, T., Klei, L., Kosmicki, J., Shih-Chen, F., Aleksic, B., Biscaldi, M., Bolton, P. F., Brownfeld, J. M., Cai, J., ... Buxbaum, J. D. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), 209-215. <https://doi.org/10.1038/nature13772>
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, *25*(10), 1010-1022. <https://doi.org/10.1101/gad.2037511>

Derijck, A., van der Heijden, G., Giele, M., Philippens, M., & de Boer, P. (2008). DNA double-strand break repair in parental chromatin of mouse zygotes, the first cell cycle as an origin of de novo mutation. *Human Molecular Genetics*, 17(13), 1922-1937. <https://doi.org/10.1093/hmg/ddn090>

D’Gama, A. M. (2021). Somatic Mosaicism and Autism Spectrum Disorder. *Genes*, 12(11), 1699. <https://doi.org/10.3390/genes12111699>

Di, X., Azeez, A., Li, X., Haque, E., & Biswal, B. B. (2018). Disrupted focal white matter integrity in autism spectrum disorder: A voxel-based meta-analysis of diffusion tensor imaging studies. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 82, 242-248. <https://doi.org/10.1016/j.pnpbp.2017.11.007>

Dogan, M. V., Shields, B., Cutrona, C., Gao, L., Gibbons, F. X., Simons, R., Monick, M., Brody, G. H., Tan, K., Beach, S. R., & Philibert, R. A. (2014). The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics*, 15, 151. <https://doi.org/10.1186/1471-2164-15-151>

Dong, S., Walker, M. F., Carriero, N. J., DiCola, M., Willsey, A. J., Ye, A. Y., Waqar, Z., Gonzalez, L. E., Overton, J. D., Frahm, S., Keaney, J. F., Teran, N. A., Dea, J., Mandell, J. D., Hus Bal, V., Sullivan, C. A., DiLullo, N. M., Khalil, R. O., Gockley, J., ... Sanders, S. J. (2014). De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. *Cell Reports*, 9(1), 16-23. <https://doi.org/10.1016/j.celrep.2014.08.068>

Donovan, A. P. A., & Basson, M. A. (2017). The neuroanatomy of autism – a developmental perspective. *Journal of Anatomy*, 230(1), 4-15. <https://doi.org/10.1111/joa.12542>

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, 11, 587. <https://doi.org/10.1186/1471-2105-11-587>

Durand, C. M., Betancur, C., Boeckers, T. M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I. C., Anckarsäter, H., Sponheim, E., Goubran-Botros, H., Delorme, R., Chabane, N., Mouren-Simeoni, M.-C., de Mas, P., Bieth, E., Rogé, B., Héron, D., ... Bourgeron, T. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genetics*, 39(1), 25-27. <https://doi.org/10.1038/ng1933>

English, M. C. W., Maybery, M. T., & Visser, T. A. W. (2017). Reduced Pseudoneglect for Physical Space, but not Mental Representations of Space, for Adults with Autistic Traits. *Journal of Autism and Developmental Disorders*, 47(7), 1956-1965. <https://doi.org/10.1007/s10803-017-3113-5>

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: A systematic literature review. *European Child & Adolescent Psychiatry*, 22(6), 329-340. <https://doi.org/10.1007/s00787-013-0375-0>

Ferri, S. L., Abel, T., & Brodtkin, E. S. (2018). Sex Differences in Autism Spectrum Disorder: A Review. *Current Psychiatry Reports*, 20(2), 9. <https://doi.org/10.1007/s11920-018-0874-2>

Floris, D. L., Wolfers, T., Zabihi, M., Holz, N. E., Zwiers, M. P., Charman, T., Tillmann, J., Ecker, C., Dell’Acqua, F., Banaschewski, T., Moessnang, C., Baron-Cohen, S., Holt, R., Durston, S., Loth, E., Murphy, D. G. M., Marquand, A., Buitelaar, J. K., Beckmann, C. F., & EU-AIMS Longitudinal European Autism Project Group. (2021). Atypical Brain Asymmetry in Autism-A Candidate for Clinically Meaningful Stratification. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, 6(8), 802-812. <https://doi.org/10.1016/j.bpsc.2020.08.008>

Fombonne, E. (2003). Epidemiological surveys of autism and other pervasive developmental disorders: An update. *Journal of Autism and Developmental Disorders*, 33(4), 365-382. <https://doi.org/10.1023/a:1025054610557>

Foroutan, A., Haghshenas, S., Bhai, P., Levy, M. A., Kerkhof, J., McConkey, H., Niceta, M., Ciolfi, A., Pedace, L., Miele, E., Genevieve, D., Heide, S., Alders, M., Zampino, G., Merla, G., Fradin, M., Bieth, E., Bonneau, D., Dieterich, K., ... Lebre, A.-S. (2022). Clinical Utility of a Unique Genome-Wide DNA Methylation Signature for

KMT2A-Related Syndrome. *International Journal of Molecular Sciences*, 23(3), 1815. <https://doi.org/10.3390/ijms23031815>

Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., Greenwood, C. M., & Hansen, K. D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology*, 15(11), 503. <https://doi.org/10.1186/s13059-014-0503-2>

Fournier, A., Sasai, N., Nakao, M., & Defossez, P.-A. (2012). The role of methyl-binding proteins in chromatin organization and epigenome maintenance. *Briefings in Functional Genomics*, 11(3), 251-264. <https://doi.org/10.1093/bfgp/elr040>

Fransquet, P. D., Wrigglesworth, J., Woods, R. L., Ernst, M. E., & Ryan, J. (2019). The epigenetic clock as a predictor of disease and mortality risk : A systematic review and meta-analysis. *Clinical Epigenetics*, 11(1), 62. <https://doi.org/10.1186/s13148-019-0656-7>

Frazier, T. W., Thompson, L., Youngstrom, E. A., Law, P., Hardan, A. Y., Eng, C., & Morris, N. (2014). A Twin Study of Heritable and Shared Environmental Contributions to Autism. *Journal of autism and developmental disorders*, 44(8), 2013-2025. <https://doi.org/10.1007/s10803-014-2081-2>

Frazier, T. W., Youngstrom, E. A., Kubu, C. S., Sinclair, L., & Rezai, A. (2008). Exploratory and confirmatory factor analysis of the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders*, 38(3), 474-480. <https://doi.org/10.1007/s10803-007-0415-z>

Freed, D., & Pevsner, J. (2016). The Contribution of Mosaic Variants to Autism Spectrum Disorder. *PLoS Genetics*, 12(9), e1006245. <https://doi.org/10.1371/journal.pgen.1006245>

Friedman, J. H., Baskett, F., & Shustek, L. J. (1975). An Algorithm for Finding Nearest Neighbors. *IEEE Transactions on Computers*, C-24(10), 1000-1006. <https://doi.org/10.1109/T-C.1975.224110>

Friedman, J. M. (2022). Neurofibromatosis 1. In *GeneReviews® [Internet]*. University of Washington, Seattle. <https://www.ncbi.nlm.nih.gov/books/NBK1109/>

Fu, J. M., Satterstrom, F. K., Peng, M., Brand, H., Collins, R. L., Dong, S., Wamsley, B., Klei, L., Wang, L., Hao, S. P., Stevens, C. R., Cusick, C., Babadi, M., Banks, E., Collins, B., Dodge, S., Gabriel, S. B., Gauthier, L., Lee, S. K., ... Talkowski, M. E. (2022). Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nature Genetics*, 54(9), Art. 9. <https://doi.org/10.1038/s41588-022-01104-0>

Gardener, H., Spiegelman, D., & Buka, S. L. (2011). Perinatal and neonatal risk factors for autism : A comprehensive meta-analysis. *Pediatrics*, 128(2), 344-355. <https://doi.org/10.1542/peds.2010-1036>

Gardner, R. M., Lee, B. K., Magnusson, C., Rai, D., Frisell, T., Karlsson, H., Idring, S., & Dalman, C. (2015). Maternal body mass index during early pregnancy, gestational weight gain, and risk of autism spectrum disorders : Results from a Swedish total population and discordant sibling study. *International Journal of Epidemiology*, 44(3), 870-883. <https://doi.org/10.1093/ije/dyv081>

Garg, P., Jadhav, B., Rodriguez, O. L., Patel, N., Martin-Trujillo, A., Jain, M., Metsu, S., Olsen, H., Paten, B., Ritz, B., Kooy, R. F., Gecz, J., & Sharp, A. J. (2020). A Survey of Rare Epigenetic Variation in 23,116 Human Genomes Identifies Disease-Relevant Epivariations and CGG Expansions. *American Journal of Human Genetics*, 107(4), 654-669. <https://doi.org/10.1016/j.ajhg.2020.08.019>

Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity*, 105(1), 4-13. <https://doi.org/10.1038/hdy.2010.54>

Girirajan, S., Johnson, R. L., Tassone, F., Balciuniene, J., Katiyar, N., Fox, K., Baker, C., Srikanth, A., Yeoh, K. H., Khoo, S. J., Nauth, T. B., Hansen, R., Ritchie, M., Hertz-Picciotto, I., Eichler, E. E., Pessah, I. N., & Selleck, S. B. (2013). Global increases in both common and rare copy number load associated with autism. *Human Molecular Genetics*, 22(14), 2870-2880. <https://doi.org/10.1093/hmg/ddt136>

Goldmann, J. M., Veltman, J. A., & Gilissen, C. (2019). De Novo Mutations Reflect Development and Aging of the Human Germline. *Trends in Genetics*, 35(11), 828-839. <https://doi.org/10.1016/j.tig.2019.08.005>

Goldmann, J. M., Wong, W. S. W., Pinelli, M., Farrah, T., Bodian, D., Stittrich, A. B., Glusman, G., Vissers, L. E. L. M., Hoischen, A., Roach, J. C., Vockley, J. G., Veltman, J. A., Solomon, B. D., Gilissen, C., & Niederhuber, J. E. (2016). Parent-of-origin-specific signatures of de novo mutations. *Nature Genetics*, *48*(8), 935-939. <https://doi.org/10.1038/ng.3597>

González-Ortega, G., Llamas-Velasco, S., Arteche-López, A., Quesada-Espinosa, J. F., Puertas-Martín, V., Gómez-Grande, A., López-Álvarez, J., Saiz Díaz, R. A., Lezana-Rosales, J. M., Villarejo-Galende, A., & González de la Aleja, J. (2021). Early-Onset Dementia Associated with a Heterozygous, Nonsense, and de novo Variant in the MBD5 Gene. *Journal of Alzheimer's Disease: JAD*, *84*(1), 73-78. <https://doi.org/10.3233/JAD-210648>

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age : Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), Art. 6. <https://doi.org/10.1038/nrg.2016.49>

Gopinathan, G., & Diekwisch, T. G. H. (2022). Epigenetics and Early Development. *Journal of Developmental Biology*, *10*(2), 26. <https://doi.org/10.3390/jdb10020026>

Grant, O. A., Wang, Y., Kumari, M., Zabet, N. R., & Schalkwyk, L. (2022). Characterising sex differences of autosomal DNA methylation in whole blood using the Illumina EPIC array. *Clinical Epigenetics*, *14*(1), 62. <https://doi.org/10.1186/s13148-022-01279-7>

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., ... Børghlum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, *51*(3), 431-444. <https://doi.org/10.1038/s41588-019-0344-8>

Guthrie, S., Am, W., & C, Lord. (2013). Comparison of DSM-IV and DSM-5 factor structure models for toddlers with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*(8). <https://doi.org/10.1016/j.jaac.2013.05.004>

Haghshenas, S., Bhai, P., Aref-Eshghi, E., & Sadikovic, B. (2020). Diagnostic Utility of Genome-Wide DNA Methylation Analysis in Mendelian Neurodevelopmental Disorders. *International Journal of Molecular Sciences*, *21*(23), 9303. <https://doi.org/10.3390/ijms21239303>

Haghshenas, S., Levy, M. A., Kerkhof, J., Aref-Eshghi, E., McConkey, H., Balci, T., Siu, V. M., Skinner, C. D., Stevenson, R. E., Sadikovic, B., & Schwartz, C. (2021). Detection of a DNA Methylation Signature for the Intellectual Developmental Disorder, X-Linked, Syndromic, Armfield Type. *International Journal of Molecular Sciences*, *22*(3), 1111. <https://doi.org/10.3390/ijms22031111>

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., Miller, J., Fedele, A., Collins, J., Smith, K., Lotspeich, L., Croen, L. A., Ozonoff, S., Lajonchere, C., Grether, J. K., & Risch, N. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of General Psychiatry*, *68*(11), 1095-1102. <https://doi.org/10.1001/archgenpsychiatry.2011.76>

Hansen, A. W., Murugan, M., Li, H., Khayat, M. M., Wang, L., Rosenfeld, J., Andrews, B. K., Jhangiani, S. N., Coban Akdemir, Z. H., Sedlazeck, F. J., Ashley-Koch, A. E., Liu, P., Muzny, D. M., Task Force for Neonatal Genomics, Davis, E. E., Katsanis, N., Sabo, A., Posey, J. E., Yang, Y., ... Gibbs, R. A. (2019). A Genocentric Approach to Discovery of Mendelian Disorders. *American Journal of Human Genetics*, *105*(5), 974-986. <https://doi.org/10.1016/j.ajhg.2019.09.027>

Harris, H. K., Sideridis, G. D., Barbaresi, W. J., & Harstad, E. (2020). Pathogenic Yield of Genetic Testing in Autism Spectrum Disorder. *Pediatrics*, *146*(4), e20193211. <https://doi.org/10.1542/peds.2019-3211>

Havdahl, A., Niarchou, M., Starnawska, A., Uddin, M., van der Merwe, C., & Warriar, V. (2021). Genetic contributions to autism spectrum disorder. *Psychological Medicine*, *51*(13), 2260-2273. <https://doi.org/10.1017/S0033291721000192>

Hayeems, R. Z., Babul-Hirji, R., Hoang, N., Weksberg, R., & Shuman, C. (2016). Parents' Experience with Pediatric Microarray : Transferrable Lessons in the Era of Genomic Counseling. *Journal of Genetic Counseling*, *25*(2), 298-304. <https://doi.org/10.1007/s10897-015-9871-3>

Hazlett, H. C., Poe, M. D., Gerig, G., Smith, R. G., & Piven, J. (2006). Cortical gray and white brain tissue volume in adolescents and adults with autism. *Biological Psychiatry, 59*(1), 1-6. <https://doi.org/10.1016/j.biopsych.2005.06.015>

Hendrich, B., & Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and Cellular Biology, 18*(11), 6538-6547. <https://doi.org/10.1128/MCB.18.11.6538>

Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T. M., Nordentoft, M., & Glenthøj, B. (2018). Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biological Psychiatry, 83*(6), 492-498. <https://doi.org/10.1016/j.biopsych.2017.08.017>

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor Structure, Reliability and Criterion Validity of the Autism-Spectrum Quotient (AQ): A Study in Dutch Population and Patient Groups. *Journal of Autism and Developmental Disorders, 38*(8), 1555-1566. <https://doi.org/10.1007/s10803-008-0538-x>

Hoekstra, R. A., Bartels, M., Verweij, C. J. H., & Boomsma, D. I. (2007). Heritability of autistic traits in the general population. *Archives of Pediatrics & Adolescent Medicine, 161*(4), 372-377. <https://doi.org/10.1001/archpedi.161.4.372>

Holliday, R., & Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science (New York, N.Y.), 187*(4173), 226-232.

