

Université de Montréal

Effet du dosage génique sur les phénotypes cognitifs et psychiatriques

Par

Catherine Proulx

Département de biochimie et médecine moléculaire, Faculté de médecine

Mémoire présenté en vue de l'obtention du grade de Maîtrise
en Bio-informatique

Avril 2022

© Catherine Proulx, 2022

Université de Montréal

Unité académique : Département de biochimie et médecine moléculaire, Faculté de médecine

Ce mémoire (ou cette thèse) intitulé(e)

Effet du dosage génique sur les phénotypes cognitifs et psychiatriques

Présenté par

Catherine Proulx

A été évalué(e) par un jury composé des personnes suivantes

Sarah Gagliano-Taliun

Président-rapporteur

Sébastien Jacquemont

Directeur de recherche

Guillaume Bourque

Membre du jury

Résumé

Les troubles neurodéveloppementaux (TND) constituent un problème de santé majeur, puisque 14% de la population en souffre, dont 48% sont des cas sévères. 10-15% des patients référés ont au moins un CNV considéré comme causal. Cependant, la majorité d'entre eux sont non-récurrents, c'est-à-dire observés uniquement chez un ou peu de patients. Par conséquent, il est difficile d'estimer leur contribution clinique.

Depuis plusieurs années, l'équipe du Dr Jacquemont a développé un modèle statistique permettant d'estimer l'impact des CNV autosomiques sur l'intelligence générale et le risque d'autisme. Le modèle utilise la somme du score d'intolérance aux pertes de fonction des gènes (score LOEUF) présents séparément pour chaque type de CNV chez l'individu. Ce modèle n'est cependant pas adapté aux chromosomes sexuels et manque de résolution au niveau des gènes. Notre étude vise dans un premier temps à appliquer ce modèle sur le chromosome X. Nous allons ensuite développer une CNV GWAS (genome-wide association study) en appliquant le modèle spécifiquement sur chaque gène présent dans les CNV.

La mesure de l'impact des CNV du chromosome X sur l'intelligence n'a pas été concluante, mais a révélé un potentiel problème de mosaïcisme dans la cohorte UKBiobank. Notre CNV GWAS a permis l'identification de nouveaux gènes associés positivement ou négativement à l'intelligence générale, au risque de TSA, ou aux deux phénotypes. Ce modèle nous a permis de mieux comprendre la dynamique pléiotropique des gènes sur ces phénotypes. Ces travaux contribueront à l'interprétation clinique des CNV chez les patients affectés de TDN.

Mots-clés : Troubles neurodéveloppementaux, Variation du nombre de copies, Cognition, Intelligence générale, Psychiatrie, Trouble du spectre autistique.

Abstract

Neurodevelopmental disorders (NDD) represent a significant health burden, with 14% of the population suffering from them, 48% of which are severe cases. 10-15% of referred patients have at least one CNVs considered causal. However, the majority of them are non-recurrent, *i.e.* observed only in one or a few patients. Therefore, it is difficult to estimate their clinical contribution.

For several years, Dr. Jacquemont's team has been developing a statistical model to estimate the impact of autosomal CNV on general intelligence and autism risk. The model uses the sum of the loss-of-function gene score (LOEUF score) present separately for each type of CNV in the individual. However, this model is not adapted to sex chromosomes and lacks resolution at the gene level. Our study aims first to apply this model to the X chromosome. We will then develop a CNV GWAS (genome-wide association study) by applying the model specifically to each gene present in the CNVs.

Measuring the impact of X chromosome CNVs on intelligence was inconclusive but revealed a potential mosaicism problem in the UKBiobank cohort. Our CNV GWAS allowed the identification of new genes positively or negatively associated with general intelligence, ASD risk, or both phenotypes. This model has allowed us to better understand the pleiotropic dynamics of genes on these phenotypes. This work will contribute to the clinical interpretation of CNVs in patients with NDD.

Keywords : Neurodevelopmental disorders, Copy number variant, Cognition, General intelligence, Psychiatry, Autism spectrum disorder.

Table des matières

Résumé.....	3
Abstract	5
Table des matières	7
Liste des tableaux.....	11
Liste des figures.....	13
Liste des sigles et abréviations	15
Remerciements	21
Chapitre 1 – Introduction	23
Chapitre 2 – Recension des écrits.....	25
2.1 Troubles neurodéveloppementaux	25
2.1.1 Définition	25
2.1.2 Comorbidité des TND avec la déficience intellectuelle	26
2.2 Causes des TND	27
2.2.1 Facteurs environnementaux.....	27
2.2.2 Facteurs génétiques	28
2.2.2.1 Bases génétiques	28
2.2.2.1 CNV impliqués dans les TND.....	28
2.3 Identification des variants génétiques associés aux TND.....	32
2.3.1 Études sur les jumeaux.....	32
2.3.2 Caryotypage.....	33
2.3.3 Analyse chromosomique sur puce à ADN.....	34
2.3.4 Séquençage de l'exome et du génome entier	36

2.4 Algorithmes de détection des CNV	37
2.5 Mesure de l'impact des CNV sur la cognition	39
2.5.1 Vers un modèle quantitatif	39
2.5.2 Vers un modèle résolusif	45
Chapitre 3 – Problématique et hypothèses de travail	48
3.1 Problématique et hypothèse	48
3.1.1 Autosomes vs Chromosome X	48
3.1.2 Modèle additif vs modèle résolusif	50
3.2 Objectifs	50
3.2.1 Objectif 1 : Impact d'un CNV dans le chromosome X sur la cognition	50
3.2.2 Objectif 2 : Impact des gènes individuels sur la cognition et le risque de TSA	50
Chapitre 4 – Méthodologie	51
4.1 Cohortes	51
4.2 Mesure de l'intelligence générale	51
4.3 Mesure de l'information génétique	55
4.3.1 Détection et filtrage des CNV	55
4.4 Analyses statistiques	56
4.4.1 Modélisation de l'impact des CNV sur l'intelligence	56
4.4.1.1 Correction du score Z de l'intelligence par l'âge et le sexe	56
4.4.1.2 Impact des gènes individuels sur le quotient intellectuel	56
4.4.1.3 Impact des gènes tolérants sur l'intelligence générale	57
4.4.1.4 Analyses de sensibilité sur l'âge	58
4.4.1.5 Analyses sur le chromosome X	59
4.4.2 Modélisation de l'impact des CNV sur le risque de TSA	59

Chapitre 5 – Résultats.....	61
5.1 Normalisation des mesures cognitives	61
5.2 Objectif 1 : Impact des CNV du chromosome X sur la cognition	65
5.2.1 Application du modèle additif sur les CNV du chromosome X	65
5.2.2 Investigation du mosaïcisme	66
5.3 Objectif 2a : Impact des gènes individuels sur la cognition.....	68
5.3.1 Impact des gènes tolérants sur l’intelligence générale	75
5.3.2 Analyse de sensibilité sur l’âge	78
5.4 Objectif 2b : Impact des gènes individuels sur le risque de TSA.....	81
5.5 Pléiotropie TSA - Intelligence générale.....	93
Chapitre 6 – Discussion	97
6.1 Normalisation des mesures cognitives	97
6.2 Objectif 1 : Impact des CNV du chromosome X sur la cognition	98
6.3 Objectif 2 : Impact du dosage génique sur la cognition et le risque de TSA.....	99
6.3.1 Impact du dosage génique sur l’intelligence générale	99
6.3.1.1 Impact des gènes tolérants sur l’intelligence générale	100
6.3.1.2 Analyse de sensibilité sur l’âge	101
6.3.2 Impact du dosage génique sur le risque de TSA	103
6.3.3 Pléiotropisme sur l’intelligence générale et le risque de TSA.....	104
6.4 Perspectives générales	106
Chapitre 7 – Conclusion.....	107
Références bibliographiques	108
Annexes	119
Tableaux supplémentaires	119

A.1 Clusters de gènes associés à une réduction de l'intelligence générale si délétés.....	119
A.2 Clusters de gènes associés à une réduction de l'intelligence générale si dupliques....	121
A.3 Gènes inclus dans les CNV dont la présence augmente avec l'âge	123
A.4 Clusters de gènes associés à une augmentation du risque de TSA si délétés	125
A.5 Clusters de gènes associés à une augmentation du risque de TSA si dupliques	130
A.6 Gènes inclus dans la région 6p22.3 associée à l'augmentation du risque de TSA	139

Liste des tableaux

Tableau 1. – Niveaux de sévérité de la déficience intellectuelle.....	27
Tableau 2. – Caractéristiques distinguant les CNV bénins des CNV pathogéniques.	29
Tableau 3. – Émergence des scores d’intolérance à la perte de fonction des gènes.....	42
Tableau 4. – Description des cohortes pour les analyses sur le risque de TSA.	52
Tableau 5. – Description des cohortes pour les analyses sur la cognition.	53
Tableau 6. – Gènes inclus dans la CNV GWAS sur l’intelligence générale.....	69
Tableau 7. – Gènes qui augmentent l’intelligence générale si affectés par le dosage génique....	73
Tableau 8. – Gènes inclus dans la fenêtre de LOEUF =]1,44:1,85] où les gènes ont un effet positif sur l’intelligence générale dans la sliding-window.	75
Tableau 9. – Tests d’hétérogénéité sur les gènes tolérants de LOEUF =]1,4:1,85].	76
Tableau 10. – Gènes dont le LOEUF =]1,44:1,85] ayant un effet significatif sur l’intelligence générale en fonction du modèle utilisé.....	77
Tableau 11. – Impact de la somme du 1/LOEUF des délétions et des duplications sur l’âge.	80
Tableau 12. – Gènes inclus dans la CNV GWAS sur le risque de TSA.	82
Tableau 13. – Gènes totalement emportés par la région 6p22.3 associée à l’augmentation du risque de TSA.....	87
Tableau 14. – Fréquence des gènes de la région 6p22.3 associée à l’augmentation du risque de TSA dans les différentes cohortes du jeu de données.....	87
Tableau 15. – Analyse préliminaire sur les gènes partiellement inclus dans les CNV de la région 6p22.3 associée à l’augmentation du risque de TSA.	88
Tableau 16. – Gènes totalement emportés par la région 6p22.2 associée à l’augmentation du risque de TSA. La région comprend 28 gènes dont 26 encodent des histones.....	90
Tableau 17. – Fréquence des gènes de la région 6p22.2 associée à l’augmentation du risque de TSA dans les différentes cohortes du jeu de données.....	91
Tableau 18. – Gènes associés à un effet protecteur sur le risque de TSA.	92
Tableau 19. – Fréquence des gènes associés à un effet protecteur sur le risque de TSA.	92
Tableau 20. – Effet pléiotropique des gènes sur l’intelligence générale et le risque de TSA.....	93

Tableau 21. – Gènes inclus dans la CNV GWAS sur le risque de TSA covarié avec le score Z de l’intelligence générale ajusté selon l’âge.....	94
Tableau supplémentaire 1. – Délétions réduisant l’intelligence générale.	120
Tableau supplémentaire 2. – Duplications réduisant l’intelligence générale.	122
Tableau supplémentaire 3. – Gènes inclus dans les CNV qui sont plus fréquents chez les individus âgés.	124
Tableau supplémentaire 4. – Délétions augmentant le risque de TSA.	127
Tableau supplémentaire 5. – Délétions augmentant le risque de TSA lorsque covarié avec l’intelligence générale.	129
Tableau supplémentaire 6. – Duplications augmentant le risque de TSA.....	134
Tableau supplémentaire 7. – Duplications augmentant le risque de TSA lorsque covarié avec l’intelligence générale.	138
Tableau supplémentaire 8. – Fonctions des gènes inclus dans la région 6p22.3 associée à l’augmentation du risque de TSA.	139

Liste des figures

Figure 1. – Mécanismes biologiques à l'origine des CNV.....	31
Figure 2. – Estimation de l'héritabilité des TSA à travers les années.....	33
Figure 3. – Représentation graphique des informations données par les puces à ADN.....	38
Figure 4. – Modèle additif des scores d'intolérance des gènes.	43
Figure 5. – Effet du pLI sur l'intelligence générale pour les délétions et les duplications.	44
Figure 6. – Concordance entre les prédictions du modèle additif des scores d'intolérance sur l'intelligence et l'effet de CNV rapportés dans la littérature.....	45
Figure 7. – Effet de grandeur des gènes individuels sur l'intelligence générale lorsqu'ils sont affectés par le dosage génique.....	47
Figure 8. – Mécanismes d'inactivation du chromosome X.	49
Figure 9. – Déclin de l'intelligence générale avec le vieillissement dans la population générale.	62
Figure 10. – Comparaison entre les types de correction du score Z de l'intelligence générale sur l'âge et le sexe.	63
Figure 11. – Comparaison de l'effet du dosage génique sur l'intelligence générale en fonction du type de correction sur le phénotype.	64
Figure 12. – Effet des CNV du chromosome X sur le score Z de l'intelligence générale.	65
Figure 13. – Distribution des âges dans UKBB (n=238 176) selon le sexe.....	67
Figure 14. – Filtre sur les individus ou les CNV pouvant générer du bruit dans le jeu de données.....	67
Figure 15. – Miami plot des gènes associés à une réduction de l'intelligence générale lorsqu'affectés par le dosage génique.....	70
Figure 16. – Miami plot des gènes associés à une augmentation de l'intelligence générale lorsqu'affectés par le dosage génique.....	71
Figure 17. – Catégorisation des gènes avec une association significative dans la CNV GWAS sur l'intelligence générale pour A. les délétions et B. les duplications.	72
Figure 18. – Régions qui augmentent l'intelligence générale lorsque délétée ou dupliquée.....	74

Figure 19. – Volcano plots de l’impact des gènes tolérant sur le score Z du QI..... 77

Figure 20. – Distribution du score du modèle additif en fonction de l’âge pour (A et B) les porteurs de délétions et (C et D) les porteurs de duplications. 79

Figure 21. – CNV GWAS sur l’âge. A. CNV associés à une augmentation de l’âge. B. CNV associés à une baisse de l’âge. 81

Figure 22. – Miami plot des gènes associés à une augmentation du risque de TSA lorsqu’affectés par le dosage génique. 83

Figure 23. – Miami plot des gènes associés à un effet protecteur sur le risque de TSA lorsqu’affectés par le dosage génique..... 84

Figure 24. – Catégorisation des gènes avec une association significative dans la CNV GWAS sur le risque de TSA pour (A) les délétions et (B) les duplications. 85

Figure 25. – Duplication 6p22.3 : nouvelle région candidate augmentant le risque de TSA. 86

Figure 26. – Duplication 6p22.2 : nouvelle région candidate augmentant le risque de TSA. 89

Figure 27. – Miami plot des gènes associés à une augmentation du risque de TSA lorsqu’affectés par le dosage génique dans le modèle covarié avec le score Z de l’intelligence général ajusté par l’âge. 95

Figure 28. – Intersections des gènes associés au risque de TSA, covarié ou non avec l’intelligence générale, et des gènes associés à l’intelligence générale pour A. les délétions et B. les duplications. 96

Liste des sigles et abréviations

%HI : Probabilité d'haploinsuffisance

ACPA : Analyse chromosomique sur puce à ADN

ADN : Acide désoxyribonucléique

ANTXRL : ANTXR Like

ANXA8L1 : Annexin A8 Like 1

BAF : Fréquence de l'allèle B

CaG : CartaGene

CAP2 : Cyclase Associated Actin Cytoskeleton Regulatory Protein 2

CGH : Hybridation comparative génomique

CHD8 : Chromodomain helicase DNA binding protein 8

CNV : Variation du nombre de copies

DAS-II : Échelle d'habiletés différentielle

DSM-5 : Manuel de diagnostic et de statistique des maladies mentales-Édition 5

DYRK1A : Dual specificity tyrosine phosphorylation regulated kinase 1A

ExAC : Exome Aggregation Consortium

FAM8A1 : Family With Sequence Similarity 8 Member A1

FDR : Taux de découverte fausse

FI : Intelligence fluide

FISH : Hybridation in situ en fluorescence

FMR1 : Fragile X mental retardation 1

GP : Processus gaussien

GRIN2B : Glutamate ionotropic receptor NMDA type subunit 2B

G-Scot : Generation Scotland

GWAS : Études d'association pangénomiques

HFE : Homeostatic Iron Regulator

HMM : Modèle de Markov caché

KIF13A : Kinesin Family Member 13A

LBC : Lothian birth cohort
LCR : Répétitions à faible nombre de copies
LOEUF : Borne supérieure de la perte de fonction observée/attendue
LRR : Ratio du log R
MALL : Mal, T Cell Differentiation Protein Like
NAHR : Recombinaison homologue non allélique
NHEJ : Jonction d'extrémités non homologues
NLGN3/NLGN4X : Neuroligin 3/4 X-Linked
NPHP1 : Nephrocystin 1
NRXN1 : Neurexin 1
NUP153 : Nucleoporin 153
PCA : Analyse en composantes principales
pLI : Probabilité d'intolérance à la perte de fonction
PSG1/2/6/7/11 : Pregnancy Specific Beta-1-Glycoprotein 1/2/6/7/11
PTEN : Phosphatase and tensin homolog
QI : Quotient intellectuel
SD : Écart type
SFARI : Simons foundation autism research initiative
SHANK3 : SH3 and multiple ankyrin repeat domains 3
SNP : Polymorphisme d'un seul nucléotide
SNV : Variant d'un seul nucléotide
SPARK : Simons foundation powering autism research
SSC : Simon simplex collection
SYS : Saguenay youth study
TBL1XR : Transducin beta like 1 X-linked receptor 1
TBR1 : T-box brain transcription factor 1
TDAH : Trouble du déficit de l'attention avec ou sans hyperactivité
TND : Troubles neurodéveloppementaux
TSA : Troubles du spectre autistique

UKBB : UKBiobank

VUS : Variation ayant une signification clinique incertaine

VWA8 : Von Willebrand Factor A Domain Containing 8

WASI-I/II : Échelle d'intelligence de Wechsler pour adultes-version I/II

WES : Séquençage de l'exome entier

WGS : Séquençage du génome entier

WISC-III/IV/V : Échelle d'intelligence de Wechsler pour enfants-version III/IV/V

WPPSI-IV : Échelle d'intelligence de Wechsler pour pré-scolaire et primaire-version IV

À mon conjoint, ma famille

Remerciements

J'aimerais exprimer toute ma gratitude envers mon directeur de maîtrise, Dr Sébastien Jacquemont, sans qui ce projet n'aurait pu voir le jour. Je te suis réellement reconnaissante de m'avoir donné l'opportunité de travailler dans un domaine qui m'est très cher. Merci également pour ton enthousiasme envers ce projet, ton soutien et tes conseils.

Un énorme merci à Guillaume Huguet, qui tout au long de ce projet m'a guidée, m'a appris et m'a aidé non seulement avec mes analyses, mais aussi avec la gestion de mes multiples inquiétudes. Je te suis extrêmement reconnaissante pour ta grande disponibilité tout au long de ma maîtrise pour répondre à mes questions, pour ton partage de connaissances et ton enseignement avec des analogies toujours incroyablement bien pensées. Merci pour ta patience et pour tout le temps que tu as consacré dans le cadre de mon projet.

J'aimerais également remercier spécialement Elise Douard : tu as été pour moi une grande source d'inspiration tout au long de mon parcours à la maîtrise. Merci encore pour ton aide, spécialement durant ces derniers mois. Mes remerciements à Cécile Poulain, Mame Seynabou Diop et Rackeb Tesfaye, ainsi que tous les autres membres du laboratoire Jacquemont. Vous êtes vraiment une équipe extraordinaire, travailler avec vous a été un honneur.

Un merci spécial à Patrick Dion, qui m'a permis de faire le lien avec cette formidable équipe.

Toute ma reconnaissance à mon conjoint d'avoir pris le temps de m'écouter parler de ce projet durant les deux dernières années et de m'avoir rassurée lorsque j'en avais besoin. Merci également à ma famille et à mes amis pour leur soutien moral tout au long de cette aventure.

Chapitre 1 – Introduction

Les troubles neurodéveloppementaux (TND) forment un groupe hétérogène de maladies psychiatriques. Ils incluent notamment la déficience intellectuelle et les troubles du spectre autistique (TSA), qui ont une prévalence respective de 1,3% (1–3) et 1,7% (4). Ces troubles ont une forte contribution génétique : l'héritabilité de l'intelligence générale est estimée à 50%-80%, alors que celle des TSA est estimée à 51%-97% (5,6).

Depuis les dernières décennies, l'évolution des technologies de génotypage, telle l'analyse chromosomique sur puce à ADN (ACPA), ont permis l'identification de nombreux variants génétiques associés aux TND. Ainsi, depuis la découverte en 1959 du premier variant associé à un TND caractérisé par une déficience intellectuelle, le syndrome de Down (7), de nombreux variants ont été associés à la déficience intellectuelle et aux TSA. Ces nouvelles technologies ont permis de mettre en évidence les Variations du Nombre de Copies (CNV), qui correspondent à des délétions ou duplications de fragments génomiques de d'au moins 1000 paires de bases. La trisomie 21, identifiée des années plus tôt, est d'ailleurs un exemple extrême d'un CNV, où le chromosome 21 est dupliqué en entier.

Aujourd'hui, on sait que les CNV sont associés à de nombreuses maladies génétiques, dont les TND. Il est observé que chez 10 à 15% des patients référés pour un TND, au moins un CNV considéré comme causal est présent (8). Certains CNV sont récurrents, c'est-à-dire qu'ils sont observables chez plusieurs patients non-apparentés. De ce fait, il est possible de réaliser des études d'associations sur ces variants. Néanmoins, très peu de CNV récurrents sont formellement associés à la déficience intellectuelle et aux TSA ($N \sim 16$) (9), car la majorité des CNV sont non-récurrents et ainsi très rares. On ne peut donc pas les étudier par études cas-contrôles, et dès lors, aucune méthode à l'heure actuelle n'est adaptée pour les étudier. Par conséquent, il est difficile pour les généticiens et les cliniciens d'estimer leur contribution aux symptômes d'un patient donné. De plus, les mécanismes qui sous-tendent aux TND sont difficiles à identifier,

puisque les CNV peuvent affecter un grand nombre de gènes (estimé à plus de 1000 pour les TSA), répertoriés dans les bases de données (*e.g.*, NCBI (10), DECIPHER (11), SFARI (12)).

Pour répondre à cette problématique, nous allons développer des outils statistiques permettant d'étudier ces variants génétiques non récurrents. En outre, nous allons développer une méthode pour quantifier les effets du dosage génétiques sur des phénotypes cognitifs (l'intelligence générale) et psychiatriques (les TSA).

Chapitre 2 – Recension des écrits

2.1 Troubles neurodéveloppementaux

2.1.1 Définition

Les TND constituent un problème de santé majeur, étant notamment de plus en plus reconnus comme une cause importante de morbidité psychiatrique chez les enfants (13). La prévalence des TND dans la population est de 13,87%, tandis que les troubles sévères associés à un délai développemental ont une prévalence de 3,65% (14). Cette prévalence varie cependant en fonction du trouble. Par exemple, dans le cas des TSA, elle s'étend entre 2,3% et 10,3%, alors que pour le trouble du déficit de l'attention avec hyperactivité (TDAH), une étude rapporte une prévalence allant jusqu'à 44,5% (13). Les TND sont des maladies psychiatriques hérissables qui comprennent un large éventail de symptômes de léger à sévère. Ils regroupent des troubles qui apparaissent souvent tôt durant le développement de l'enfant et entraînent des déficits du fonctionnement personnel, social, scolaire ou professionnel (15). Les TND comprennent également des symptômes qui ne sont pas cognitifs, comme des malformations (exemples : au niveau du visage ou cardiovasculaire) (16). Ils englobent un groupe hétérogène de troubles, dont les manifestations cliniques et la sévérité varient en fonction du TND et de l'individu qui est atteint (16). Ils comprennent notamment la déficience intellectuelle, les TSA, les troubles de la communication, les troubles de l'apprentissage et le TDAH (16). D'autres troubles ayant une trajectoire neurodéveloppementale, comme la schizophrénie et le trouble maniaco-dépressif, sont également inclus dans cette catégorie (16).

Les TND ont également une implication importante au niveau économique : notamment, une étude estime que la prise en charge tout au long de la vie d'un patient atteint d'un TSA en plus d'une déficience intellectuelle peut atteindre un coût total de 2,4 millions de dollars aux États-Unis (17). Ce coût pour un patient atteint de TSA seulement (sans déficience intellectuelle) atteint

quant à lui 1,4 millions de dollars (17). Ces chiffres mettent en évidence l'importance des études sur les TND, qui, en plus d'être une cause de souffrance significative chez les patients et leur famille, représentent un coût sanitaire important pour la société (13).

2.1.2 Comorbidité des TND avec la déficience intellectuelle

Il existe beaucoup de comorbidités entre les TND et des altérations cognitives, la plus fréquente étant la déficience intellectuelle (13). La déficience intellectuelle est un trouble qui se caractérise par plusieurs limitations au niveau cognitif. Elle affecte des compétences nécessaires à la vie de tous les jours, notamment au niveau de l'apprentissage, de la résolution de problèmes, du raisonnement et de la planification (15). Elle est caractérisée par un quotient intellectuel (QI) inférieur à 70 dans les cas les plus légers, et un QI inférieur à 20 dans les cas les plus graves (18). La sévérité des déficits qu'elle entraîne varie selon l'individu, se divisant en quatre niveaux : léger (85% des cas, QI entre 50-69), modéré (10% des cas, QI entre 36-49), sévère (3,5% des cas, QI entre 20-35) et profond (1,5% des cas, QI inférieur à 20) (Tableau 1) (18). Elle résulte en des difficultés intellectuelles et d'adaptation qui sont supérieures à celles de leurs pairs du même genre et du même âge, avec une origine linguistique et sociale comparable (15). Sa prévalence dans la population générale est estimée à 1,3% (1–3). Elle représente un défi médical et économique considérable pour la société, étant le diagnostic le plus coûteux dans les pays industrialisés (19).

La déficience intellectuelle est particulièrement importante dans les TSA. Il s'agit d'un spécificateur du diagnostic et elle est diagnostiquée dans 19 à 48% des cas (18,20–22). Les TSA sont définis par des intérêts restreints et stéréotypés avec des troubles de la communication et de l'interaction sociale présents dès le développement précoce et nuisant au fonctionnement de l'individu dans son quotidien (15). Ils sont caractérisés par une grande hétérogénéité clinique avec un large spectre de sévérité des troubles identifiés, mais aussi génétique et biologique. Les comorbidités neurodéveloppementales sont présentes chez 70% des enfants atteints d'autisme et son héritabilité est de 60% à 90% (23).

Niveau	QI	Critère du DSM-5	Distribution des cas
Léger	50–69	Peut vivre indépendamment avec un minimum de soutien, tels que ceux disponibles dans les foyers de groupe.	85%
Modéré	36–49	Peut vivre indépendamment avec des niveaux modérés de soutien.	10%
Sévère	20–35	A besoin d'assistance quotidienne pour les activités d'auto-soins et d'une supervision de la sécurité.	3,5%
Profond	<20	A besoin d'assistance en tout temps.	1.5%

Tableau 1. – Niveaux de sévérité de la déficience intellectuelle.

Adapté de Boat et al. (2015) (18).

Ainsi, l'intelligence générale, aussi appelée l'habileté cognitive générale (24), est un critère d'une grande importance dans le diagnostic des TND (25). Telle la fièvre des troubles psychiatriques, elle n'est pas spécifique à un TND en particulier, mais lorsqu'elle est altérée, il s'agit d'un bon indicateur de la présence d'un problème. L'intelligence générale est un trait héritable dont la part génétique est estimée entre 50% à 80% (26). Le QI est généralement utilisé pour l'estimer. Il est mesuré à l'aide de tests standardisés qui sont adaptés au niveau cognitif de l'individu. Puisque les individus atteints de TSA sont parfois affectés par des déficits du langage, le QI non verbal est généralement choisi pour mesurer leur intelligence. En moyenne dans la population, les individus obtiennent un score de 100 points de QI avec un écart-type de 15.

2.2 Causes des TND

2.2.1 Facteurs environnementaux

Les TND sont causés par un développement et un fonctionnement anormal du cerveau (16). Des causes environnementales peuvent moduler les TND. Ces facteurs peuvent être de nature prénatale ou postnatale. Avant la naissance (en prénatal), puisque le placenta lie l'enfant à la mère qui le porte, la nutrition, la physiologie et le style de vie de la mère (comme la

consommation d'alcool ou de cigarette) ont un impact sur le développement du fœtus (16). Après la naissance (en postnatal), la nutrition de l'enfant, son niveau de stress, l'alcool, la cigarette et la qualité des pratiques parentales (notamment des styles parentaux excessivement permissifs, autoritaires ou désengagés) sont également des facteurs environnementaux qui peuvent avoir une influence sur le développement de TND chez l'enfant (16).

2.2.2 Facteurs génétiques

2.2.2.1 Bases génétiques

L'ADN est le support physique de l'information génétique (27). C'est une macromolécule composée de deux brins de nucléotides reliés les uns aux autres, formant une double hélice. Chaque nucléotide est constitué d'un sucre, le désoxyribose, d'un groupement phosphate et d'une base : l'adénine, la cytosine, la thymine ou la guanine. L'ensemble de l'information génétique, notre génome, se divise en 23 paires de chromosomes. Chez l'humain, les chromosomes se divisent en 2 catégories : 22 paires d'autosomes et 1 paires de chromosomes sexuels. Les hommes ont un chromosome X et un chromosome Y, alors que la femme possède deux chromosomes X. L'altération de l'information génétique, causée par différents types de variants génétiques, entraîne l'altération des gènes qui y sont présents, ce qui peut altérer leur fonction biologique. De telles altérations peuvent mener à différents troubles, comme les TND. La contribution génétique des TND est de 50% à 80% (23,28–30).

2.2.2.1 CNV impliqués dans les TND

Les CNV ont une grande importance dans la causalité des TND. Ils correspondent à des délétions ou duplications de fragments génomiques d'au moins 1000 paires de base. Un exemple extrême de duplication est la trisomie 21, où celui-ci est dupliqué en entier. Les individus ont en moyenne plus de 1000 CNV dans leur génome, résultant en une différence de 4 millions de paires de bases en comparaison avec le génome de référence (0.1% de la variation génétique au niveau

structurel). Les délétions et les duplications peuvent affecter le dosage génique, qui correspond au nombre de copies d'un gène dans un génome. Les CNV ne sont pas tous pathogéniques. Les CNV se catégorisent en trois groupes : les CNV bénins, pathogéniques ou ayant une signification clinique incertaine (VUS) (31). La dangerosité des CNV dépend de quatre facteurs : la taille, le contenu, l'origine et la fréquence dans la population (Tableau 2).

Caractéristique	CNV bénin	CNV pathogénique
Taille	Petit (< 3 MB)	Grand (>3-5 MB)
Contenu	Absence de gène important	Gènes importants
Origine	Hérité	<i>De novo</i>
Fréquence	Commun	Rare

Tableau 2. – Caractéristiques distinguant généralement les CNV bénins des CNV pathogéniques.

Il existe deux types de CNV : les CNV récurrents (Figure 1A) et les CNV non-récurrents (Figure 1B), qui émergent de deux mécanismes de recombinaison de l'ADN différents. Les CNV récurrents sont des variants apparaissant à des positions similaires dans le génome chez des individus non-apparentés. La structure particulière du génome, comme les séquences répétées qui y sont présentes, peut favoriser les réarrangements chromosomiques. En effet, lorsqu'une région est flanquée par des répétitions à faible nombre de copies (LCR, low-copy repeats), comme les duplications segmentaires, cela peut mener à une recombinaison désalignée. La recombinaison est un échange de segments homologues entre deux molécules d'ADN (27). Elle permet l'échange de matériel génétique entre deux molécules d'ADN homologues et est très précise : lors de l'échange des brins, il n'y a ni perte, ni gain de nucléotides. Néanmoins, si une région est flanquée par des duplications segmentaires paralogues, la résultante de ce réarrangement fait en sorte qu'une des deux copies d'ADN porte une délétion, tandis que l'autre copie porte une duplication; d'où l'apparition d'un CNV. Ce mécanisme de recombinaison homologue non allélique (NAHR) (Figure 1C) mène à la formation de CNV récurrents (Figure 1A), c'est-à-dire qu'ils ont toujours les

mêmes points de rupture et donc la même taille et le même contenu génique, et ils sont observables chez différents patients non-apparentés. Puisque ces CNV sont observés chez plusieurs individus, leurs effets sur de nombreux phénotypes ont pu être étudiés via des études d'association. Par exemple, les régions 22q11.2 et 16p11.2 sont des délétions récurrentes pour lesquelles on estime respectivement une diminution moyenne de QI de -1,9 (32) et -1,5 déviations standards (33).