Hood, R. L., Schenkel, L. C., Nikkel, S. M., Ainsworth, P. J., Pare, G., Boycott, K. M., Bulman, D. E., & Sadikovic, B. (2016). The defining DNA methylation signature of Floating-Harbor Syndrome. *Scientific Reports, 6*, 38803. <https://doi.org/10.1038/srep38803>

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology, 14*(10), R115. <https://doi.org/10.1186/gb-2013-14-10-r115>

Horvath, S., Oshima, J., Martin, G. M., Lu, A. T., Quach, A., Cohen, H., Felton, S., Matsuyama, M., Lowe, D., Kabacik, S., Wilson, J. G., Reiner, A. P., Maierhofer, A., Flunkert, J., Aviv, A., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., ... Raj, K. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and *ex vivo* studies. *Aging, 10*(7), 1758-1775. <https://doi.org/10.18632/aging.101508>

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., Wiencke, J. K., & Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics, 13*, 86. <https://doi.org/10.1186/1471-2105-13-86>

Howrigan, D. P., Rose, S. A., Samocha, K. E., Fromer, M., Cerrato, F., Chen, W. J., Churchhouse, C., Chambert, K., Chandler, S. D., Daly, M. J., Dumont, A., Genovese, G., Hwu, H.-G., Laird, N., Kosmicki, J. A., Moran, J. L., Roe, C., Singh, T., Wang, S.-H., ... Neale, B. M. (2020). Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nature Neuroscience, 23*(2), 185-193. <https://doi.org/10.1038/s41593-019-0564-3>

Huguet, G., Ey, E., & Bourgeron, T. (2013). The genetic landscapes of autism spectrum disorders. *Annual Review of Genomics and Human Genetics, 14*, 191-213. <https://doi.org/10.1146/annurev-genom-091212-153431>

Husson, T., Lecoquierre, F., Cassinari, K., Charbonnier, C., Quenez, O., Goldenberg, A., Guerrot, A.-M., Richard, A.-C., Drouin-Garraud, V., Brehin, A.-C., Soleimani, M., Taton, R., Rotharmel, M., Rosier, A., Chambon, P., Le Meur, N., Joly-Helas, G., Saugier-Verber, P., Boland, A., ... Champion, D. (2020). Rare genetic susceptibility variants assessment in autism spectrum disorder: Detection rate and practical use. *Translational Psychiatry, 10*(1), 77. <https://doi.org/10.1038/s41398-020-0760-7>

Hutton, J. (2016). Does Rubella Cause Autism : A 2015 Reappraisal? *Frontiers in Human Neuroscience, 10*, 25. <https://doi.org/10.3389/fnhum.2016.00025>

Ichino, L., Boone, B. A., Strauskulage, L., Harris, C. J., Kaur, G., Gladstone, M. A., Tan, M., Feng, S., Jami-Alahmadi, Y., Duttke, S. H., Wohlschlegel, J. A., Cheng, X., Redding, S., & Jacobsen, S. E. (2021). MBD5 and

MBD6 couple DNA methylation to gene silencing through the J-domain protein SILENZIO. *Science*, 372(6549), 1434-1439. <https://doi.org/10.1126/science.abg6130>

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931-945. <https://doi.org/10.1038/nature03001>

Iossifov, I., O’Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., Stessman, H. A., Witherspoon, K. T., Vives, L., Patterson, K. E., Smith, J. D., Paepier, B., Nickerson, D. A., Dea, J., Dong, S., Gonzalez, L. E., Mandell, J. D., Mane, S. M., Murtha, M. T., ... Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515(7526), Art. 7526. <https://doi.org/10.1038/nature13908>

Jacquemont, S., Coe, B. P., Hersch, M., Duyzend, M. H., Krumm, N., Bergmann, S., Beckmann, J. S., Rosenfeld, J. A., & Eichler, E. E. (2014). A higher mutational burden in females supports a « female protective model » in neurodevelopmental disorders. *American Journal of Human Genetics*, 94(3), 415-425. <https://doi.org/10.1016/j.ajhg.2014.02.001>

Jaffe, A. E., Murakami, P., Lee, H., Leek, J. T., Fallin, M. D., Feinberg, A. P., & Irizarry, R. A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *International Journal of Epidemiology*, 41(1), 200-209. <https://doi.org/10.1093/ije/dyr238>

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A., Batzoglu, S., Sanders, S. J., & Farh, K. K.-H. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, 176(3), 535-548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>

Jiao, C., Zhang, C., Dai, R., Xia, Y., Wang, K., Giase, G., Chen, C., & Liu, C. (2018). Positional effects revealed in Illumina methylation array and the impact on analysis. *Epigenomics*, 10(5), 643-659. <https://doi.org/10.2217/epi-2017-0105>

Jo, H., Eckel, S. P., Chen, J.-C., Cockburn, M., Martinez, M. P., Chow, T., Lurmann, F. W., Funk, W. E., Xiang, A. H., & McConnell, R. (2019). Gestational diabetes mellitus, prenatal air pollution exposure, and autism spectrum disorder. *Environment International*, 133(Pt A), 105110. <https://doi.org/10.1016/j.envint.2019.105110>

Johannessen, J., Nærland, T., Hope, S., Torske, T., Høyland, A. L., Strohmaier, J., Heiberg, A., Rietschel, M., Djurovic, S., & Andreassen, O. A. (2017). Parents’ Attitudes toward Clinical Genetic Testing for Autism Spectrum Disorder—Data from a Norwegian Sample. *International Journal of Molecular Sciences*, 18(5), 1078. <https://doi.org/10.3390/ijms18051078>

Jones, P. A. (2012). Functions of DNA methylation : Islands, start sites, gene bodies and beyond. *Nature Reviews. Genetics*, 13(7), 484-492. <https://doi.org/10.1038/nrg3230>

Jónsson, H., Sulem, P., Kehr, B., Kristmundsdóttir, S., Zink, F., Hjartarson, E., Hardarson, M. T., Hjorleifsson, K. E., Eggertsson, H. P., Gudjonsson, S. A., Ward, L. D., Arnadóttir, G. A., Helgason, E. A., Helgason, H., Gylfason, A., Jonasdóttir, A., Jonasdóttir, A., Rafnar, T., Frigge, M., ... Stefansson, K. (2017). Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*, 549(7673), 519-522. <https://doi.org/10.1038/nature24018>

Just, M. A., Cherkassky, V. L., Keller, T. A., & Minshew, N. J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism : Evidence of underconnectivity. *Brain: A Journal of Neurology*, 127(Pt 8), 1811-1821. <https://doi.org/10.1093/brain/awh199>

Kadam, A., Patni, B., Pandit, A., & Patole, S. (2021). Stability of the Initial Diagnosis of Autism Spectrum Disorder by DSM-5 in Children : A Short-Term Follow-Up Study. *Journal of Tropical Pediatrics*, 67(1), fmaa104. <https://doi.org/10.1093/tropej/fmaa104>

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217-250.

Kaplanis, J., Samocha, K. E., Wiel, L., Zhang, Z., Arvai, K. J., Eberhardt, R. Y., Gallone, G., Lelieveld, S. H., Martin, H. C., McRae, J. F., Short, P. J., Torene, R. I., de Boer, E., Danecsek, P., Gardner, E. J., Huang, N., Lord,

J., Martincorena, I., Pfundt, R., ... Retterer, K. (2020). Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature*, 586(7831), 757-762. <https://doi.org/10.1038/s41586-020-2832-5>

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., ... MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), Art. 7809. <https://doi.org/10.1038/s41586-020-2308-7>

Karolis, V. R., Corbetta, M., & Thiebaut de Schotten, M. (2019). The architecture of functional lateralisation and its relationship to callosal connectivity in the human brain. *Nature Communications*, 10(1), Art. 1. <https://doi.org/10.1038/s41467-019-09344-1>

Kerkhof, J., Squeo, G. M., McConkey, H., Levy, M. A., Piemontese, M. R., Castori, M., Accadia, M., Biamino, E., Della Monica, M., Di Giacomo, M. C., Gervasini, C., Maitz, S., Melis, D., Milani, D., Piccione, M., Prontera, P., Selicorni, A., Sadikovic, B., & Merla, G. (2022). DNA methylation episingature testing improves molecular diagnosis of Mendelian chromatinopathies. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 24(1), 51-60. <https://doi.org/10.1016/j.gim.2021.08.007>

Kernohan, K. D., Cigana Schenkel, L., Huang, L., Smith, A., Pare, G., Ainsworth, P., Care4Rare Canada Consortium, Boycott, K. M., Warman-Chardon, J., & Sadikovic, B. (2016). Identification of a methylation profile for DNMT1-associated autosomal dominant cerebellar ataxia, deafness, and narcolepsy. *Clinical Epigenetics*, 8, 91. <https://doi.org/10.1186/s13148-016-0254-x>

Kim, I. B., Lee, T., Lee, J., Kim, J., Lee, S., Koh, I. G., Kim, J. H., An, J.-Y., Lee, H., Kim, W. K., Ju, Y. S., Cho, Y., Yu, S. J., Kim, S. A., Oh, M., Han, D. W., Kim, E., Choi, J. K., Yoo, H. J., & Lee, J. H. (2022). Non-coding de novo mutations in chromatin interactions are implicated in autism spectrum disorder. *Molecular Psychiatry*. <https://doi.org/10.1038/s41380-022-01697-2>

King, M., & Bearman, P. (2009). Diagnostic change and the increased prevalence of autism. *International Journal of Epidemiology*, 38(5), 1224-1234. <https://doi.org/10.1093/ije/dyp261>

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, 46(3), 310-315. <https://doi.org/10.1038/ng.2892>

Klei, L., McClain, L. L., Mahjani, B., Panayidou, K., De Rubeis, S., Grahnat, A.-C. S., Karlsson, G., Lu, Y., Melhem, N., Xu, X., Reichenberg, A., Sandin, S., Hultman, C. M., Buxbaum, J. D., Roeder, K., & Devlin, B. (2021). How rare and common risk variation jointly affect liability for autism spectrum disorder. *Molecular Autism*, 12(1), 66. <https://doi.org/10.1186/s13229-021-00466-2>

Kleinhans, N. M., Müller, R.-A., Cohen, D. N., & Courchesne, E. (2008). Atypical functional lateralization of language in autism spectrum disorders. *Brain Research*, 1221, 115-125. <https://doi.org/10.1016/j.brainres.2008.04.080>

Koyama, R., & Ikegaya, Y. (2015). Microglia in the pathogenesis of autism spectrum disorders. *Neuroscience Research*, 100, 1-5. <https://doi.org/10.1016/j.neures.2015.06.005>

Kozuki, N., & Walker, N. (2013). Exploring the association between short/long preceding birth intervals and child mortality: Using reference birth interval children of the same mother as comparison. *BMC Public Health*, 13 Suppl 3, S6. <https://doi.org/10.1186/1471-2458-13-S3-S6>

Krawczak, M., Ball, E. V., & Cooper, D. N. (1998). Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *American Journal of Human Genetics*, 63(2), 474-488. <https://doi.org/10.1086/301965>

Krumm, N., Turner, T. N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B. P., Stessman, H. A., He, Z.-X., Leal, S. M., Bernier, R., & Eichler, E. E. (2015). Excess of rare, inherited truncating mutations in autism. *Nature Genetics*, 47(6), 582-588. <https://doi.org/10.1038/ng.3303>