La plupart des CNV apparaissent néanmoins de façon aléatoire dans le génome et sont donc très rares. Ces CNV sont dits non récurrents et sont générés par des mécanismes différents (Figure 1B). Ils peuvent partager une région commune entre patients, mais leurs points de rupture sont dispersés. Ils peuvent apparaître via des mécanismes réplicatifs, notamment par des erreurs de réplication par glissement, mais aussi par des mécanismes non-réplicatifs, comme la jonction d'extrémités non homologues (NHEJ, non-homologous end joining) (Figure 1D). Ce dernier est un mécanisme de réparation de cassure double-brin de l'ADN. Si le réappariement est décalé, soit en amont ou en aval, il en résulte respectivement d'une perte ou d'un gain d'une partie de la séquence de l'ADN (34). Comme ils sont excessivement rares, on ne peut étudier l'effet de ces CNV par des études d'association cas-contrôle de la même façon que pour les CNV récurrents. Leur effet est donc mal connu.

Les CNV sont possiblement les variants qui ont le plus d'impact dans la pathogenèse des TND, puisqu'ils peuvent altérer plusieurs gènes en même temps. Grâce à l'introduction de l'ACPA en clinique pédiatrique, il a été observé que chez 10 à 15% des enfants référés pour un TND, au moins un CNV pathogénique considéré comme causal est présent (8). Les CNV pathogéniques sont définis comme étant (a) *de novo*, (b) rare, grande taille et parfois (c) récurrents (Tableau 2) (8). Plusieurs CNV récurrents sont associés à la déficience intellectuelle, aux TSA et à la schizophrénie, comme les régions 22q11.2 et 16p11.2, qui ont un effet pléiotropique (25). La pléiotropie réfère à l'influence d'un même locus ou variant génétique sur plus d'un phénotypes ou groupes diagnostics (37).

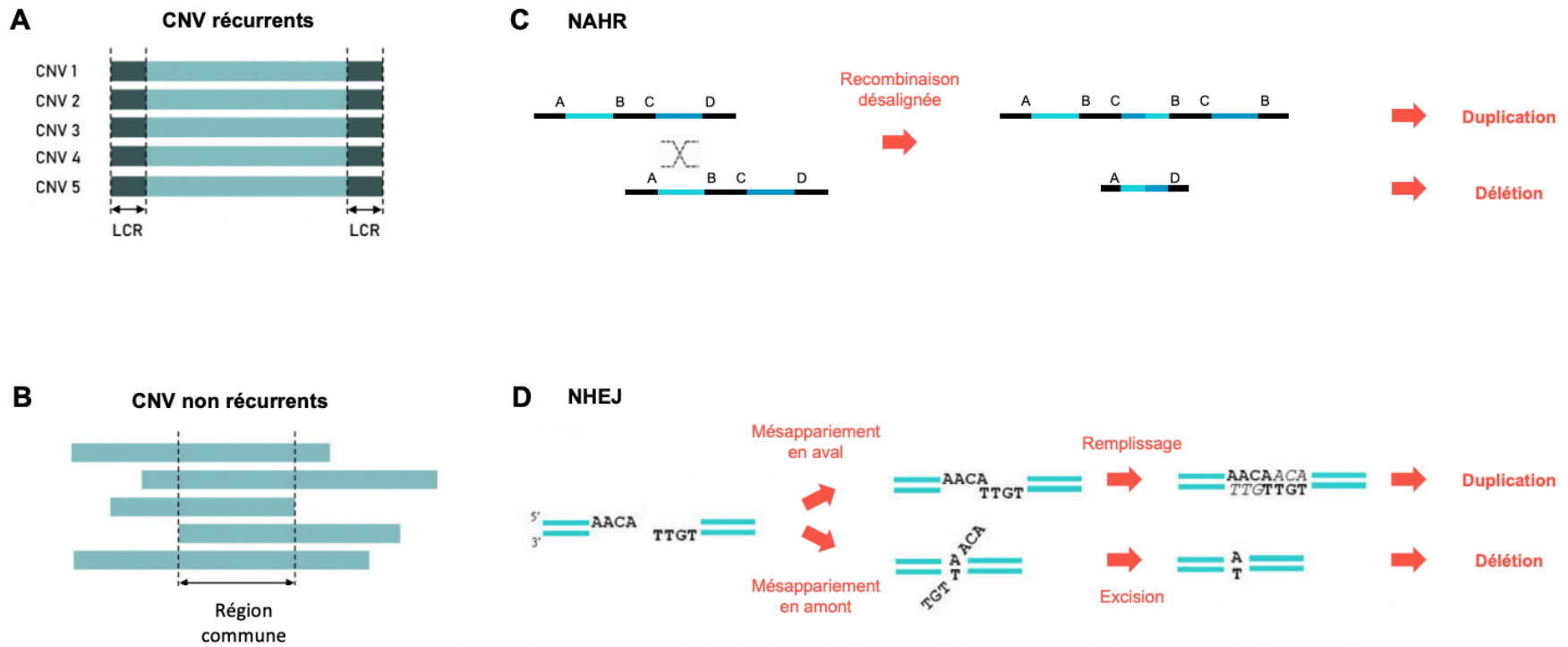


Figure 1. – Mécanismes biologiques à l’origine des CNV. A. La recombinaison homologue non-allélique (NAHR) mène à une recombinaison désalignée qui produit une copie avec une duplication et l’autre avec une délétion. Elle a lieu lorsqu’une région est flanquée par des répétitions à faible nombre de copies (LCR) et elle mène à la formation de CNV récurrents (B). C. La jonction d’extrémités non homologues (NHEJ) est un mécanisme de réparation de cassure double-brin de l’ADN. Si un mésappariement a lieu en aval, il produira une duplication, alors qu’en amont, il produit une délétion. D. Ce mécanisme mène à la formation de CNV non récurrents, qui n’ont pas les mêmes bornes chez les individus non apparentés. Adapté de Chen et al. (2014) (35) (A), de Pâques et Haber (1999) (34) (C) et de de Pös et al. (2021) (36) (B et D).

2.3 Identification des variants génétiques associés aux TND

2.3.1 Études sur les jumeaux

Les premières études sur l'évaluation de la contribution génétique des TND remontent au début du 20e siècle avec les études sur les jumeaux. Cette méthodologie repose sur le fait que les jumeaux monozygotes ont un bagage génétique identique, ainsi qu'un environnement similaire, particulièrement en prénatal où cet environnement est identique. En comparant les traits du jumeaux monozygotes à ceux des jumeaux dizygotes, il serait donc possible d'estimer la part génétique d'un trouble en particulier. La première étude utilisant cette méthodologie et révélant alors une forte contribution de la génétique sur les TND est celle de Smith (1930) sur la déficience intellectuelle (38). D'autres études sur les jumeaux ont suivi des années plus tard avec l'étude de Kallmann et Roth sur la schizophrénie précoce (1956), puis celle de Folstein et Rutter sur l'autisme (1977), appuyant également un apport considérable de la génétique sur ces TND (39,40). À travers les années, l'estimation de l'héritabilité des TND a beaucoup fluctué. Par exemple, pour les TSA, depuis 1977, cette héritabilité a évolué entre 26% à 96% (Figure 2) (41). Cette variation est observée en raison de plusieurs facteurs, comme la définition du phénotype par les différentes études, la taille de l'échantillon utilisé et la présence d'autres TND comorbides avec les TSA comme le TDAH. Depuis, l'héritabilité des TND, estimée aujourd'hui à 49%, a bien été établie par les études sur les jumeaux (42) malgré les limites de cette méthode, qui ne prend pas en compte l'interaction de la génétique avec l'environnement (43).

De plus, cette méthodologie ne permet pas d'identifier les gènes impliqués dans cette héritabilité. La compréhension des mécanismes menant à cette conclusion est restée floue jusqu'au développement de nouvelles technologies qui ont permis l'observation du support matériel du bagage génétique : l'ADN.

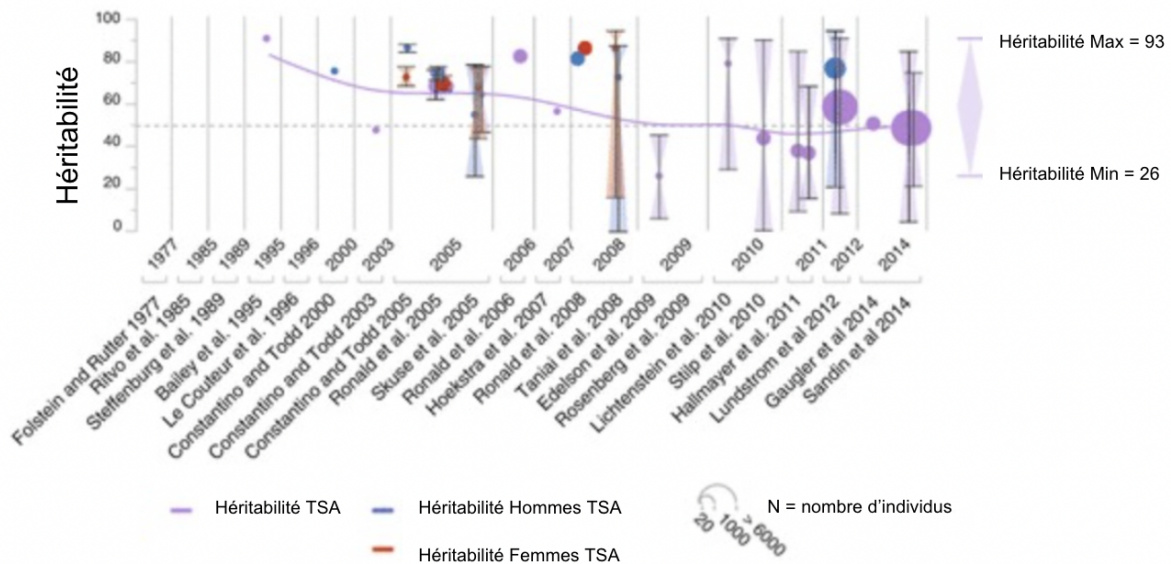


Figure 2. – Estimation de l'héritabilité des TSA à travers les années. La moyenne de l'héritabilité pondérée par la taille de l'échantillon est représentée sur la droite de la figure. Adapté de Huguet et al. (2016) (41).

2.3.2 Caryotypage

Depuis les années 1920, plusieurs expériences ont montré que l'ADN est le support de notre matériel génétique, notamment grâce aux expériences de Griffith (1928), d'Avery (1944) et de Hershey et Chase (1952) (44), suivies de la découverte de la structure en double hélice de l'ADN en 1953 par Watson et Crick (45). L'évolution des technologies d'analyse du génome dans les 50 dernières années a ensuite fortement contribué à notre compréhension des causes génétiques des TND (46). C'est en 1959 qu'a été découverte la première mutation génétique avec l'identification de la cause du syndrome de Down, qui est en fait dû à une trisomie du chromosome 21 (7). Par la suite, dans les années 1970s, des techniques de caryotypage des chromosomes par bandes ont été développées. Elles ont permis la détection d'une grande quantité de microdélétions, de microduplications et de translocations (46). Les premiers variants génétiques liés aux TSA (incluant les marqueurs Xq27, Xp22, 16q23, 6q26, ainsi que des chromosomes Y longs) ont d'ailleurs été détectés par des études cytogénétiques en 1985 (47).

En 1969, dix ans après la découverte de la trisomie 21, un nouveau syndrome a été identifié via un autre type de mutation vue par caryotypage : il s'agit du syndrome du X fragile, visible en raison de la condensation d'une portion du chromosome X qui est causée par des modifications épigénétiques (48). Ce n'est cependant qu'en 1991 que l'expansion d'une répétition de séquence de nucléotide CGG dans le promoteur du gène FMR1 a été identifiée comme étant la cause du X fragile, qui est la forme héritée la plus commune de déficience intellectuelle (49–51). Cette altération, causée par une anomalie de méthylation qui inactive le gène, est aussi associée notamment à un retard du langage et à des comportements répétitifs (16). Il s'agit également du gène qui, à lui seul, est la plus fréquente cause des TSA, étant responsable de ces troubles dans 1% à 6% des cas selon les études (52,53).

D'autres technologies permettant la détection de pertes et de gains dans le génome ont été développées par la suite, notamment l'hybridation in situ en fluorescence (FISH, fluorescence in situ hybridization) (46). Toutefois, la résolution de ces méthodes, limitée à quelques mégabases, était trop faible pour associer des gènes spécifiques aux TND (41). Notamment, dans les TSA, la prévalence des aberrations chromosomiques de grande taille est seulement de 2% : ces techniques ne sont donc pas suffisamment précises pour identifier la majorité des causes génétiques de ces troubles (41).

2.3.3 Analyse chromosomique sur puce à ADN

Depuis 1990, des efforts internationaux ont été fournis dans le cadre du projet « Génome humain » qui avait pour but d'obtenir la première carte des régions euchromatiques du génome humain. C'est en 2003 que le projet a abouti, permettant d'avoir la première carte du génome humain avec une couverture de ~99% de ces régions, qui inclues 20 000 à 25 000 gènes codants (54). Ces travaux, ayant mis en évidence de nombreux Polymorphisme d'un Seul Nucléotide (SNP, single-nucleotide polymorphism), ont permis le développement des techniques d'analyse de génome entier telles que l'ACPA.

Il existe deux types principaux d'ACPA : l'hybridation comparative génomique (CGH) et les SNP arrays. Ces deux technologies ont leurs limites respectives. Pour la CGH, son interprétation dépend d'un génome de référence, auquel est hybridé l'ADN avec des colorants fluorescents. De plus, elle est plus dispendieuse et une plus grande quantité d'ADN est nécessaire pour l'utiliser en comparaison avec les autres méthodes d'analyse. Cette technique serait néanmoins plus fiable en comparaison avec les SNP arrays. Ces dernières, comme leur nom l'indique, avaient originellement pour but de faire la détection des SNP. La distribution et la densité des marqueurs sur ce type de puce n'est donc pas uniforme et rend donc la détection des CNV de petite taille limitée par cette méthode.

Pour les TSA et la déficience intellectuelle, l'ACPA possède un meilleur rendement diagnostique (15%-20%) que le caryotypage (3% en excluant la trisomie 21 et les autres syndromes chromosomiques facilement reconnaissables) (8). Les études utilisant cette technologie ont permis d'identifier plusieurs CNV associés à des TND, dont certaines régions récurrentes. Notamment, les régions chromosomiques 3q29 (55), 16p11.2 (56), 1q21.2 (57,58), 15q13.3 (59), and 17q12 (60) sont associées aux TND lorsqu'elles sont délétées ou dupliquées. L'ACPA a permis de déterminer que les grands réarrangements chromosomiques et les CNV augmentent le risque de TSA dans 5%-10% des cas (61-63). Ces techniques ont également rendu possible, grâce à leur plus grande résolution, l'identification de nombreux gènes candidats associés à aux TND via la mise en évidence de CNV plus petits. Grâce à leur grande résolution, certaines plateformes d'ACPA permettent l'identification de CNV allant jusqu'à 50 kilobases (64).

Dans les années qui ont suivi, de nouvelles technologies telles que le séquençage de l'exome et du génome entier, sujets de la prochaine section, ont quant à elles permis l'identification de gènes associés aux TND et l'observation des CNV de très petite taille.

2.3.4 Séquençage de l'exome et du génome entier

Plusieurs gènes ont été associés aux TND par des études fonctionnelles et génétiques sur des gènes candidats au début des années 2000. C'est notamment le cas de *NLGN3*, *NLGN4X*, *SHANK3* et *NRXN1*, qui sont associés aux TSA (65–67). Un peu plus tard, le développement d'une nouvelle technologie, le séquençage de nouvelle génération, a permis l'association de gènes aux TND à travers tout le génome. Cette nouvelle technologie permet d'obtenir la séquence exhaustive du génome d'un individu en alignant son génome ou son exome, segmenté en petites séquences d'environ 300 paires de bases (communément appelés « reads »), au génome de référence (73). Le séquençage de nouvelle génération a l'avantage de permettre l'identification des bornes d'un CNV au nucléotide près (73). Cette technologie peut également être utilisée pour détecter des variants d'un seul nucléotide (SNV). En plus de *FMR1*, six autres gènes ont été rapportés précédemment par O'Roak et al. (2011) par séquençage de l'exome entier (WES) pour les TSA : *CHD8*, *DYRK1A*, *GRIN2B*, *TBR1*, *PTEN*, et *TBL1XR*. L'altération de ces gènes cause une croissance et un développement anormal du cerveau. Récemment, Satterstrom et al. (2020) ont identifié une liste de 102 gènes impliqués dans le risque de TSA par WES (68). Majoritairement, ces gènes ont un rôle au niveau de la synapse ou dans la régulation d'autres gènes qui y ont une fonction au niveau du neurone ou de la synapse.

Aujourd'hui, on sait que les TND sont des maladies complexes causées par la combinaison de plusieurs variants génétiques, dont la taille d'effet individuelle est moindre (69,70). La majorité des variants qui contribuent à l'héritabilité des TND sont des variants communs. Ceux-ci peuvent être révélés par les études d'association pangénomiques (GWAS), généralement réalisées sur les SNP. La plus grande GWAS sur le risque d'autisme à ce jour a identifié 93 marqueurs dans trois différents loci (71). De plus, plusieurs GWAS ont permis de mettre en évidence une association positive entre les variants communs qui augmentent le risque de TSA et ceux qui augmentent le QI (71,72).

Le déséquilibre de liaison est un phénomène qui peut créer du bruit dans les GWAS. Il s'agit de la tendance avec laquelle deux ou plusieurs allèles sont hérités ensemble plus ou moins souvent que ce qui aurait été attendu si la transmission avait été aléatoire. Elle est généralement causée par la proximité entre deux gènes sur un même chromosome. Lorsqu'une association génotype-phénotype est obtenue dans une GWAS, cette association peut donc être causée par n'importe quel SNP en déséquilibre de liaison avec le SNP trouvé dans l'analyse. Le déséquilibre de liaison peut être mesuré par le r^2 , qui représente la fréquence ou le degré de corrélation avec lequel deux allèles sont hérités simultanément. Conventionnellement, deux allèles sont considérés en déséquilibre de liaison lorsque la corrélation de Pearson au carré (r^2) entre les porteurs de ces deux allèles est supérieure à 0,8. Cette mesure permet de repérer les SNP en déséquilibre de liaison et de prévenir le génotypage des SNP produisant une redondance de l'information.

2.4 Algorithmes de détection des CNV

Les puces à ADN donnent deux informations qui sont importantes pour la détection des CNV. La première est le ratio du log R (LRR), qui représente la quantité de matière normalisée pour chaque sonde (73). Ainsi, l'intensité totale du signal de fluorescence est donnée par $R = a + b$, où a et b représentent respectivement l'intensité du signal normalisé des allèles A et B, et $LRR = \log_2(R_{\text{observé}} / R_{\text{attendu}})$, où R_{attendu} est donné via des échantillons de références. Puisqu'une délétion ne contient qu'une seule copie d'ADN (versus deux copies à l'état normal), le LRR sera plus faible aux positions affectées par un tel CNV (Figure 3b). Le scénario inverse se produit pour les duplications : une troisième copie d'ADN est présente, et donc le LRR sera plus élevé (Figure 3b).

La deuxième information donnée par les puces à ADN est la fréquence de l'allèle B (BAF) (73). Il s'agit du ratio normalisé de $\theta = \arctan((a/b)/(\pi/2))$ pour chaque sonde. Elle représente le ratio de l'allèle B à une position donnée, sachant que deux allèles sont possibles : A ou B. Pour une position donnée, un individu dans l'état normal (ayant donc deux copies d'ADN à chaque position de son génome) a donc trois génotypes possibles : AA (BAF = 0,0), BB (BAF = 1,0) ou AB

(BAF = 0,5) (Figure 3a) (3). Un individu portant une délétion n'aura que deux options possibles : A ou B, n'ayant qu'une copie à cette position. Un individu avec une duplication aura 4 génotypes possibles : AAA, BBB, ABB ou AAB.

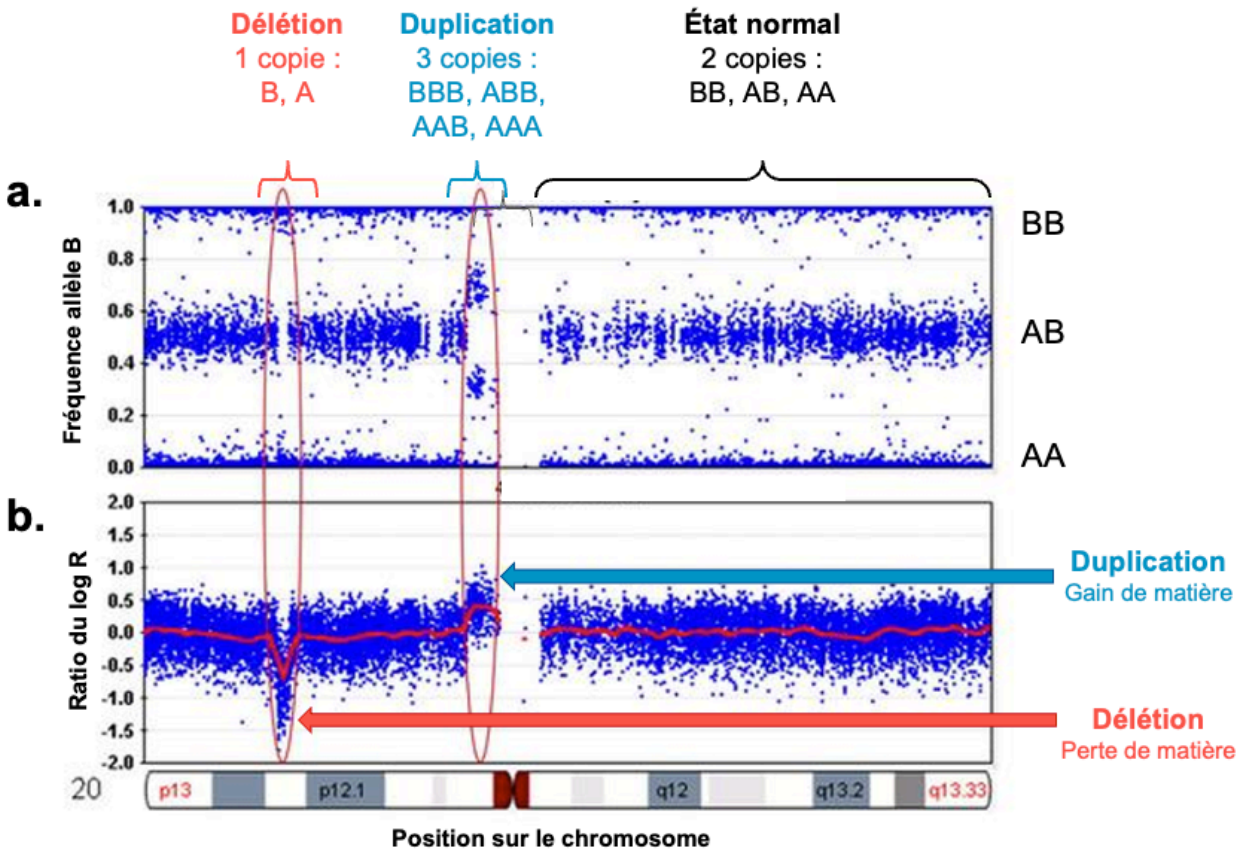


Figure 3. – Représentation graphique des informations données par les puces à ADN. a. Fréquence de l'allèle B (BAF). À l'état normal, pour deux allèles (soient A et B), nous avons deux copies d'ADN, et donc trois génotypes possibles (BB : BAF=1,0; AB : BAF=0,5; AA : BAF=0,0). Une duplication représente une copie de plus que l'état normal, soit trois copies avec 4 génotypes possibles (BBB : BAF=1,0; ABB : BAF=0,6; AAB : BAF=0,3; AAA : BAF=0,0) alors qu'une délétion représente une copie de moins, soit une copie d'ADN et seulement 2 génotypes possibles (B : BAF=1,0; A : BAF=0,0). b. Le ratio du log R (LRR) représente l'intensité normalisée du signal. Il augmente par rapport à l'état normal en présence d'une duplication, alors qu'en présence d'une délétion, il diminue. Adapté de Scientific protocols (74).

Autant pour la CGH que les SNP arrays, la détection des CNV via ce type de données se fait principalement par des algorithmes qui utilisent des modèles statistiques. Le modèle le plus utilisé pour faire la détection des CNV est le modèle de Markov caché (HMM) (75).

Les modèles de Markov sont des modèles probabilistes qui permettent de déterminer les états d'une séquence, où chaque état est déterminé en se basant uniquement sur l'état qui le précède. Dans le cas des HMM, les états sont inconnus de l'utilisateur et sont déterminés via les probabilités d'émission et de transition pour chaque état. L'intensité des signaux des puces à ADN, c'est-à-dire le LRR et la BAF sont utilisés dans ce type de modèle. Ces signaux permettent de déterminer l'état de la position de chaque SNP (normal, délétion ou duplication) le plus probable par programmation dynamique. L'utilisation des HMM pour détecter les CNV via l'ACPA est peu affectée par les faux positifs et est rentable au niveau du coût, mais les faux négatifs, souvent des CNV omis en raison de leur petite taille et donc identifiable par un nombre limité de sondes, y sont abondants (73,75). Il est donc recommandé d'utiliser plusieurs outils différents sur les données afin d'obtenir un résultat plus fiable.

2.5 Mesure de l'impact des CNV sur la cognition

2.5.1 Vers un modèle quantitatif

En clinique, la dangerosité des CNV sur la cognition et les troubles psychiatriques est principalement mesurée de façon qualitative (*i.e.*, bénins, pathogéniques ou VUS) (31). Les patients référés en génétique pédiatrique pour un TND sont porteurs d'un CNV pathogénique dans 15% des cas (8). Parmi eux, 90% ne sont observés que chez un ou peu de patients (25). Ils ne sont donc que très peu, ou pas du tout étudiés dans la littérature par manque d'observations et donc de puissance statistique : leur effet reste donc mal connu. Par conséquent, il est difficile pour les généticiens et les cliniciens d'estimer la contribution de ces variants aux symptômes d'un patient donné. Ceci qui pose un problème majeur, car de tels CNV sont reportés quotidiennement chez les patients référés en génétique clinique pour des TND. De plus, puisque les patients atteints de TND présentent des symptômes inclus dans un large spectre, tant au niveau cognitif que dans leur sévérité, le fait de classifier un CNV comme étant bénin, pathogénique ou VUS est limitant, car il ne renseigne pas le clinicien quant aux tailles d'effet de ce CNV sur ces différents phénotypes

cliniques. Malheureusement, le peu d'études qui documentent l'effet quantitatif des CNV sur les traits cognitifs sont limitées aux CNV les plus fréquents (*e.g.*, 22q11.2, 16p11.2), et lorsque cet effet est documenté, il est souvent surestimé, car il provient d'études sur des patients suivis en clinique (76).

En 2012, un premier modèle tentant de mesurer quantitativement l'impact des CNV sur le QI a été développé par Girirajan et al. (77). Ce modèle montre une association entre le nombre de gènes inclus dans les CNV et le QI. Toutefois, cette stratégie n'a pas montré d'estimé robuste et elle ne permet pas de faire de prédiction de l'impact des CNV sur les traits cognitifs.

Depuis les 10 dernières années, plusieurs scores d'haploinsuffisance ont été développés pour mesurer la pathogénicité des gènes (Tableau 3). L'haploinsuffisance est la situation où un gène n'est présent qu'en une seule copie (la deuxième étant délétée ou inactivée), et où le produit de cette unique copie n'est pas suffisant pour permettre d'accomplir adéquatement la fonction du gène (27). En 2010, Huang, Lee, Marcotte, & Hurlise et al. ont développé un algorithme déterminant le score d'haploinsuffisance, %HI, d'un gène (78). Cet algorithme utilise l'analyse discriminante linéaire afin d'établir la probabilité qu'un gène soit tolérant ou non à la perte de fonction en se basant sur la conservation de séquence du gène chez l'humain comparée à celle du macaque, sur la conservation du promoteur, l'expression embryonnaire et le niveau de ressemblance avec les gènes dont l'haploinsuffisance est connue. Ils ont montré que ce score permet de mieux prédire la pathogénicité des CNV que leur taille ou encore le nombre de gènes qu'ils emportent. Ce score linéaire s'étend de 0% à 100%, et les gènes avec un score inférieur à 10% sont considérés comme étant intolérants à la perte de fonction. Cependant, ce score est basé sur des CNV qui sont observés chez des patients en cliniques (provenant de la base de données DECIPHER (11)). Il existe donc un biais lorsque le score est appliqué à des CNV provenant de la population générale.

En 2016, un nouveau score d'haploinsuffisance a été inventé par Lek et al. (2016) : le score de probabilité d'intolérance à la perte de fonction (pLI). Ce score est basé sur le séquençage de

l'exome de plus de 60 706 individus provenant du Exome Aggregation Consortium (ExAC). Il représente le ratio des observations de variants délétères (défini) sur ce qui est attendu pour un gène donné par rapport à une population de référence. Dès lors, si on observe moins de variants délétères par rapport à ce qui avait été estimé, cela signifie qu'il y a une pression de sélection négative sur ce gène, réduisant la fitness (définie par la « mesure de l'habileté d'un individu à survivre et à se reproduire ») des porteurs de tels variants. Autrement dit, comme son nom l'indique, le pLI représente la probabilité qu'un gène soit intolérant à la perte de fonction. Il s'étend de 0 à 1 : les gènes avec un petit pLI ($\leq 0,1$) sont tolérants à la perte de fonction des protéines, alors que les gènes avec un grand pLI ($\geq 0,9$) sont intolérants à cette perte de fonction. Il s'agit d'un score ayant essentiellement une distribution binaire.

En 2020, un autre score utile pour estimer la taille d'effet des CNV a été développé, le LOEUF, qui représente la borne supérieure du ratio entre la perte de fonction observée et la perte de fonction prédite dans la population générale (79). Ce score, qui s'étend de 0 à 2, est linéaire, ce qui est plus ou moins le cas du pLI, dont l'interprétation est principalement binaire. Un petit LOEUF ($< 0,35$) représente des gènes intolérants à la perte de fonction des protéines, tandis qu'un grand LOEUF ($\geq 0,35$) représente des gènes tolérants (80). Puisque cette échelle est à l'inverse du pLI au niveau de son interprétation, l'inverse du LOEUF ($1/\text{LOEUF}$) est parfois utilisé. Une limite de ce score est que pour les petits gènes, le nombre de mutations attendues dans la population générale est très petit. Le ratio de variants observés/attendus est donc peu précis et peu informatif pour ces gènes (79).

Depuis plusieurs années, l'équipe du Dr Jacquemont a développé un modèle statistique qui permet d'estimer l'impact de tous les CNV autosomiques sur les déficits cognitifs (basés sur le QI) et sur le risque d'autisme (25,76,81). L'avantage d'utiliser les CNV plutôt que les SNP dans le modèle est qu'il n'y a pas d'ambiguïté en ce qui concerne leur effet : les délétions diminuent toujours le nombre de transcrits, tandis que les duplications l'augmentent (82). De plus, comparés au SNP, les CNV montrent moins d'ambiguïté dans l'identification des gènes touchés en particulier pour les délétions. Le modèle statistique développé se base sur un modèle de

régression linéaire qui utilise la somme des pLI ou des LOEUF scores des gènes présents dans les CNV. En 2018, Hugué et al. ont montré que le pLI estime et prédit le mieux la taille d'effet des délétions sur le QI en comparaison à 9 autres scores génétiques et fonctionnels. Les modèles testés avaient été entraînés sur 4 073 individus de la population générale (*i.e.*, Imagen (83) et Saguenay Youth Study (84)). Le modèle établi utilise la somme de ce score d'intolérance des gènes présents séparément pour les délétions et les duplications chez l'individu (Figure 4). Ils ont montré que pour les délétions, chaque point de pLI perdu était associé à une perte de 2,74 ($p=8 \times 10^{-5}$) points de QI de performance. Récemment, le modèle a été répliqué sur 8 cohortes provenant de la population générale et de cohortes autistiques (*i.e.*, IMAGEN (83), Saguenay Youth Study (SYS) (84), CARTaGENE (CaG) (85), Lothian Birth Cohort (LBC) (86), Generation Scotland (G-Scot) (87), Simons Simplex Collection (SSC) (88), MSSNG (89) et une cohorte neurodéveloppemental de Ste-Justine), pour un total de 24 092 individus. Cela représente une

Score	Description	Limites	Échelle	Seuil d'intolérance	Année de parution
HI%	Analyse discriminante linéaire basée sur la conservation humain-macaque, la conservation du promoteur, l'expression embryonnaire et la ressemblance avec les gènes dont l'haploinsuffisance est connue.	Basé sur des données cliniques, peu applicable à la population générale.	0-100	≤ 10	2010
pLI	Ratio des observations de variants délétères sur ce qui est attendu pour un gène donné par rapport à une population de référence.	Score binaire.	0-1	$\geq 0,9$	2016
LOEUF	Borne supérieure de la perte de fonction observée/attendue dans la population générale.	Peu précis pour les petits gènes.	0-2	$\leq 0,35$	2020

Tableau 3. – Émergence des scores d'intolérance à la perte de fonction des gènes.

cohorte 9 fois plus grande que celle de la première étude. Pour les délétions, la méta-analyse montre une perte de 2,70 points de QI non-verbal par point de pLI, soit -0,18 de z-score (IC : -0,21; -0,14) (Figure 5a). De plus, ils observent pour la première fois un signal significatif pour les duplications, soit une perte de 0,75 points d'intelligence générale est observée par point de pLI, ce qui représente -0,05 de z-score (IC : -0.07; -0.04) (Figure 5b). Cela représente une taille d'effet trois fois plus faible que pour les délétions. Une fois les cohortes regroupées en une seule analyse, on obtient une perte de 2,64 ($p=1,25 \times 10^{-28}$) points de QI par point de pLI pour les délétions et 0,81 ($p=1.90 \times 10^{-9}$) points de QI pour les duplications. Les résultats n'étaient pas différents entre les cohortes provenant de la population générale versus celles provenant de cohortes avec TSA; et restent similaires en utilisant le score LOEUF.

Les analyses de concordance ont démontré une précision de 78% de ces modèles pour estimer l'effet de 47 CNV sur le QI en comparaison à des mesures empiriques déjà publiées (Figure 6).

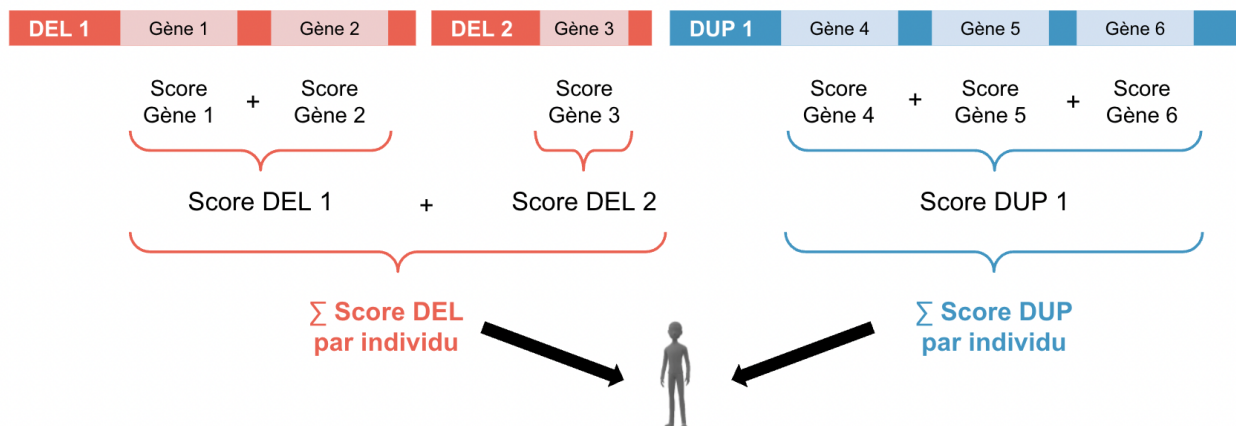


Figure 4. – Modèle additif des scores d'intolérance des gènes. Deux scores sont attribués pour chaque individu : la somme d'intolérance des gènes pour les délétions (rouge) et celle pour les duplications (bleu). Ce score représente la somme du score (pLI ou LOEUF) de chaque gène emporté par les délétions ou, séparément, pour les duplications d'un individu. Adapté de Douard et al. (2021) (81).

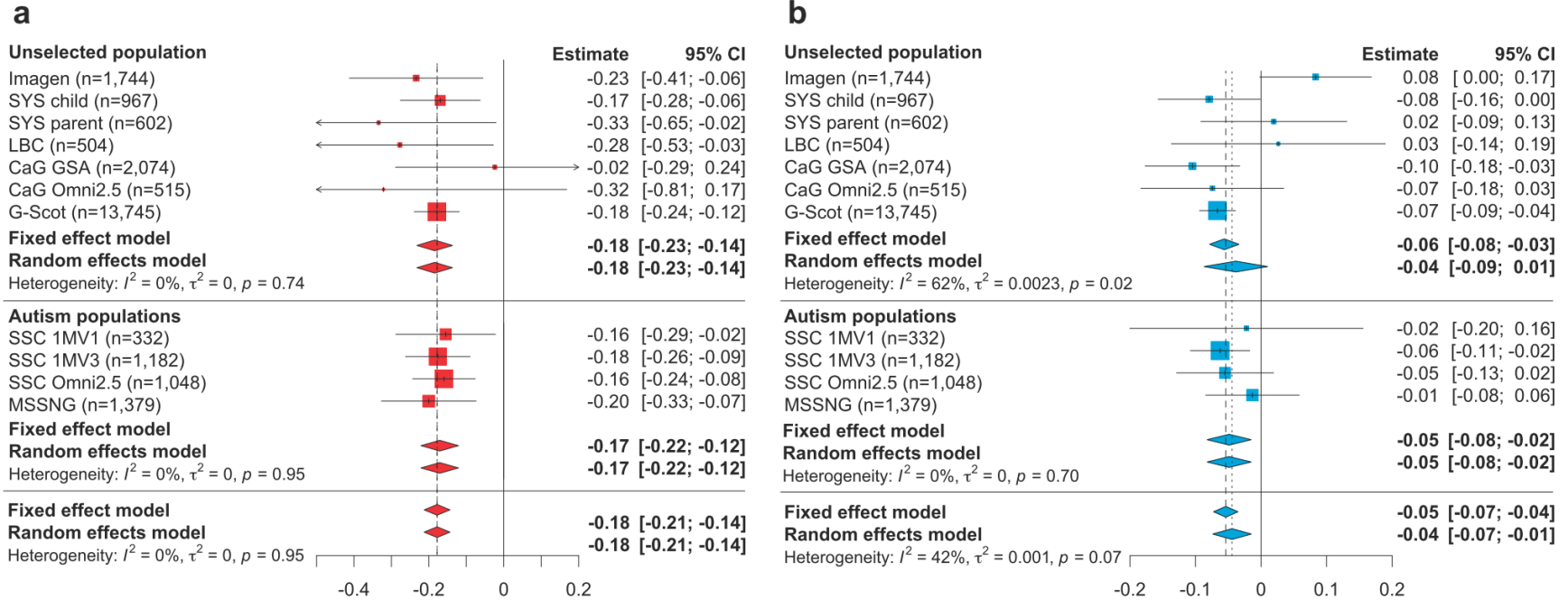


Figure 5. – Effet du pLI sur l'intelligence générale pour les délétions et les duplications. Méta-analyse de l'estimation de l'impact des délétions en rouge (a) et des duplications en bleu (b) estimée via la somme du pLI des gènes inclus dans les CNV sur l'intelligence générale. La perte d'un point de pLI abaisse l'intelligence générale de 0,18 z-score pour les délétions, soit 2,70 points de QI, et de 0,05 z-scores, soit 0,75 points de QI pour les duplications. L'axe des X représente le z-score de l'intelligence générale. Les carrés représentent la taille d'effet pour chaque cohorte. Les losanges représentent l'effet global entre cohortes. Figure tirée de Huguet et al. (2021) (76).

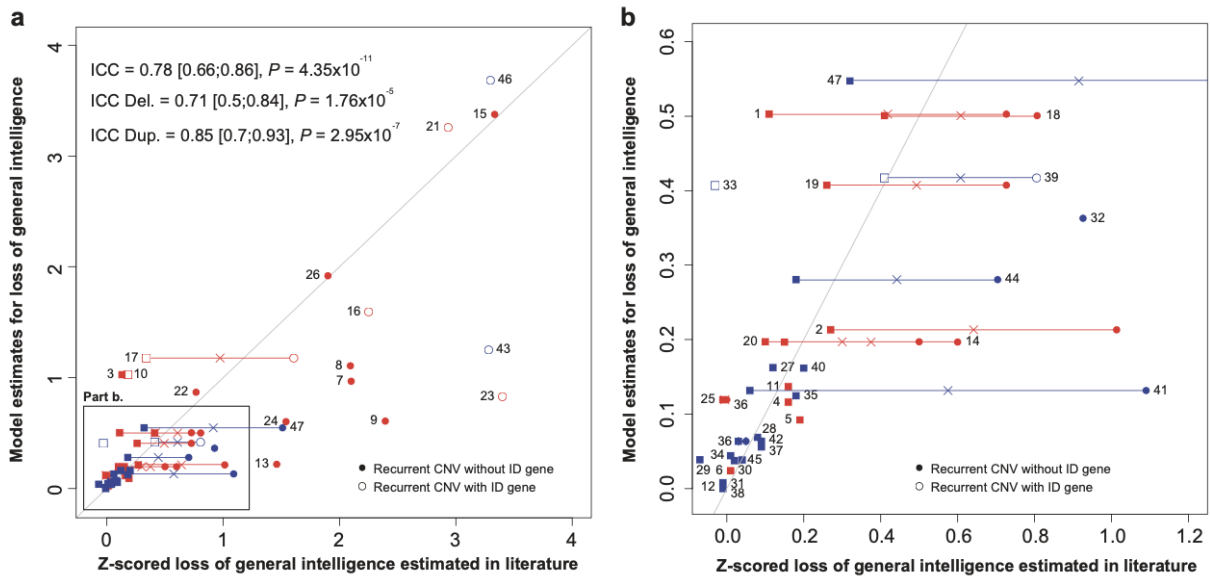


Figure 6. – Concordance entre les prédictions du modèle additif des scores d’intolérance sur l’intelligence et l’effet de CNV rapportés dans la littérature. a. L’axe des X et des Y représentent respectivement la taille d’effet des CNV sur le score z de l’intelligence générale. Les carrés représentent les CNV vus dans les cohortes cliniques et les cercles représentent les CNV provenant de la cohorte UKBB. Le rouge et le bleu représentent respectivement les délétions et les duplications. Les croix représentent la taille moyenne de la perte de QI en score z pour 13 CNV récurrents vus dans les cohortes cliniques et dans UKBB. b. Agrandissement du rectangle en bas à gauche de la Figure 6a. Tiré de Huguet et al. 2021 (76).

2.5.2 Vers un modèle résolutif

Une limite du modèle additif établi par l’équipe du Dr Jacquemont est qu’il manque de recul au niveau des effets individuels de chaque gène. En effet, en observant l’effet additif de l’intolérance des gènes présents dans un CNV, il est possible que le signal de certains gènes soit écrasé par celui des autres gènes qui y sont également présents. Par exemple, si un CNV contient un gène très intolérant, le score de ce gène risque d’être si élevé que l’impact des autres gènes sera négligeable et ainsi masqué. De façon similaire, si des gènes ont un effet opposé sur le phénotype lorsqu’ils sont emportés par des CNV, cet effet ne sera pas visible par le modèle, qui est défini dans une seule dimension. En outre, le modèle suppose que l’effet des gènes est linéaire, alors

que ce n'est pas nécessairement le cas. Finalement, le modèle se limite aux autosomes, n'incluant pas les CNV des chromosomes sexuels.

Pour surmonter les limites du modèle additif, une vision plus résolutive du dosage génique sur la cognition et les troubles psychiatriques est nécessaire. Depuis quelques années, plusieurs stratégies ont été développées pour quantifier les effets spécifiques des gènes emportés par les CNV sur les phénotypes psychiatriques. En 2010, Pinto et al. ont publié une GWAS évaluant l'impact des CNV rares sur l'autisme (62). Cette étude a permis non seulement de révéler l'impact de CNV récurrents sur l'autisme, mais aussi celui des CNV rares, ce qui n'avait pas été fait auparavant. En 2017, Marshall et al. ont publié une étude d'association pangénomique évaluant l'impact des gènes inclus dans les CNV sur la schizophrénie (90). Cette approche a permis de révéler de nouvelles régions impliquées dans la schizophrénie, en plus de régions candidates dont l'effet serait protecteur. Le fait d'utiliser une approche pangénomique par gène permet de mieux identifier les mécanismes biologiques qui mènent au phénotype. Dans le cas de la schizophrénie, les gènes emportés par des CNV étaient enrichis pour la fonction synaptique et les comportements neurodéveloppementaux chez la souris (90). À ce jour, il n'existe pas de travaux publiés utilisant une approche pangénomique sur les CNV sur l'intelligence générale.

En plus des GWAS, une approche plus résolutive pour mesurer l'impact du dosage génique sur l'intelligence générale est l'utilisation d'une sliding-window sur les différentes fenêtres d'intolérance des gènes (Figure 7). Cette approche préliminaire développée par le laboratoire du Dr Jacquemont consiste à réaliser un modèle linéaire pour chaque fenêtre de LOEUF, avec deux prédicteurs sur l'intelligence générale : le nombre de gènes à l'intérieur de la fenêtre versus le nombre de gènes à l'extérieur de la fenêtre. Cette méthodologie permet de capturer l'effet spécifique des gènes sur l'intelligence générale en fonction de leur catégorie d'intolérance. Ainsi, si l'effet est expliqué par un sous-groupe de gènes (*e.g.*, les gènes intolérants), cette méthodologie plus résolutive est en mesure de le capturer.

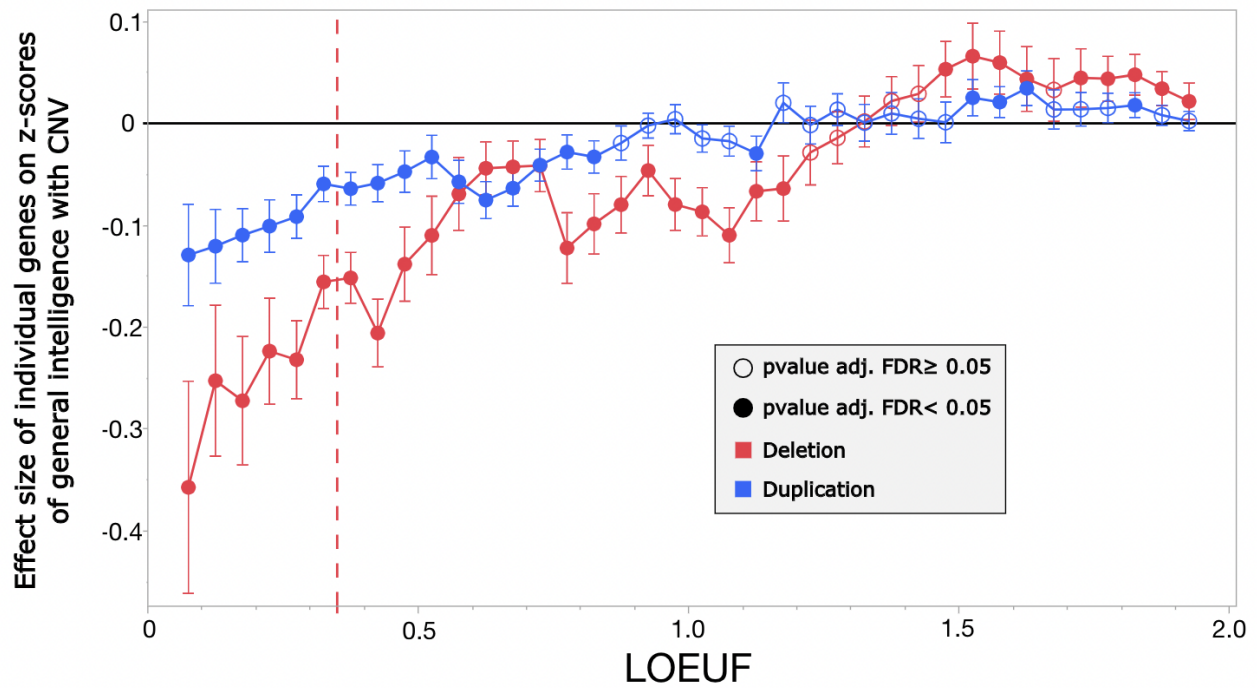


Figure 7. – Effet de grandeur des gènes individuels sur l’intelligence générale lorsqu’ils sont affectés par le dosage génique. Le modèle comprend deux prédicteurs sur l’intelligence générale : le nombre de gènes à l’intérieur de la fenêtre versus le nombre de gènes à l’extérieur de la fenêtre. Les points rouges représentent les modèles sur les délétions et les bleus représentent ceux sur les duplications. Les points remplis représentent les tests significatifs ($FDR < 0,05$). La ligne pointillée rouge représente le seuil d’intolérance des gènes ($LOEUF < 0,35$). Figure préliminaire non publiée réalisée par Huguet G. (2022).

Chapitre 3 – Problématique et hypothèses de travail

3.1 Problématique et hypothèse

3.1.1 Autosomes vs Chromosome X

Comme mentionné ci-dessus, un modèle de régression linéaire utilisant des scores d'intolérances a déjà été établi dans le but d'estimer l'impact des CNV sur le QI. Néanmoins, ce modèle n'est pas adapté pour les CNV présents sur le chromosome X pour deux raisons.

Premièrement, le modèle ne considère pas l'aspect du sexe des individus pour estimer l'effet de ces CNV. Dès lors, cet aspect doit être pris en compte, puisqu'il n'y a qu'un seul chromosome X chez l'homme, alors que la femme en a deux. Chez cette dernière, l'un des deux chromosomes X est inactivé durant le développement précoce de l'embryon (91). Normalement, l'inactivation du chromosome X est aléatoire (Figure 8a), mais trois mécanismes ont montré qu'elle peut aussi être spécifique, soient l'inactivation non-aléatoire primaire (Figure 8b), primaire biaisée (Figure 8c) ou secondaire (Figure 8d). L'inactivation non-aléatoire secondaire mène à une inactivation préférentielle pour les chromosomes X affectés par des variants réduisant la fitness de l'individu. Il s'agit donc d'un mécanisme protecteur présent uniquement chez la femme.

La deuxième raison est que l'analyse du chromosome X nécessite une cohorte de très grande taille. En effet, l'humain n'a qu'un ou deux chromosomes X (respectivement pour l'homme et la femme), versus les 44 autosomes étudiés par le modèle déjà établi. En comparant la somme de la taille de tous les autosomes versus la taille du chromosome X, pour un même nombre d'individus, cela représente un échantillon qui inclut 18 à 37 fois moins de matière (respectivement pour un échantillon ne contenant que des femmes ou que des hommes) pour le chromosome X. L'échantillon utilisé pour tester le modèle sur le chromosome X devrait donc être considérablement plus grand que celui utilisé pour les autosomes pour obtenir une puissance statistique similaire.

Notre hypothèse est que l'impact des CNV du chromosome X sur la cognition est proportionnel à la somme de l'intolérance des gènes qui y sont présents pour les femmes (sans mécanisme d'inactivation spécifique), tout comme pour les autosomes, alors que pour les hommes, cet effet serait plus sévère en raison de l'absence des mécanismes d'inactivation du chromosome X qui ont un effet protecteur chez la femme.

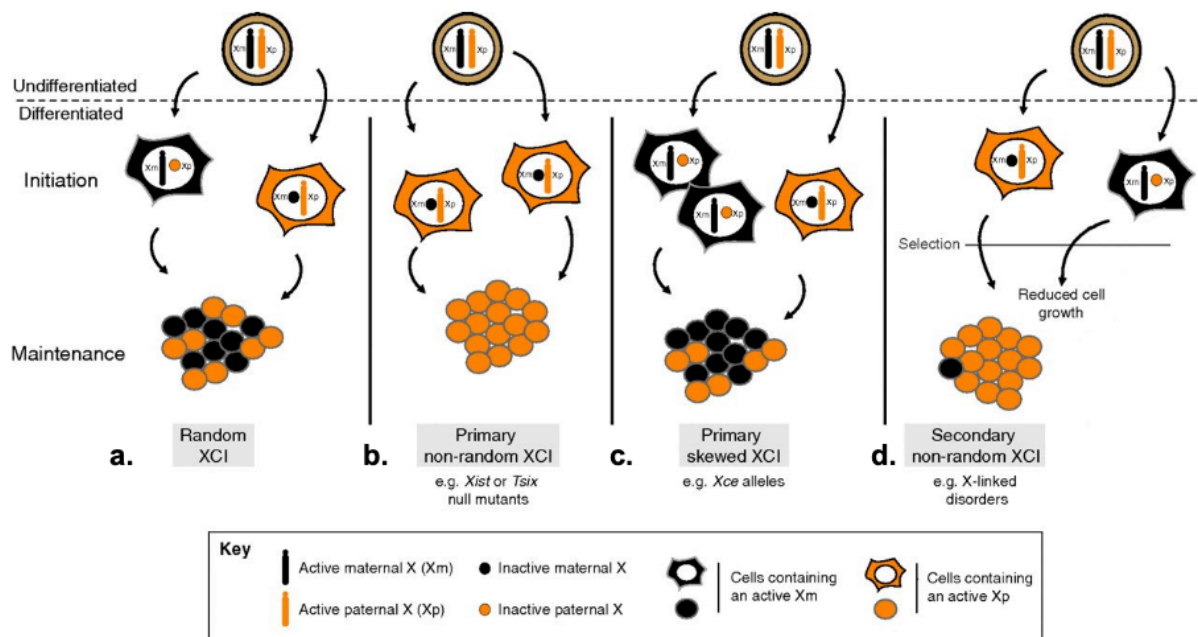


Figure 8. – Mécanismes d'inactivation du chromosome X. XCI : Inactivation du chromosome X. Xm : chromosome X maternel, Xp : chromosome X paternel. a. L'inactivation aléatoire mène à une population mosaïque de cellules où soit Xm ou Xp est actif. b. L'inactivation non aléatoire primaire est liée à une perturbation de gènes impliqués dans la l'inactivation du X (Xist et Tsix) et mène à inactivation préférentielle complète. c. L'inactivation non aléatoire primaire biaisée est liée à une perturbation de régions régulatrices de l'inactivation et mène à un nombre inégal de cellules ayant inactivé l'un des deux chromosomes X. d. L'inactivation non aléatoire secondaire inactive préférentiellement les chromosomes X qui portent une mutation qui n'est pas liée directement à cette inactivation, mais qui affecte la fitness de l'individu. Il s'agit d'un mécanisme de sélection qui peut être protecteur, puisque l'inactivation préférentielle réduit l'impact du variant délétère. Adapté de Gendrel et Heard 2011 (91).

3.1.2 Modèle additif vs modèle résolutif

Une deuxième problématique est qu'en observant l'effet additif de l'intolérance des gènes présents dans un CNV, il est possible que le signal de certains gènes soit surestimé ou sous-estimé puisque le modèle estime l'effet moyen de l'ensemble des gènes. Dès lors, le modèle ne permet pas de capturer l'impact spécifique des gènes emportés par un CNV.

Notre hypothèse est qu'en étudiant l'effet des gènes individuellement selon leurs caractéristiques biologiques, le modèle sera plus résolutif et permettra de cibler certains gènes spécifiques ayant un impact sur l'intelligence générale et sur le risque de TSA lorsqu'ils sont emportés par un CNV.

3.2 Objectifs

3.2.1 Objectif 1 : Impact d'un CNV dans le chromosome X sur la cognition

Notre premier objectif est d'estimer avec la même stratégie que pour les autosomes l'impact des CNV dans le chromosome X sur la cognition. Les modèles statistiques utilisés seront ceux déjà établis par le laboratoire.

3.2.2 Objectif 2 : Impact des gènes individuels sur la cognition et le risque de TSA

Notre deuxième objectif est d'appliquer les modèles statistiques déjà établis par le laboratoire individuellement sur chaque gène présent dans les CNV des individus étudiés. Comme mentionné ci-dessus, les modèles sont déjà établis. Ils seront modifiés afin d'établir l'impact d'un gène sur l'intelligence générale ou le risque de TSA pour les individus qui portent ce gène dans un CNV par rapport à ceux qui ne portent pas le gène en question dans un CNV. Particulièrement, nous porterons une attention particulière aux gènes tolérants qui, lorsqu'ils sont inclus dans une délétion ou une duplication, ont un effet positif sur le QI.

Chapitre 4 – Méthodologie

4.1 Cohortes

Le projet inclut six cohortes provenant de la population générale ainsi que deux cohortes provenant de populations autistiques, pour un total de 457 398 individus non-apparentés d'origine européenne après contrôles de qualité (Tableau 4). Pour les analyses sur l'intelligence générale, 263 396 individus provenant de ces cohortes avec une mesure de l'intelligence disponible après contrôles de qualité ont été utilisés (Tableau 5). Le consentement éclairé à été donné pour tous les participants ou leur tuteur légal inclus dans ce projet.

4.2 Mesure de l'intelligence générale

L'intelligence générale a été mesurée à l'aide de trois mesures différentes : le QI a été utilisé lorsqu'il était disponible, sinon, le g-facteur et l'intelligence fluide (FI, fluid intelligence) ont été utilisés. Ces mesures ont été normalisées en score Z afin de les rendre comparables. À noter que pour le QI, la transformation en score Z est faite sur les valeurs standardisées du QI (moyenne de 100 et écart-type de 15). Pour les autres mesures, les valeurs de références sont sur les cohortes en question.

Le QI dans ce projet fait référence au QI non verbal. Il a été mesuré à l'aide de différents tests psychométriques adaptés au niveau développemental du participant. Pour les cohortes provenant de la population générale (Imagen, SYS et LBC), l'échelle d'intelligence de Wechsler pour enfants (WISC-III, WISC-IV) et le Test Moray House ont été utilisés. Pour les cohortes autistiques (SSC et SPARK), les tests qui ont été utilisés sont le Mullen Scales of Early Learning, le Leiter International Performance Scale, le Raven progressive matrices, le Stanford-Binet Intelligence Scale, le WISC-IV et le WISC-V, le Wechsler Abbreviated Scale of Intelligence (WASI-I, WASI-II), le Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV), le Leiter-reviewed et le Differential Ability Scales (DAS-II).

Cohorte	Technologie	n=457 398	n Femme (%)	Âge moyen (SD)	Phénotype (% avec mesure)	Score Z aj. (SD)
Cohorte de la population générale (n=450 187)						
Imagen	610Kq-660Wq	1 669	846 (51)	14,4 (0,4)	QI (97)	0,46 (0,97)
SYS (parents)	HOE-12V	600	320 (53)	49,5 (4,9)	G-facteur (100)	0,006 (0,94)
LBC	610Kq	551	270 (49)	70,0 (0,0)	QI (90)	0,05 (0,96)
CaG	GSA	4 908	2 905 (59)	55,2 (7,8)	G-facteur (40)	0,36 (0,97)
	Omni2,5	624	341 (55)	53,6 (8,9)	G-facteur (79)	-0,02 (0,95)
G-Scot	610Kq	5 673	3 148 (55)	48,3 (15,2)	G-facteur (99)	-0,26 (1,01)
UKBB	Affymetrix	290 067	157 627 (54)	65,2 (7,6)	G-facteur (10)	0,09 (0,91)
		146 095	79 285 (54)	58,1 (8,2)	FI (100)	0,08 (0,98)
Cohortes autistiques (n=7 211)						
SSC	1MV1	279	37 (13)	9,8 (3,3)	QI (100)	-0,49 (1,61)
	1MV3	916	124 (13)	8,9 (3,6)	QI (100)	-0,91 (1,64)
	Omni2,5	791	105 (13)	9,4 (3,8)	QI (99)	-1,08 (1,83)
SPARK	GSA	5 225	1 089 (21)	10,9 (6,1)	QI (35)	-0,50 (1,94)

Tableau 4. – Description des cohortes pour les analyses sur le risque de TSA. Seuls les individus d'origine européenne sont inclus dans cette analyse. Imagen : cohorte d'adolescents non reliés (83). Saguenay Youth Study (SYS) : cohorte d'adolescents canadiens-français et leurs parents (84). Lothian Birth Cohort (LBC) : cohorte incluant presque tous les enfants nés en 1936 en Écosse (86). CartaGene (CaG) : cohorte de canadiens-français non apparentés. Generation Scotland (G-Scot) : cohorte de familles provenant du Royaume-Uni. UKBiobank (UKBB) : cohorte adulte provenant du Royaume-Unis (92). Simon Simplexe Collection (SSC) : cohorte de familles avec au moins un enfant autiste (88). Simons Foundation Powering Autism Research (SPARK) : cohorte de familles avec au moins un enfant autiste (93). n : nombre d'individus après les filtres de contrôles de qualité; SD : Écart-type; Score Z aj. : score Z de l'intelligence ajusté (âge et âge au carré); QI : quotient intellectuel; FI : intelligence fluide.

Cohorte	Technologie	n=263 396	n Femme (%)	Âge moyen (SD)	Phénotype	Score Z aj. (SD)
Cohorte de la population générale (n=258 292)						
Imagen	610Kq-660Wq	1 744	891 (51)	14,4 (0,4)	QI	0,44 (0,98)
SYS (enfants)	610Kq; HOE-12V	967	505 (52)	15,0 (1,8)	QI	0,30 (0,87)
SYS (parents)		598	319 (53)	49,5 (4,9)	G-facteur	0,34 (0,93)
LBC	610Kq	503	246 (49)	70,0 (-)	QI	0,05 (0,95)
CaG	GSA	2 074	1 094 (53)	54,3 (7,6)	G-facteur	0,36 (0,97)
	Omni2,5	515	281 (55)	52,4 (8,6)		0,24 (0,96)
G-Scot	610Kq	13 715	8 081 (59)	46,7 (15,0)	G-facteur	0,34 (0,97)
UKBB G-facteur	Affymetrix	30 794	15 788 (51)	65,1 (7,6)	G-facteur	0,51 (0,89)
UKBB G-facteur online		36 875	20 452 (55)	61,6 (7,5)		0,32 (0,93)
UKBB FI		156 734	85 201 (54)	57,8 (8,2)	FI	0,30 (0,97)
UKBB FI online		13 773	8 284 (60)	64,2 (7,7)		0,34 (0,97)
Cohortes autistiques (n= 5 104)						
SSC	1MV1	332	44 (13)	9,5 (3,2)	QI	-0,60 (1,56)
	1MV3	1 181	157 (13)	8,8 (3,5)		-0,98 (1,64)
	Omni2,5	1 048	140 (13)	9,2 (3,7)		-1,23 (1,83)
SPARK	GSA	2 543	540 (21)	12,4 (6,2)	QI	-0,62 (1,96)

Tableau 5. – Description des cohortes pour les analyses sur la cognition. Imagen : cohorte d'adolescents non reliés (83). Saguenay Youth Study (SYS) : cohorte d'adolescents canadiens-français et leurs parents (84). Lothian Birth Cohort (LBC) : cohorte incluant presque tous les enfants nés en 1936 en Écosse (86). CartaGene (CaG) : cohorte de canadiens-français non apparentés (85). Generation Scotland (G-Scot) : cohorte de familles provenant du Royaume-Uni (87). UKBiobank (UKBB) : cohorte adulte provenant du Royaume-Unis (92). UKBB online : individus de UKBB dont le test cognitif a été fait en ligne. Simon Simplexe Collection (SSC) : cohorte de familles avec au moins un enfant autiste (88). Simons Foundation Powering Autism Research (SPARK) : cohorte de familles avec au moins un enfant autiste (93). n : nombre d'individus après les filtres de contrôles de qualité; SD : Écart-type; Score Z aj. : score Z de l'intelligence ajusté (origine ancestrale (PC 1-10), sexe, âge, âge au carré, type de test cognitif); QI : quotient intellectuel; FI : intelligence fluide.

Le g-facteur est une mesure indirecte de l'intelligence. Il représente la première composante de l'analyse en composantes principales (PCA) de différents tests cognitifs standardisés. Il a été calculé préalablement dans Huguet et al. (2021) (76), séparément pour les différentes cohortes, qui n'incluent pas nécessairement les mêmes tests cognitifs (voir Huguet et al. 2021). La FI, utilisée pour 170 507 individus de UKBB, représente l'un de ces tests cognitifs, qui est la somme de bonnes réponses sur 13 questions de raisonnement et de résolution de problèmes. Les études précédentes utilisant ces données ont montré que le QI et le g-facteur sont très corrélés et définissent les mêmes amplitudes d'impact des CNV sur les habiletés cognitives (25,76). Ainsi, il s'agit d'une mesure robuste des habiletés cognitives, qui n'est pas sensible aux différents tests cognitifs utilisés (76).

Deux méthodes ont été employées pour ajuster l'intelligence en fonction de l'âge et du sexe. La première, décrite dans les précédents travaux de Huguet et al. (2021) (76), consiste à extraire les résidus du modèle linéaire sur le phénotype (le g-facteur ou la FI) ajusté pour l'âge en covariables, séparément pour chaque cohorte. La mesure extraite correspond au g-facteur ou à la FI prédits par le modèle pour l'âge moyen de la cohorte + les résidus pour chaque individu. La même stratégie est utilisée pour faire la correction selon le sexe.

La deuxième méthode qui a été testée pour ajuster le phénotype en fonction de l'âge consiste à utiliser une régression par processus gaussien (GP) (94). Il s'agit d'une approche d'apprentissage machine qui vise à trouver la fonction décrivant le mieux les données en assumant la présence de bruit gaussien. L'avantage de cette méthode est qu'elle ne prend pas pour acquis que la correction est linéaire. Cette approche semble donc adaptée à l'intelligence, dont le déclin s'accroît avec le vieillissement. La librairie Python PyNM (95) a été utilisée pour appliquer cet ajustement.