- Kryukov, G. V., Pennacchio, L. A., & Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. *American Journal of Human Genetics*, *80*(4), 727-739. <https://doi.org/10.1086/513473>
- Kulage, K. M., Goldberg, J., Usseglio, J., Romero, D., Bain, J. M., & Smaldone, A. M. (2020). How has DSM-5 Affected Autism Diagnosis? A 5-Year Follow-Up Systematic Literature Review and Meta-analysis. *Journal of Autism and Developmental Disorders*, *50*(6), 2102-2127. <https://doi.org/10.1007/s10803-019-03967-5>
- Laget, S., Joulie, M., Le Masson, F., Sasai, N., Christians, E., Pradhan, S., Roberts, R. J., & Defossez, P.-A. (2010). The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS One*, *5*(8), e11982. <https://doi.org/10.1371/journal.pone.0011982>
- Lai, M.-C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., Szatmari, P., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *The Lancet. Psychiatry*, *6*(10), 819-829. [https://doi.org/10.1016/S2215-0366\(19\)30289-5](https://doi.org/10.1016/S2215-0366(19)30289-5)
- Laird, P. W. (2010). Principles and challenges of genomewide DNA methylation analysis. *Nature Reviews. Genetics*, *11*(3), 191-203. <https://doi.org/10.1038/nrg2732>
- Lalani, S. R., Safiullah, A. M., Fernbach, S. D., Harutyunyan, K. G., Thaller, C., Peterson, L. E., McPherson, J. D., Gibbs, R. A., White, L. D., Hefner, M., Davenport, S. L. H., Graham, J. M., Bacino, C. A., Glass, N. L., Towbin, J. A., Craigen, W. J., Neish, S. R., Lin, A. E., & Belmont, J. W. (2006). Spectrum of CHD7 Mutations in 110 Individuals with CHARGE Syndrome and Genotype-Phenotype Correlation. *American Journal of Human Genetics*, *78*(2), 303-314.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., ... The Wellcome Trust: (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), Art. 6822. <https://doi.org/10.1038/35057062>
- Lane, C., Milne, E., & Freeth, M. (2017). Characteristics of Autism Spectrum Disorder in Sotos Syndrome. *Journal of Autism and Developmental Disorders*, *47*(1), 135-143. <https://doi.org/10.1007/s10803-016-2941-z>
- Lappalainen, T., Scott, A. J., Brandt, M., & Hall, I. M. (2019). Genomic analysis in the age of human genome sequencing. *Cell*, *177*(1), 70-84. <https://doi.org/10.1016/j.cell.2019.02.032>
- Law, P.-P., & Holland, M. L. (2019). DNA methylation at the crossroads of gene and environment interactions. *Essays in Biochemistry*, *63*(6), 717-726. <https://doi.org/10.1042/EBC20190031>
- Le Couteur, A., Bailey, A., Goode, S., Pickles, A., Robertson, S., Gottesman, I., & Rutter, M. (1996). A broader phenotype of autism: The clinical spectrum in twins. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *37*(7), 785-801. <https://doi.org/10.1111/j.1469-7610.1996.tb01475.x>
- Le Couteur, A., Haden, G., Hammal, D., & McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: The ADI-R and the ADOS. *Journal of Autism and Developmental Disorders*, *38*(2), 362-372. <https://doi.org/10.1007/s10803-007-0403-3>
- Lee, B. K., Arver, S., Widman, L., Gardner, R. M., Magnusson, C., Dalman, C., & Kosidou, K. (2017). Maternal hirsutism and autism spectrum disorders in offspring. *Autism Research: Official Journal of the International Society for Autism Research*, *10*(9), 1544-1546. <https://doi.org/10.1002/aur.1797>
- Lee, B. K., Magnusson, C., Gardner, R. M., Blomström, Å., Newschaffer, C. J., Burstyn, I., Karlsson, H., & Dalman, C. (2015). Maternal hospitalization with infection during pregnancy and risk of autism spectrum disorders. *Brain, Behavior, and Immunity*, *44*, 100-105. <https://doi.org/10.1016/j.bbi.2014.09.001>
- Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature reviews. Genetics*, *11*(10), 10.1038/nrg2825. <https://doi.org/10.1038/nrg2825>

Lehne, B., Drong, A. W., Loh, M., Zhang, W., Scott, W. R., Tan, S.-T., Afzal, U., Scott, J., Jarvelin, M.-R., Elliott, P., McCarthy, M. I., Kooner, J. S., & Chambers, J. C. (2015). A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biology*, *16*(1), 37. <https://doi.org/10.1186/s13059-015-0600-x>

Lejeune, J., Gautier, M., Lafourcade, J., Berger, R., & Turpin, R. (1964). [PARTIAL DELETION OF THE SHORT ARM OF THE 5 CHROMOSOME: THE 5TH CASE OF CRYING CAT SYNDROME]. *Annales De Genetique*, *7*, 7-10.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., ... MacArthur, D. G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285-291. <https://doi.org/10.1038/nature19057>

Leppa, V. M., Kravitz, S. N., Martin, C. L., Andrieux, J., Le Caignec, C., Martin-Coignard, D., DyBuncio, C., Sanders, S. J., Lowe, J. K., Cantor, R. M., & Geschwind, D. H. (2016). Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in Multiplex Families. *American Journal of Human Genetics*, *99*(3), 540-554. <https://doi.org/10.1016/j.ajhg.2016.06.036>

Leshem, R., Bar-Oz, B., Diav-Citrin, O., Gbaly, S., Soliman, J., Renoux, C., & Matok, I. (2021). Selective Serotonin Reuptake Inhibitors (SSRIs) and Serotonin Norepinephrine Reuptake Inhibitors (SNRIs) During Pregnancy and the Risk for Autism spectrum disorder (ASD) and Attention deficit hyperactivity disorder (ADHD) in the Offspring: A True Effect or a Bias? A Systematic Review & Meta-Analysis. *Current Neuropharmacology*, *19*(6), 896-906. <https://doi.org/10.2174/1570159X19666210303121059>

Levy, D., Ronemus, M., Yamrom, B., Lee, Y., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., Buja, A., Krieger, A., Yoon, S., Troge, J., Rodgers, L., Iossifov, I., & Wigler, M. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*, *70*(5), 886-897. <https://doi.org/10.1016/j.neuron.2011.05.015>

Levy, M. A., McConkey, H., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., Bralo, M. P., Cappuccio, G., Ciolfi, A., Clarke, A., DuPont, B. R., Elting, M. W., Faivre, L., Fee, T., Fletcher, R. S., Cherk, F., Foroutan, A., Friez, M. J., Gervasini, C., ... Sadikovic, B. (2022). Novel diagnostic DNA methylation epesignatures expand and refine the epigenetic landscapes of Mendelian disorders. *HGG Advances*, *3*(1), 100075. <https://doi.org/10.1016/j.xhgg.2021.100075>

Levy, M. A., Relator, R., McConkey, H., Prankeviciene, E., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., Palomares Bralo, M., Cappuccio, G., Ciolfi, A., Clarke, A., DuPont, B. R., Elting, M. W., Faivre, L., Fee, T., Ferilli, M., Fletcher, R. S., Cherk, F., ... Sadikovic, B. (2022). Functional correlation of genome-wide DNA methylation profiles in genetic neurodevelopmental disorders. *Human Mutation*, *43*(11), 1609-1628. <https://doi.org/10.1002/humu.24446>

Li, D., Xie, Z., Le Pape, M., & Dye, T. (2015). An evaluation of statistical methods for DNA methylation microarray data analysis. *BMC Bioinformatics*, *16*, 217. <https://doi.org/10.1186/s12859-015-0641-x>

Li, Q., Becker, B., Jiang, X., Zhao, Z., Zhang, Q., Yao, S., & Kendrick, K. M. (2019). Decreased interhemispheric functional connectivity rather than corpus callosum volume as a potential biomarker for autism spectrum disorder. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *119*, 258-266. <https://doi.org/10.1016/j.cortex.2019.05.003>

Li, S., & Tollefsbol, T. O. (2021). DNA methylation methods: Global DNA methylation and methylomic analyses. *Methods (San Diego, Calif.)*, *187*, 28-43. <https://doi.org/10.1016/j.ymeth.2020.10.002>

Libby, P. (2021). Inflammation in Atherosclerosis-No Longer a Theory. *Clinical Chemistry*, *67*(1), 131-142. <https://doi.org/10.1093/clinchem/hvaa275>

Lichtenstein, P., Carlström, E., Råstam, M., Gillberg, C., & Anckarsäter, H. (2010). The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *The American Journal of Psychiatry*, *167*(11), 1357-1363. <https://doi.org/10.1176/appi.ajp.2010.10020223>

- 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., Lek, M., Liu, L., Flannick, J., Ripke, S., Nagaswamy, U., Muzny, D., Reid, J. G., Hawes, A., Newsham, I., ... Daly, M. J. (2013). Rare complete knockouts in humans: Population distribution and significant role in autism spectrum disorders. *Neuron*, 77(2), 235-242. <https://doi.org/10.1016/j.neuron.2012.12.029>
- Lim, E. T., Uddin, M., De Rubeis, S., Chan, Y., Kamumbu, A. S., Zhang, X., D'Gama, A., Kim, S. N., Hill, R. S., Goldberg, A. P., Poultney, C., Minshew, N. J., Kushima, I., Aleksic, B., Ozaki, N., Parellada, M., Arango, C., Penzol, M. J., Carracedo, A., ... Walsh, C. A. (2017). Rates, Distribution, and Implications of Post-zygotic Mosaic Mutations in Autism Spectrum Disorder. *Nature neuroscience*, 20(9), 1217-1224. <https://doi.org/10.1038/nn.4598>
- Lindeboom, R. G. H., Vermeulen, M., Lehner, B., & Supek, F. (2019). The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nature Genetics*, 51(11), 1645-1651. <https://doi.org/10.1038/s41588-019-0517-5>
- Lindell, A. K., & Hudry, K. (2013). Atypicalities in cortical structure, handedness, and functional lateralization for language in autism spectrum disorders. *Neuropsychology Review*, 23(3), 257-270. <https://doi.org/10.1007/s11065-013-9234-5>
- Liu, C., Zhai, H., Su, S., Song, S., Chen, G., & Jiang, Y. (2021). Visuospatial Bias in Children with Autism Spectrum Disorder: Evidence from Line Bisection Tasks. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-021-05350-9>
- Llera, A., Brammer, M., Oakley, B., Tillmann, J., Zabihi, M., Amelink, J. S., Mei, T., Charman, T., Ecker, C., Dell'Acqua, F., Banaschewski, T., Moessnang, C., Baron-Cohen, S., Holt, R., Durston, S., Murphy, D., Loth, E., Buitelaar, J. K., Floris, D. L., & Beckmann, C. F. (2022). Evaluation of data imputation strategies in complex, deeply-phenotyped data sets: The case of the EU-AIMS Longitudinal European Autism Project. *BMC Medical Research Methodology*, 22(1), 229. <https://doi.org/10.1186/s12874-022-01656-z>
- Logue, M. W., Smith, A. K., Wolf, E. J., Maniates, H., Stone, A., Schichman, S. A., McGlinchey, R. E., Milberg, W., & Miller, M. W. (2017). The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. *Epigenomics*, 9(11), 1363-1371. <https://doi.org/10.2217/epi-2017-0078>
- Long, H. K., King, H. W., Patient, R. K., Odom, D. T., & Klose, R. J. (2016). Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Research*, 44(14), 6693-6706. <https://doi.org/10.1093/nar/gkw258>
- Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, 56(6), 466-474. <https://doi.org/10.1016/j.jaac.2017.03.013>
- Lukmanji, S., Manji, S. A., Kadhim, S., Sauro, K. M., Wirrell, E. C., Kwon, C.-S., & Jetté, N. (2019). The co-occurrence of epilepsy and autism: A systematic review. *Epilepsy & Behavior: E&B*, 98(Pt A), 238-248. <https://doi.org/10.1016/j.yebeh.2019.07.037>
- Luo, X., Wang, F., Wang, G., & Zhao, Y. (2020). Identification of methylation states of DNA regions for Illumina methylation BeadChip. *BMC Genomics*, 21(Suppl 1), 672. <https://doi.org/10.1186/s12864-019-6019-0>
- Lykke-Andersen, S., & Jensen, T. H. (2015). Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. *Nature Reviews. Molecular Cell Biology*, 16(11), 665-677. <https://doi.org/10.1038/nrm4063>
- Lynch, C. J., Uddin, L. Q., Supekar, K., Khouzam, A., Phillips, J., & Menon, V. (2013). Default mode network in childhood autism: Posteromedial cortex heterogeneity and relationship with social deficits. *Biological Psychiatry*, 74(3), 212-219. <https://doi.org/10.1016/j.biopsych.2012.12.013>
- Maher, G. M., O'Keeffe, G. W., Kearney, P. M., Kenny, L. C., Dinan, T. G., Mattsson, M., & Khashan, A. S. (2018). Association of Hypertensive Disorders of Pregnancy With Risk of Neurodevelopmental Disorders in

Offspring: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, 75(8), 809-819. <https://doi.org/10.1001/jamapsychiatry.2018.0854>

Maksimovic, J., Phipson, B., & Oshlack, A. (2017). A cross-package Bioconductor workflow for analysing methylation array data. *F1000Research*, 5, 1281. <https://doi.org/10.12688/f1000research.8839.3>

Mandy, W. P. L., & Skuse, D. H. (2008). Research review: What is the association between the social-communication element of autism and repetitive interests, behaviours and activities? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 49(8), 795-808. <https://doi.org/10.1111/j.1469-7610.2008.01911.x>

Mansur, F., Teles E Silva, A. L., Gomes, A. K. S., Magdalon, J., de Souza, J. S., Griesi-Oliveira, K., Passos-Bueno, M. R., & Sertié, A. L. (2021). Complement C4 Is Reduced in iPSC-Derived Astrocytes of Autism Spectrum Disorder Subjects. *International Journal of Molecular Sciences*, 22(14), 7579. <https://doi.org/10.3390/ijms22147579>

Marchal, C., & Miotto, B. (2015). Emerging concept in DNA methylation: Role of transcription factors in shaping DNA methylation patterns. *Journal of Cellular Physiology*, 230(4), 743-751. <https://doi.org/10.1002/jcp.24836>

Markou, P., Ahtam, B., & Papadatou-Pastou, M. (2017). Elevated Levels of Atypical Handedness in Autism: Meta-Analyses. *Neuropsychology Review*, 27(3), 258-283. <https://doi.org/10.1007/s11065-017-9354-4>

Marwaha, S., Knowles, J. W., & Ashley, E. A. (2022). A guide for the diagnosis of rare and undiagnosed disease: Beyond the exome. *Genome Medicine*, 14, 23. <https://doi.org/10.1186/s13073-022-01026-w>

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities*, 30(6), 1107-1114. <https://doi.org/10.1016/j.ridd.2009.06.003>

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutayavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., ... Stamatoyannopoulos, J. A. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science (New York, N.Y.)*, 337(6099), 1190-1195. <https://doi.org/10.1126/science.1222794>

Mbadiwe, T., & Millis, R. M. (2013). Epigenetics and autism. *Autism Research and Treatment*, 2013, 826156. <https://doi.org/10.1155/2013/826156>

McKavanagh, R., Buckley, E., & Chance, S. A. (2015). Wider minicolumns in autism: A neural basis for altered processing? *Brain: A Journal of Neurology*, 138(Pt 7), 2034-2045. <https://doi.org/10.1093/brain/awv110>

Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences*, 15(10), 483-506. <https://doi.org/10.1016/j.tics.2011.08.003>

Michaelson, J. J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., Wu, W., Corominas, R., Peoples, A., Koren, A., Gore, A., Kang, S., Lin, G. N., Estabillio, J., Gadomski, T., ... Sebat, J. (2012). Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell*, 151(7), 1431-1442. <https://doi.org/10.1016/j.cell.2012.11.019>

Miles, J. H., Takahashi, T. N., Bagby, S., Sahota, P. K., Vaslow, D. F., Wang, C. H., Hillman, R. E., & Farmer, J. E. (2005). Essential versus complex autism: Definition of fundamental prognostic subtypes. *American Journal of Medical Genetics. Part A*, 135(2), 171-180. <https://doi.org/10.1002/ajmg.a.30590>

Miller, L. E., Dai, Y. G., Fein, D. A., & Robins, D. L. (2021). Characteristics of toddlers with early versus later diagnosis of autism spectrum disorder. *Autism: The International Journal of Research and Practice*, 25(2), 416-428. <https://doi.org/10.1177/1362361320959507>

Min, J. L., Hemani, G., Davey Smith, G., Relton, C., & Suderman, M. (2018). Meffil: Efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics (Oxford, England)*, 34(23), 3983-3989. <https://doi.org/10.1093/bioinformatics/bty476>

Mirza-Schreiber, N., Zech, M., Wilson, R., Brunet, T., Wagner, M., Jech, R., Boesch, S., Škorvánek, M., Necpál, J., Weise, D., Weber, S., Mollenhauer, B., Trenkwalder, C., Maier, E. M., Borggraefe, I., Vill, K., Hackenberg, A., Pilshofer, V., Kotzaeridou, U., ... Oexle, K. (2022). Blood DNA methylation provides an accurate biomarker of KMT2B-related dystonia and predicts onset. *Brain: A Journal of Neurology*, *145*(2), 644-654. <https://doi.org/10.1093/brain/awab360>

Miyake, K., Hirasawa, T., Koide, T., & Kubota, T. (2012). Epigenetics in autism and other neurodevelopmental diseases. *Advances in Experimental Medicine and Biology*, *724*, 91-98. https://doi.org/10.1007/978-1-4614-0653-2_7

Moradimanesh, Z., Khosrowabadi, R., Eshaghi Gordji, M., & Jafari, G. R. (2021). Altered structural balance of resting-state networks in autism. *Scientific Reports*, *11*(1), 1966. <https://doi.org/10.1038/s41598-020-80330-0>

Mouga, S., Duarte, I. C., Café, C., Sousa, D., Duque, F., Oliveira, G., & Castelo-Branco, M. (2022). Parahippocampal deactivation and hyperactivation of central executive, saliency and social cognition networks in autism spectrum disorder. *Journal of Neurodevelopmental Disorders*, *14*(1), 9. <https://doi.org/10.1186/s11689-022-09417-1>

Mraz, K. D., Green, J., Dumont-Mathieu, T., Makin, S., & Fein, D. (2007). Correlates of head circumference growth in infants later diagnosed with autism spectrum disorders. *Journal of Child Neurology*, *22*(6), 700-713. <https://doi.org/10.1177/0883073807304005>

Mullegama, S. V., & Elsea, S. H. (2016). Clinical and Molecular Aspects of MBD5-Associated Neurodevelopmental Disorder (MAND). *European Journal of Human Genetics*, *24*(9), 1235-1243. <https://doi.org/10.1038/ejhg.2016.35>

Mullegama, S. V., Klein, S. D., Williams, S. R., Innis, J. W., Probst, F. J., Haldeman-Englert, C., Martinez-Agosto, J. A., Yang, Y., Tian, Y., Elsea, S. H., & Ezashi, T. (2021). Transcriptome analysis of MBD5-associated neurodevelopmental disorder (MAND) neural progenitor cells reveals dysregulation of autism-associated genes. *Scientific Reports*, *11*, 11295. <https://doi.org/10.1038/s41598-021-90798-z>

Mullegama, S. V., Rosenfeld, J. A., Orellana, C., van Bon, B. W. M., Halbach, S., Repnikova, E. A., Brick, L., Li, C., Dupuis, L., Rosello, M., Aradhya, S., Stavropoulos, D. J., Manickam, K., Mitchell, E., Hodge, J. C., Talkowski, M. E., Gusella, J. F., Keller, K., Zonana, J., ... Elsea, S. H. (2014). Reciprocal deletion and duplication at 2q23.1 indicates a role for MBD5 in autism spectrum disorder. *European Journal of Human Genetics: EJHG*, *22*(1), 57-63. <https://doi.org/10.1038/ejhg.2013.67>

Murr, R. (2010). Interplay between different epigenetic modifications and mechanisms. *Advances in Genetics*, *70*, 101-141. <https://doi.org/10.1016/B978-0-12-380866-0.60005-8>

Myers, S. M., Challman, T. D., Bernier, R., Bourgeron, T., Chung, W. K., Constantino, J. N., Eichler, E. E., Jacquemont, S., Miller, D. T., Mitchell, K. J., Zoghbi, H. Y., Martin, C. L., & Ledbetter, D. H. (2020). Insufficient Evidence for "Autism-Specific" Genes. *American Journal of Human Genetics*, *106*(5), 587-595. <https://doi.org/10.1016/j.ajhg.2020.04.004>

Naeem, H., Wong, N. C., Chatterton, Z., Hong, M. K. H., Pedersen, J. S., Corcoran, N. M., Hovens, C. M., & Macintyre, G. (2014). Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. *BMC Genomics*, *15*, 51. <https://doi.org/10.1186/1471-2164-15-51>

Nakatochi, M., Kushima, I., & Ozaki, N. (2021). Implications of germline copy-number variations in psychiatric disorders: Review of large-scale genetic studies. *Journal of Human Genetics*, *66*(1), Art. 1. <https://doi.org/10.1038/s10038-020-00838-1>

Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., Lee, C., Turner, E. H., Smith, J. D., Rieder, M. J., Yoshiura, K.-I., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D. A., ... Shendure, J. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature Genetics*, *42*(9), 790-793. <https://doi.org/10.1038/ng.646>

Nomi, J. S., Molnar-Szakacs, I., & Uddin, L. Q. (2019). Insular function in autism: Update and future directions in neuroimaging and interventions. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 89, 412-426. <https://doi.org/10.1016/j.pnpbp.2018.10.015>

Nomi, J. S., & Uddin, L. Q. (2015). Developmental changes in large-scale network connectivity in autism. *NeuroImage. Clinical*, 7, 732-741. <https://doi.org/10.1016/j.nicl.2015.02.024>

Nordenbæk, C., Jørgensen, M., Kyvik, K. O., & Bilenberg, N. (2014). A Danish population-based twin study on autism spectrum disorders. *European Child & Adolescent Psychiatry*, 23(1), 35-43. <https://doi.org/10.1007/s00787-013-0419-5>

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonze, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., ... Phillippy, A. M. (2022). The complete sequence of a human genome. *Science (New York, N.Y.)*, 376(6588), 44-53. <https://doi.org/10.1126/science.abj6987>

O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., Turner, E. H., Stanaway, I. B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J. M., Borenstein, E., Rieder, M. J., ... Eichler, E. E. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, 485(7397), 246-250. <https://doi.org/10.1038/nature10989>

Orrico, A., Galli, L., Rossi, M., Cortesi, A., Mazzi, M., & Caterino, E. (2021). The Variable Expression of a Novel MBD5 Gene Frameshift Mutation in an Italian Family. *Neuropediatrics*, 52(2), 138-141. <https://doi.org/10.1055/s-0040-1715633>

Ostrowski, P. J., Zachariou, A., Loveday, C., Belezá-Meireles, A., Bertoli, M., Dean, J., Douglas, A. G. L., Ellis, I., Foster, A., Graham, J. M., Hague, J., Hilhorst-Hofstee, Y., Hoffer, M., Johnson, D., Josifova, D., Kant, S. G., Kini, U., Lachlan, K., Lam, W., ... Tatton-Brown, K. (2019). The CHD8 overgrowth syndrome: A detailed evaluation of an emerging overgrowth phenotype in 27 patients. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, 181(4). <https://doi.org/10.1002/ajmg.c.31749>

Ouldin, K., Samri, I., Bouguenouch, L., Hamdaoui, H., Otmani, I. E., Hbib, M., Chaouki, S., & Hida, M. (2012). Le syndrome de Cri du Chat : A propos d'une observation. *The Pan African Medical Journal*, 11, 4.

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., Chawarska, K., Macari, S. L., Messinger, D., Stone, W. L., Zwaigenbaum, L., & Iosif, A.-M. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: A baby siblings research consortium study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 56(9), 988-998. <https://doi.org/10.1111/jcpp.12421>

Padmanabhan, A., Lynch, C. J., Schaer, M., & Menon, V. (2017). The Default Mode Network in Autism. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, 2(6), 476-486. <https://doi.org/10.1016/j.bpsc.2017.04.004>

Pan, P.-Y., Bölte, S., Kaur, P., Jamil, S., & Jonsson, U. (2021). Neurological disorders in autism: A systematic review and meta-analysis. *Autism: The International Journal of Research and Practice*, 25(3), 812-830. <https://doi.org/10.1177/1362361320951370>

Park, B. Y., Yao, R., Tierney, E., Brucato, M., Hong, X., Wang, G., Ji, Y., Pearson, C., Fallin, M. D., Wang, X., & Volk, H. (2021). The association between maternal lipid profile after birth and offspring risk of autism spectrum disorder. *Annals of Epidemiology*, 53, 50-55.e1. <https://doi.org/10.1016/j.annepidem.2020.08.009>

Peixoto, P., Cartron, P.-F., Serandour, A. A., & Hervouet, E. (2020). From 1957 to Nowadays: A Brief History of Epigenetics. *International Journal of Molecular Sciences*, 21(20), 7571. <https://doi.org/10.3390/ijms21207571>

Pelegí-Sisó, D., de Prado, P., Ronkainen, J., Bustamante, M., & González, J. R. (2021). methylclock: A Bioconductor package to estimate DNA methylation age. *Bioinformatics (Oxford, England)*, 37(12), 1759-1760. <https://doi.org/10.1093/bioinformatics/btaa825>

Pérez, R. F., Santamarina, P., Tejedor, J. R., Urduño, R. G., Álvarez-Pitti, J., Redon, P., Fernández, A. F., Fraga, M. F., & Lurbe, E. (2019). Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes. *Journal of Translational Medicine*, *17*(1), 15. <https://doi.org/10.1186/s12967-018-1751-9>

Peters, W., & Matson, J. (2020). Comparing Rates of Diagnosis Using DSM-IV-TR Versus DSM-5 Criteria for Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *50*(6). <https://doi.org/10.1007/s10803-019-03941-1>

Petersen, B.-S., Fredrich, B., Hoepfner, M. P., Ellinghaus, D., & Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genetics*, *18*, 14. <https://doi.org/10.1186/s12863-017-0479-5>

Pham, C., Symeonides, C., O'Hely, M., Sly, P. D., Knibbs, L. D., Thomson, S., Vuillermin, P., Saffery, R., Ponsonby, A.-L., & Barwon Infant Study Investigator Group. (2022). Early life environmental factors associated with autism spectrum disorder symptoms in children at age 2 years: A birth cohort study. *Autism: The International Journal of Research and Practice*, *13623613211068224*. <https://doi.org/10.1177/13623613211068224>

Picci, G., Gotts, S. J., & Scherf, K. S. (2016). A theoretical rut: Revisiting and critically evaluating the generalized under/over-connectivity hypothesis of autism. *Developmental Science*, *19*(4), 524-549. <https://doi.org/10.1111/desc.12467>

Pidsley, R., Y Wong, C. C., Volta, M., Lunnon, K., Mill, J., & Schalkwyk, L. C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*, *14*, 293. <https://doi.org/10.1186/1471-2164-14-293>

Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., Van Dijk, S., Muhlhäuser, B., Stirzaker, C., & Clark, S. J. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology*, *17*, 208. <https://doi.org/10.1186/s13059-016-1066-1>