4.3 Mesure de l'information génétique

4.3.1 Détection et filtrage des CNV

Les données utilisées proviennent toutes de Huguet et al. (2021) (76). Les paramètres de détection et de filtrage des CNV y sont détaillés. Ces étapes ont déjà été réalisées de façon préalable sur ce projet. Les qualités des puces utilisées devaient être pour le call rate $\geq 95\%$; écart-type du LRR < 0.35 ; écart-type de la BAF $< 0,08$; |facteur d'ondulation| $< 0,05$.

Les outils utilisés pour faire la détection des CNV via les données de génotypage sont PennCNV (96) et QuantiSNP (97), deux algorithmes utilisant les HMM, avec les critères de détection suivants pour les CNV : ≥ 3 sondes consécutives, $\geq 1\text{Kb}$ et le score de confiance du CNV ≥ 15 . Les résultats des deux outils ont ensuite été combinés par l'algorithme CNVision (98).

Les CNV détectés ont ensuite été filtrés en fonction de leur qualité, selon les critères suivants : taille $> 50\text{ kb}$, type de CNV similaire entre les algorithmes, chevauchement $< 50\%$ avec les régions HLA, les duplications segmentaires ou les régions centromériques, sondes consécutives ≥ 10 à travers l'ensemble des technologies et score de confiance ≥ 30 . Puis, un modèle d'apprentissage en profondeur entraîné pour définir les vrai CNV sur une base de plus de 34 000 CNV visualisés manuellement à travers les différentes technologies de puce a été appliqué pour retirer le plus d'artéfacts de CNV possible.

Toutes les positions mentionnées des CNV sont données selon la version hg19 de la carte du génome et l'annotation des gènes via la base de données ENSEMBL (<https://grch37.ensembl.org/index.html>) (99).

4.4 Analyses statistiques

4.4.1 Modélisation de l'impact des CNV sur l'intelligence

4.4.1.1 Correction du score Z de l'intelligence par l'âge et le sexe

Le premier modèle a été réalisé afin de comparer les deux types de corrections (correction par les résidus du modèle linéaire vs correction par GP) sur les valeurs d'habileté intellectuelle. Pour ce faire, nous avons regardé l'impact des délétions et des duplications via la somme du 1/LOEUF sur le score Z de l'intelligence générale ajusté (âge et sexe) :

Modèle pour les délétions : $Score\ Z\ intell.\ aj.\ lm \sim \beta_{0; DEL} + \beta_{1; DEL} \times \sum_{g\grave{e}me} \frac{1}{LOEUF}$

Modèle pour les duplications : $Score\ Z\ intell.\ aj.\ GP \sim \beta_{0; DEL} + \beta_{1; DEL} \times \sum_{g\grave{e}me} \frac{1}{LOEUF}$

4.4.1.2 Impact des gènes individuels sur le quotient intellectuel

Une CNV GWAS sur l'intelligence générale a été réalisée sur la base d'un modèle de régression linéaire où chaque gène complètement inclus dans des CNV et présents chez au moins 10 individus différents ont été analysés. Le modèle a été appliqué séparément pour les délétions et les duplications. Le prédicteur du modèle pour les délétions est la présence (ou l'absence) du gène dans une délétion pour chaque individu. Les individus possédant le gène testé dans une duplication ont été exclus. Le même modèle a été appliqué pour les duplications, en excluant les porteurs des gènes inclus dans une délétion, le cas échéant.

Pour que le modèle ne capture que l'impact des CNV sur l'intelligence générale, le phénotype a été ajusté au préalable. Pour le G-facteur des différentes cohortes et FI, l'âge, l'âge au carré, le sexe ont été utilisés pour ajuster ces mesures d'intelligence, ce qui n'est pas le cas pour les QI, puisqu'ils sont déjà ajustés. Les 10 premières composantes principales de l'origine ancestrale des participants (PC 1-10) et le type de test cognitif (QI, G-facteur des différentes cohortes et FI)

ont également été incluses dans l'ajustement du phénotype, afin de corriger tout effet de la stratification de la population dans les données pour toutes les mesures d'intelligence, sans exception. Comme le jeu de données utilisé contient des individus apparentés, un effet aléatoire sur l'identificateur de chaque famille a été ajouté au modèle (REML).

Modèle pour les délétions : $Score\ Z\ intell.\ aj. \sim \beta_{0,DEL} + \beta_{1,DEL} \times G\grave{e}ne$

Le seuil de significativité utilisé est une correction FDR appliquée uniquement sur les groupes de gènes indépendants. En effet, comme certains gènes se regroupent dans les mêmes CNV, ceux-ci peuvent être en déséquilibre de liaison. Les gènes consécutifs dont les porteurs ont une similarité définie par un $r^2 > 0,8$ sont donc regroupés pour ne représenter qu'un seul test.

4.4.1.3 Impact des gènes tolérants sur l'intelligence générale

Les gènes testés dans ces analyses concernent ceux inclus dans la fenêtre de LOEUF = [1,44:1,85]. Cette fenêtre montre des impacts positifs potentiels sur l'intelligence générale dans la sliding-window réalisée précédemment (Figure 7), autant pour les délétions que les duplications. Pour tester la présence d'une diversité d'impact des gènes dans cet intervalle, nous avons réalisé un test d'hétérogénéité sur l'impact de ces gènes pris individuellement sur l'intelligence à l'aide de la librairie R meta (100).

Pour les analyses subséquentes sur ces fenêtres de LOEUF, le même modèle que celui utilisé pour la CNV GWAS sur l'intelligence générale décrite dans la section 4.4.1.2 a été utilisé. La correction FDR est donc la même, soit celle sur le gène le plus significatif de chaque clusters (incluant tous les clusters du génome et non pas seulement celui des gènes inclus dans la fenêtre d'intérêt).

Un second modèle a également été testé pour cette partie. Ce modèle est identique à celui décrit pour la CNV GWAS sur l'intelligence générale décrite dans la section 4.4.1.2, mais il inclut également comme covariable le poids des délétions dans le génome de chaque individu, défini

par la somme du score 1/LOEUF de chaque gène inclus dans une délétion pour chaque individu. Le même modèle a été appliqué séparément pour les duplications.

$$\text{Modèle pour les délétions : } \textit{Score Z intell. aj.} \sim \beta_{0,DEL} + \beta_{1,DEL} \times \textit{Gène} \\ + \beta_{2,DEL} \times \sum_{\textit{gène}} \frac{1}{\textit{LOEUF}}$$

4.4.1.4 Analyses de sensibilité sur l'âge

Pour avoir une représentation de la distribution de la somme du 1/LOEUF avec l'âge, nous avons réalisé une sliding-window sur l'âge moyen des porteurs de CNV. Les fenêtres ont une taille de 10 ans, sans chevauchement entre elles.

Nous avons également réalisé un modèle linéaire sur l'âge avec comme prédicteurs la somme du 1/LOEUF pour les délétions et les duplications, et comme covariables, le sexe et les 10 premières composantes principales de l'origine ancestrale des participants (PC1-10).

$$\text{Modèle pour les délétions : } \hat{\textit{Age}} \sim \beta_{0,DEL} + \beta_{1,DEL} \times \sum_{\textit{gène}} \frac{1}{\textit{LOEUF}} \\ + \beta_{2,DEL} \times \textit{sexe} \\ + \beta_{3,DEL} \times \textit{PC1} + \dots + \beta_{13,DEL} \times \textit{PC10}$$

La CNV GWAS présentée dans cette section porte sur l'intelligence générale ou le risque de TSA. Plutôt que de faire un test sur chaque gène affecté par le dosage génique, les tests lancés dans cette section sont sur les CNV uniques dans le jeu de données. Pour déterminer les CNV uniques dans le jeu de données, nous avons joint tous les CNV ayant un chevauchement réciproque de 70% entre eux. Le plus petit START et le plus grand STOP ont été conservés pour chaque région.

Un modèle de régression linéaire sur l'âge a ensuite été réalisé sur chacune de ces régions, séparément pour les délétions et les duplications. Le prédicteur du modèle pour les délétions est

la présence d'une délétion chevauchant à 70% le CNV testé pour chaque individu. Le même modèle a été appliqué pour les duplications.

4.4.1.5 Analyses sur le chromosome X

Le modèle additif a été testé sur les CNV du chromosome X. L'impact de la somme du 1/LOEUF pour chaque gène inclus dans les délétions (ou les duplications) sur l'intelligence générale ajustée pour l'âge, l'âge au carré, le sexe, les principales composantes 1 à 10 pour l'origine ancestrale et le type du test cognitif a été mesuré à l'aide d'un modèle linéaire. Les hommes et les femmes ont été séparés en deux groupes pour ces analyses.

Modèle pour les délétions : $Score\ Z\ intell.\ aj. \sim \beta_{0;DEL} + \beta_{1;DEL} \times \sum_{gène} \frac{1}{LOEUF}$

Des analyses de sensibilité ont été faites, ainsi la même analyse a été réalisée avec deux variantes. La première est que les individus de >55 ans ont été exclus. La deuxième est que tous les CNV des individus chevauchant à 70% les régions potentiellement mosaïques dans le chromosome X chez les individus de UKBB, définis par Loh et al. 2018 (101), ont été exclus. Les CNV dans les régions télomériques, qui sont également propices à l'apparition de CNV mosaïque, ont été exclus par visualisation manuelle.

4.4.2 Modélisation de l'impact des CNV sur le risque de TSA

Une CNV GWAS a été réalisée sur le risque d'autisme. La même stratégie décrite dans la section 4.4.1.2 pour l'intelligence a été appliquée pour étudier l'impact des CNV sur le risque d'autisme, mais via un modèle de régression logistique, séparément pour les gènes emportés par les délétions et les duplications.

Modèle pour les délétions : $TSA \sim \beta_{0;DEL} + \beta_{1;DEL} \times Gène + \beta_{2;DEL} \times sexe$

Une analyse de sensibilité préliminaire a également été réalisée. Le modèle logistique est exactement le même que celui décrit ci-dessus, mais il inclut toutefois tous les gènes dont un moins un nucléotide codant est emporté par les CNV (et non exclusivement ceux qui sont emportés dans leur intégralité).

Un modèle incluant également le score Z de l'intelligence ajusté par l'âge et l'âge au carré a également été utilisé comme covariable :

Modèle pour les délétions : $TSA \sim \beta_{0;DEL} + \beta_{1;DEL} \times Gène$
 $+ \beta_{2;DEL} \times sexe + \beta_{3;DEL} \times Intell. aj.$

Pour les analyses sur le risque de TSA, seuls les individus non-apparentés d'origine européenne ont été conservés (n=7 211 cas, n=450 187 contrôles). L'effet aléatoire de l'identifiant familial n'a donc pas été inclus pour ce modèle. Le sexe a été inclus comme covariable.

Chapitre 5 – Résultats

5.1 Normalisation des mesures cognitives

Avant de regarder l'impact du dosage génique sur le score Z de l'intelligence, nous avons d'abord cherché à savoir si la correction du phénotype par l'âge et le sexe pouvait être mieux capturée en utilisant le GP plutôt que la correction par les résidus du modèle linéaire, précédemment utilisée (25,76). En effet, le déclin de l'âge est linéaire jusqu'à environ 60 ans, mais celui-ci chute ensuite de façon plus drastique (Figure 9). La correction par GP étant très lourde au niveau de la mémoire et du temps d'exécution, pour cette section, seuls les 30 794 individus de UKBB avec un g-facteur ont été inclus dans cette analyse. De plus, comme UKBB est une population adulte plus âgée en moyenne que les autres cohortes (Tableau 5), l'utilité de cette correction par rapport à celle des résidus du modèle linéaire est d'autant plus pertinente pour corriger le déclin de l'âge qui s'accroît avec le vieillissement (Figure 10A).

La correction par le sexe permet quant à elle de corriger la différence de QI moyen entre les hommes et les femmes, celui des hommes étant en moyenne plus élevé de 3,3 à 5,0 points de QI comparé à la moyenne de celui des femmes (102). Il est donc nécessaire de faire cet ajustement, que ce soit pour l'une ou l'autre des corrections utilisées.

En comparant les deux types de corrections, on constate que le comportement de la prédiction est très similaire pour les hommes et les femmes (Figure 10B). Les scores Z de l'intelligence générale des deux types de correction corréleront à 0,99 (coefficient de Pearson), et lorsque l'on normalise à nouveau la prédiction des résidus du modèle linéaire en score Z, on constate un chevauchement presque parfait entre les deux corrections (Figure 10D). Il est d'ailleurs à noter que le décalage de la correction par les résidus du modèle linéaire par rapport au 0 sur l'axe des y est dû au fait que le score Z a été calculé sur la totalité de la cohorte, alors qu'ici, seule une

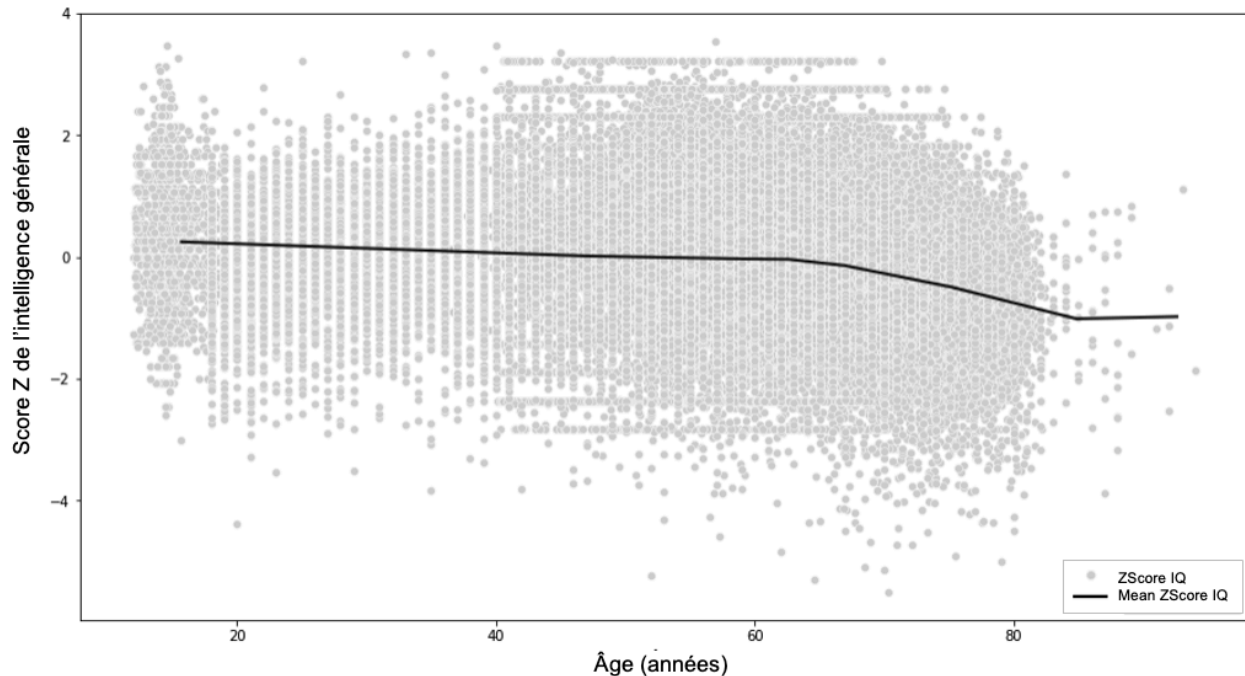


Figure 9. – Déclin de l’intelligence générale avec le vieillissement dans la population générale. À partir d’environ 60 ans, le déclin de l’âge s’accroît. Les points gris représentent le score Z de l’intelligence générale sans ajustement pour les individus de toutes les cohortes de la population générale dont l’âge est disponible (n=207 644). Les lignes verticales qui apparaissent entre 20 et 40 ans sont dues au fait que dans la cohorte G-Scot, l’âge est mesuré en années plutôt qu’en mois, ce qui représente une mesure en quantile plutôt qu’une mesure continue. La ligne noire représente le score Z de l’intelligence générale moyen dans une sliding-window sur l’âge où les fenêtres utilisées ont une taille de 10 ans sans chevauchement.

portion de la cohorte UKBB est incluse, dont le score Z moyen de l’intelligence générale est 0,51 (SD: 0,89) (Tableau 5).

Afin de tester la pertinence de la correction par le GP par rapport à celle des résidus du modèle linéaire, nous avons regardé l’effet du dosage génique, mesurée par la somme du 1/LOEUF, sur l’intelligence générale ajustée avec les deux types de correction (Figure 11). Pour le modèle appliqué sur les gènes délétés (Figure 11A), on constate que les estimés obtenus avec les deux types de correction sont similaires, soient $-0,029$ (SD=0,001; $p=1,45 \times 10^{-56}$) et $-0,027$ (SD=0,002; $p=1,40 \times 10^{-53}$) respectivement pour la correction avec les résidus du modèle linéaire et celle avec

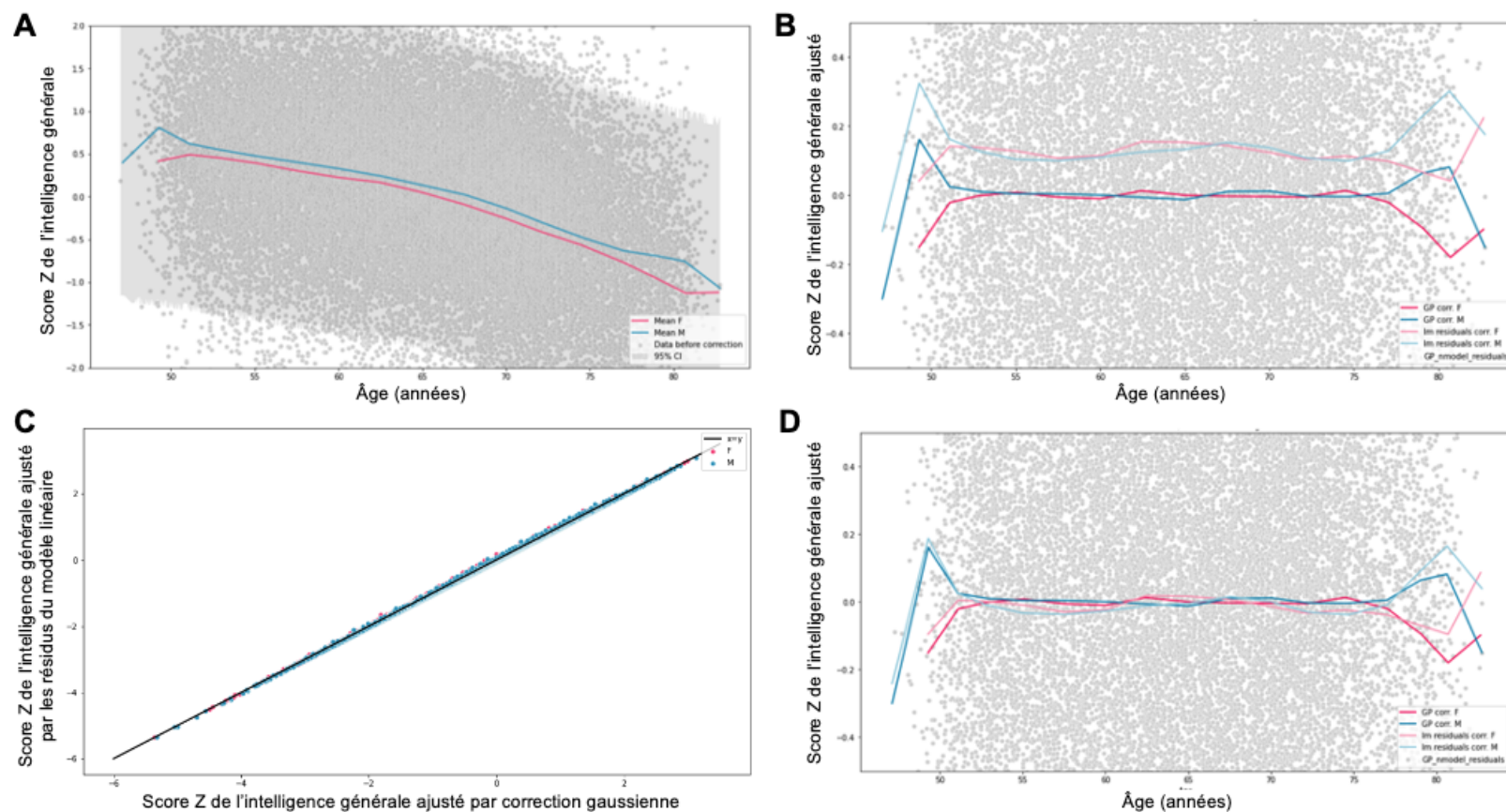


Figure 10. – Comparaison entre les types de correction du score Z de l'intelligence générale sur l'âge et le sexe. Les données utilisées sont les individus de la cohorte UKBB avec un g-facteur (n=30 242) A. Score Z du g-facteur moyen avant correction en fonction de l'âge pour les hommes (bleu) et les femmes (rose). Les fenêtres utilisées pour calculer le score Z du g-facteur moyen au fil des années ont une taille de 5 ans avec un chevauchement de 2,5 ans entre elles. B. Ajustement du score Z du g-facteur par correction gaussienne (lignes foncées) et par les résidus du modèle linéaire (lignes pâles) pour les hommes (bleu) et les femmes (rose). C. Corrélation entre le score Z du g-facteur corrigé par correction gaussienne et par les résidus du modèle linéaire (coefficient de Pearson = 0,99). D. Ajustement du score Z du g-facteur par correction gaussienne (lignes foncées) et par les résidus du modèle linéaire en score Z (lignes pâles) pour les hommes (bleu) et les femmes (rose).

le GP. Nous obtenons une significativité légèrement meilleure avec la correction avec les résidus du modèle linéaire. Pour les duplications, le même scénario est observable : les estimés sont proches (résidus du modèle linéaire : -0,011; SD=0,001; $p=3,69 \times 10^{-24}$; GP : -0,010; SD=0,001; $p=6,65 \times 10^{-21}$), avec un test légèrement plus significatif avec le modèle linéaire (Figure 11B).

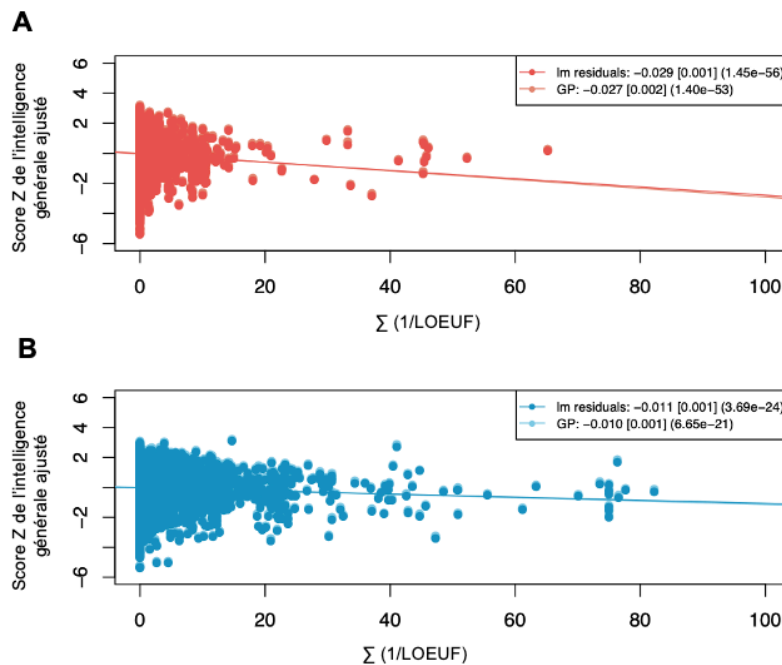


Figure 11. – Comparaison de l'effet du dosage génique sur l'intelligence générale en fonction du type de correction sur le phénotype. A. Correction de l'intelligence générale pour les délétions en rouge foncé pour les résidus du modèle linéaire et en rouge pâle pour la correction gaussienne. Le coefficient de régression linéaire est décrit avec : Estimé [SD] (p -value). B. Correction de l'intelligence générale pour les délétions en bleu foncé pour les résidus du modèle linéaire et en bleu pâle pour la correction gaussienne. N.B. : on ne voit pas beaucoup les points correspondant à la correction gaussienne en raison de la grande corrélation entre les deux types de correction, et donc leur chevauchement. GP : Correction gaussienne; lm residuals : Correction par les résidus du modèle linéaire.

Aux vues de ces résultats, comme la correction avec le GP est beaucoup plus lourde au niveau computationnel et que cette correction ne semble pas amener plus de robustesse au signal, les analyses subséquentes ont été réalisées avec le score Z de l'intelligence générale ajusté avec les résidus du modèle linéaire, comme dans nos précédentes publications (25,76).

5.2 Objectif 1 : Impact des CNV du chromosome X sur la cognition

5.2.1 Application du modèle additif sur les CNV du chromosome X

Puisque le laboratoire a désormais à sa disposition la cohorte UKBB (n=238 176), qui est à elle seule presque 10 fois plus grande que l'ensemble des cohortes précédemment utilisées pour mesurer l'impact des CNV sur la cognition, nous avons désormais le potentiel d'appliquer le modèle additif d'intolérance des gènes sur les CNV du chromosome X.

Pour ce faire, les hommes (n=108 451) et les femmes (n=129 725) ont été séparés en deux groupes différents, puisque leur nombre de copies du chromosome X diffère. De ce fait, l'interprétation de l'impact des CNV du chromosome X pourrait être différent en fonction du sexe.

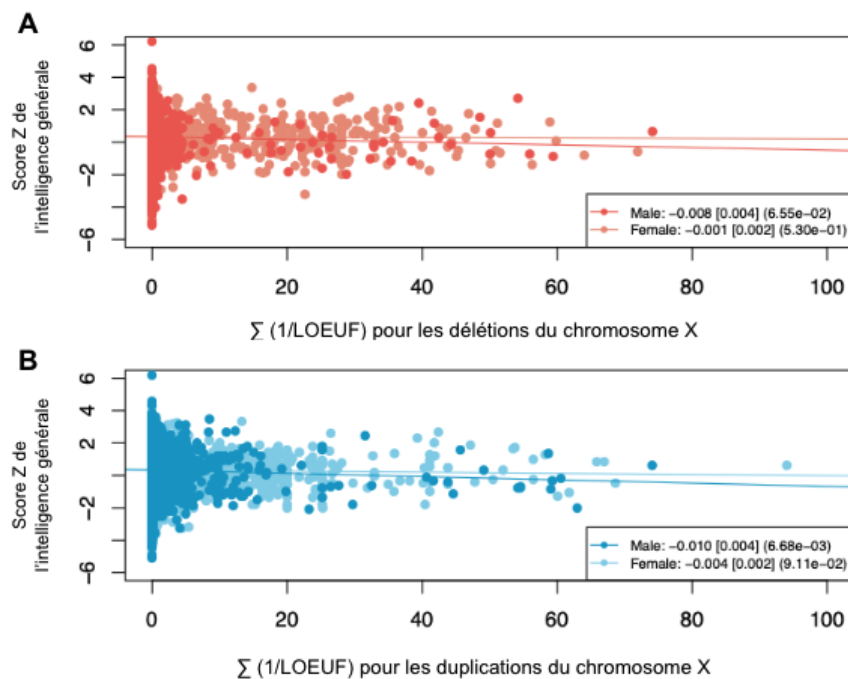


Figure 12. – Effet des CNV du chromosome X sur le score Z de l'intelligence générale. A. Effet des gènes délétés sur le score Z de l'intelligence pour les hommes (rouge foncé) et les femmes (rouge pâle). Le coefficient de régression linéaire est décrit avec: Estimé [SD] (p-valeur). B. Effet des gènes dupliqués sur le score Z de l'intelligence pour les hommes (bleu foncé) et les femmes (bleu pâle).

L'application du modèle classique d'intolérance des gènes sur les CNV du chromosome X ne montre pas de signal significatif pour les délétions (Figure 12A), ni de signal pour les duplications chez les femmes (Figure 12B). Seules les duplications chez les hommes montrent un effet significatif sur l'intelligence générale ($-0,013$, $p=1,50 \times 10^{-3}$). L'une des différences entre les données utilisées sur les autosomes (76) et celle de notre analyse est l'âge moyen du jeu de données. En effet, UKBB est une population adulte dont la moyenne d'âge est supérieure à celle des autres cohortes utilisées. Une étude récente indique qu'avec l'avancement de l'âge, la proportion d'altération chromosomale mosaïque détectée dans le sang augmente considérablement, particulièrement chez les hommes (103). De plus, les chromosomes sexuels ont une grande instabilité. Notamment, les hommes ont tendance à avoir une perte mosaïque de leur chromosome Y, et les femmes ont tendance à avoir une perte mosaïque de l'un de leur chromosome X, tendances qui augmentent drastiquement avec l'âge. Comme les échantillons biologiques de UKBB sont des échantillons sanguins, ce point est très important pour nos analyses, particulièrement pour les CNV du chromosome X.

Nous avons donc investigué les données à l'aide de filtres supplémentaires afin d'atténuer le bruit potentiellement causé par les CNV mosaïques dans notre jeu de données. Toutes les aneuploïdies ont déjà été retirées du jeu de données initial.

5.2.2 Investigation du mosaïcisme

Afin de tenter l'élimination des CNV mosaïques, nous avons d'abord reproduit les analyses en ajoutant un filtre sur les individus de plus de 55 ans (Figure 13), puisque le mosaïcisme augmente drastiquement avec l'âge (103,104). Après filtres, il reste 26 092 hommes et 33 664 femmes dans le jeu de données.

Après l'application du filtre sur l'âge, le signal n'est cependant pas plus significatif que sans le filtre (Figure 14A et B). On perd même le signal chez les duplications pour les hommes.

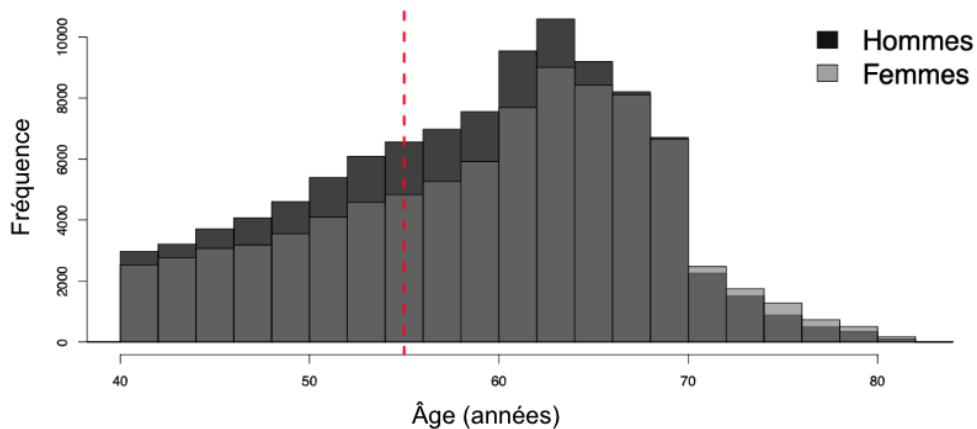


Figure 13. – Distribution des âges dans UKBB (n=238 176) selon le sexe. L’histogramme noir représente les hommes et le gris représente les femmes. La ligne pointillée rouge représente 55 ans.

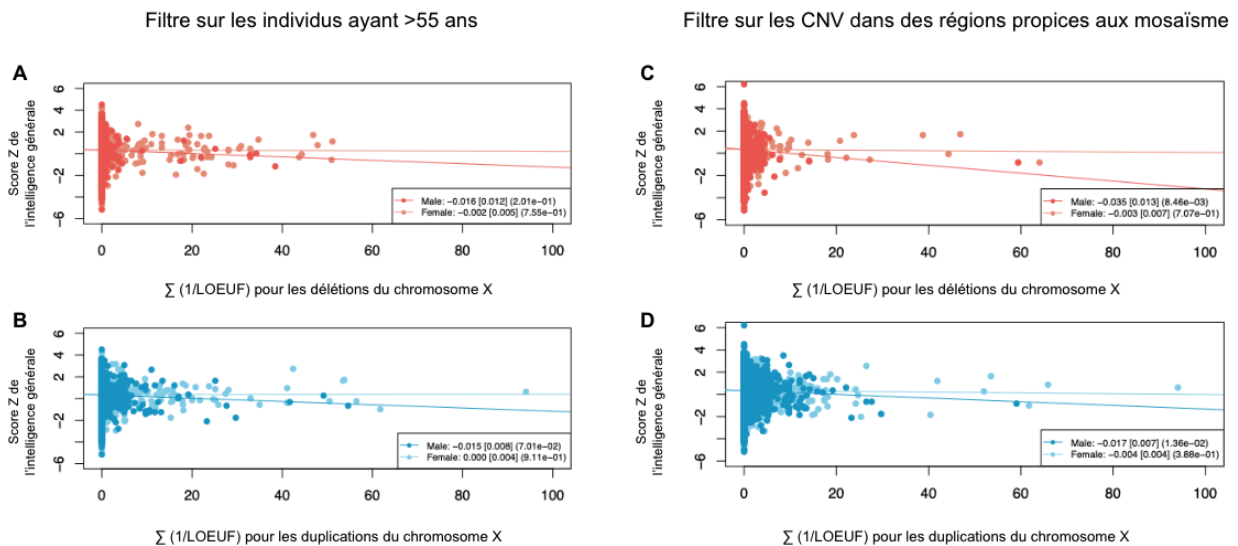


Figure 14. – Filtre sur les individus ou les CNV pouvant générer du bruit dans le jeu de données. Le coefficient de régression linéaire est décrit avec: Estimé [SD] (p-valeur). A. Effet des gènes délétés ou B. dupliqués sur le score Z de l’intelligence pour les hommes (rouge foncé) et les femmes (rouge pâle) de 55 ans ou moins. C. Effet des gènes délétés ou D. dupliqués avec retrait des CNV chevauchant les régions propices aux mosaïcisme (*DXZ4* et *DXZ1*, et télomères).