Pierce, K., Gazestani, V. H., Bacon, E., Barnes, C. C., Cha, D., Nalabolu, S., Lopez, L., Moore, A., Pence-Stophaeros, S., & Courchesne, E. (2019). Evaluation of the Diagnostic Stability of the Early Autism Spectrum Disorder Phenotype in the General Population Starting at 12 Months. *JAMA Pediatrics*, *173*(6), 578-587. <https://doi.org/10.1001/jamapediatrics.2019.0624>

Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W., & Albertson, D. G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genetics*, *20*(2), 207-211. <https://doi.org/10.1038/2524>

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T. R., Correia, C., Abrahams, B. S., Almeida, J., Bacchelli, E., Bader, G. D., Bailey, A. J., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Bölte, S., ... Betancur, C. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, *466*(7304), 368-372. <https://doi.org/10.1038/nature09146>

Pös, O., Radvanszky, J., Buglyó, G., Pös, Z., Rusnakova, D., Nagy, B., & Szemes, T. (2021). DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical Journal*, *44*(5), 548-559. <https://doi.org/10.1016/j.bj.2021.02.003>

Postema, M. C., van Rooij, D., Anagnostou, E., Arango, C., Auzias, G., Behrmann, M., Filho, G. B., Calderoni, S., Calvo, R., Daly, E., Deruelle, C., Di Martino, A., Dinstein, I., Duran, F. L. S., Durston, S., Ecker, C., Ehrlich, S., Fair, D., Fedor, J., ... Francks, C. (2019). Altered structural brain asymmetry in autism spectrum disorder in a study of 54 datasets. *Nature Communications*, *10*(1), 4958. <https://doi.org/10.1038/s41467-019-13005-8>

Power, R. A., Kyaga, S., Uher, R., MacCabe, J. H., Långström, N., Landen, M., McGuffin, P., Lewis, C. M., Lichtenstein, P., & Svensson, A. C. (2013). Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry*, *70*(1), 22-30. <https://doi.org/10.1001/jamapsychiatry.2013.268>

Pugsley, K., Scherer, S. W., Bellgrove, M. A., & Hawi, Z. (2022). Environmental exposures associated with elevated risk for autism spectrum disorder may augment the burden of deleterious de novo mutations among probands. *Molecular Psychiatry*, 27(1), 710-730. <https://doi.org/10.1038/s41380-021-01142-w>

Quenez, O., Cassinari, K., Coutant, S., Lecoquierre, F., Le Guennec, K., Rousseau, S., Richard, A.-C., Vasseur, S., Bouvignies, E., Bou, J., Lienard, G., Manase, S., Fourneaux, S., Drouot, N., Nguyen-Viet, V., Vezain, M., Chambon, P., Joly-Helas, G., Le Meur, N., ... Nicolas, G. (2021). Detection of copy-number variations from NGS data using read depth information: A diagnostic performance evaluation. *European Journal of Human Genetics: EJHG*, 29(1), 99-109. <https://doi.org/10.1038/s41431-020-0672-2>

Radio, F. C., Pang, K., Ciolfi, A., Levy, M. A., Hernández-García, A., Pedace, L., Pantaleoni, F., Liu, Z., de Boer, E., Jackson, A., Bruselles, A., McConkey, H., Stellacci, E., Lo Cicero, S., Motta, M., Carrozzo, R., Dentici, M. L., McWalter, K., Desai, M., ... Tartaglia, M. (2021). SPEN haploinsufficiency causes a neurodevelopmental disorder overlapping proximal 1p36 deletion syndrome with an epistatue of X chromosomes in females. *American Journal of Human Genetics*, 108(3), 502-516. <https://doi.org/10.1016/j.ajhg.2021.01.015>

Randall, M., Egberts, K. J., Samtani, A., Scholten, R. J., Hooft, L., Livingstone, N., Sterling-Levis, K., Woolfenden, S., & Williams, K. (2018). Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *The Cochrane Database of Systematic Reviews*, 7, CD009044. <https://doi.org/10.1002/14651858.CD009044.pub2>

Rane, P., Cochran, D., Hodge, S. M., Haselgrove, C., Kennedy, D., & Frazier, J. A. (2015). Connectivity in Autism: A review of MRI connectivity studies. *Harvard review of psychiatry*, 23(4), 223-244. <https://doi.org/10.1097/HRP.0000000000000072>

Redcay, E., & Courchesne, E. (2005). When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biological Psychiatry*, 58(1), 1-9. <https://doi.org/10.1016/j.biopsych.2005.03.026>

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015a). Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine: official journal of the American College of Medical Genetics*, 17(5), 405-424. <https://doi.org/10.1038/gim.2015.30>

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015b). Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine: official journal of the American College of Medical Genetics*, 17(5), 405-424. <https://doi.org/10.1038/gim.2015.30>

Riggs, A. D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenetics and Cell Genetics*, 14(1), 9-25. <https://doi.org/10.1159/000130315>

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., Cook, E. H., Leventhal, B. L., & Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45(9), 1094-1103. <https://doi.org/10.1097/01.chi.0000227880.42780.0e>

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47. <https://doi.org/10.1093/nar/gkv007>

Ritvo, E. R., Freeman, B. J., Mason-Brothers, A., Mo, A., & Ritvo, A. M. (1985). Concordance for the syndrome of autism in 40 pairs of afflicted twins. *The American Journal of Psychiatry*, 142(1), 74-77. <https://doi.org/10.1176/ajp.142.1.74>

Robinson, E. B., St. Pourcain, B., Anttila, V., Kosmicki, J. A., Bulik-Sullivan, B., Grove, J., Maller, J., Samocha, K. E., Sanders, S. J., Ripke, S., Martin, J., Hollegaard, M. V., Werge, T., Hougaard, D. M., Neale, B. M., Evans, D. M., Skuse, D., Mortensen, P. B., Børglum, A. D., ... Daly, M. J. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature genetics*, 48(5), 552-555. <https://doi.org/10.1038/ng.3529>

- Roesser, J. (2011). Diagnostic yield of genetic testing in children diagnosed with autism spectrum disorders at a regional referral center. *Clinical Pediatrics*, 50(9), 834-843. <https://doi.org/10.1177/0009922811406261>
- Ronemus, M., Iossifov, I., Levy, D., & Wigler, M. (2014). The role of de novo mutations in the genetics of autism spectrum disorders. *Nature Reviews. Genetics*, 15(2), 133-141. <https://doi.org/10.1038/nrg3585>
- Rooney, K., Levy, M. A., Haghshenas, S., Kerkhof, J., Rogaia, D., Tedesco, M. G., Imperatore, V., Mencarelli, A., Squeo, G. M., Di Venere, E., Di Cara, G., Verrotti, A., Merla, G., Tedder, M. L., DuPont, B. R., Sadikovic, B., & Prontera, P. (2021). Identification of a DNA Methylation Episignature in the 22q11.2 Deletion Syndrome. *International Journal of Molecular Sciences*, 22(16), 8611. <https://doi.org/10.3390/ijms22168611>
- Rosenberg, R. E., Law, J. K., Yenokyan, G., McGready, J., Kaufmann, W. E., & Law, P. A. (2009). Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Archives of Pediatrics & Adolescent Medicine*, 163(10), 907-914. <https://doi.org/10.1001/archpediatrics.2009.98>
- Rosenfeld, J. A., Coe, B. P., Eichler, E. E., Cuckle, H., & Shaffer, L. G. (2013). Estimates of penetrance for recurrent pathogenic copy-number variations. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 15(6), 478-481. <https://doi.org/10.1038/gim.2012.164>
- Ross, J. P., van Dijk, S., Phang, M., Skilton, M. R., Molloy, P. L., & Oytam, Y. (2022). Batch-effect detection, correction and characterisation in Illumina HumanMethylation450 and MethylationEPIC BeadChip array data. *Clinical Epigenetics*, 14, 58. <https://doi.org/10.1186/s13148-022-01277-9>
- Rouxel, F., Relator, R., Kerkhof, J., McConkey, H., Levy, M., Dias, P., Barat-Houari, M., Bednarek, N., Boute, O., Chatron, N., Cherik, F., Delahaye-Duriez, A., Doco-Fenzy, M., Faivre, L., Gauthier, L. W., Heron, D., Hildebrand, M. S., Lesca, G., Lespinasse, J., ... Genevieve, D. (2022). CDK13-related disorder: Report of a series of 18 previously unpublished individuals and description of an epigenetic signature. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 24(5), 1096-1107. <https://doi.org/10.1016/j.gim.2021.12.016>
- Sacco, R., Gabriele, S., & Persico, A. M. (2015). Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. *Psychiatry Research*, 234(2), 239-251. <https://doi.org/10.1016/j.psychres.2015.08.016>
- Sadikovic, B., Aref-Eshghi, E., Levy, M. A., & Rodenhiser, D. (2019). DNA methylation signatures in mendelian developmental disorders as a diagnostic bridge between genotype and phenotype. *Epigenomics*, 11(5), 563-575. <https://doi.org/10.2217/epi-2018-0192>
- Sadikovic, B., Levy, M. A., & Aref-Eshghi, E. (2020). Functional annotation of genomic variation: DNA methylation episignatures in neurodevelopmental Mendelian disorders. *Human Molecular Genetics*, 29(R1), R27-R32. <https://doi.org/10.1093/hmg/ddaa144>
- Sadikovic, B., Levy, M. A., Kerkhof, J., Aref-Eshghi, E., Schenkel, L., Stuart, A., McConkey, H., Henneman, P., Venema, A., Schwartz, C. E., Stevenson, R. E., Skinner, S. A., DuPont, B. R., Fletcher, R. S., Balci, T. B., Siu, V. M., Granadillo, J. L., Masters, J., Kadour, M., ... Alders, M. (2021). Clinical epigenomics: Genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 23(6), 1065-1074. <https://doi.org/10.1038/s41436-020-01096-4>
- Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., Kosmicki, J. A., Rehnström, K., Mallick, S., Kirby, A., Wall, D. P., MacArthur, D. G., Gabriel, S. B., DePristo, M., Purcell, S. M., Palotie, A., Boerwinkle, E., Buxbaum, J. D., Cook, E. H., ... Daly, M. J. (2014). A framework for the interpretation of de novo mutation in human disease. *Nature Genetics*, 46(9), 944-950. <https://doi.org/10.1038/ng.3050>
- Sanchez, C. E., Barry, C., Sabhlok, A., Russell, K., Majors, A., Kollins, S. H., & Fuemmeler, B. F. (2018). Maternal pre-pregnancy obesity and child neurodevelopmental outcomes: A Meta-analysis. *Obesity reviews: an official journal of the International Association for the Study of Obesity*, 19(4), 464-484. <https://doi.org/10.1111/obr.12643>
- Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., Murtha, M. T., Bal, V. H., Bishop, S. L., Dong, S., Goldberg, A. P., Jinlu, C., Keane, J. F., Klei, L., Mandell, J. D., Moreno-De-

Luca, D., Poultney, C. S., Robinson, E. B., Smith, L., ... State, M. W. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, 87(6), 1215-1233. <https://doi.org/10.1016/j.neuron.2015.09.016>

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., Walker, M. F., Ober, G. T., Teran, N. A., Song, Y., El-Fishawy, P., Murtha, R. C., Choi, M., Overton, J. D., Bjornson, R. D., ... State, M. W. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485(7397), 237-241. <https://doi.org/10.1038/nature10945>

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *JAMA*, 311(17), 1770-1777. <https://doi.org/10.1001/jama.2014.4144>

Santos, M., Uppal, N., Butti, C., Wicinski, B., Schmeidler, J., Giannakopoulos, P., Heinsen, H., Schmitz, C., & Hof, P. R. (2011). Von Economo neurons in autism: A stereologic study of the frontoinsular cortex in children. *Brain Research*, 1380, 206-217. <https://doi.org/10.1016/j.brainres.2010.08.067>

Sato, W., & Uono, S. (2019). The atypical social brain network in autism: Advances in structural and functional MRI studies. *Current Opinion in Neurology*, 32(4), 617-621. <https://doi.org/10.1097/WCO.0000000000000713>

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., ... Buxbaum, J. D. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, 180(3), 568-584.e23. <https://doi.org/10.1016/j.cell.2019.12.036>

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., ... Buxbaum, J. D. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, 180(3), 568-584.e23. <https://doi.org/10.1016/j.cell.2019.12.036>

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), 1412-1417. <https://doi.org/10.1073/pnas.0510310103>

Schenkel, L. C., Aref-Eshghi, E., Rooney, K., Kerkhof, J., Levy, M. A., McConkey, H., Rogers, R. C., Phelan, K., Sarasua, S. M., Jain, L., Pauly, R., Boccutto, L., DuPont, B., Cappuccio, G., Brunetti-Pierri, N., Schwartz, C. E., & Sadikovic, B. (2021). DNA methylation epi-signature is associated with two molecularly and phenotypically distinct clinical subtypes of Phelan-McDermid syndrome. *Clinical Epigenetics*, 13(1), 2. <https://doi.org/10.1186/s13148-020-00990-7>

Schenkel, L. C., Kernohan, K. D., McBride, A., Reina, D., Hodge, A., Ainsworth, P. J., Rodenhiser, D. I., Pare, G., Bérubé, N. G., Skinner, C., Boycott, K. M., Schwartz, C., & Sadikovic, B. (2017). Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome. *Epigenetics & Chromatin*, 10, 10. <https://doi.org/10.1186/s13072-017-0118-4>

Schumann, C. M., Barnes, C. C., Lord, C., & Courchesne, E. (2009). Amygdala enlargement in toddlers with autism related to severity of social and communication impairments. *Biological Psychiatry*, 66(10), 942-949. <https://doi.org/10.1016/j.biopsych.2009.07.007>

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.-H., Hicks, J., Spence, S. J., Lee, A. T., Puura, K., Lehtimäki, T., ... Wigler, M. (2007). Strong association of de novo copy number mutations with autism. *Science (New York, N.Y.)*, 316(5823), 445-449. <https://doi.org/10.1126/science.1138659>

Shattuck, P. T. (2006). The contribution of diagnostic substitution to the growing administrative prevalence of autism in US special education. *Pediatrics*, 117(4), 1028-1037. <https://doi.org/10.1542/peds.2005-1516>

Shen, M. D., Nordahl, C. W., Young, G. S., Wootton-Gorges, S. L., Lee, A., Liston, S. E., Harrington, K. R., Ozonoff, S., & Amaral, D. G. (2013). Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. *Brain: A Journal of Neurology*, *136*(Pt 9), 2825-2835. <https://doi.org/10.1093/brain/awt166>

Shen, Y., Dies, K. A., Holm, I. A., Bridgemohan, C., Sobeih, M. M., Caronna, E. B., Miller, K. J., Frazier, J. A., Silverstein, I., Picker, J., Weissman, L., Raffalli, P., Jeste, S., Demmer, L. A., Peters, H. K., Brewster, S. J., Kowalczyk, S. J., Rosen-Sheidley, B., McGowan, C., ... Autism Consortium Clinical Genetics/DNA Diagnostics Collaboration. (2010). Clinical genetic testing for patients with autism spectrum disorders. *Pediatrics*, *125*(4), e727-735. <https://doi.org/10.1542/peds.2009-1684>

Shenouda, J., Barrett, E., Davidow, A. L., Halperin, W., Silenzio, V. M. B., & Zahorodny, W. (2022). Prevalence of autism spectrum disorder in a large, diverse metropolitan area: Variation by sociodemographic factors. *Autism Research: Official Journal of the International Society for Autism Research*, *15*(1), 146-155. <https://doi.org/10.1002/aur.2628>

Shiah, Y.-J., Fraser, M., Bristow, R. G., & Boutros, P. C. (2017). Comparison of pre-processing methods for Infinium HumanMethylation450 BeadChip array. *Bioinformatics (Oxford, England)*, *33*(20), 3151-3157. <https://doi.org/10.1093/bioinformatics/btx372>

Shuster, J., A, P., J, B., & Me, T. (2014). Review of factor analytic studies examining symptoms of autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(1). <https://doi.org/10.1007/s10803-013-1854-3>

Simonoff, E., Kent, R., Stringer, D., Lord, C., Briskman, J., Lukito, S., Pickles, A., Charman, T., & Baird, G. (2020). Trajectories in Symptoms of Autism and Cognitive Ability in Autism From Childhood to Adult Life: Findings From a Longitudinal Epidemiological Cohort. *Journal of the American Academy of Child and Adolescent Psychiatry*, *59*(12), 1342-1352. <https://doi.org/10.1016/j.jaac.2019.11.020>

Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J. D., Bass, N., Bigdeli, T. B., Breen, G., Bromet, E. J., Buckley, P. F., Bunney, W. E., Bybjerg-Grauholm, J., Byerley, W. F., Chapman, S. B., Chen, W. J., Churchhouse, C., Craddock, N., Cusick, C. M., ... Daly, M. J. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*, *604*(7906), Art. 7906. <https://doi.org/10.1038/s41586-022-04556-w>

Singh, T., Walters, J. T. R., Johnstone, M., Curtis, D., Suvisaari, J., Torniainen, M., Rees, E., Iyegbe, C., Blackwood, D., McIntosh, A. M., Kirov, G., Geschwind, D., Murray, R. M., Di Forti, M., Bramon, E., Gandal, M., Hultman, C. M., Sklar, P., Palotie, A., ... Barrett, J. C. (2017). The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nature Genetics*, *49*(8), Art. 8. <https://doi.org/10.1038/ng.3903>

Siu, M. T., Butcher, D. T., Turinsky, A. L., Cytrynbaum, C., Stavropoulos, D. J., Walker, S., Caluseriu, O., Carter, M., Lou, Y., Nicolson, R., Georgiades, S., Szatmari, P., Anagnostou, E., Scherer, S. W., Choufani, S., Brudno, M., & Weksberg, R. (2019). Functional DNA methylation signatures for autism spectrum disorder genomic risk loci: 16p11.2 deletions and CHD8 variants. *Clinical Epigenetics*, *11*(1), 103. <https://doi.org/10.1186/s13148-019-0684-3>

Siu, M. T., & Weksberg, R. (2017). Epigenetics of Autism Spectrum Disorder. *Advances in Experimental Medicine and Biology*, *978*, 63-90. https://doi.org/10.1007/978-3-319-53889-1_4

Smith, I. C., Reichow, B., & Volkmar, F. R. (2015). The Effects of DSM-5 Criteria on Number of Individuals Diagnosed with Autism Spectrum Disorder: A Systematic Review. *Journal of Autism and Developmental Disorders*, *45*(8), 2541-2552. <https://doi.org/10.1007/s10803-015-2423-8>

Solomon, O., MacIsaac, J., Quach, H., Tindula, G., Kobor, M. S., Huen, K., Meaney, M. J., Eskenazi, B., Barcellos, L. F., & Holland, N. (2018). Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-old children. *Epigenetics*, *13*(6), 655-664. <https://doi.org/10.1080/15592294.2018.1497386>

Squarcina, L., Nosari, G., Marin, R., Castellani, U., Bellani, M., Bonivento, C., Fabbro, F., Molteni, M., & Brambilla, P. (2021). Automatic classification of autism spectrum disorder in children using cortical thickness and support vector machine. *Brain and Behavior*, *11*(8), e2238. <https://doi.org/10.1002/brb3.2238>

- Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I. C., Jakobsson, G., & Bohman, M. (1989). A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 30(3), 405-416. <https://doi.org/10.1111/j.1469-7610.1989.tb00254.x>
- Stessman, H. A. F., Xiong, B., Coe, B. P., Wang, T., Hoekzema, K., Fenckova, M., Kvarnung, M., Gerdts, J., Trinh, S., Cosemans, N., Vives, L., Lin, J., Turner, T. N., Santen, G., Ruivenkamp, C., Kriek, M., van Haeringen, A., Aten, E., Friend, K., ... Eichler, E. E. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nature Genetics*, 49(4), 515-526. <https://doi.org/10.1038/ng.3792>
- Stoner, R., Chow, M. L., Boyle, M. P., Sunkin, S. M., Mouton, P. R., Roy, S., Wynshaw-Boris, A., Colamarino, S. A., Lein, E. S., & Courchesne, E. (2014). Patches of disorganization in the neocortex of children with autism. *The New England Journal of Medicine*, 370(13), 1209-1219. <https://doi.org/10.1056/NEJMoa1307491>
- Strong, E., Butcher, D. T., Singhanian, R., Mervis, C. B., Morris, C. A., De Carvalho, D., Weksberg, R., & Osborne, L. R. (2015). Symmetrical Dose-Dependent DNA-Methylation Profiles in Children with Deletion or Duplication of 7q11.23. *American Journal of Human Genetics*, 97(2), 216-227. <https://doi.org/10.1016/j.ajhg.2015.05.019>
- Sun, X., Allison, C., Wei, L., Matthews, F. E., Auyeung, B., Wu, Y. Y., Griffiths, S., Zhang, J., Baron-Cohen, S., & Brayne, C. (2019). Autism prevalence in China is comparable to Western prevalence. *Molecular Autism*, 10, 7. <https://doi.org/10.1186/s13229-018-0246-0>
- Supek, F., Lehner, B., & Lindeboom, R. G. H. (2021). To NMD or Not To NMD: Nonsense-Mediated mRNA Decay in Cancer and Other Genetic Diseases. *Trends in Genetics: TIG*, 37(7), 657-668. <https://doi.org/10.1016/j.tig.2020.11.002>
- Szatmari, P., Chawarska, K., Dawson, G., Georgiades, S., Landa, R., Lord, C., Messinger, D. S., Thurm, A., & Halladay, A. (2016). Prospective Longitudinal Studies of Infant Siblings of Children With Autism: Lessons Learned and Future Directions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55(3), 179-187. <https://doi.org/10.1016/j.jaac.2015.12.014>
- Talkowski, M. E., Mullegama, S. V., Rosenfeld, J. A., van Bon, B. W. M., Shen, Y., Repnikova, E. A., Gastier-Foster, J., Thrush, D. L., Kathiresan, S., Ruderfer, D. M., Chiang, C., Hanscom, C., Ernst, C., Lindgren, A. M., Morton, C. C., An, Y., Astbury, C., Brueton, L. A., Lichtenbelt, K. D., ... Elsea, S. H. (2011). Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. *American Journal of Human Genetics*, 89(4), 551-563. <https://doi.org/10.1016/j.ajhg.2011.09.011>
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews. Genetics*, 20(8), 467-484. <https://doi.org/10.1038/s41576-019-0127-1>
- Tammimies, K., Marshall, C. R., Walker, S., Kaur, G., Thiruvahindrapuram, B., Lionel, A. C., Yuen, R. K. C., Uddin, M., Roberts, W., Weksberg, R., Woodbury-Smith, M., Zwaigenbaum, L., Anagnostou, E., Wang, Z., Wei, J., Howe, J. L., Gazzellone, M. J., Lau, L., Sung, W. W. L., ... Fernandez, B. A. (2015). Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. *JAMA*, 314(9), 895-903. <https://doi.org/10.1001/jama.2015.10078>
- Taniai, H., Nishiyama, T., Miyachi, T., Imaeda, M., & Sumi, S. (2008). Genetic influences on the broad spectrum of autism: Study of proband-ascertained twins. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 147B(6), 844-849. <https://doi.org/10.1002/ajmg.b.30740>
- Taylor, L. E., Swerdfeger, A. L., & Eslick, G. D. (2014). Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies. *Vaccine*, 32(29), 3623-3629. <https://doi.org/10.1016/j.vaccine.2014.04.085>
- Tenesa, A., & Haley, C. S. (2013). The heritability of human disease: Estimation, uses and abuses. *Nature Reviews. Genetics*, 14(2), 139-149. <https://doi.org/10.1038/nrg3377>

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijdsdijk, F. (2016). Heritability of autism spectrum disorders : A meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 57(5), 585-595. <https://doi.org/10.1111/jcpp.12499>

Turner, L. M., & Stone, W. L. (2007). Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 48(8), 793-802. <https://doi.org/10.1111/j.1469-7610.2007.01744.x>

Turner, T. N., Coe, B. P., Dickel, D. E., Hoekzema, K., Nelson, B. J., Zody, M. C., Kronenberg, Z. N., Hormozdiari, F., Raja, A., Pennacchio, L. A., Darnell, R. B., & Eichler, E. E. (2017). Genomic Patterns of De Novo Mutation in Simplex Autism. *Cell*, 171(3), 710-722.e12. <https://doi.org/10.1016/j.cell.2017.08.047>

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, 16(1), 55-61. <https://doi.org/10.1038/nrn3857>

Urduingio, R. G., Torró, M. I., Bayón, G. F., Álvarez-Pitti, J., Fernández, A. F., Redon, P., Fraga, M. F., & Lurbe, E. (2016). Longitudinal study of DNA methylation during the first 5 years of life. *Journal of Translational Medicine*, 14(1), 160. <https://doi.org/10.1186/s12967-016-0913-x>

van Bon, B. W. M., Koolen, D. A., Brueton, L., McMullan, D., Lichtenbelt, K. D., Adès, L. C., Peters, G., Gibson, K., Moloney, S., Novara, F., Pramparo, T., Dalla Bernardina, B., Zocante, L., Balottin, U., Piazza, F., Pecile, V., Gasparini, P., Guerci, V., Kets, M., ... de Vries, B. B. A. (2010). The 2q23.1 microdeletion syndrome : Clinical and behavioural phenotype. *European Journal of Human Genetics: EJHG*, 18(2), 163-170. <https://doi.org/10.1038/ejhg.2009.152>

van Rooij, D., Anagnostou, E., Arango, C., Auzias, G., Behrmann, M., Busatto, G. F., Calderoni, S., Daly, E., Deruelle, C., Di Martino, A., Dinstein, I., Duran, F. L. S., Durston, S., Ecker, C., Fair, D., Fedor, J., Fitzgerald, J., Freitag, C. M., Gallagher, L., ... Buitelaar, J. K. (2018). Cortical and Subcortical Brain Morphometry Differences Between Patients With Autism Spectrum Disorder and Healthy Individuals Across the Lifespan : Results From the ENIGMA ASD Working Group. *The American Journal of Psychiatry*, 175(4), 359-369. <https://doi.org/10.1176/appi.ajp.2017.17010100>

Varley, K. E., Gertz, J., Bowling, K. M., Parker, S. L., Reddy, T. E., Pauli-Behn, F., Cross, M. K., Williams, B. A., Stamatoyannopoulos, J. A., Crawford, G. E., Absher, D. M., Wold, B. J., & Myers, R. M. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Research*, 23(3), 555-567. <https://doi.org/10.1101/gr.147942.112>

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., ... Zhu, X. (2001). The Sequence of the Human Genome. *Science*, 291(5507), 1304-1351. <https://doi.org/10.1126/science.1058040>

Verdugo, R. A., Deschepper, C. F., Muñoz, G., Pomp, D., & Churchill, G. A. (2009). Importance of randomization in microarray experimental designs with Illumina platforms. *Nucleic Acids Research*, 37(17), 5610-5618. <https://doi.org/10.1093/nar/gkp573>

Verhoeven, W., Egger, J., Kipp, J., Verheul-Aan de Wiel, J., Ockeloen, C., Kleefstra, T., & Pfundt, R. (2019). A novel MBD5 mutation in an intellectually disabled adult female patient with epilepsy : Suggestive of early onset dementia? *Molecular Genetics & Genomic Medicine*, 7(8), e849. <https://doi.org/10.1002/mgg3.849>