Nous avons ensuite tenté de retirer les régions propices au mosaïcisme. Les régions retirées correspondent à celles identifiées sur le chromosome X par Loh et al. 2018 (101) et comprennent les gènes *DXZ4* et *DXZ1*. Nous avons également retiré les CNV situés dans les télomères des CNV. Nous avons retiré tous les CNV chevauchant ces régions (652 CNV exonique potentiellement mosaïques). Aucun signal n'est toutefois significatif suite à l'application de ce filtre (Figure 14C et D).

Il semble y avoir un problème de mosaïcisme dans la cohorte de UKBB. Les filtres qui ont été mis en place ne parviennent pas à éliminer systématiquement les CNV qui posent problème. La poursuite de ces analyses nécessite donc une autre méthode pour filtrer les CNV mosaïques, sans quoi il est impossible d'obtenir un signal non-bruité.

5.3 Objectif 2a : Impact des gènes individuels sur la cognition

L'un de nos principaux objectifs était de mesurer l'impact du dosage génique sur le score Z de l'intelligence générale pour tous les gènes inclus dans un CNV avec au moins 10 porteurs. Nous avons réalisé une CNV GWAS incluant 263 396 individus (Tableau 5). Au total, 15 508 gènes différents étaient inclus dans au moins un CNV (Tableau 6). Parmi eux, 805 et 2 080 gènes étaient inclus respectivement dans une délétion ou dans une duplication pour au moins 10 porteurs, pour un total de 2 350 gènes différents (Figure 15 et 16). Parmi ces tests, 168 délétions (20,9%) et 244 (11,7%) duplications étaient significativement associées à l'intelligence générale après correction FDR. Les gènes en déséquilibre de liaison en raison de leur présence dans le même CNV ($r^2 > 0,8$) ont ensuite été regroupés en clusters. Pour chaque cluster, la p-value du gène le plus significatif a été conservée. Les 805 gènes délétés se regroupent en 390 clusters de gènes différents, dont 41 clusters sont significatifs (10,5%), et les 2 080 gènes dupliqués se regroupent en 1 023 clusters, dont 46 clusters significatifs (4,5%).

TYPE	Délétion	Duplication
Nombre de gène	15 508	
Nombre de gène avec ≥ 10 porteurs	805 FDR<0,05 : 168	2 080 FDR<0,05 : 244
Nombre de clusters	390 FDR<0,05 : 41	1 023 FDR<0,05 : 46

Tableau 6. – Gènes inclus dans la CNV GWAS sur l'intelligence générale.

Parmi ces gènes significativement associés à l'intelligence générale, 85,7% de ceux délétés et 76,6% de ceux dupliqués sont également rapportés dans la littérature, soit dans la base de données SFARI (12), dans Satterstrom et al. (2020) (68) ou ce sont des gènes emportés par des CNV récurrents associés aux TND (Figure 17). Les autres gènes (14,3% et 23,4% pour les délétions et les duplications respectivement) sont de nouveaux gènes candidats montrant une association positive ou négative avec l'intelligence générale. Ces gènes ne sont ni à proximité (± 250 kb) des gènes et CNV récurrents déjà rapportés dans la littérature et répertoriés dans Huguet et al. (2021) (76), et ils ne sont pas dans un cluster incluant une de ces régions connues. Ces gènes se regroupent dans 13 et 18 clusters différents pour les délétions et les duplications respectivement.

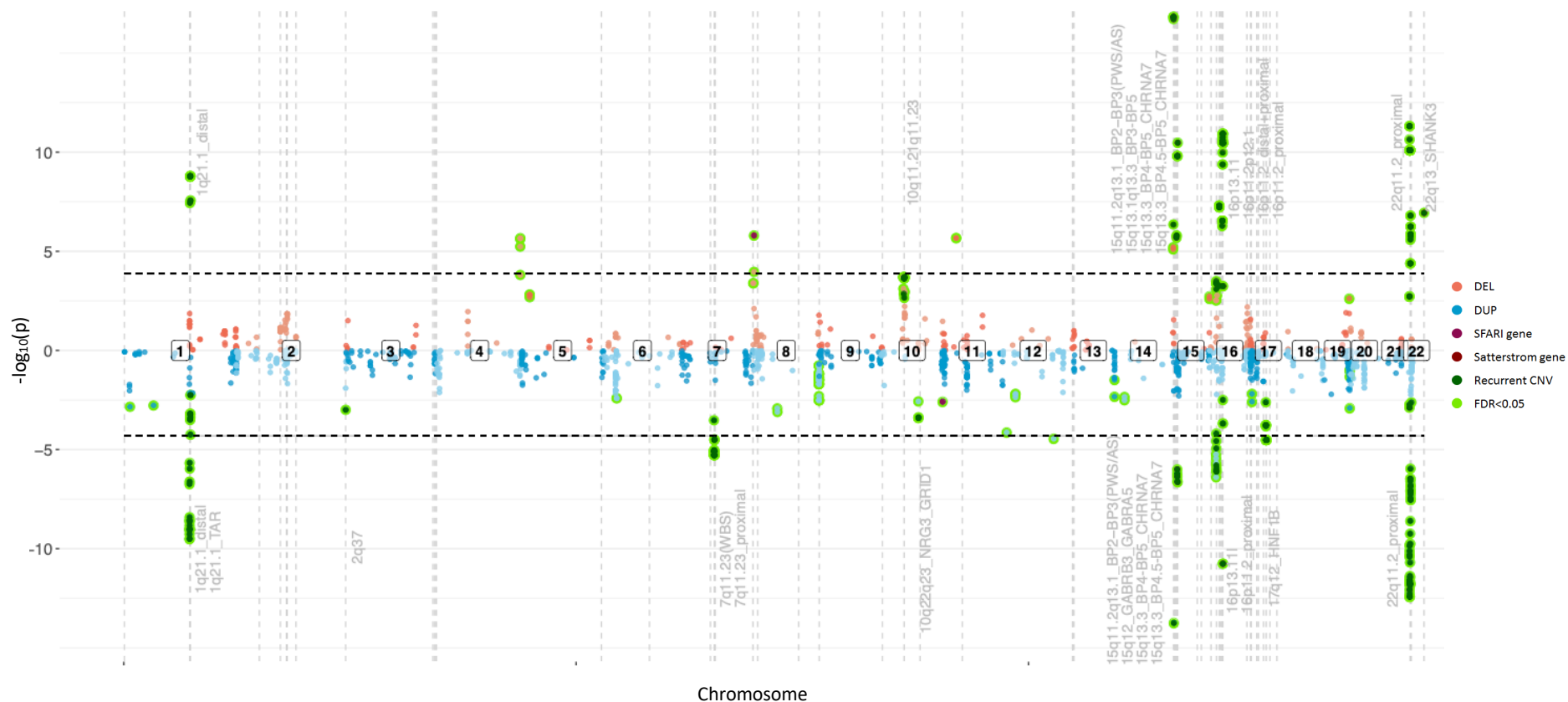


Figure 15. – Miami plot des gènes associés à une réduction de l’intelligence générale lorsqu’ affectés par le dosage génique. L’axe des Y représente la p-value non-ajustée des tests, transformée en $-\log_{10}$ (plus un gène emporté par une délétion a un score élevé sur cet axe, plus le test est significatif; les duplications sont représentées en miroirs sur la partie négative de l’axe). Les lignes noires horizontales pointillées représentent la correction de Bonferroni sur le nombre de clusters sur pour les délétions ($p < 1,28 \times 10^{-4}$) et les duplications ($p < 4,89 \times 10^{-5}$). Les points rouges représentent les gènes délétés et les bleus, les gènes dupliqués. Les point pourpres et bordeaux sont des gènes déjà connu dans la littérature (*i.e.*, SFARI et Satterstrom et al. (2020)) comme étant associés au risque de TSA. Les points vert foncé sont des gènes inclus dans des CNV récurrents. Les points entourés de vert représentent les tests qui passent la correction FDR. Les lignes grises verticales représentent les CNV récurrents, et leur nom est affiché si un gène inclus dans une telle région est significatif.

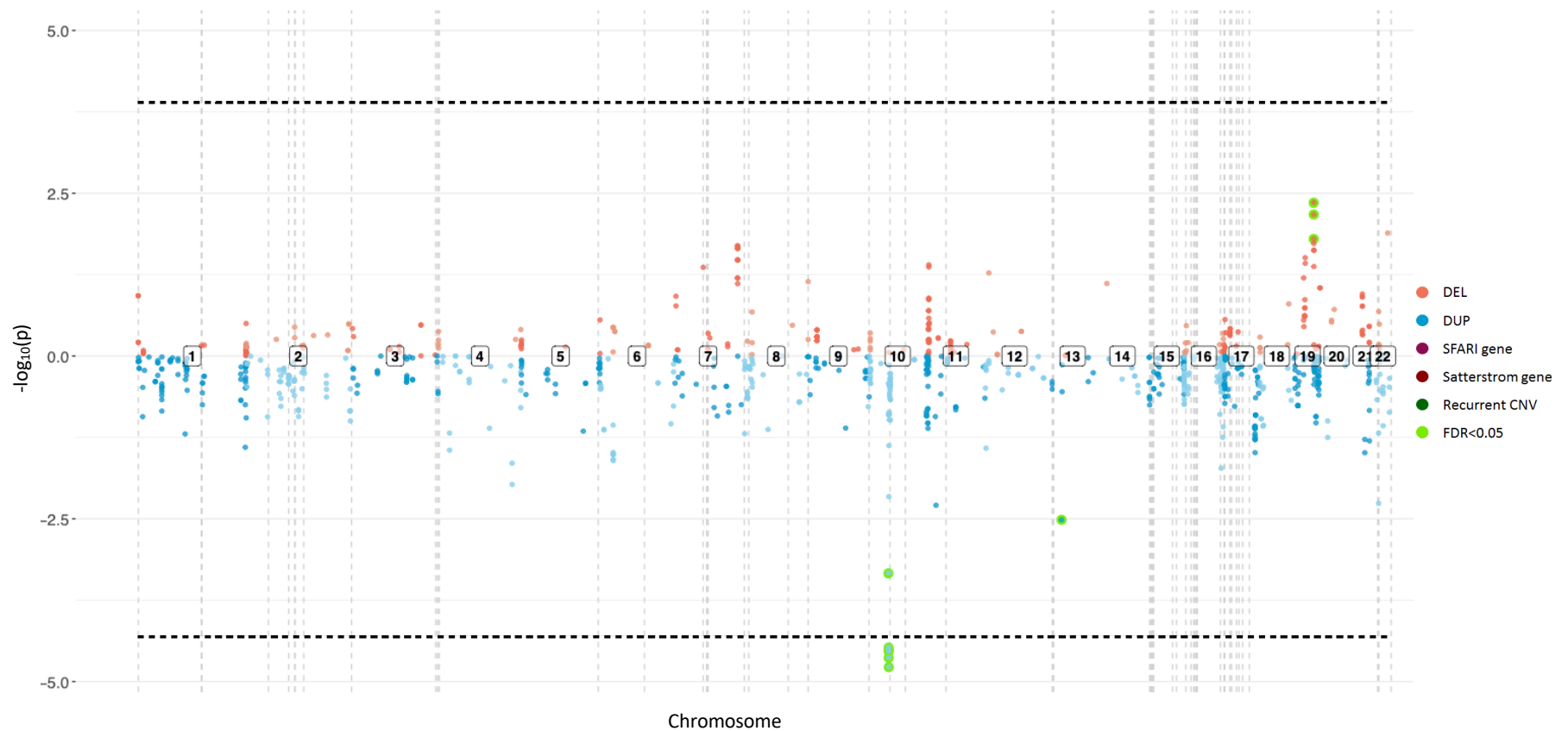


Figure 16. – Miami plot des gènes associés à une augmentation de l’intelligence générale lorsqu’affectés par le dosage génique. L’axe des Y représente la p-value non-ajustée des tests, transformée en $-\log_{10}$ (plus un gène emporté par une délétion a un score élevé sur cet axe, plus le test est significatif; les duplications sont représentées en miroirs sur la partie négative de l’axe). Les lignes noires horizontales pointillées représentent la correction de Bonferroni sur le nombre de clusters sur pour les délétions ($p < 1,28 \times 10^{-4}$) et les duplications ($p < 4,89 \times 10^{-5}$). Les points rouges représentent les gènes délévés et les bleus, les gènes dupliqués. Les point pourpres et bordeaux sont des gènes déjà connu dans la littérature (*i.e.*, SFARI et Satterstrom et al. (2020)) comme étant associés au risque de TSA. Les points vert foncé sont des gènes inclus dans des CNV récurrents. Les points entourés de vert représentent les tests qui passent la correction FDR. Les lignes grises verticales représentent les CNV récurrents, et leur nom est affiché si un gène inclus dans une telle région est significatif.

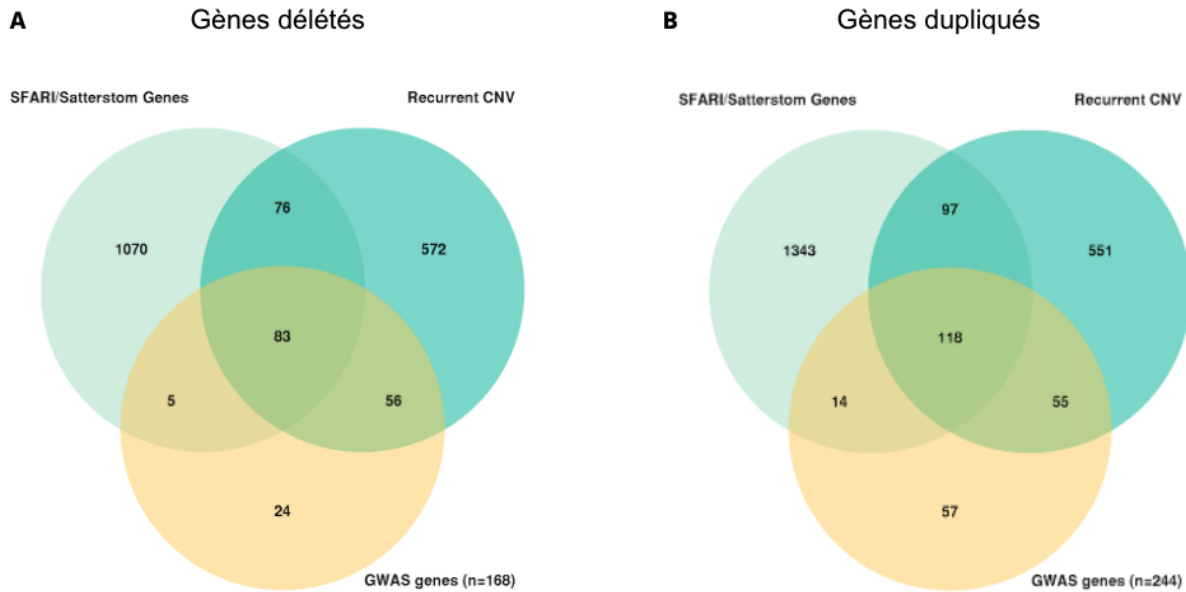


Figure 17. – Catégorisation des gènes avec une association significative dans la CNV GWAS sur l’intelligence générale pour A. les délétions et B. les duplications. Le cercle de couleur menthe représente les gènes associés au risque de TSA répertoriés dans la base de données SFARI et ceux rapportés par Satterstrom et al. (2020), et les gènes situés à moins de 250 kb de ces gènes. Le cercle turquoise représente les gènes emportés par des CNV récurrents associés aux troubles TND, et les gènes situés à moins de 250 kb de ces CNV. Le cercle jaune représente les associations significatives obtenus dans la CNV GWAS sur l’intelligence générale.

Alors que la grande majorité (99,9% pour les délétions et les duplications) des clusters de gènes significativement associés à l’intelligence générale ont un estimé négatif (Tableau supplémentaire 1 et 2), ce qui représente une diminution du QI, une petite partie des gènes l’augmente (Tableau 7). Ces 10 gènes augmentant l’intelligence générale se regroupent en un seul cluster pour les délétions et 3 clusters différents pour les duplications.

Gene	TYPE	CHR	START	STOP	LOEUF	n	Est.	SD	P	FDR
<i>GPRIN2</i>			46994087	47000573	1,49	257	0,22	0,06	4,62E-04	
<i>NPY4R</i>			47083457	47088320	1,75	287	0,24	0,06	3,34E-05	
<i>ANXA8L1*</i>	DUP	10	47157983	47174093	1,84	342	0,23	0,05	2,34E-05	3,53E-04
<i>FAM25B</i>			47177204	47181681		344	0,22	0,05	2,97E-05	
<i>AGAP10</i>			47191844	47239738	1.80	344	0,22	0,05	2,97E-05	
<i>ANTXRL*</i>	DUP	10	47657581	47701443	0,93	313	0,24	0,06	1,67E-05	2,53E-04
<i>VWA8*</i>	DUP	13	42140971	42535256	1,15	12	0,85	0,29	3,04E-03	3,23E-02
<i>PSG1</i>			43370616	43383974	1,99	119	0,22	0,09	1,59E-02	
<i>PSG6</i>	DEL	19	43406238	43423715	1,84	124	0,24	0,09	6,71E-03	4,16E-02
<i>PSG7*</i>			43428284	43441359		135	0,24	0,09	4,38E-03	

Tableau 7. – Gènes qui augmentent l’intelligence générale si affectés par le dosage génique.

Le gène de référence du cluster est indiqué par une étoile (*). n : nombre de porteurs du gène; Est : Estimé; SD : Écart-type; P : p-value.

La plupart des CNV incluant le cluster du gène *ANXA8L1* incluent également le cluster de *ANTXRL* (Figure 18A). Toutefois, il existe également beaucoup de petits CNV incluant seulement le cluster de *ANTXRL*, sans toucher à celui de *ANXA8L1*; d’où le fait que ce sont deux clusters différents. C’est une région avec énormément de duplications segmentaires qui définissent les bornes de ces CNV. Tous les gènes de cette région sont tolérants. Le gène tolérant *VWA8* est quant à lui seul dans son cluster (Figure 18B). La délétion de la région 19q13.3-q13.31 inclut 3 gènes : *PSG1*, *PSG6* et *PSG7* (Figure 18C). Ce dernier n’a pas de LOEUF défini. Le cluster voisin, qui inclut *PSG11* et *PSG2*, est borderline : sa p-value est <0,05, mais le test ne passe pas la correction FDR.

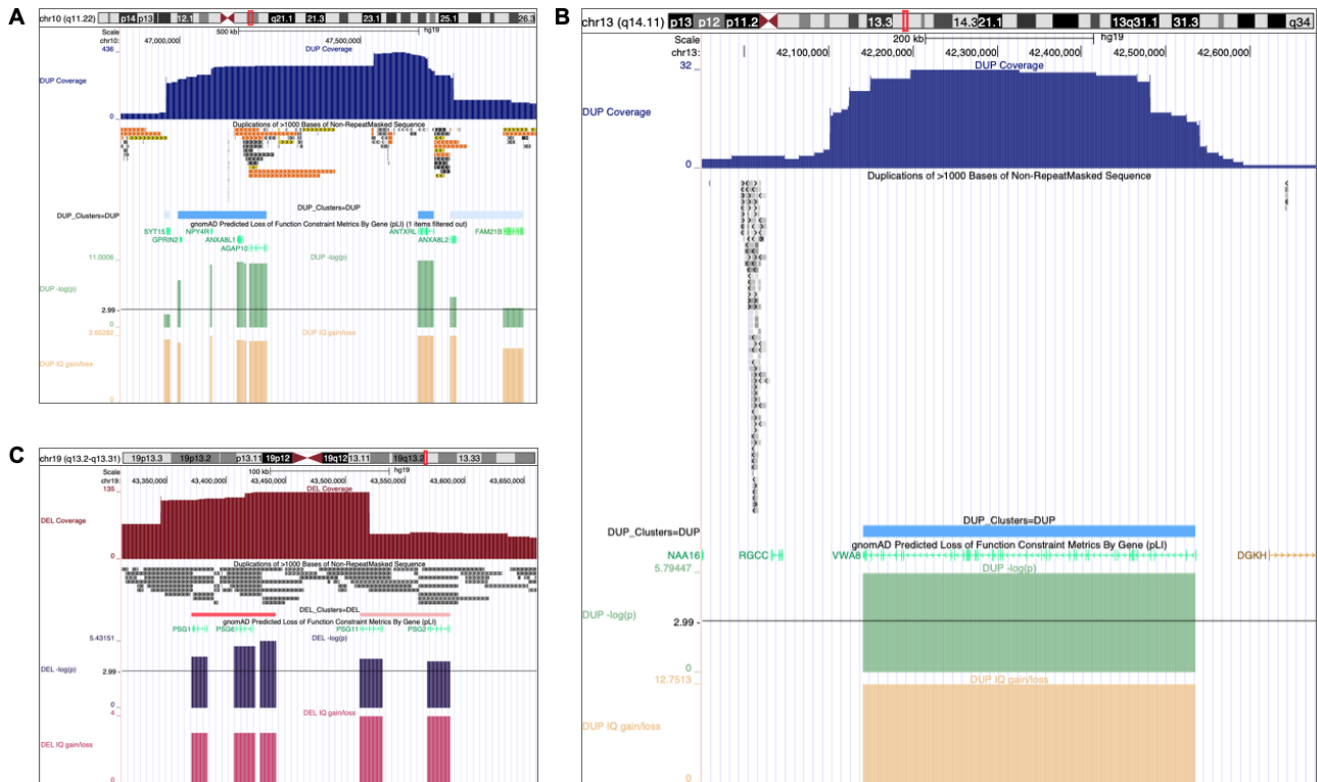


Figure 18. – Régions qui augmentent l’intelligence générale lorsque délétee ou dupliquée. A. Duplications de la région 10q11.22. B. Duplications de la région 13q14.11. C. Délétion de la région 19q13.3-q13.31. La couverture des duplications (DUP Coverage) et des délétions (DEL Coverage) sont respectivement illustrées en bleu foncé et en rouge foncé. Les clusters de gènes significatifs (FDR<0,05) dupliqués (DUP Cluster) et délétés (DEL Cluster) sont illustrés respectivement en bleu ou en rouge. Les clusters bleu pâle ou rouge pâle représentent les tests qui ne sont pas significatifs. Les duplications segmentaires (Duplications of > 1000 Bases of Non-RepeatMasked Sequence) sont représentées en gris. La tolérance des gènes (gnomadAD Predicted Loss of Function Constraint Metrics By Gene (pLI)) est représentée par le pLI via le gradient de couleur vert (gène tolérant) - rouge (gène intolérant). La p-value du test de la CNV GWAS transformé en -log10 est illustré en vert pour les duplications ou en mauve pour les délétions, la ligne noire correspondant à une p-value non ajustée<0,05. Le gain ou perte de QI sont représentés en jaune pour les duplications et en rose pour les délétions. Figure générée via UCSC.

5.3.1 Impact des gènes tolérants sur l'intelligence générale

Aux vues des résultats de la CNV GWAS (Tableau 7) et ceux obtenus précédemment avec la sliding-window sur la taille d'effet des gènes de LOEUF similaires sur le score Z de l'intelligence générale (Figure 7), nous avons voulu investiguer davantage les gènes de la fenêtre de LOEUF =]1,44:1,85]. Cette fenêtre, qui représente des gènes tolérants à la perte de fonction, a la particularité d'avoir un effet moyen positif significatif sur le score Z de l'intelligence, autant pour les délétions que les duplications (Figure 7). Elle inclut 474 gènes différents ayant au moins 10 porteurs, dont 211 se retrouvent dans des délétions et 390 dans des duplications (Tableau 8).

TYPE	Tous (n=601)	Filtre gènes olfactifs (n=488)
Délétion	211	153
Duplication	390	335

Tableau 8. – Gènes inclus dans la fenêtre de LOEUF =]1,44:1,85] où les gènes ont un effet positif sur l'intelligence générale dans la sliding-window. n : nombre de tests réalisés.

Pour avoir une meilleure idée de la dynamique des gènes se catégorisant dans cette fenêtre de LOEUF, nous avons fait un test d'hétérogénéité sur les gènes qui y sont inclus (Tableau 9). Nous avons également réalisé la même analyse, mais sans les gènes olfactifs, qui sont abondants dans le génome et n'ont pas d'effet direct sur la cognition : ils occasionnent donc généralement du bruit. L'hétérogénéité est grande et significative ($p < 1,00 \times 10^{-04}$) pour les délétions ($I^2 = 69,7\%$) et les duplications ($I^2 = 62,4\%$), avec ou sans les gènes olfactifs (Tableau 9). En retirant les gènes olfactifs, l'hétérogénéité augmente ($I^2 = 74,5\%$ et $I^2 = 65,6\%$ respectivement pour les délétions et les duplications). Il y a donc bien une constellation d'effets à la fois positifs et négatifs pour les gènes dans ces fenêtres de LOEUF.

TYPE	Tous les gènes		Filtre sur les gènes olfactifs	
	I ²	P	I ²	P
Délétions	69,7%	< 1,00x10 ⁻⁰⁴	74,5%	< 1,00x10 ⁻⁰⁴
Duplications	62,4%	< 1,00x10 ⁻⁰⁴	65,6%	< 1,00x10 ⁻⁰⁴

Tableau 9. – Tests d’hétérogénéité sur les gènes tolérants de LOEUF =]1,4:1,85]. P : p-value. La p-value est considérée comme significative quand $p < 0,1$ pour les tests d’hétérogénéité.

L’effet des gènes de cette fenêtre est donc sur un spectre : certains ont un impact significatif négatif sur l’intelligence, et d’autres un effet positif (Tableau 9, Figure 19A et C). On remarque néanmoins une majorité de gènes ayant un impact négatif sur l’intelligence, ce qui est contradictoire avec ce que l’on voit dans la sliding-window (Figure 7). Une explication possible est que notre modèle ne prédit pas l’impact des autres gènes inclus dans le génome des individus. Afin de contrôler cet effet, nous avons reproduit le modèle en ajoutant comme covariable la somme de l’intolérance des délétions ou des duplications (Figure 19B et D). Pour les délétions, on remarque que les gènes dont l’estimé est négatif sont moins significatifs ou ne le sont plus du tout (Tableau 10), et leur impact prédit sur l’intelligence est moins grand. Pour les duplications, la majorité des gènes qui avaient un impact négatif sur l’intelligence ne sont plus significatifs, et on constate une augmentation du nombre de gènes qui ont un effet significatif positif sur l’intelligence.

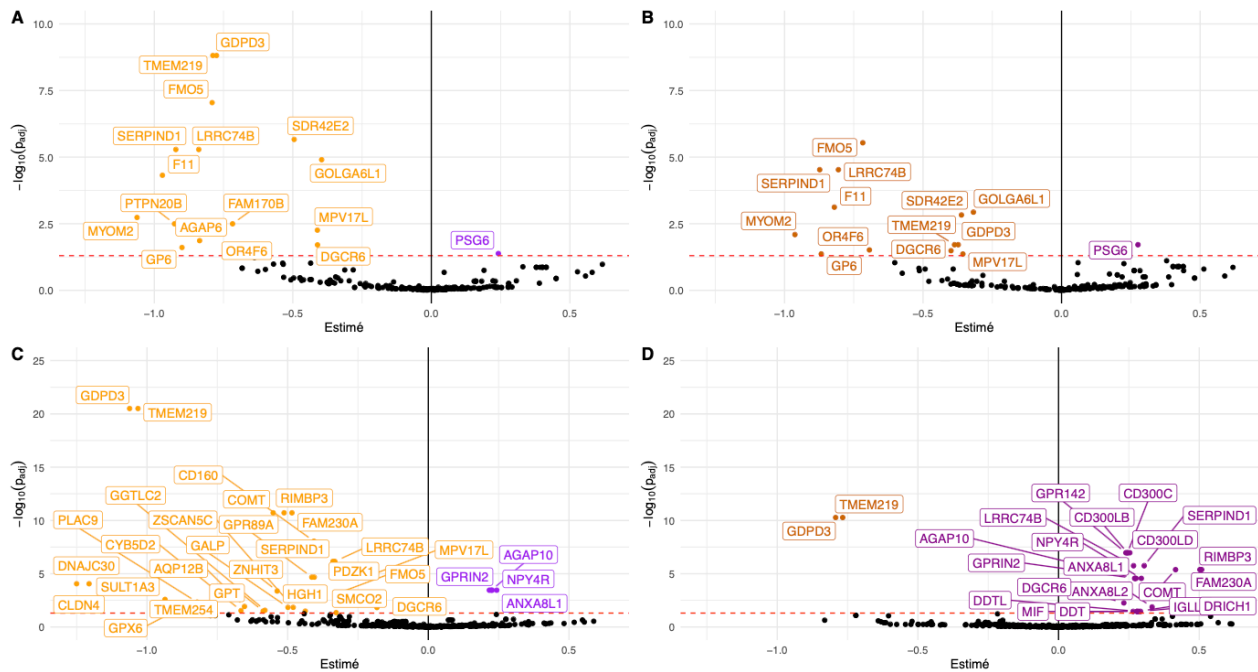


Figure 19. – Volcano plots de l’impact des gènes tolérant sur le score Z du QI. Les points mauves et oranges représentent des gènes ayant un effet significatif respectivement positif ou négatif sur le score Z du QI lorsqu’ils sont affectés par le dosage génique. Le score Z du QI a été ajusté selon l’âge au carré, le sexe et les 10 premières principales composantes de la PCA sur l’origine ancestrale des participants. La correction FDR a été appliquée sur le nombre de clusters de gènes ayant au moins 10 porteurs, séparément pour les délétions et les duplications. A. Tests sur les gènes emportés par des délétions. B. Tests sur les gènes emportés par les délétions avec le poids de l’intolérance des autres gènes délétés comme covariable. C. Tests sur les gènes emportés par des duplications. D. Tests sur les gènes emportés par les duplications avec le poids de l’intolérance des autres gènes dupliqués comme convariable. p_{adj} : p-value ajustée par correction FDR.

TYPE	Modèle classic		Modèle avec le poids des autres gène délétés	
	Estimé Négatif	Estimé Positif	Estimé Négatif	Estimé Positif
Délétion	20	1	17	1
Duplication	28	4	2	20

Tableau 10. – Gènes dont le LOEUF =]1,44:1,85] ayant un effet significatif sur l’intelligence générale en fonction du modèle utilisé.

5.3.2 Analyse de sensibilité sur l'âge

Dans de précédentes analyses, nous avons remarqué qu'en filtrant les cohortes en fonction de l'âge, l'impact des CNV mesuré par la somme du 1/LOEUF sur l'intelligence générale donnait des résultats différents. Particulièrement, l'impact des délétions est de plus en plus sévère à mesure que l'on retire les individus les plus âgés du jeu de données. Autrement dit, l'impact du dosage génique semble moins sévère chez les individus les plus âgés.

Il est à noter que les résultats présentés dans cette section sont des résultats préliminaires. En effet, l'âge utilisé pour ces analyses est celui auquel les individus des différentes cohortes ont passé leur test cognitif. L'âge pertinent dans le cadre de ces analyses est plutôt celui durant lequel le test biologique a été réalisé, mais cette mesure n'était pas disponible au moment de la réalisation de ces analyses.

Pour investiguer cet effet, nous avons d'abord regardé la distribution de la somme du LOEUF en fonction de l'âge pour les délétions et les duplications (Figure 20). Deux jeux de données différents ont été utilisés : le premier est UKBB (n=187 618) et le deuxième est composé des individus des autres cohortes de la population générale (CaG, G-Scot, Imagen, LBC, SYS; n=23 855). Dans les deux cas, et pour les délétions autant que pour les duplications, on constate un appauvrissement des individus dont la somme du LOEUF est élevée avec le vieillissement.