Veroniki, A. A., Rios, P., Cogo, E., Straus, S. E., Finkelstein, Y., Kealey, R., Reynen, E., Soobiah, C., Thavorn, K., Hutton, B., Hemmelgarn, B. R., Yazdi, F., D'Souza, J., MacDonald, H., & Tricco, A. C. (2017). Comparative safety of antiepileptic drugs for neurological development in children exposed during pregnancy and breast feeding : A systematic review and network meta-analysis. *BMJ Open*, 7(7), e017248. <https://doi.org/10.1136/bmjopen-2017-017248>

Vidal, S., Brandi, N., Pacheco, P., Maynou, J., Fernandez, G., Xiol, C., Pascual-Alonso, A., Pineda, M., Rett Working Group, & Armstrong, J. (2019). The most recurrent monogenic disorders that overlap with the phenotype of Rett syndrome. *European Journal of Paediatric Neurology: EJPN: Official Journal of the European Paediatric Neurology Society*, 23(4), 609-620. <https://doi.org/10.1016/j.ejpn.2019.04.006>

Vissers, L. E. L. M., Bhatt, S. S., Janssen, I. M., Xia, Z., Lalani, S. R., Pfundt, R., Derwinska, K., de Vries, B. B. A., Gilissen, C., Hoischen, A., Nesteruk, M., Wisniewiecka-Kowalik, B., Smyk, M., Brunner, H. G., Cheung, S. W., van Kessel, A. G., Veltman, J. A., & Stankiewicz, P. (2009). Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Human Molecular Genetics*, 18(19), 3579-3593. <https://doi.org/10.1093/hmg/ddp306>

Vogel, A. C., Gutmann, D. H., & Morris, S. M. (2017). Neurodevelopmental disorders in children with neurofibromatosis type 1. *Developmental Medicine and Child Neurology*, 59(11), 1112-1116. <https://doi.org/10.1111/dmnc.13526>

von Ehrenstein, O. S., Ling, C., Cui, X., Cockburn, M., Park, A. S., Yu, F., Wu, J., & Ritz, B. (2019). Prenatal and infant exposure to ambient pesticides and autism spectrum disorder in children: Population based case-control study. *The BMJ*, 364, 1962. <https://doi.org/10.1136/bmj.1962>

Wadhawan, I., Hai, Y., Foyouzi Yousefi, N., Guo, X., Graham, J. M., & Rosenfeld, J. A. (2020). De novo copy number variants and parental age: Is there an association? *European Journal of Medical Genetics*, 63(4), 103829. <https://doi.org/10.1016/j.ejmg.2019.103829>

Walz, K., & Young, J. I. (2014). The methyl binding domain containing protein MBD5 is a transcriptional regulator responsible for 2q23.1 deletion syndrome. *Rare Diseases (Austin, Tex.)*, 2(1), e967151. <https://doi.org/10.4161/2167549X.2014.967151>

Wan, H., Zhang, C., Li, H., Luan, S., & Liu, C. (2018). Association of maternal diabetes with autism spectrum disorders in offspring: A systemic review and meta-analysis. *Medicine*, 97(2), e9438. <https://doi.org/10.1097/MD.0000000000009438>

Wang, B., Zeng, H., Liu, J., & Sun, M. (2021). Effects of Prenatal Hypoxia on Nervous System Development and Related Diseases. *Frontiers in Neuroscience*, 15, 755554. <https://doi.org/10.3389/fnins.2021.755554>

Wang, C., Geng, H., Liu, W., & Zhang, G. (2017). Prenatal, perinatal, and postnatal factors associated with autism: A meta-analysis. *Medicine*, 96(18), e6696. <https://doi.org/10.1097/MD.0000000000006696>

Wang, D., Liu, X., Zhou, Y., Xie, H., Hong, X., Tsai, H.-J., Wang, G., Liu, R., & Wang, X. (2012). Individual variation and longitudinal pattern of genome-wide DNA methylation from birth to the first two years of life. *Epigenetics*, 7(6), 594-605. <https://doi.org/10.4161/epi.20117>

Wang, H., Ma, Z.-H., Xu, L.-Z., Yang, L., Ji, Z.-Z., Tang, X.-Z., Liu, J.-R., Li, X., Cao, Q.-J., & Liu, J. (2022). Developmental brain structural atypicalities in autism: A voxel-based morphometry analysis. *Child and Adolescent Psychiatry and Mental Health*, 16(1), 7. <https://doi.org/10.1186/s13034-022-00443-4>

Wang, T., Guo, H., Xiong, B., Stessman, H. A. F., Wu, H., Coe, B. P., Turner, T. N., Liu, Y., Zhao, W., Hoekzema, K., Vives, L., Xia, L., Tang, M., Ou, J., Chen, B., Shen, Y., Xun, G., Long, M., Lin, J., ... Eichler, E. E. (2016). De novo genic mutations among a Chinese autism spectrum disorder cohort. *Nature Communications*, 7, 13316. <https://doi.org/10.1038/ncomms13316>

Warrier, V., Toro, R., Won, H., Leblond, C. S., Cliquet, F., Delorme, R., De Witte, W., Bralten, J., Chakrabarti, B., Børghlum, A. D., Grove, J., Poelmans, G., Hinds, D. A., Bourgeron, T., & Baron-Cohen, S. (2019). Social and non-social autism symptoms and trait domains are genetically dissociable. *Communications Biology*, 2, 328. <https://doi.org/10.1038/s42003-019-0558-4>

Warrier, V., Zhang, X., Reed, P., Havdahl, A., Moore, T. M., Cliquet, F., Leblond, C. S., Rolland, T., Rosengren, A., Rowitch, D. H., Hurles, M. E., Geschwind, D. H., Børghlum, A. D., Robinson, E. B., Grove, J., Martin, H. C., Bourgeron, T., & Baron-Cohen, S. (2022). Genetic correlates of phenotypic heterogeneity in autism. *Nature Genetics*, 1-12. <https://doi.org/10.1038/s41588-022-01072-5>

Wegiel, J., Kuchna, I., Nowicki, K., Imaki, H., Wegiel, J., Marchi, E., Ma, S. Y., Chauhan, A., Chauhan, V., Bobrowicz, T. W., de Leon, M., Louis, L. A. S., Cohen, I. L., London, E., Brown, W. T., & Wisniewski, T. (2010). The neuropathology of autism: Defects of neurogenesis and neuronal migration, and dysplastic changes. *Acta Neuropathologica*, 119(6), 755-770. <https://doi.org/10.1007/s00401-010-0655-4>

Wei, S., Tao, J., Xu, J., Chen, X., Wang, Z., Zhang, N., Zuo, L., Jia, Z., Chen, H., Sun, H., Yan, Y., Zhang, M., Lv, H., Kong, F., Duan, L., Ma, Y., Liao, M., Xu, L., Feng, R., ... Jiang, Y. (2021). Ten Years of EWAS. *Advanced Science* (Weinheim, Baden-Wuerttemberg, Germany), 8(20), e2100727. <https://doi.org/10.1002/adv.202100727>

Weiner, D. J., Wigdor, E. M., Ripke, S., Walters, R. K., Kosmicki, J. A., Grove, J., Samocha, K. E., Goldstein, J. I., Okbay, A., Bybjerg-Grauholm, J., Werge, T., Hougaard, D. M., Taylor, J., iPSYCH-Broad Autism Group, Psychiatric Genomics Consortium Autism Group, Skuse, D., Devlin, B., Anney, R., Sanders, S. J., ... Robinson, E. B. (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nature Genetics*, 49(7), 978-985. <https://doi.org/10.1038/ng.3863>

Werling, D. M., Brand, H., An, J.-Y., Stone, M. R., Zhu, L., Glessner, J. T., Collins, R. L., Dong, S., Layer, R. M., Markenscoff-Papadimitriou, E., Farrell, A., Schwartz, G. B., Wang, H. Z., Currall, B. B., Zhao, X., Dea, J., Duhn, C., Erdman, C. A., Gilson, M. C., ... Sanders, S. J. (2018). An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nature Genetics*, 50(5), 727-736. <https://doi.org/10.1038/s41588-018-0107-y>

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377-399. <https://doi.org/10.1002/sim.4067>

Wieckowski, A. T., Hamner, T., Nanovic, S., Porto, K. S., Coulter, K. L., Eldeeb, S. Y., Chen, C.-M. A., Fein, D. A., Barton, M. L., Adamson, L. B., & Robins, D. L. (2021). Early and Repeated Screening Detects Autism Spectrum Disorder. *The Journal of Pediatrics*, 234, 227-235. <https://doi.org/10.1016/j.jpeds.2021.03.009>

Wigdor, E. M., Weiner, D. J., Grove, J., Fu, J. M., Thompson, W. K., Carey, C. E., Baya, N., Merwe, C. van der, Walters, R. K., Satterstrom, F. K., Palmer, D. S., Rosengren, A., Bybjerg-Grauholm, J., Hougaard, D. M., Mortensen, P. B., Daly, M. J., Talkowski, M. E., Sanders, S. J., Bishop, S. L., ... Robinson, E. B. (2022). The female protective effect against autism spectrum disorder. *Cell Genomics*, 2(6). <https://doi.org/10.1016/j.xgen.2022.100134>

Wiggins, L. D., Rice, C. E., Barger, B., Soke, G. N., Lee, L.-C., Moody, E., Edmondson-Pretzel, R., & Levy, S. E. (2019). DSM-5 criteria for autism spectrum disorder maximizes diagnostic sensitivity and specificity in preschool children. *Social Psychiatry and Psychiatric Epidemiology*, 54(6), 693-701. <https://doi.org/10.1007/s00127-019-01674-1>

Williams, J. G., Higgins, J. P. T., & Brayne, C. E. G. (2006). Systematic review of prevalence studies of autism spectrum disorders. *Archives of Disease in Childhood*, 91(1), 8-15. <https://doi.org/10.1136/adc.2004.062083>

Wing, L. (1981). Asperger's syndrome: A clinical account. *Psychological Medicine*, 11(1), 115-129. <https://doi.org/10.1017/s0033291700053332>

Wolff, M., Johannesen, K. M., Hedrich, U. B. S., Masnada, S., Rubboli, G., Gardella, E., Lesca, G., Ville, D., Milh, M., Villard, L., Afenjar, A., Chantot-Bastaraud, S., Mignot, C., Lardenois, C., Nava, C., Schwarz, N., Gérard, M., Perrin, L., Doummar, D., ... Møller, R. S. (2017). Genetic and phenotypic heterogeneity suggest therapeutic implications in SCN2A-related disorders. *Brain: A Journal of Neurology*, 140(5), 1316-1336. <https://doi.org/10.1093/brain/awx054>

Woods, N. T., Baskin, R., Golubeva, V., Jhuraney, A., De-Gregoriis, G., Vaclova, T., Goldgar, D. E., Couch, F. J., Carvalho, M. A., Iversen, E. S., & Monteiro, A. N. (2016). Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. *NPJ Genomic Medicine*, 1, 16001. <https://doi.org/10.1038/npjgenmed.2016.1>

Xu, D., Wang, C., Kiryluk, K., Buxbaum, J. D., & Ionita-Laza, I. (2020). Co-localization between Sequence Constraint and Epigenomic Information Improves Interpretation of Whole-Genome Sequencing Data. *American Journal of Human Genetics*, 106(4), 513-524. <https://doi.org/10.1016/j.ajhg.2020.03.003>

Yankowitz, L. D., Herrington, J. D., Yerys, B. E., Pereira, J. A., Pandey, J., & Schultz, R. T. (2020). Evidence against the « normalization » prediction of the early brain overgrowth hypothesis of autism. *Molecular Autism*, 11(1), 51. <https://doi.org/10.1186/s13229-020-00353-2>

- Yao, S., Zhou, M., Zhang, Y., Zhou, F., Zhang, Q., Zhao, Z., Jiang, X., Xu, X., Becker, B., & Kendrick, K. M. (2021). Decreased homotopic interhemispheric functional connectivity in children with autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *14*(8), 1609-1620. <https://doi.org/10.1002/aur.2523>
- Yerys, B. E., Gordon, E. M., Abrams, D. N., Satterthwaite, T. D., Weinblatt, R., Jankowski, K. F., Strang, J., Kenworthy, L., Gaillard, W. D., & Vaidya, C. J. (2015). Default mode network segregation and social deficits in autism spectrum disorder: Evidence from non-medicated children. *NeuroImage. Clinical*, *9*, 223-232. <https://doi.org/10.1016/j.nicl.2015.07.018>
- Yoon, S., Munoz, A., Yamrom, B., Lee, Y., Andrews, P., Marks, S., Wang, Z., Reeves, C., Winterkorn, L., Krieger, A. M., Buja, A., Pradhan, K., Ronemus, M., Baldwin, K. K., Levy, D., Wigler, M., & Iossifov, I. (2021). Rates of contributory de novo mutation in high and low-risk autism families. *Communications Biology*, *4*, 1026. <https://doi.org/10.1038/s42003-021-02533-z>
- Young, H., Oreve, M.-J., & Speranza, M. (2018). Clinical characteristics and problems diagnosing autism spectrum disorder in girls. *Archives De Pediatrie: Organe Officiel De La Societe Francaise De Pediatrie*, *25*(6), 399-403. <https://doi.org/10.1016/j.arcped.2018.06.008>
- Yousefi, P., Huen, K., Davé, V., Barcellos, L., Eskenazi, B., & Holland, N. (2015). Sex differences in DNA methylation assessed by 450 K BeadChip in newborns. *BMC Genomics*, *16*(1), 911. <https://doi.org/10.1186/s12864-015-2034-y>
- Zeidan, J., Fombonne, E., Scorch, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, *15*(5), 778-790. <https://doi.org/10.1002/aur.2696>
- Zhang, X. (2014). Exome sequencing greatly expedites the progressive research of Mendelian diseases. *Frontiers of Medicine*, *8*(1). <https://doi.org/10.1007/s11684-014-0303-9>
- Zhou, J., Park, C. Y., Theesfeld, C. L., Wong, A. K., Yuan, Y., Scheckel, C., Fak, J. J., Funk, J., Yao, K., Tajima, Y., Packer, A., Darnell, R. B., & Troyanskaya, O. G. (2019). Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. *Nature Genetics*, *51*(6), 973-980. <https://doi.org/10.1038/s41588-019-0420-0>
- Zhou, W., Laird, P. W., & Shen, H. (2017). Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Research*, *45*(4), e22. <https://doi.org/10.1093/nar/gkw967>
- Zhou, X., Feliciano, P., Shu, C., Wang, T., Astrovskaya, I., Hall, J. B., Obiajulu, J. U., Wright, J. R., Murali, S. C., Xu, S. X., Brueggeman, L., Thomas, T. R., Marchenko, O., Fleisch, C., Barns, S. D., Snyder, L. G., Han, B., Chang, T. S., Turner, T. N., ... Chung, W. K. (2022). Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nature Genetics*, *54*(9), Art. 9. <https://doi.org/10.1038/s41588-022-01148-2>
- Zhuang, J., Widschwendter, M., & Teschendorff, A. E. (2012). A comparison of feature selection and classification methods in DNA methylation studies using the Illumina Infinium platform. *BMC Bioinformatics*, *13*, 59. <https://doi.org/10.1186/1471-2105-13-59>
- Zindler, T., Frieling, H., Neyazi, A., Bleich, S., & Friedel, E. (2020). Simulating ComBat: How batch correction can lead to the systematic introduction of false positive results in DNA methylation microarray studies. *BMC Bioinformatics*, *21*(1), 271. <https://doi.org/10.1186/s12859-020-03559-6>

Annexe 1. Ensemble des syndromes pour lesquels une épisignature est décrite dans la littérature. Tiré de (M. A. Levy, McConkey, et al., 2022) ; *épisignature différente selon la localisation de la variation sur le gène.

Syndrome	Gène ou région d'intérêt	OMIM	N échantillons
X-linked alpha-thalassemia/mental retardation syndrome (ATRX)	ATRX	301040	22
Arboleda-Tham syndrome (ARTHS)	KAT6A	616268	18
Autism, susceptibility to, 18 (AUTS18)	CHD8	615032	28
Beck-Fahrner syndrome (BEFAHRS)	TET3	618798	16
Blepharophimosis Intellectual disability SMARCA2 syndrome	SMARCA2	619293	5
Börjeson-Forssman-Lehmann syndrome (BFLS)	PHF6	301900	16

Cerebellar ataxia, deafness, and narcolepsy, autosomal dominant (ADCADN)	DNMT1	604121	5
CHARGE syndrome	CHD7	214800	65
Chr16p11.2 deletion syndrome, 593-KB	Chr16p11.2 deletion	611913	18
Coffin-Siris syndrome-1,2 (CSS1,2)	ARID1B ARID1A	135900 614607	4
Coffin-Siris syndrome-1,2,3,4 (CSS1,2,3,4) Nicolaides-Baraitser syndrome (NCBRS)	ARID1B ARID1A SMARCB1 SMARCA4 SMARCA2	135900 614607 614608 614609 601358	97
Coffin-Siris syndrome-4 (CSS4)	SMARCA4	614609	3
Coffin-Siris syndrome-9 (CSS9)	SOX11	615866	10
Cohen-Gibson syndrome (COGIS); Weaver syndrome (WVS)	EED ; EZH2	617561 277590	7

Cornelia de Lange syndromes 1,2,3,4 (CDLS1,2,3,4)	NIPBL	122470	57
	SMC1A	300590	
	SMC3	610759	
	RAD21	614701	
Down syndrome	Chr21 trisomy	190685	40
Dystonia 28, childhood-onset (DYT28)	KMT2B	617284	11
Epileptic encephalopathy, childhood-onset (EEOC)	CHD2	615369	8
Floating Harbor syndrome (FLHS)	SRCAP	136140	20
Gabriele-de Vries syndrome (GADEVS)	YY1	617557	10
Genitopatellar syndrome (see also Ohdo syndrome, SBBYSS variant) (KAT6B)	KAT6B	606170	4
Helsmoortel-van der Aa syndrome (HVDAS)*	ADNP	615873	13
Helsmoortel-van der Aa syndrome (HVDAS)*	ADNP	615873	23

Hunter McAlpine craniosynostosis syndrome	Chr5q35-qter duplication	601379	4
Immunodeficiency-centromeric instability-facial anomalies syndrome 1 (ICF1)	DNMT3B	242860	8
Immunodeficiency-centromeric instability-facial anomalies syndromes 2,3,4 (ICF2,3,4)	ZBTB24 CDCA7 HELLS	614069 616910 616911	7
Intellectual developmental disorder with seizures and language delay (IDDSELD)	SETD1B	619000	10
Kabuki syndromes 1,2 (KABUK1,2)	KMT2D KDM6A	147920 300867	149
KDM2B-related syndrome	KDM2B	/	9
Autosomal dominant intellectual developmental disorder-65 (MRD65)	KDM4B	619320	6
Kleefstra syndrome 1 (KLEFS1)	EHMT1	610253	32
Koolen de Vreis syndrome (KDVS)	KANSL1	610443	11

Luscan-Lumish syndrome (LLS)	SETD2	616831	4
Menke-Hennekam syndromes 1,2 (MKHK1,2)	CREBBP EP300	618332 618333	13
Intellectual developmental disorder, X-linked, syndromic, Armfield type (MRXSA)	FAM50A	300261	6
Mental retardation, autosomal dominant 23 (MRD23)	SETD5	615761	25
Mental retardation, autosomal dominant 51 (MRD51)	KMT5B	617788	7
Intellectual developmental disorder, X-linked 93 (MRX93)	BRWD3	300659	11
Intellectual developmental disorder, X-linked 97 (MRX97)	ZNF711	300803	15
Intellectual developmental disorder, X-linked syndromic, Nascimento-type (MRXSN)	UBE2A	300860	4
Intellectual developmental disorder, X-linked, Snyder-Robinson type (MRXSSR)	SMS	309583	17

Intellectual developmental disorder, X-linked, syndromic, Claes-Jensen type (MRXSCJ)	KDM5C	300534	49
Myopathy, lactic acidosis, and sideroblastic anemia 2 (MLASA2)	YARS2	613561	11
Ohdo syndrome, SBBYSS variant (SBBYSS)	KAT6B	603736	10
Phelan-McDermid syndrome (PHMDS)	Chr22q13.3 deletion	606232	11
Rahman syndrome (RMNS)	HIST1H1E	617537	8
Renpenning syndrome (RENS1)*	PQBP1	309500	8
Rubinstein-Taybi syndrome 1 (RSTS1)*	CREBBP	180849	37
Rubinstein-Taybi syndromes 1,2 (RSTS1,2)*	CREBBP EP300	180849 613684	39
Rubinstein-Taybi syndrome 2 (RSTS2)*	EP300	613684	29
Sotos syndrome 1 (SOTOS1)	NSD1	117550	69

Tatton-Brown-Rahman syndrome (TBRS)	DNMT3A	615879	27
--	--------	--------	----

Annexe 2. Données cliniques et de biologie moléculaire des patients présentant une variation du gène

MBD5. *Apparentés.

ID	Batch	Sex	Age at blood sampling	Variant	Variant type	Consequences	ACMG	Inheritance	Clinical features	Cluster
2116-1*	1	F	4	arr 2q23.1(148,731,026-148,799,306)x1	Partial del	Del E1-2 <i>MBD5</i> (+ E1-5 <i>ORC4</i>)	5	DN	NA	Yes
2116-2*	1	F	4	arr 2q23.1(148,731,026-148,799,306)x1	Partial del	Del E1-2 <i>MBD5</i> (+ E1-5 <i>ORC4</i>)	5	DN	NA	Yes
2117	1	F	4	arr 2q23.1(148,731,026-148,799,306)x1	Partial del	Del E1-2 <i>MBD5</i> (+ E1-5 <i>ORC4</i>)	5	DN	NA	Yes
2200	1+2	F	6	arr 2q23.1(148,731,026-148,869,728)x1	Del	Del E1-2 <i>MBD5</i> (+ E1-4 <i>ORC4</i>)	5	DN	ASD, Global delay without ID, abnormality of limbs	Yes

2201	1+2	F	7	arr 2q23.1(148,755,023-149,053,084)x1	Del	Del E1-4 <i>MBD5</i> (+ E1 <i>ORC4</i>)	5	DN	ADHD, ID, dysmorphic features	Yes
2088	2	F	26	NM_018328.4c.459delT	Frameshift	p.His153Glnfs*8	5	DN	Epileptic encephalopathy	Yes
2132-1*	1	F	2.5	NM_018328.4c.3395delA	Frameshift	p.Asp1132Alafs*3	5	DN	Epilepsy, global delay	No
2132-2*	1	F	2.5	NM_018328.4c.3395delA	Frameshift	p.Asp1132Alafs*3	5	DN	Epilepsy, global delay	No
2101	1+2	F	4	NM_018328.4c.3619C>T	Stop	p.Arg1207*	5	DN	Ataxia	No
2115	1+2	M	10	NM_018328.4c.440C>G	Stop	p.Ser147*	5	DN	NA	No
2152	2	F	6	NM_018328.4c.604A>T	Stop	p.Arg202*	5	DN	Global delay, ataxia congenital, ventriculomegaly	No
563	1+2	F	14	NM_018328.5c.180C>A	Stop	p.Cys60*	5	DN	ASD, ADHD, dyspraxia without ID, dysmorphic features, ventricular ectopic beats, postural symmetry	No
2199	1+2	F	3	arr 2q23.1(148,705,701-148,879,681)x1	Partial del	Del E1-2 <i>MBD5</i> (+ E1-10 <i>ORC4</i>)	4	Not maternal	Global delay, epilepsy, dysmorphia	Yes

2323	2	M	2	arr 2q23.1(148,715,861-148,934,846)x1	Partial del	Del E1-2 <i>MBD5</i> (+ <i>ORC4</i>)	4	DN	NA	Yes
2102	1+2	M	6	arr 2q23.1(148,764,203-148,786,336)x1	Partial del	Del E1-2 <i>MBD5</i> (+ E1 <i>ORC4</i>)	4	DN	NA	Yes
2103	1+2	F	11	arr 2q23.1(148,771,763-148,881,451)x1	Partial del	Del E1-2 <i>MBD5</i> (+ E1 <i>ORC4</i>)	4	DN	NA	Yes
2099	1	M	/	NM_018328.5c.1A>G	Start loss	p.Met1Val	4	DN	Epileptic encephalopathy	Yes
2302	2	F	4	arr 2q22.3q23.1(148,453,079-149,000,244)x1	Partial del	Del E1-3 <i>MBD5</i> (+ <i>ORC4, ACVR2A</i>)	4	DN	Epilepsy	No
2294	2	F	3.6	arr 2q22.3q23.2(147,381,806-149,696,752)x1	Del	Del <i>MBD5</i> (+ <i>ORC4, KIF5</i>)	4	DN	ID, epilepsy, dysmorphia	No
2198	1+2	F	4	arr 2q23.1(148,755,023-148,955,720)x1	Partial del	Del E1-3 <i>MBD5</i> (+ E1 <i>ORC4</i>)	4	DN	Disruptive behavior, Global delay without ID	No

2105	1	M	2	NM_018328.4c.2274delC	Frameshift	p.Val759Cysfs*25	4	DN	NA	No
2325	2	M	10	arr 2q23.1(148,829,318-149,020,865)x1	Partial del	Del E2-3 <i>MBD5</i>	3	Paternal	ADHD, learning disorder	NA
2204	1	F	17	arr 2q23.1(148,829,318-149,041,202)x1	Partial del	Del <i>MBD5</i> E2-3	3	Maternal	ASD, ID with developmental regression, dysmorphic features	NA
2202	1	F	4	arr 2q23.1(148,934,587-149,060,904)x1	Partial del	Del E3 <i>MBD5</i>	3	Maternal	Disruptive behaviour, ID, dysmorphia, cerebellar vermis hypoplasia	NA
2205	1	F	1	arr 2q23.1(149,041,143-149,434,604)x3	Partial dup	Dup E5-15 <i>MBD5</i> (+ E1 <i>EPC2</i>)	3	DN	Global delay	NA
2326	2	M	4	arr 2q23.1(149,093,832-149,198,502)x1	Partial del	Del E4 <i>MBD5</i>	3	Maternal	ASD	NA
2203	1	F	16	arr 2q23.1(149,263,045-153,281,464)x1	Partial del	Del E15 <i>MBD5</i> (+16 genes including <i>EPC2</i>)	3	DN	ADHD, ID, dysmorphic features, abnormality of limbs	NA
2128	1	F	7	arr2q23.1(149,228,342-149,414,104)x3	Partial dup	Dup E10-15 <i>MBD5</i> (+ E1 <i>EPC2</i>)	3	DN	NA	NA

2338	2	F	5	NM_018328.4:c.3288delC	Frameshift	p.Ser1097Alafs*24	3	Paternal	ID	NA
2268	2	M	2.3	NM_018328.4:c.544T>C	Missense	p.Tyr182His	3	Paternal	ID, hypoplasia of the corpus callosum	NA