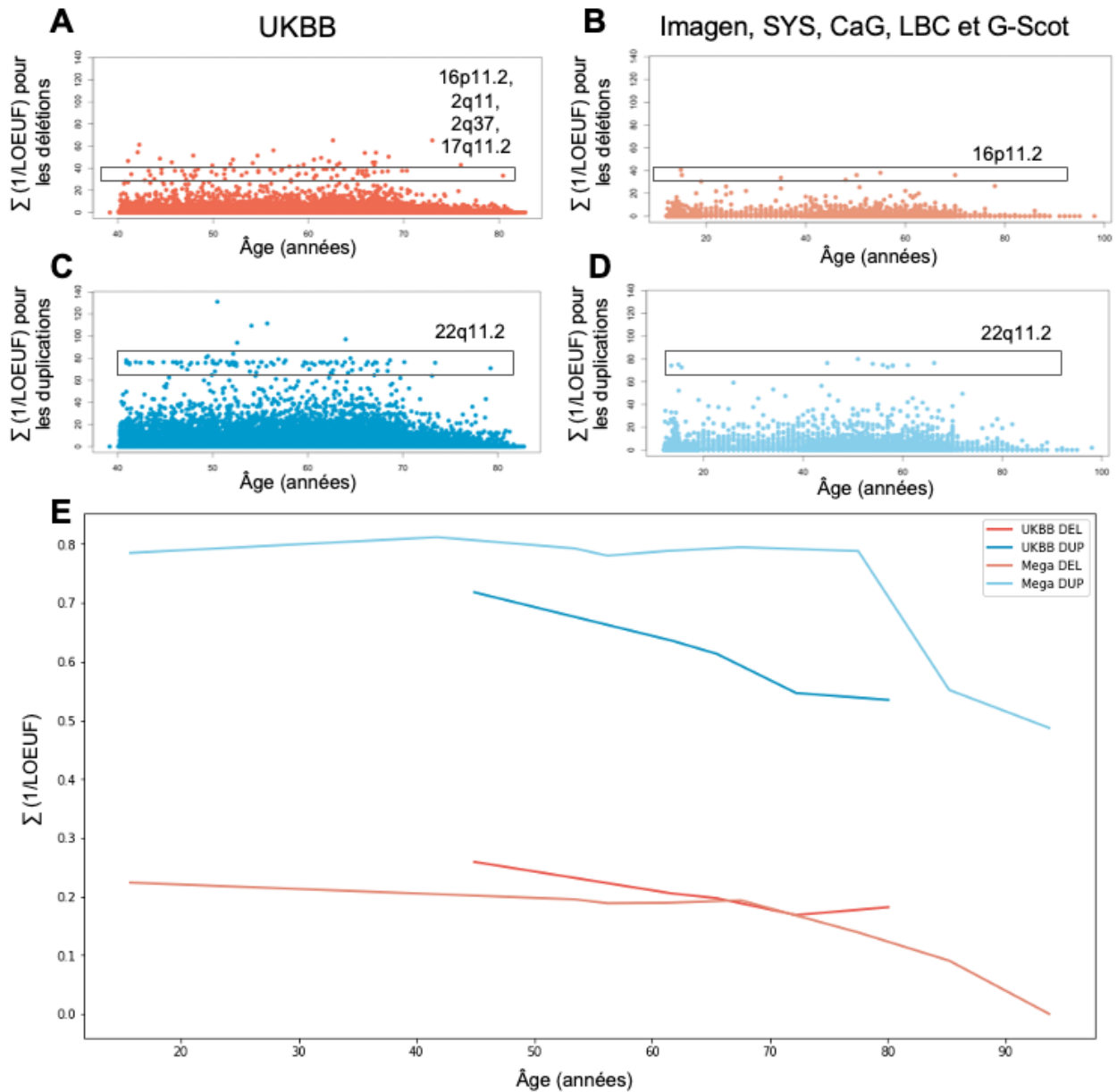


Figure 20. – Distribution du score du modèle additif en fonction de l'âge pour (A et B) les porteurs de délétions et (C et D) les porteurs de duplications. L'analyse a été réalisée sur deux jeu de données : UKBB (A et C) et sur les cohortes de la population générale (Imagen, SYS, CaG, LBC et G-Scot). Les points encadrés alignés sur l'axe de Y représentent des porteurs de CNV récurrents. E. Sliding-window sur l'âge moyen des porteurs de CNV. Les fenêtres ont une taille de 10 ans, sans chevauchement entre elles. Les lignes rouge foncé et rouge pâle représentent les délétions respectivement pour UKBB et les cohortes de la population générale. Les lignes bleu foncé et bleu pâle représentent les duplications respectivement pour UKBB et les cohortes de la population générale.

Pour en savoir plus, nous avons réalisé un modèle linéaire sur l'âge avec comme prédicteurs la somme du LOEUF pour les délétions et les duplications (Tableau 11). Pour UKBB, peu importe la tranche d'âge investiguée, une somme élevée de LOEUF pour les délétions ou les duplications entraîne un effet négatif sur l'âge. L'effet n'est pas significatif pour le second jeu de données.

Age subset (years)	TYPE	UKBB (N=187,618)				Imagen, CaG, G-Scot, LBC, SYS (N=23,835)			
		Est.	SD	CI 95%	P	Est.	SD	CI 95%	P
All	DEL	-0,09	0,02	-0,12 ; -0,06	1,00E-08	-0,09	0,08	-0.25 ; 0,07	2,84e-01
	DUP	-0,04	0,01	-0,05 ; -0,02	7,68E-09	-0,03	0,03	-0.09 ; 0,04	3,78e-01
≤60	DEL	-0,04	0,01	-0,07 ; -0,02	8,35E-04	-0,16	0,09	-0.33 ; 0,02	7,46e-02
	DUP	-0,01	0,01	-0,02 ; 0,00	4,69E-02	-0,03	0,03	-0.09 ; 0,04	4,05e-01
>60	DEL	-0,03	0,01	-0,06 ; -0,01	1,15E-03	0,01	0,04	-0.07 ; 0,09	7,74e-01
	DUP	-0,01	0,00	-0,02 ; -0,00	2,71E-02	0,00	0,02	-0.05 ; 0,04	8,86e-01

Tableau 11. – Impact de la somme du 1/LOEUF des délétions et des duplications sur l'âge. UKBB : UKBiobank; CaG: CartaGene, G-Scot : Generation Scotland, LBC : Lothian Birth Cohort, SYS : Saguenay Youth Study; Age subset : Sous-groupe d'âge inclus; Est : Estimé; SD : Écart-type; CI 95% : Intervalle de confiance à 95%; P : p-value.

Pour avoir un aperçu de quels CNV étaient enrichis ou appauvris avec l'âge, nous avons également réalisé une CNV GWAS sur les 56 095 CNV autosomiques uniques de UKBB (n=187 618). On constate que deux régions, l'une dans le chromosome 10 entre les positions 46 962 114 et 48 005 276 et l'autre dans le chromosome 14 entre les positions 19 264 875 et 20 444 607, sont associées à une augmentation de l'âge lorsque délétées (Figure 21A). Les gènes inclus dans ces deux régions sont décrits en annexe (Tableau supplémentaire 3). À l'opposé, on voit également que la région récurrente 16p11.2 est associée à une diminution de l'âge lorsque dupliquée (Figure 21B).

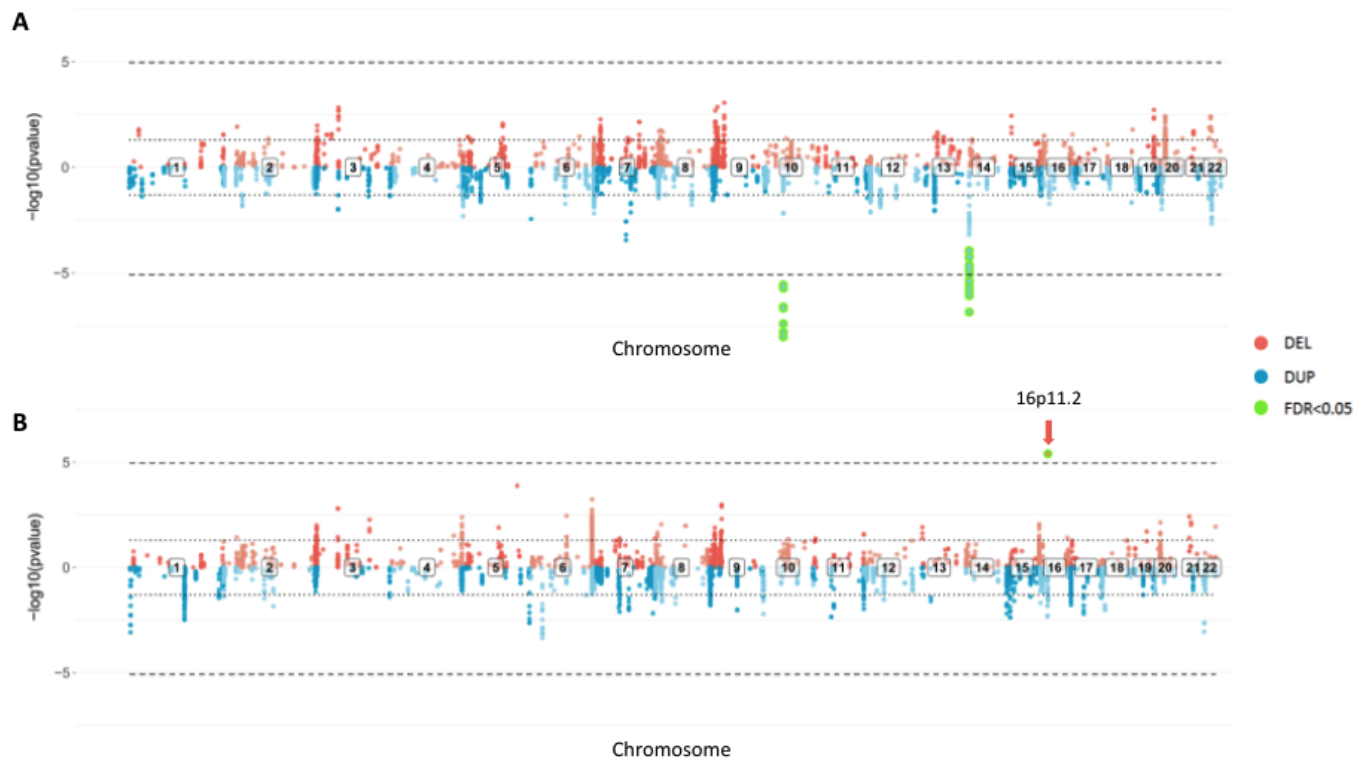


Figure 21. – CNV GWAS sur l'âge. A. CNV associés à une augmentation de l'âge. B. CNV associés à une baisse de l'âge. L'axe des Y représente la p-value non-ajustée des tests, transformée en $-\log_{10}$ (plus une délétion a un score élevé sur cet axe, plus le test est significatif; les duplications sont représentées en miroirs sur la partie négative de l'axe). Les lignes noires horizontales pointillées représentent la correction de Bonferroni sur le nombre de clusters sur pour les délétions ($p < 1,28 \times 10^{-04}$) et les duplications ($p < 4,89 \times 10^{-05}$). Les points rouges représentent les délétions et les bleus, les duplications. Les points entourés de vert représentent les tests qui passent la correction FDR.

5.4 Objectif 2b : Impact des gènes individuels sur le risque de TSA

Nous avons également comme objectif de mesurer l'impact du dosage génique sur le risque de TSA. Nous avons donc réalisé une CNV GWAS sur tous les individus non-apparentés de notre jeu de données ($n=457\ 398$), incluant 7 211 cas et 450 187 individus de la population générale (Tableau 4). Il est à noter qu'au moment de la réalisation de ces analyses, les individus et UKBB ayant réalisé un test cognitif en ligne n'étaient pas disponibles et ne sont donc pas inclus. Dans ce jeu de données, 17 630 gènes différents étaient inclus dans les CNV identifiés (Tableau 12).

Après avoir retiré les gènes ayant moins de 10 porteurs, 1 337 gènes emportés par des délétions et 3 465 gènes emportés par des duplications ont été analysés, pour un total de 3 888 gènes différents (Figure 22 et 23). Ces gènes se regroupent en 661 clusters pour les délétions et 1 682 clusters pour les duplications. Parmi les tests réalisés, 292 associations significatives (21,8%) se regroupant en 85 clusters distincts (12,9%) ont été trouvées pour les délétions. Pour les duplications, 461 associations significatives (13,3%) représentées par 161 clusters (9,6%) ont été trouvées.

TYPE	Délétion	Duplication
Nombre de gène	17 630	
Nombre de gène avec ≥ 10 porteurs	1 337 FDR<0,05 : 292	3 465 FDR<0,05 : 461
Nombre de clusters	661 FDR<0,05 : 85	1 682 FDR<0,05 : 161

Tableau 12. – Gènes inclus dans la CNV GWAS sur le risque de TSA.

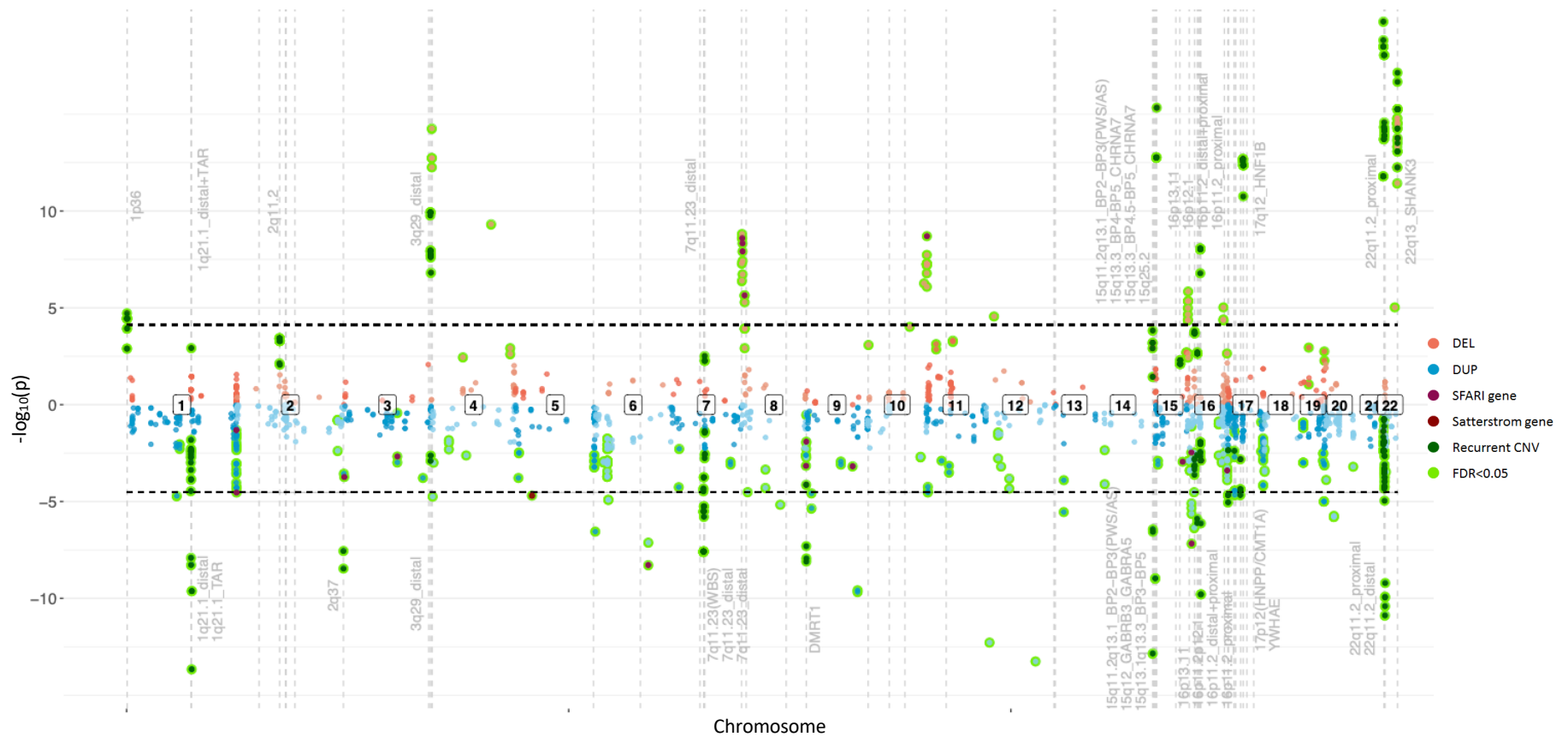


Figure 22. – Miami plot des gènes associés à une augmentation du risque de TSA lorsqu’ affectés par le dosage génique. L’axe des Y représente la p-value non-ajustée des tests, transformée en $-\log_{10}$ (plus un gène emporté par une délétion a un score élevé sur cet axe, plus le test est significatif; les duplications sont représentées en miroirs sur la partie négative de l’axe). Les lignes noires horizontales pointillées représentent la correction de Bonferroni sur le nombre de clusters pour les délétions ($p < 7.56 \times 10^{-05}$) et les duplications ($p < 2.97 \times 10^{-05}$). Les points rouges représentent les gènes délétés et les bleus, les gènes dupliqués. Les points pourpres et bordeaux sont des gènes déjà connus dans la littérature (*i.e.*, SFARI et Satterstrom et al. (2020)) comme étant associés au risque de TSA. Les points vert foncé sont des gènes inclus dans des CNV récurrents. Les points entourés de vert représentent les tests qui passent la correction FDR. Les lignes grises verticales représentent les CNV récurrents, et leur nom est affiché si un gène inclus dans une telle région est significatif.

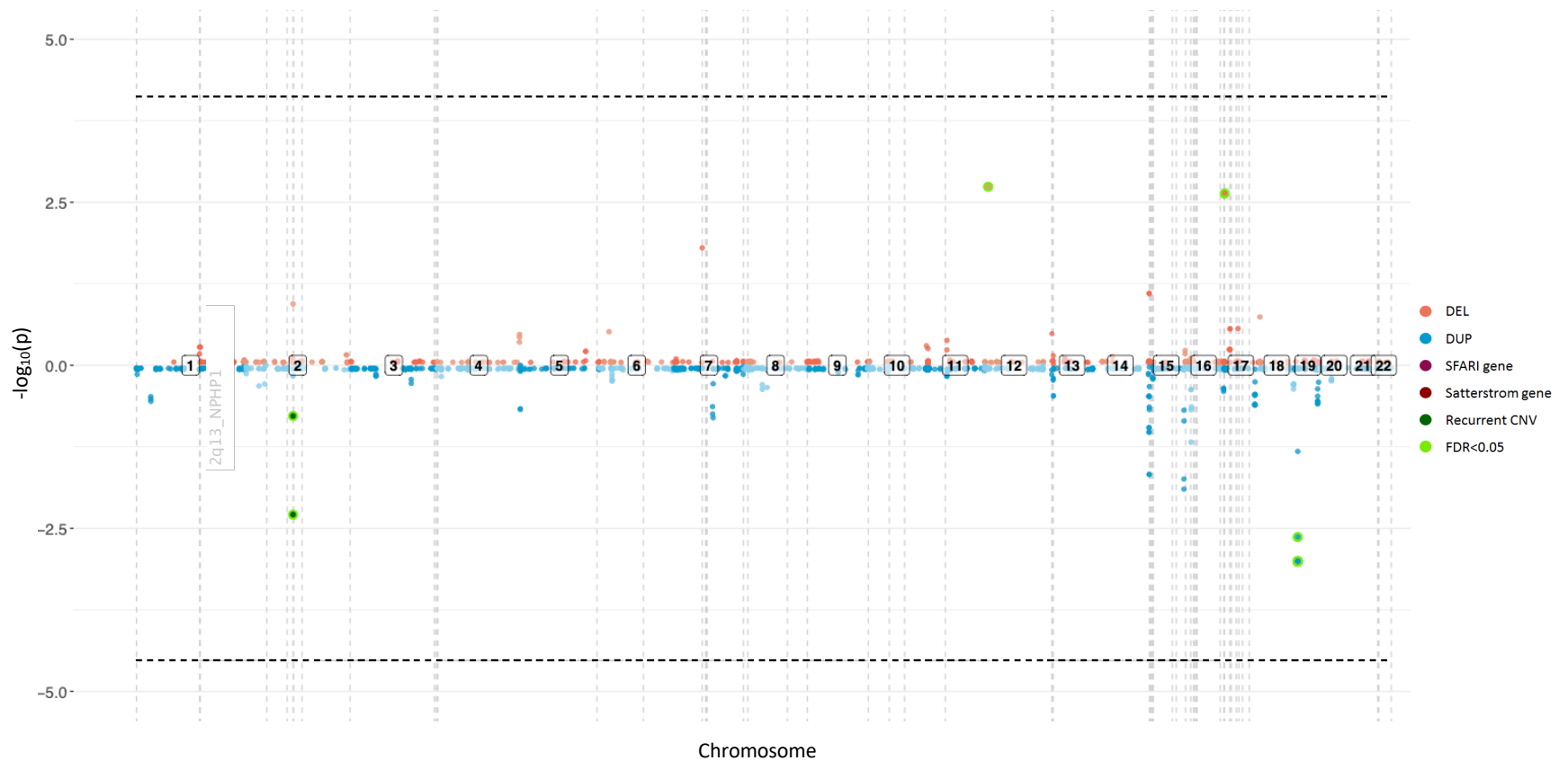


Figure 23. – Miami plot des gènes associés à un effet protecteur sur le risque de TSA lorsqu’ affectés par le dosage génique. L’axe des Y représente la p-value non-ajustée des tests, transformée en $-\log_{10}$ (plus un gène emporté par une délétion a un score élevé sur cet axe, plus le test est significatif; les duplications sont représentées en miroirs sur la partie négative de l’axe). Les lignes noires horizontales pointillées représentent la correction de Bonferroni sur le nombre de clusters pour les délétions ($p < 7.56 \times 10^{-5}$) et les duplications ($p < 2.97 \times 10^{-5}$). Les points rouges représentent les gènes délétés et les bleus, les gènes dupliqués. Les point pourpres et bordeaux sont des gènes déjà connu dans la littérature (*i.e.*, SFARI et Satterstrom et al. (2020)) comme étant associés au risque de TSA. Les points vert foncé sont des gènes inclus dans des CNV récurrents. Les points entourés de vert représentent les tests qui passent la correction FDR. Les lignes grises verticales représentent les CNV récurrents, et leur nom est affiché si un gène inclus dans une telle région est significatif.

La majorité des gènes identifiés affectant le risque de TSA (87,0% des gènes délétés et 66,6% des gènes dupliqués) sont déjà répertoriés dans la littérature (SFARI ou Satterstrom et al. (2020)), ou sont dans un CNV récurrent associé à la psychiatrie (Figure 24). Les autres gènes (13% des délétions et 33,3% des duplications) sont de nouveaux gènes candidats associés au risque de TSA.

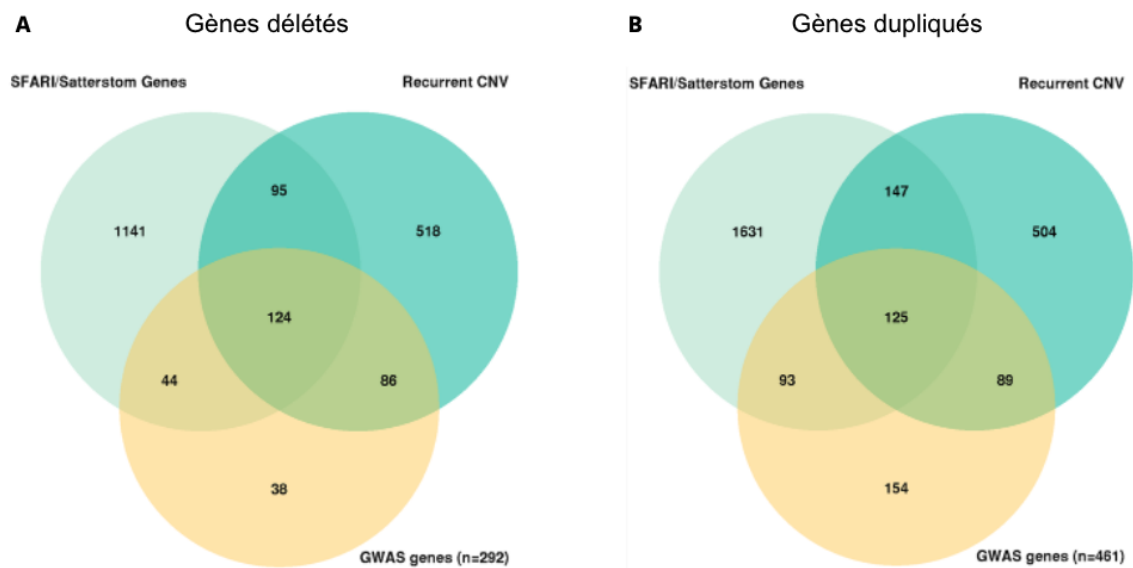


Figure 24. – Catégorisation des gènes avec une association significative dans la CNV GWAS sur le risque de TSA pour (A) les délétions et (B) les duplications. Le cercle de couleur menthe représente les gènes associés au risque de TSA répertoriés dans la base de données SFARI et ceux rapportés par Satterstrom et al. (2020), et les gènes situés à moins de 250 kb de ces gènes. Le cercle turquoise représente les gènes emportés par des CNV récurrents associés aux troubles TND, et les gènes situés à moins de 250 kb de ces CNV. Le cercle jaune représente les associations significatives obtenus dans la CNV GWAS sur le risque de TSA.

Parmi les nouveaux gènes candidats augmentant le risque de TSA, plusieurs régions intéressantes ont été identifiées. Notamment, les régions 6p22.3 (Figure 25) et 6p22.2 (Figure 26), toutes les deux situées sur le chromosome 6.

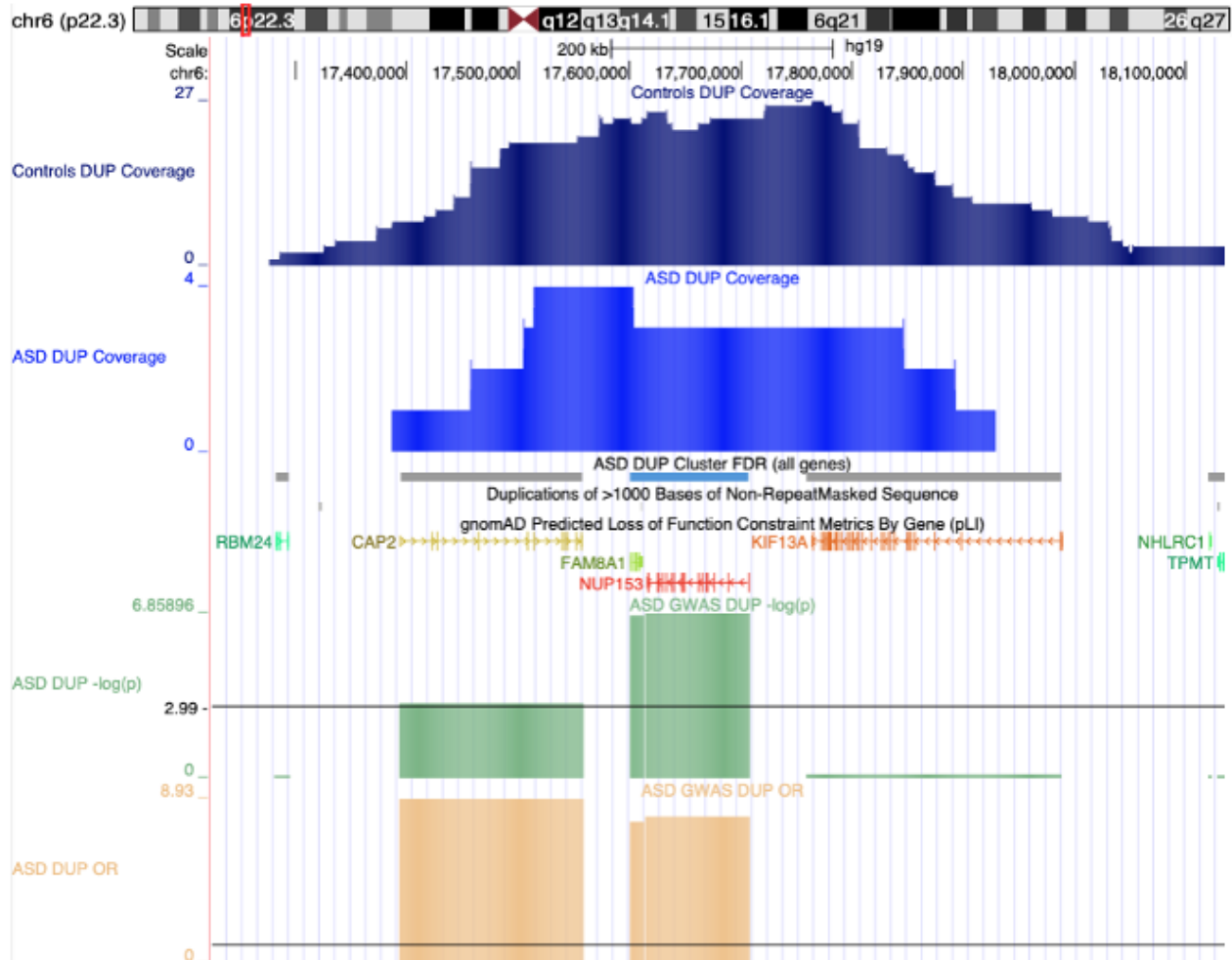


Figure 25. – Duplication 6p22.3 : nouvelle région candidate augmentant le risque de TSA. La couverture des duplications des individus contrôles (Controls DUP Coverage) et TSA (ASD DUP Coverage) sont respectivement illustrées en bleu foncé et en bleu. Les clusters de gènes (ASD DUP Cluster FDR) sont illustrés en bleu lorsque les gènes inclus ont au moins 10 porteurs et en gris lorsqu'ils ont moins de 10 porteurs. Les clusters bleu foncé ou gris foncé représentent les tests significatifs (FDR<0,05). Les duplications segmentaires (Duplications of > 1000 Bases of Non-RepeatMasked Sequence) sont représentées en gris. La tolérance des gènes (gnomadAD Predicted Loss of Function Constraint Metrics By Gene (pLI)) est représentée par le pLI via le gradient de couleur vert (gène tolérant) - rouge (gène intolérant). La p-value du test de la CNV GWAS transformé en $-\log_{10}$ est illustré en vert, la ligne noire correspondant à une p-value non ajustée<0,05. Le odds ratio est représenté en jaune, la ligne noire correspondant à une augmentation du risque de TSA (OR>1). Figure générée via UCSC.

Le cluster de la région 6p22.3 inclut deux gènes complètement emportés par les CNV de notre jeu de données (*FAM8A1*, *NUP153*) (Tableau 13, Figure 25). Le gène *NUP153* (OR : 7,67; $p=1,42 \times 10^{-02}$) est un gène intolérant (LOEUF=0,17) qui régule le transport des macromolécules entre le noyau et le cytoplasme de la cellule. Peu d'informations sont disponibles sur le gène tolérant (LOEUF=0,69) *FAM8A1* (OR : 7,93; $p=1,42 \times 10^{-02}$). Ces deux gènes se retrouvent dans trois cohortes distinctes, soient les cohortes de la population générale UKBB et CaG, ainsi que la cohorte autistique SPARK (Tableau 14). La fréquence avec laquelle on retrouve ces gènes dans une duplication est similaire dans les différentes cohortes. L'absence de ce variant dans les autres cohortes peut s'expliquer par leur taille trop petite pour contenir un variant aussi rare.

gene	CHR	START	STOP	LOEUF	n	OR	P	FDR
<i>FAM8A1</i>	6	17600533	17611946	0,69	26	7,67	1,20E-03	1,42E-02
<i>NUP153*</i>	6	17615266	17707156	0,17	24	7,93	1,05E-03	

Tableau 13. – Gènes totalement emportés par la région 6p22.3 associée à l'augmentation du risque de TSA. Le gène de référence du cluster est indiqué par une étoile (*). CHR : chromosome; n : nombre de porteurs du gène; OR : odds ratio; P : p-value.

gene	Freq. total	Freq. CaG	Freq. G-scot	Freq. Imagen	Freq. LBC	Freq. SPARK	Freq. SSC	Freq. SYS	Freq. UKBB
<i>FAM8A1</i>	0,00568	0,01808	0	0	0	0,05742	0	0	0,00504
<i>NUP153*</i>	0,00525	0,01808	0	0	0	0,05742	0	0	0,00459

Tableau 14. – Fréquence des gènes de la région 6p22.3 associée à l'augmentation du risque de TSA dans les différentes cohortes du jeu de données. Le gène de référence du cluster est indiqué par une étoile (*). Freq. : Fréquence; CaG : CartaGene; G-Scot : Generation Scotland; LBC : Lothian Birth Cohort; SPARK : Simons Foundation Powering Autism Research; SYS : Saguenay Youth Study; UKBB : UKBiobank.

Nous avons remarqué que la duplication de la région 6p22.3 inclut aussi partiellement deux autres gènes : *CAP2* et *KIF13A* (Tableau 15, Figure 25). Ces gènes n'ont pas été inclus dans nos analyses, qui ne comprennent que les gènes complètement emportés par les CNV. Cependant, dans le cadre d'analyses préliminaires sur tous les gènes partiellement emportés par les CNV, nous avons

trouvé une association significative augmentant le risque de TSA associés à ces quatre gènes (Tableau 15). Cette association est plus significative que dans les analyses ne comprenant que les gènes complètement inclus dans les CNV, et les odds ratio sont également plus élevés.

gene	CHR	START	STOP	LOEUF	n	OR	P	FDR
<i>CAP2*</i>	6	17393736	17558021	0,41	25	12,53	8,16E-06	7,57E-05
<i>FAM8A1*</i>	6	17600533	17611946	0,69	29	9,33	5,49E-05	4,25E-04
<i>NUP153</i>		17615266	17707156	0,17	30	6,23	3,17E-03	
<i>KIF13A</i>		17759414	17987866	0,23	30	5,96	3,92E-03	

Tableau 15. – Analyse préliminaire sur les gènes partiellement inclus dans les CNV de la région 6p22.3 associée à l’augmentation du risque de TSA. Le gène de référence du cluster est indiqué par une étoile (*). CHR : chromosome; n : nombre de porteurs du gène; OR : odds ratio; P : p-value.

Le deuxième cluster associé à une augmentation du risque de TSA est la duplication 6p22.2. Ce cluster inclut 26 gènes encodant des histones et deux autres gènes (*HFE*, *AL031777.2*) (Tableau 16, Figure 26).

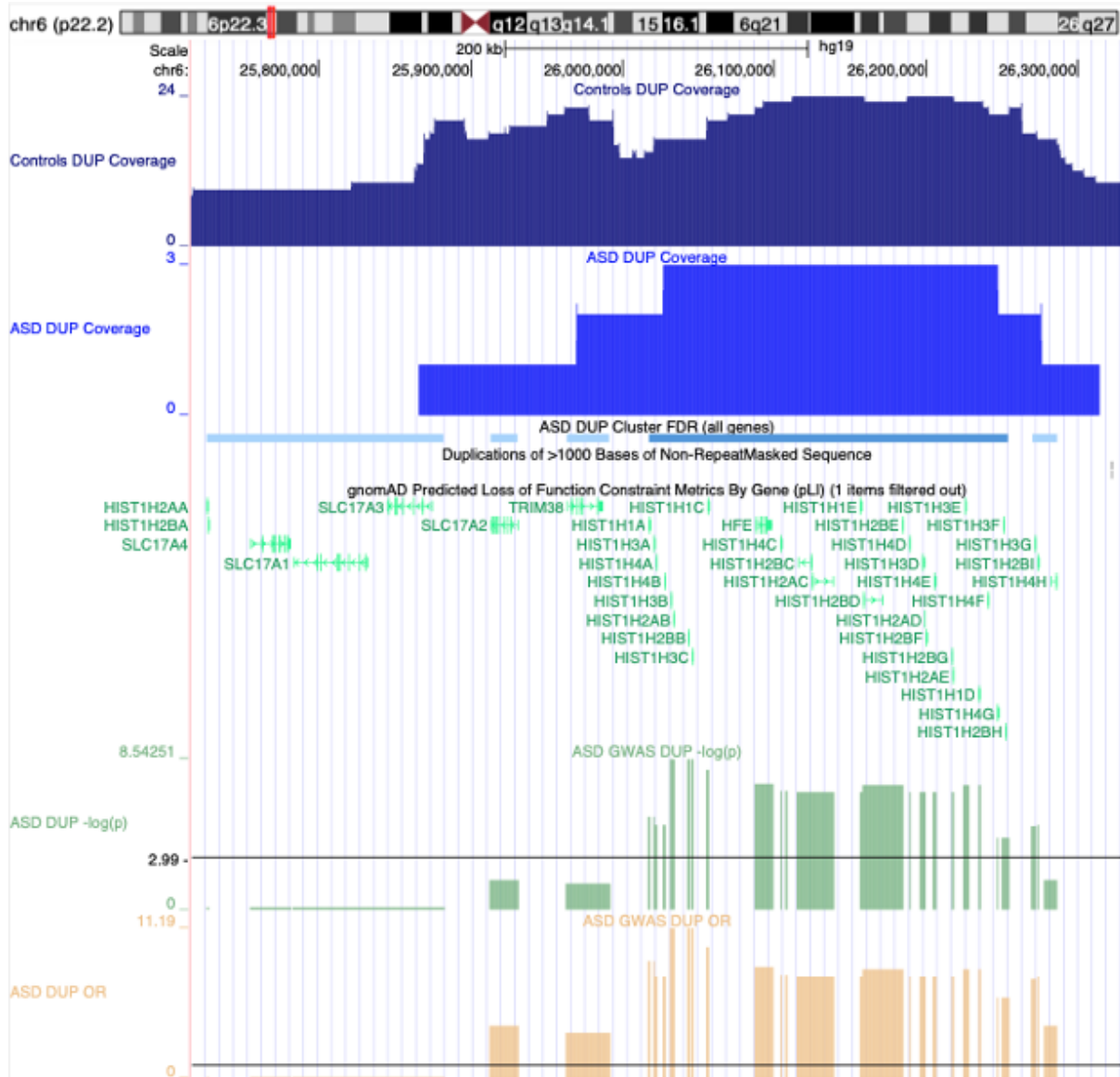


Figure 26. – Duplication 6p22.2 : nouvelle région candidate augmentant le risque de TSA. La couverture des duplications des individus contrôles (Controls DUP Coverage) et TSA (ASD DUP Coverage) sont respectivement illustrées en bleu foncé et en bleu. Les clusters de gènes (ASD DUP Cluster FDR) sont illustrés en bleu lorsque les gènes inclus ont au moins 10 porteurs et en gris lorsqu'ils ont moins de 10 porteurs. Les clusters bleu foncé ou gris foncé représentent les tests significatifs (FDR<0,05). Les duplications segmentaires (Duplications of > 1000 Bases of Non-RepeatMasked Sequence) sont représentées en gris. La tolérance des gènes (gnomadAD Predicted Loss of Function Constraint Metrics By Gene (pLI)) est représentée par le pLI via le gradient de couleur vert (gène tolérant) - rouge (gène intolérant). La p-value du test de la CNV GWAS transformé en $-\log_{10}$ est illustré en vert, la ligne noire correspondant à une p-value non ajustée<0,05. Le odds ratio est représenté en jaune, la ligne noire correspondant à une augmentation du risque de TSA (OR>1). Figure générée via UCSC.

gene	CHR	START	STOP	LOEUF	n	OR	P	FDR
<i>H1-1</i>	6	26017260	26018015		17	8,69	5,27E-03	3,68E-03
<i>H3C1</i>	6	26020679	26021186		17	8,69	5,27E-03	
<i>H4C1</i>	6	26021877	26022278		19	7,57	8,30E-03	
<i>HIST1H4B</i>	6	26027124	26027480	1,94	19	7,57	8,30E-03	
<i>H3C2*</i>	6	26031817	26032327		20	11,19	1,95E-04	
<i>H2AC4</i>	6	26033320	26033846		20	11,19	1,95E-04	
<i>H2BC3</i>	6	26043455	26043941		20	11,19	1,95E-04	
<i>H3C3</i>	6	26045612	26046097		20	11,19	1,95E-04	
<i>H1-2</i>	6	26055968	26056698		23	9,73	3,70E-04	
<i>HFE</i>	6	26087509	26098571	1,19	25	8,26	8,35E-04	
<i>H4C3</i>	6	26104161	26104565		26	7,67	1,20E-03	
<i>H1-6</i>	6	26107640	26108363		26	7,67	1,20E-03	
<i>H2BC4</i>	6	26115101	26124154		27	7,55	1,28E-03	
<i>H2AC6</i>	6	26124373	26139344		27	7,55	1,28E-03	
<i>H1-4</i>	6	26156557	26157343		27	7,55	1,28E-03	
<i>H2BC5</i>	6	26158374	26171577		26	8,11	8,96E-04	
<i>H2BC6</i>	6	26172287	26184883		26	8,11	8,96E-04	
<i>H4C4</i>	6	26188938	26189340		27	7,55	1,28E-03	
<i>H3C4</i>	6	26197012	26197514		27	7,55	1,28E-03	
<i>AL031777.2</i>	6	26195823	26199521		27	7,55	1,28E-03	
<i>H2AC7</i>	6	26199012	26199521		27	7,55	1,28E-03	
<i>H4C5</i>	6	26204780	26206266		27	7,55	1,28E-03	
<i>H2AC8</i>	6	26217203	26217711		27	7,55	1,28E-03	
<i>H3C6</i>	6	26224427	26227701		26	8,11	8,96E-04	
<i>H1-3</i>	6	26234440	26235215		26	8,11	8,96E-04	
<i>H4C7</i>	6	26246839	26247224		23	6,03	1,75E-02	
<i>H3C7</i>	6	26250370	26250863		23	6,03	1,75E-02	
<i>H2BC9</i>	6	26251879	26253938		23	6,03	1,75E-02	

Tableau 16. – Gènes totalement emportés par la région 6p22.2 associée à l'augmentation du risque de TSA. La région comprend 28 gènes dont 26 encodent des histones. Le gène de référence du cluster est indiqué par une étoile (*). CHR : chromosome; n : nombre de porteurs du gène; OR : odds ratio; P : p-value.

Ces gènes se retrouvent dans quatre cohortes : les deux cohortes autistiques SPARK et SSC, ainsi que dans UKBB et CaG (Tableau 17). Comme il s'agit d'un variant rare, l'absence de ce variant dans les autres cohortes peut s'expliquer par leur taille plus petite.

gene	Freq. total	Freq. CaG	Freq. G-scot	Freq. Imagen	Freq. LBC	Freq. SPARK	Freq. SSC	Freq. SYS	Freq. UKBB
<i>H1-1</i>	3,72E-03	1,81E-02	0	0	0	1,91E-02	5,04E-02	0	3,21E-03
<i>H3C1</i>	3,72E-03	1,81E-02	0	0	0	1,91E-02	5,04E-02	0	3,21E-03
<i>H4C1</i>	4,15E-03	1,81E-02	0	0	0	1,91E-02	5,04E-02	0	3,67E-03
<i>HIST1H4B</i>	4,15E-03	1,81E-02	0	0	0	1,91E-02	5,04E-02	0	3,67E-03
<i>H3C2*</i>	4,37E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	3,67E-03
<i>H2AC4</i>	4,37E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	3,67E-03
<i>H2BC3</i>	4,37E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	3,67E-03
<i>H3C3</i>	4,37E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	3,67E-03
<i>H1-2</i>	5,03E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	4,36E-03
<i>HFE</i>	5,47E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	4,81E-03
<i>H4C3</i>	5,68E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,04E-03
<i>H1-6</i>	5,68E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,04E-03
<i>H2BC4</i>	5,90E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,27E-03
<i>H2AC6</i>	5,90E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,27E-03
<i>H1-4</i>	5,90E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,27E-03
<i>H2BC5</i>	5,68E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,04E-03
<i>H2BC6</i>	5,68E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,04E-03
<i>H4C4</i>	5,90E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,27E-03
<i>AL031777.2</i>	5,90E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,27E-03
<i>H3C4</i>	5,90E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,27E-03
<i>H2AC7</i>	5,90E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,27E-03
<i>H4C5</i>	5,90E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,27E-03
<i>H2AC8</i>	5,90E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,27E-03
<i>H3C6</i>	5,68E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,04E-03
<i>H1-3</i>	5,68E-03	1,81E-02	0	0	0	3,83E-02	5,04E-02	0	5,04E-03
<i>H4C7</i>	5,03E-03	1,81E-02	0	0	0	1,91E-02	5,04E-02	0	4,59E-03
<i>H3C7</i>	5,03E-03	1,81E-02	0	0	0	1,91E-02	5,04E-02	0	4,59E-03
<i>H2BC9</i>	5,03E-03	1,81E-02	0	0	0	1,91E-02	5,04E-02	0	4,59E-03

Tableau 17. – Fréquence des gènes de la région 6p22.2 associée à l'augmentation du risque de TSA dans les différentes cohortes du jeu de données. Le gène de référence du cluster est indiqué par une étoile (*). Freq. : Fréquence; CaG : CartaGene; G-Scot : Generation Scotland; LBC : Lothian Birth Cohort; SPARK : Simons Foundation Powering Autism Research; SYS : Saguenay Youth Study; UKBB : UKBiobank.

En plus de ces nouveaux gènes candidats augmentant le risque de TSA (Tableau supplémentaire 4 et 6), les analyses ont également révélé 10 gènes associés significativement à un effet protecteur sur le risque de TSA (Tableau 18). Ces gènes se regroupent en trois clusters différents : un emporte des gènes délétés et deux emportent des gènes dupliqués. Un point notable est que l'un de ces clusters (incluant les gènes *MALL* et *NPHP1*) fait partie d'une région récurrente, le 2q13.

Gene	TYPE	CHR	START	STOP	LOEUF	n (Total)	n (nonASD)	n (ASD)	OR	CI95%		P	FDR	REC
<i>MALL</i> *	DUP	2	110841447	110873599	1,41	888	886	2	0,14	0,02	0,43	5,14E-03	4,79E-02	2q13 NPHP1
<i>NPHP1</i>			110879888	110962643	0,98	952	942	10	0,64	0,32	1,14	1,67E-01		2q13 NPHP1
<i>SLC2A3</i> *	DEL	12	8071826	8088783	0,65	3597	3564	33	0,58	0,40	0,80	1,83E-03	1,68E-02	-
<i>MBD3L2B</i> *	DUP	19	7018980	7021461	-	2559	2540	19	0,47	0,29	0,71	9,91E-04	1,39E-02	-
<i>MBD3L5</i>			7030589	7033022	1,93	2558	2539	19	0,47	0,29	0,71	9,94E-04		-
<i>MBD3L4</i>			7037759	7040190	1,89	2558	2539	19	0,47	0,29	0,71	9,94E-04		-
<i>MBD3L2</i>			7049332	7051746	-	2558	2539	19	0,47	0,29	0,71	9,94E-04		-
<i>MBD3L3</i>			7056220	7058651	1,89	2558	2539	19	0,47	0,29	0,71	9,94E-04		-
<i>ZNF557</i>			7069714	7087979	0,93	2402	2383	19	0,49	0,30	0,75	2,34E-03		-

Tableau 18. – Gènes associés à un effet protecteur sur le risque de TSA. Le gène de référence du cluster est indiqué par une étoile (*). CHR : chromosome; n : nombre de porteurs du gène; ASD : trouble du spectre autistique; OR : odds ratio; P : p-value.

gene	Freq. total	Freq. CaG	Freq. G-scot	Freq. Imagen	Freq. LBC	Freq. SPARK	Freq. SSC	Freq. SYS	Freq. UKBB
<i>MALL</i> *	0,19	0,00	0,00	0,12	0,00	0,02	0,05	0,00	0,20
<i>NPHP1</i>	0,21	0,33	0,18	0,18	0,18	0,06	0,35	0,17	0,21
<i>SLC2A3</i> *	0,79	0,40	0,32	0,42	0,54	0,27	0,96	0,83	0,80
<i>MBD3L2B</i> *	0,56	0,25	0,23	0,60	1,09	0,25	0,30	0,17	0,57
<i>MBD3L5</i>	0,56	0,24	0,23	0,60	1,09	0,25	0,30	0,17	0,57
<i>MBD3L4</i>	0,56	0,24	0,23	0,60	1,09	0,25	0,30	0,17	0,57
<i>MBD3L2</i>	0,56	0,24	0,23	0,60	1,09	0,25	0,30	0,17	0,57
<i>MBD3L3</i>	0,56	0,24	0,23	0,60	1,09	0,25	0,30	0,17	0,57
<i>ZNF557</i>	0,53	0,20	0,19	0,42	0,91	0,25	0,30	0,00	0,54

Tableau 19. – Fréquence des gènes associés à un effet protecteur sur le risque de TSA. Le gène de référence du cluster est indiqué par une étoile (*). Freq. : Fréquence; CaG : CartaGene; G-Scot : Generation Scotland; LBC : Lothian Birth Cohort; SPARK : Simons Foundation Powering Autism Research; SYS : Saguenay Youth Study; UKBB : UKBiobank.

5.5 Pléiotropie TSA - Intelligence générale

Comme il existe un débat actuel sur l'effet potentiellement préférentiel des CNV sur l'intelligence générale ou sur le risque de TSA, nous avons également voulu comparer les résultats de la CNV GWAS sur ces deux phénotypes.

Au total, la CNV GWAS sur l'intelligence générale et le risque de TSA ont trouvé 349 associations significatives avec l'un des deux (ou les deux) phénotypes pour les délétions et 547 pour les duplications (Tableau 20). Parmi elles, respectivement pour les délétions et pour les duplications, 51,9% et 55,4% des gènes sont seulement associés au risque de TSA; 16,3% et 15,7% des gènes sont seulement associés à l'intelligence générale; 31,2% et 28,9% sont associés aux deux phénotypes (Tableau 20).

TYPE	Délétions (%)	Duplications (%)
N associations significatives		
Intelligence générale	57 (16,3)	86 (15,7)
TSA	181 (51,9)	303 (55,4)
Intelligence générale + TSA	111 (31,2)	158 (28,9)
Total	349 (100)	547 (100)

Tableau 20. – Effet pléiotropique des gènes sur l'intelligence générale et le risque de TSA.

Nous avons également fait un deuxième modèle de régression logistique sur le risque de TSA, mais qui inclut cette fois-ci en plus le score Z de l'intelligence générale comme covariable.

Parmi les 1 730 gènes investigués, 622 et 1 297 gènes emportés respectivement par une délétion ou une duplication avaient au moins 10 porteurs, dont 224 (36,0%) et 338 (26,1%) associations significatives, respectivement (Tableau 21, Figure 27). Ces gènes se regroupent dans un total de 260 clusters de délétions, dont 58 significatifs (22,3%) (Tableau supplémentaire 5), et 581 clusters

de duplications, dont 130 significatifs (22,3%) (Tableau supplémentaire 7). Aucun de ces gènes ou clusters n'ont un effet significatif protecteur sur le risque d'autisme.

TYPE	Délétion	Duplication
Nombre de gène	17 630	
Nombre de gène avec ≥ 10 porteurs	622 FDR<0,05 : 224	1 297 FDR<0,05 : 338
Nombre de clusters	260 FDR<0,05 : 58	581 FDR<0,05 : 130

Tableau 21. – Gènes inclus dans la CNV GWAS sur le risque de TSA covarié avec le score Z de l'intelligence générale ajusté selon l'âge.

En covariant le risque de TSA avec l'intelligence générale, on voit bien que certains gènes ont un effet spécifique sur le risque de TSA, sans être affecté par l'intelligence générale (Figure 27).

Pour les délétions (Figure 28A), 91,5% de ces gènes étaient significativement associés aux TSA, dont 88,1% sont aussi associés à l'intelligence générale. 8,5% n'ont pas été rapportés dans ni l'une ni l'autre des analyses précédentes. Aucune délétion associée strictement à l'intelligence générale n'a été révélée par cette analyse.

Pour les duplications (Figure 28B), 78,7% de ces gènes étaient significativement associés aux TSA, dont 58,3% sont aussi associés à l'intelligence générale. 21,3% n'ont pas été rapportés dans ni l'une ni l'autre des analyses précédentes. Aucune duplication associée strictement à l'intelligence générale n'a été révélée par cette analyse.

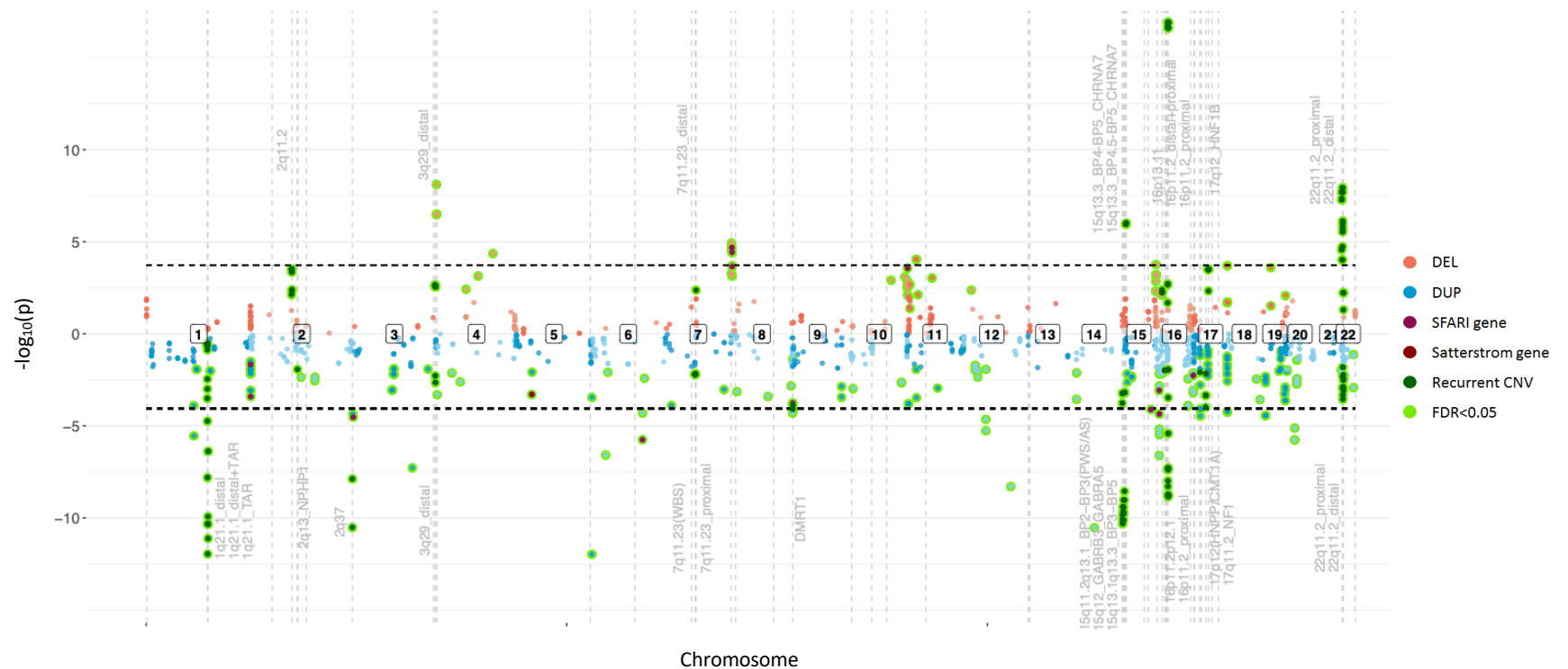


Figure 27. – Miami plot des gènes associés à une augmentation du risque de TSA lorsqu’ affectés par le dosage génique dans le modèle covarié avec le score Z de l’intelligence général ajusté par l’âge. L’axe des X représente la position des gènes sur chaque chromosome. L’axe des Y représente la p-value non-ajustée des tests, transformée en $-\log_{10}$ (plus un gène emporté par une délétion a un score élevé sur cet axe, plus le test est significatif; les duplications sont représentées en miroirs sur la partie négative de l’axe). Les lignes noires horizontales pointillées représentent la correction de Bonferroni sur pour les délétions ($p < 1,92 \times 10^{-04}$) et les duplications ($p < 8,61 \times 10^{-05}$). Les points rouges représentent les gènes délétés et les bleus, les gènes dupliqués. Les point pourpres et bordeaux sont des gènes déjà connu dans la littérature (*i.e.*, SFARI et Satterstrom et al. 2020) comme étant associés au risque de TSA. Les points vert foncé sont des gènes inclus dans des CNV récurrents. Les points entourés de vert représentent les tests qui passent la correction FDR. Les lignes grises verticales représentent les CNV récurrents, et leur nom est affiché si un gène inclus dans une telle région est significatif.

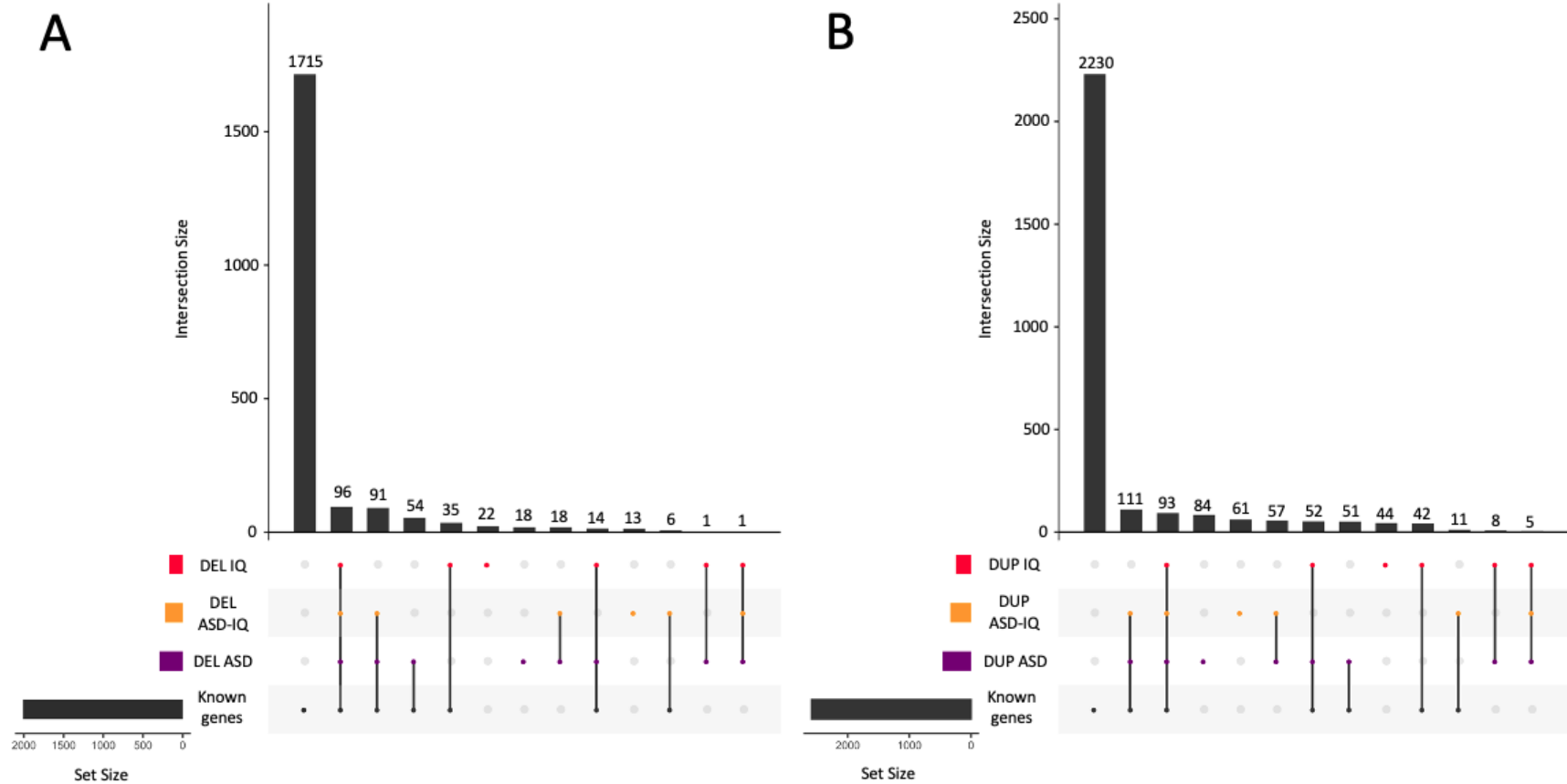


Figure 28. – Intersections des gènes associés au risque de TSA, covarié ou non avec l’intelligence générale, et des gènes associés à l’intelligence générale pour A. les délétions et B. les duplications. L’axe des x représente les différentes intersections entre les gènes provenant des CNV GWAS ou des gènes connus dans la littérature. L’axe des y représente la taille de ces intersections. La taille de chaque ensemble est représentée sur la gauche du graphe principal. Les points mauves représentent la CNV GWAS sur le risque de TSA, les rouges représentent celle sur l’intelligence générale et les oranges représentent celle sur le risque de TSA covarié avec l’intelligence générale. Les points noirs représentent des gènes déjà connus dans la littérature (*i.e.*, SFARI et Satterstrom et al. (2020)) comme étant associés au risque de TSA ou des gènes inclus dans des CNV récurrents, incluant les gènes qui sont à moins de 250 kb de ces régions connus.

Chapitre 6 – Discussion

6.1 Normalisation des mesures cognitives

Avant de plonger dans les objectifs principaux de ce projet, nous avons exploré deux corrections différentes du score Z de l'intelligence générale sur l'âge et le sexe. La première correction est la même qu'utilisée précédemment par Huguet et al. (2021) (76), réalisée via les résidus du modèle linéaire. La deuxième correction est une correction par GP (94,95). Les avantages de cette correction, contrairement à celle des résidus du modèle linéaire, est qu'elle ne prend pas pour acquis que la correction est linéaire. Nous avons comparé les deux types de corrections et avons constaté une grande corrélation entre les données corrigées par les deux méthodes ($P=0,99$, Figure 10C). Les endroits où on voit le plus de différences entre les deux corrections sont aux extrémités des courbes, là où les individus les plus jeunes et les plus vieux sont moins nombreux (Figure 10D). De surcroît, nous avons estimé l'impact du dosage génique via la somme du 1/LOEUF des gènes présents dans les CNV sur l'intelligence générale. Nous avons pu constater que le modèle le plus significatif est celui corrigeant le phénotype avec les résidus du modèle linéaire, avec des estimés similaires, mais un écart type plus restreint (résidus du modèle linéaire: $-0,029$; $SD=0,001$; $p=1,45 \times 10^{-56}$; GP: $-0,027$; $SD=0,002$; $p=1,40 \times 10^{-53}$) (Figure 11). Il semble donc que les résidus du modèle linéaire ajoutent moins de bruit dans les analyses que le GP aux vues de la différence d'écart type.

Une limite de cette analyse est que seul UKBB a été utilisé pour comparer les deux corrections, cela en raison de la grande quantité de ressources en mémoire de calcul et du temps nécessaire pour réaliser cette correction. En utilisant tout le jeu de données, incluant les cohortes dont la moyenne d'âge est plus jeune, peut-être que la différence entre le déclin de l'intelligence avant 60 ans serait mieux corrigée et permettrait d'augmenter la robustesse du modèle mesurant l'impact des CNV sur l'intelligence générale. Une perspective serait d'inclure la totalité du jeu de données afin de tester cette hypothèse en utilisant la variation stochastique du GP, désormais disponible dans la librairie PyNM (95). Cette méthode, qui n'était pas disponible lors de ces

analyses, est une version beaucoup plus efficace du GP qui utilise moins de ressources pour fonctionner.

6.2 Objectif 1 : Impact des CNV du chromosome X sur la cognition

Notre premier objectif dans le cadre de ce projet était de mesurer l'impact des CNV du chromosome X sur l'intelligence générale. Pour ce faire, nous avons utilisé la somme du 1/LOEUF pour les gènes présents dans les délétions et les duplications séparément, comme il a été fait avec succès sur les autosomes (76). De précédents résultats préliminaires provenant de travaux non publiés ont montré un effet significatif des CNV du chromosome X sur l'intelligence générale avec un petit jeu de données (n=38 756). En raison de la grande taille de la cohorte UKBB utilisée (n=238 176, Tableau 5), nous pensions avoir suffisamment de puissance statistique pour répliquer ce signal obtenu précédemment. Outre pour les duplications chez les hommes, aucun signal similaire n'a été obtenu. Nous soupçonnons qu'il s'agit d'un problème lié aux CNV mosaïques du chromosome X, qui surviennent avec l'avancement de l'âge dans les cellules sanguines (103). UKBB étant une cohorte adulte avec une moyenne d'âge élevée (Tableau 5) et les échantillons biologiques étant sanguins, ce phénomène se retrouve probablement enrichi par rapport aux précédentes cohortes, qui sont plus jeunes. Pour y remédier, nous avons tenté de filtrer les individus ayant plus de 55 ans, puisque c'est après cet âge que le mosaïcisme augmente particulièrement (104), et nous avons aussi filtré les CNV chevauchant les régions du chromosome X propices au mosaïcisme (101). Ces filtres n'ont pas été suffisants pour répliquer les résultats préliminaires. Nous ne pouvons donc conclure que l'effet est le même que pour les autosomes, soit parce que l'impact des CNV du chromosome X est différent de ces derniers, ou alors le bruit masque leur effet.

Une limite de notre approche est que nous n'avons pas investigué systématiquement tous les CNV pour y détecter leur mosaïcisme. Une méthode plus systématique de détection des CNV mosaïques comme le logiciel MoChA (101,104) serait une alternative à utiliser en perspective pour parvenir à mieux cibler les CNV mosaïque et potentiellement capturer un effet significatif de l'effet des CNV du chromosome X sur l'intelligence générale. Cela permettrait également de ne

pas filtrer systématiquement toutes les personnes âgées : le filtre retirant les individus de plus de 55 ans (soit 75% de la cohorte utilisée) réduit énormément la taille de l'échantillon, ce qui réduit la puissance statistique. En conservant les personnes âgées qui n'ont pas de CNV mosaïque via un logiciel comme MoChA, la taille de l'échantillon ne serait pas affectée.

6.3 Objectif 2 : Impact du dosage génique sur la cognition et le risque de TSA

6.3.1 Impact du dosage génique sur l'intelligence générale

Pour quantifier l'impact individuel de chaque gène présent dans les CNV sur l'intelligence générale, nous avons ensuite réalisé une CNV GWAS sur les gènes délétés ou dupliqués sur l'intelligence générale. Cette méthodologie permet d'avoir un estimé pour chaque gène sur le phénotype testé, apportant une vision plus résolutive de l'impact des CNV sur la cognition. Parmi les résultats obtenus, nous avons capturé une constellation d'effets positifs et négatifs sur l'intelligence générale, ce qui ne peut être capturé en mesurant l'effet moyen des CNV par la somme du $1/LOEUF$ des gènes qu'ils incluent. De plus, cette méthodologie nous a permis d'identifier 24 et 57 nouveaux gènes candidats qui ont un effet sur l'intelligence générale s'ils sont respectivement délétés ou dupliqués et qui ne sont pas associés au risque de TSA dans la littérature (SFARI (12), Satterstrom et al. 2020 (68)) et ne font pas partie de la liste CNV récurrents associés à la psychiatrie provenant de Huguet et al (2021) (Figure 17).

Une limite de cette analyse est que les tests sont faits sur les gènes inclus dans les CNV et non sur les CNV eux-mêmes : l'effet capturé d'un gène sur l'intelligence peut donc être celui de n'importe quel autre gène emporté par ce CNV, qu'il soit ou non dans le même cluster. En perspective, il serait pertinent de réaliser une CNV GWAS similaire à celle présentée dans l'analyse de sensibilité sur l'âge, ou les régions investiguées ne sont pas des gènes, mais des CNV. L'utilisation d'un autre type de variant génétique plus spécifique des gènes touchés, comme les variations ponctuelles

de perte de fonction, seraient également pertinentes à la compréhension des mécanismes derrière les associations trouvées.

Une seconde limite est que nous avons inclus des cohortes autistiques pour les analyses sur l'intelligence générale, pouvant faire entrer en jeu un effet pléiotropique. Le signal de certains gènes pourrait donc être observé en raison de la présence de TND chez certains individus, sachant que ces troubles sont très comorbides avec une diminution du QI.

6.3.1.1 Impact des gènes tolérants sur l'intelligence générale

Nous nous sommes ensuite intéressés aux gènes ayant un impact positif sur l'intelligence générale. En effet, trois clusters de gènes dupliqués et un cluster de gènes délétés avec un impact positif sur l'intelligence ont été identifiés dans la CNV GWAS (Tableau 7). De plus, ces clusters ont tous au moins un gène inclu dans la fenêtre de LOEUF =]1,44:1,85], où on constate dans la sliding-window (Figure 7) que les gènes ont également un effet positif sur l'intelligence générale. Cet effet est davantage visible lorsqu'on contrôle l'analyse avec le poids des autres gènes affectés par le dosage génique pour chaque individu, principalement pour les duplications : sans ce contrôle additionnel, l'effet est masqué par celui des autres gènes inclus dans les délétions ou duplications (Figure 19). En observant les estimés des gènes d'intérêt, on voit bien que pour les duplications, la majorité (90,9%) des gènes significatifs après correction FDR ont un impact positif sur l'intelligence générale (Figure 19D). Or, pour les délétions, cet effet est moins clair : on ne voit qu'un gène (5,6%) affectant positivement l'intelligence générale après correction FDR (Figure 19B, Tableau 10). Il faut aussi noter que les gènes dupliqués qui ont un impact significatif positif sur l'intelligence générale ne sont que dans 8 clusters différents : il s'agit donc principalement de gènes inclus dans les mêmes CNV, et non de multiples observations indépendantes.

De plus, le test d'hétérogénéité confirme également qu'il y a une constellation d'effets positifs et négatifs sur l'intelligence générale pour les gènes de cette fenêtre, ce qui produit du bruit dans nos résultats.

Une limite est que la fenêtre de LOEUF investiguée (LOEUF =]1,44:1,85]) dans ces analyses est très large et inclut énormément de gènes différents (n=474). De plus, cet intervalle est basé sur les résultats de la sliding-window, qui est une analyse préliminaire. L'effet que nous cherchons à capturer est donc potentiellement biaisé par celui d'autres gènes ayant un effet opposé (*i.e.*, test d'hétérogénéité). Et pour finir, nous n'avons testé uniquement les gènes qui avaient au moins 10 observations pour les délétions ou les duplications. Dès lors, il se pourrait que le signal soit la résultante de l'effet cumulé de tous les gènes avec moins de 10 porteurs. En perspective, un modèle plus résolutif incluant des mesures basées sur les fonctions biologiques des gènes ou de leur expression spécifique dans le cerveau aiderait à raffiner nos observations et à mieux comprendre les mécanismes biologiques influençant l'intelligence générale.

6.3.1.2 Analyse de sensibilité sur l'âge

Suite à des analyses de sensibilités préliminaires montrant que l'impact du dosage génique semble moins sévère chez les individus les plus âgés, nous avons réalisé un modèle mesurant l'impact de la somme du 1/LOEUF des gènes pour les délétions et les duplications sur l'âge. Peu importe la tranche d'âge étudiée (<60 ans, ≥60 ans ou tout âges confondus), l'augmentation de la somme de 1/LOEUF pour les délétions et les duplications est toujours associée à une diminution de l'âge dans UKBB (Tableau 11). Cet effet n'est cependant pas observable dans les autres cohortes de la population générale regroupées ensemble (Imagen, SYS, LBC, G-Scot), ce qui s'explique potentiellement par la moyenne d'âge moins élevée dans ces cohortes comparée à UKBB. Une publication récente montre que la charge des CNV, mesurée par leur taille ou par le nombre de gènes qu'ils emportent, diminue significativement l'espérance de vie (105). L'effet moins fort des CNV chez les plus âgés pourrait donc s'expliquer par le fait que ceux qui ont une charge élevée de CNV ont disparu de la population ou sont recrutés dans des cohortes différentes, associées à des troubles plus sévères.

En cherchant spécifiquement les régions associées à une diminution de l'âge via une CNV GWAS sur l'âge dans UKBB, nous avons constaté que la délétion de la région 16p11.2 était

significativement associée à une diminution de l'âge (Figure 21B). Or, il a été vu précédemment (106) que la délétion de la région 16p11.2 est associée à une qualification scolaire, un emploi et un salaire annuel moindre, ainsi qu'à un indice de défavorisation social élevé, par rapport aux individus sans cette délétion. En raison de ces multiples facteurs, il est possible que les personnes plus âgées portant cette délétion n'aient pas été recrutés, ou que ces personnes soient décédées à un âge précoce. Cela est également un élément pouvant expliquer pourquoi nous observons moins d'individus avec une somme élevée de 1/LOEUF des gènes dans les cohortes plus âgées.

Nous observons également deux duplications qui sont associées avec une augmentation de l'âge dans UKBB : la région du chromosome 10 entre les positions 46 962 114 et 48 005 276 et celle du chromosome 14 entre les positions 19 264 875 et 20 444 607 (Figure 21A). Une hypothèse serait que ces régions sont plus sensibles à la formation de CNV mosaïques avec l'âge, mais nos investigations préliminaires ne semblent pas pointer vers une région reconnue à cet effet. Une autre hypothèse serait la présence de gènes liés à la longévité ou au cancer. Notamment, le gène *ANXA8L1*, situé sur la duplication d'intérêt sur le chromosome 10, est associée à la leucémie myéloblastique aiguë lorsque ce gène est surexprimé (RefSeq) (107). Or, une duplication entraîne justement une augmentation du nombre de transcrits (82).

Dès lors, aux vues des observations précédentes, un double effet semble apparaître avec l'augmentation de l'âge : 1) plus nous vieillissons, moins nous observons de CNV pathogéniques dans le génome pour les raisons énumérées ci-dessus, 2) la fraction de CNV mosaïques augmente également. Ces deux effets combinés génèrent donc du bruit dans les données et doit être pris en compte lors d'analyses de CNV.

Une limite de cette analyse est que l'âge utilisé est celui auquel le phénotype, soit l'intelligence, a été mesuré chez l'individu. Cet âge peut différer de celui de la prise de sang dont l'ADN est extrait pour le génotypage, qui serait plus pertinente dans ce type d'analyses de sensibilité. En perspective, pour approfondir l'interprétation de l'association des régions du chromosome 10 et

14 associées avec l'augmentation de l'âge, l'exploration des gènes inclus dans ces régions (Tableau supplémentaire 3) serait à réaliser avec plus de profondeur.

6.3.2 Impact du dosage génique sur le risque de TSA

Pour quantifier l'impact individuel de chaque gène présent dans les CNV sur le risque de TSA, nous avons également réalisé une CNV GWAS sur ce phénotype. Celle-ci nous a permis de répliquer l'association au risque de TSA déjà identifiées dans la littérature (SFARI, Satterstrom et al. 2020 ou dans un CNV récurrent associé à la psychiatrie) pour 254 gènes pour les délétions et 307 gènes pour les duplications (Figure 24). Les autres gènes indiquent de nouvelles régions candidates affectant le risque de TSA, notamment la duplication de la région 6p22.3 (Figure 25, Tableau 13). Cette région inclut deux gènes dans leur intégralité (*NUP153* et *FAM8A1*) ainsi que deux gènes partiellement altérés (*CAP2*, *KIF13A*) (Figure 25, Tableau 15). Un point intéressant est que deux de ces gènes ont un LOEUF intolérant (*NUP153* : LOEUF=0,17; *KIF13A* : LOEUF=0,23), ce qui pourrait expliquer l'augmentation du risque d'autisme pour les porteurs de cette duplication. La fréquence de cette région est plutôt stable dans les différentes cohortes étudiées (Tableau 14). La stabilité de cette fréquence semble indiquer qu'il ne s'agit pas d'un artefact de détection lié à une technologie ou une cohorte en particulier pouvant biaiser les résultats. Une autre nouvelle région intéressante associée à l'augmentation du risque de TSA est la région 6p22.2 (Figure 26, Tableau 16). Cette région inclut 26 histones. Le risque de TSA pourrait donc être affecté par cette région via des mécanismes épigénétiques modifiant l'expression de gènes qui ne sont pas forcément inclus dans cette région spécifique, mais qui pourrait affecter le risque de TSA. Cette hypothèse est d'ailleurs cohérente avec les gènes précédemment associés au risque de TSA, qui sont enrichis dans la régulation de l'expression des régulateurs de chromatine (68,71). Comme pour la duplication du 6p22.3, la fréquence des gènes inclus dans la duplication du 6p22.2 est plutôt stable dans les différentes cohortes (Tableau 17).

Pour les effets protecteurs sur le risque de TSA, nous avons également détecté trois clusters de gènes (une délétion et deux duplications), dont un cluster (incluant les gènes *MALL* et *NPHP1*) qui

chevauche la duplication récurrente de la région 2q13 (Tableau 18). Ces associations protectrices sont peut-être dues à une surreprésentation de ces régions dans les cohortes de la population générale, telle UKBB, qui possède beaucoup de participants dans notre étude (n=436 162) comparée aux autres cohortes, et qui auraient une sensibilité plus importante à la détection de ce CNV. Aux vues de la littérature, qui rapporte un effet augmentant le risque de TSA lorsque la région 2q13 est dupliquée (108,109), ce signal est ambigu. Il est à noter que dans l'étude du risque de TSA, la vaste majorité des publications sont significatives, rares sont les études négatives. De plus, le petit nombre de participants utilisés (souvent des études de cas) dans l'étude de cette région, il est possible que ces associations au risque de TSA soient erronées. Actuellement, ce type d'effets protecteurs sur certains traits psychiatriques sont de plus en plus fréquents : c'est notamment le cas de la microduplications 22q11.1, qui a été associée à un effet protecteur pour la schizophrénie (90). Dans une récente étude, la duplication 17p12 a également été associée, bien que de façon non significative, à un effet systématiquement protecteur pour le risque de TSA, de TDAH, de dépression majeure, de schizophrénie, et de trouble bipolaire (110). Dès lors, nous posons l'hypothèse que la duplication de la région du 2q13 a un effet protecteur au TSA.

Tout comme pour la CNV GWAS sur l'intelligence générale, une limite de cette analyse est qu'elle est réalisée sur les gènes inclus dans les CNV et non les CNV en tant que tel, ce qui signifie que l'effet capturé n'est pas nécessairement lié directement aux gènes significatifs eux-mêmes, mais à la région.

6.3.3 Pléiotropisme sur l'intelligence générale et le risque de TSA

En comparant les résultats des CNV GWAS sur l'intelligence générale et du risque de TSA, nous avons remarqué que certains gènes ont un effet préférentiel sur un ou l'autre des phénotypes (Figure 28). La CNV GWAS sur le risque de TSA covarié avec l'intelligence générale montre d'ailleurs qu'il existe des gènes qui ont un effet sur le risque de TSA, sans égard pour les habiletés cognitives de leurs porteurs (délétions : n=128/224; duplications : n=268/338) (Figure 28). Cette CNV GWAS a également révélé 19 gènes délétés (dont 13 nouveau candidats non répertorié dans

SFARI (12), Satterstrom et al. 2020 (68) ou dans un CNV récurrent associé à la psychiatrie (76)) et 72 gènes dupliqués (dont 61 nouveaux candidats) associés significativement avec le risque de TSA qui n'avaient pas été repéré par la CNV GWAS sur le risque de TSA précédente.

D'autres gènes ont quant à eux un effet à la fois sur les deux phénotypes, c'est-à-dire que l'effet de ces gènes est pléiotropique. Ces gènes pléiotropiques ont toujours un effet qui augmente le risque de TSA et diminue l'intelligence générale. Il est à noter que les gènes protecteurs pour le risque de TSA n'affectent jamais l'intelligence générale.

L'étude de la pléiotropie qu'il existe entre les gènes affectant le risque de TSA et l'intelligence générale représente un défi de taille. Parmi toutes ces associations, il n'est pas trivial de déterminer quel phénotype en est responsable. En effet, près du tiers des gènes (Tableau 20) sont associés aux deux phénotypes, sans qu'il soit possible de faire la différenciation des effets spécifiques sur l'un ou l'autre. Le risque de TSA et l'intelligence générale sont étroitement liés, et la catégorisation clinique définie par les TSA et la déficience intellectuelle ne reflète pas adéquatement la réalité clinique bien plus complexe qui relie ces deux pathologies.

Il est également important de mentionner que les études covariant le risque de TSA avec une mesure continue sur l'intelligence générale sont très rares. La plupart des études, comme Satterstrom et al. (2020) (68), utilisent le diagnostic de déficience intellectuelle, qui est une mesure qualitative binaire. Or, cette mesure n'inclut que les individus ayant un QI inférieur à 70, ce qui correspond à une déficience intellectuelle modérée (Tableau 1). Les effets des pertes de QI moins sévères sont dès lors négligés, puisque les individus ciblés sont inclus dans la population générale. Le fait d'utiliser une mesure continue sur l'intelligence générale permet d'avoir une vision plus résolutive de l'impact de cette variable sur le risque de TSA.

Une limite est que cette CNV GWAS n'inclut que les individus qui ont une mesure cognitive disponible. Cette stratification peut biaiser les résultats obtenus, puisque par souci d'avoir le plus grand nombre de participants possible dans nos analyses, nous avons inclus un plus grand nombre

d'individus dans l'analyse sans ajustement pour l'intelligence générale, et donc nous n'avons pas la même puissance de détection de signal. En perspective, il faudrait réaliser une CNV GWAS sur le risque de TSA incluant les mêmes individus que celle ajustée pour l'intelligence générale.

6.4 Perspectives générales

La CNV GWAS développée dans ce projet a permis d'avoir une vision plus résolutive de l'impact du dosage génique sur l'intelligence générale et le risque de TSA. Cette méthodologie peut également être appliquée à d'autres phénotypes.

Afin d'améliorer le modèle, une prochaine étape de ce projet serait d'y ajouter d'autres variants génétiques, comme les SNV actuellement étudiés par le laboratoire. L'inclusion d'autres mesures des fonctions biologiques des gènes (*e.g.*, l'expression des gènes dans le cerveau) pourraient également ajouter de l'informativité au modèle.

Une autre perspective serait également d'utiliser notre CNV GWAS afin de valider les résultats précédemment obtenus, dans l'éventualité où le LOEUF de certains gènes sur ou sous-estimerait l'impact sur l'intelligence ou le risque de TSA ou que celui-ci ne soit pas encore défini.

Finalement, l'investigation de toutes les régions associées à l'intelligence générale ou le risque de TSA qui ont été identifiées par ce projet seraient à investiguer davantage afin de mieux comprendre leur impact sur ces phénotypes et les voies biologiques qui y sont impliquées.

Chapitre 7 – Conclusion

Cette étude a permis l'identification de nouveaux gènes candidats associés à l'intelligence générale, au risque de TSA, ou simultanément à ces deux phénotypes. L'approche en CNV GWAS a permis le développement d'une vision plus résolutive de l'impact des CNV, permettant notamment de mieux comprendre la dynamique pléiotropique des gènes sur ces phénotypes. De plus, pour la première fois, nous mettons en évidence des effets positifs des gènes tolérants sur l'intelligence générale. Les différentes analyses de sensibilité sur l'âge ont révélé une stratification des effets des CNV par rapport à leur nature mosaïque et leur dangerosité. Finalement, la mesure de l'impact des CNV du chromosome X sur l'intelligence n'a pas été fructueuse, mais a révélé un potentiel problème de mosaïcisme dans la cohorte UKBB, qui serait à investiguer dans de futures recherches.

Cette étude propose une nouvelle méthodologie permettant de quantifier l'effet des CNV sur la cognition et sur la psychiatrie, tout en étant résolutive en capturant le dynamisme des effets des gènes sur les phénotypes. Cette approche peut contribuer à l'interprétation quantitative de l'impact des CNV non-récurrents en clinique et être appliquée à d'autres traits cliniques.

Les estimations de chaque modèle seront intégrées à l'outil en ligne « Measuring the Impact on NeuroDevelopment of CNV » (<https://cnvprediction.urca.ca/>) développé par le laboratoire, qui permet déjà aux généticiens et cliniciens d'interpréter les impacts des CNV autosomiques vu en clinique sur le risque de TSA et les capacités intellectuelles. Cela permettra aux cliniciens d'avoir une vision plus résolutive des effets des gènes inclus dans les CNV de leurs patients et d'avoir plus de recul sur les effets pléiotropes potentiels des gènes entre les phénotypes.

Références bibliographiques

1. Westerinen H, Kaski M, Virta LJ, Kautiainen H, Pitkälä KH, Iivanainen M. The nationwide register-based prevalence of intellectual disability during childhood and adolescence. *J Intellect Disabil Res JIDR*. 2017 Aug;61(8):802–9.
2. Plomin R, von Stumm S. The new genetics of intelligence. *Nat Rev Genet*. 2018 Mar;19(3):148–59.
3. McGuire DO, Tian LH, Yeargin-Allsopp M, Dowling NF, Christensen DL. Prevalence of cerebral palsy, intellectual disability, hearing loss, and blindness, National Health Interview Survey, 2009–2016. *Disabil Health J*. 2019 Jul;12(3):443–51.
4. Christensen DL, Braun KVN, Baio J, Bilder D, Charles J, Constantino JN, et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR Surveill Summ*. 2018 Nov 16;65(13):1–23.
5. Bai D, Yip BHK, Windham GC, Sourander A, Francis R, Yoffe R, et al. Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort. *JAMA Psychiatry*. 2019 Oct 1;76(10):1035–43.
6. Taylor MJ, Rosenqvist MA, Larsson H, Gillberg C, D’Onofrio BM, Lichtenstein P, et al. Etiology of Autism Spectrum Disorders and Autistic Traits Over Time. *JAMA Psychiatry*. 2020 Sep;77(9):1–8.
7. Lejeune J, Turpin R, Gautier M. [Mongolism; a chromosomal disease (trisomy)]. *Bull Acad Natl Med*. 1959 Apr 7;143(11–12):256–65.
8. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. *Am J Hum Genet*. 2010 mai;86(5):749–64.
9. Sanders SJ, Sahin M, Hostyk J, Thurm A, Jacquemont S, Avillach P, et al. A framework for the investigation of rare genetic disorders in neuropsychiatry. *Nat Med*. 2019 Oct;25(10):1477–

87.

10. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2021 Dec 1;50(D1):D20–6.
11. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet.* 2009 Apr 10;84(4):524–33.
12. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism.* 2013 Oct 3;4:36.
13. Hansen BH, Oerbeck B, Skirbekk B, Petrovski BÉ, Kristensen H. Neurodevelopmental disorders: prevalence and comorbidity in children referred to mental health services. *Nord J Psychiatry.* 2018 May;72(4):285–91.
14. Boyle CA, Boulet S, Schieve LA, Cohen RA, Blumberg SJ, Yeargin-Allsopp M, et al. Trends in the Prevalence of Developmental Disabilities in US Children, 1997–2008. *Pediatrics.* 2011 juin;127(6):1034–42.
15. Diagnostic and statistical manual of mental disorders : DSM-5. Arlington, VA: American Psychiatric Association; 2013.
16. Homberg JR, Kyzar EJ, Scattoni ML, Norton WH, Pittman J, GaiKWad S, et al. Genetic and environmental modulation of neurodevelopmental disorders: Translational insights from labs to beds. *Brain Res Bull.* 2016 Jul;125:79–91.
17. Buescher AVS, Cidav Z, Knapp M, Mandell DS. Costs of Autism Spectrum Disorders in the United Kingdom and the United States. *JAMA Pediatr.* 2014 août;168(8):721–8.
18. Committee to Evaluate the Supplemental Security Income Disability Program for Children with Mental Disorders, Board on the Health of Select Populations, Board on Children, Youth, and Families, Institute of Medicine, Division of Behavioral and Social Sciences and Education, The National Academies of Sciences, Engineering, and Medicine. *Mental Disorders and Disabilities Among Low-Income Children* [Internet]. Boat TF, Wu JT, editors. Washington (DC): National Academies Press (US); 2015 [cited 2022 Mar 1]. Available from:

<http://www.ncbi.nlm.nih.gov/books/NBK332882/>

19. World Health Organization. The World health report : 2001 : Mental health : new understanding, new hope. *Rapp Sur Santé Dans Monde 2001 Santé Ment Nouv Concept Nouv Espoirs* [Internet]. 2001; Available from: <https://apps.who.int/iris/handle/10665/42390>
20. Croen LA, Zerbo O, Qian Y, Massolo ML, Rich S, Sidney S, et al. The health status of adults on the autism spectrum. *Autism*. 2015 Oct 1;19(7):814–23.
21. Postorino V, Fatta LM, Sanges V, Giovagnoli G, De Peppo L, Vicari S, et al. Intellectual disability in Autism Spectrum Disorder: Investigation of prevalence in an Italian sample of children and adolescents. *Res Dev Disabil*. 2016 Jan;48:193–201.
22. Rydzewska E, Hughes-McCormack LA, Gillberg C, Henderson A, MacIntyre C, Rintoul J, et al. Prevalence of long-term health conditions in adults with autism: observational study of a whole country population. *BMJ Open*. 2018 Sep 1;8(8):e023945.
23. Akingbuwa WA, Hammerschlag AR, Bartels M, Middeldorp CM. Systematic Review: Molecular Studies of Common Genetic Variation in Child and Adolescent Psychiatric Disorders. *J Am Acad Child Adolesc Psychiatry*. 2022 Feb;61(2):227–42.
24. Deary IJ, Yang J, Davies G, Harris SE, Tenesa A, Liewald D, et al. Genetic contributions to stability and change in intelligence from childhood to old age. *Nature*. 2012 Feb;482(7384):212–5.
25. Huguet G, Schramm C, Douard E, Jiang L, Labbe A, Tihy F, et al. Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. *JAMA Psychiatry*. 2018 May 1;75(5):447–57.
26. Posthuma D, de Geus EJC, Boomsma DI. Perceptual Speed and IQ Are Associated Through Common Genetic Factors. *Behav Genet*. 2001 Nov 1;31(6):593–602.
27. Voet D, Voet JG. *Biochimie (3e édition)*. 3rd ed. De Boeck;
28. Colvert E, Tick B, McEwen F, Stewart C, Curran SR, Woodhouse E, et al. Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry*. 2015 May;72(5):415–23.
29. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The

- Heritability of Autism Spectrum Disorder. *JAMA*. 2017 Sep 26;318(12):1182–4.
30. Hilker R, Helenius D, Fagerlund B, Skytthe A, Christensen K, Werge TM, et al. Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biol Psychiatry*. 2018 Mar 15;83(6):492–8.
 31. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med*. 2011 Jul 1;13(7):680–5.
 32. Butcher NJ, Chow EWC, Costain G, Karas D, Ho A, Bassett AS. Functional outcomes of adults with 22q11.2 deletion syndrome. *Genet Med Off J Am Coll Med Genet*. 2012 Oct;14(10):836–43.
 33. D’Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L, et al. Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. *JAMA Psychiatry*. 2016 Jan;73(1):20–30.
 34. Pâques F, Haber JE. Multiple Pathways of Recombination Induced by Double-Strand Breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev*. 1999 Jun;63(2):349–404.
 35. Chen L, Zhou W, Zhang L, Zhang F. Genome Architecture and Its Roles in Human Copy Number Variation. *Genomics Inform*. 2014 Dec 1;12:136.
 36. Pös O, Radvanszky J, Buglyó G, Pös Z, Rusnakova D, Nagy B, et al. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomed J*. 2021 Oct 1;44(5):548–59.
 37. Lee PH, Feng YCA, Smoller JW. Pleiotropy and Cross-Disorder Genetics Among Psychiatric Disorders. *Biol Psychiatry*. 2021 Jan 1;89(1):20–31.
 38. Smith, JC. Das Ursachenverhältnis Des Schwachsinn's Beleuchtet Durch Untersuchungen von Zweilligen. 1930;125:678–92.
 39. Kallmann FJ, Roth B. Genetic aspects of preadolescent schizophrenia. *Am J Psychiatry*. 1956 Feb;112(8):599–606.
 40. Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry*. 1977 Sep;18(4):297–321.
 41. Huguet G, Benabou M, Bourgeron T. The Genetics of Autism Spectrum Disorders. In:

- Sassone-Corsi P, Christen Y, editors. A Time for Metabolism and Hormones [Internet]. Cham (CH): Springer; 2016 [cited 2022 Mar 3]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK453174/>
42. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet.* 2015 Jul;47(7):702–9.
 43. Zwicker A, Denovan-Wright EM, Uher R. Gene-environment interplay in the etiology of psychosis. *Psychol Med.* 2018 Sep;48(12):1925–36.
 44. An Introduction to Genetic Analysis. Seventh Edition by Griffiths, Anthony J.F.; Lewontin, Richard C.; Miller, Jeffrey H.; Suzuki, David T.; Gelbart, William M.: Very Good Hardcover (2000) | Good Reading Secondhand Books [Internet]. [cited 2022 Mar 2]. Available from: <https://www.abebooks.com/9780716735205/Introduction-Genetic-Analysis-Seventh-Edition-0716735202/plp>
 45. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature.* 1953 Apr;171(4356):737–8.
 46. Grayton HM, Fernandes C, Rujescu D, Collier DA. Copy number variations in neurodevelopmental disorders. *Prog Neurobiol.* 2012 Oct 1;99(1):81–91.
 47. Gillberg C, Wahlström J. Chromosome Abnormalities in Infantile Autism and Other Childhood Psychoses: A Population Study of 66 Cases. *Dev Med Child Neurol.* 1985;27(3):293–304.
 48. Lubs HA. A marker X chromosome. *Am J Hum Genet.* 1969 May;21(3):231–44.
 49. Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, Caskey CT, et al. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell.* 1991 Aug 23;66(4):817–22.
 50. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell.* 1991 May 31;65(5):905–14.
 51. Kaufmann WE, Kidd SA, Andrews HF, Budimirovic DB, Esler A, Haas-Givler B, et al. Autism Spectrum Disorder in Fragile X Syndrome: Cooccurring Conditions and Current Treatment. *Pediatrics.* 2017 Jun;139(Suppl 3):S194–206.

52. Muhle R, Trentacoste SV, Rapin I. The genetics of autism. *Pediatrics*. 2004 May;113(5):e472-486.
53. Schaefer GB, Mendelsohn NJ, Professional Practice and Guidelines Committee. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. *Genet Med Off J Am Coll Med Genet*. 2013 May;15(5):399–407.
54. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004 Oct;431(7011):931–45.
55. Willatt L, Cox J, Barber J, Cabanas ED, Collins A, Donnai D, et al. 3q29 Microdeletion Syndrome: Clinical and Molecular Characterization of a New Syndrome. *Am J Hum Genet*. 2005 Jul;77(1):154–60.
56. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between Microdeletion and Microduplication at 16p11.2 and Autism. *N Engl J Med*. 2008 février;358(7):667–75.
57. Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet*. 2008 Dec;40(12):1466–71.
58. Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, et al. Recurrent Rearrangements of Chromosome 1q21.1 and Variable Pediatric Phenotypes. *N Engl J Med*. 2008 Oct 16;359(16):1685–99.
59. Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet*. 2008 Mar;40(3):322–8.
60. Moreno-De-Luca D, SGENE Consortium, Mulle JG, Simons Simplex Collection Genetics Consortium, Kaminsky EB, Sanders SJ, et al. Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am J Hum Genet*. 2010 Nov 12;87(5):618–30.
61. Vorstman J a. S, Staal WG, van Daalen E, van Engeland H, Hochstenbach PFR, Franke L. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol Psychiatry*. 2006 Jan;11(1):1, 18–28.

62. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010 Jul;466(7304):368–72.
63. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet*. 2014 May 1;94(5):677–94.
64. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol*. 2011 May 1;29(6):512–20.
65. Jamain S, Quach H, Betancur C, Råstam M, Colineaux C, Gillberg IC, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet*. 2003 May;34(1):27–9.
66. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet*. 2007 Jan;39(1):25–7.
67. Autism Genome Project Consortium, Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet*. 2007 Mar;39(3):319–28.
68. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*. 2020 Feb 6;180(3):568–584.e23.
69. Sullivan PF, Geschwind DH. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell*. 2019 Mar 21;177(1):162–83.
70. Myers SM, Challman TD, Bernier R, Bourgeron T, Chung WK, Constantino JN, et al. Insufficient Evidence for “Autism-Specific” Genes. *Am J Hum Genet*. 2020 mai;106(5):587–95.
71. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019 Mar;51(3):431–44.
72. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide

association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet.* 2018 Jul;50(7):912–9.

73. Li W, Olivier M. Current analysis platforms and methods for detecting copy number variation. *Physiol Genomics.* 2013 Jan;45(1):1–16.
74. Scientific protocols. Copy Number Variation Detection Via High-Density Snp Genotyping. 2014 Dec 30 [cited 2022 Mar 23]; Available from: <https://zenodo.org/record/13529>
75. Seiser EL, Innocenti F. Hidden Markov Model-Based CNV Detection Algorithms for Illumina Genotyping Microarrays. *Cancer Inform.* 2015 Jan 27;13(Suppl 7):77–83.
76. Huguet G, Schramm C, Douard E, Tamer P, Main A, Monin P, et al. Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability. *Mol Psychiatry* [Internet]. 2021 Jan 7 [cited 2021 Sep 14]; Available from: <http://www.nature.com/articles/s41380-020-00985-z>
77. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, et al. Phenotypic Heterogeneity of Genomic Disorders and Rare Copy-Number Variants. *N Engl J Med.* 2012 Oct 4;367(14):1321–31.
78. Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and Predicting Haploinsufficiency in the Human Genome. Schierup MH, editor. *PLoS Genet.* 2010 Oct 14;6(10):e1001154.
79. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020 May;581(7809):434–43.
80. Cummings BB, Karczewski KJ, Kosmicki JA, Seaby EG, Watts NA, Singer-Berk M, et al. Transcript expression-aware annotation improves rare variant interpretation. *Nature.* 2020 May;581(7809):452–8.
81. Douard E, Zeribi A, Schramm C, Tamer P, Loum MA, Nowak S, et al. Effect Sizes of Deletions and Duplications on Autism Risk Across the Genome. *Am J Psychiatry.* 2021 Jan;178(1):87–98.
82. Han L, Zhao X, Benton ML, Perumal T, Collins RL, Hoffman GE, et al. Functional annotation of rare structural variation in the human brain. *Nat Commun.* 2020 Jun 12;11(1):2990.
83. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Büchel C, et al. The IMAGEN

- study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol Psychiatry*. 2010 Dec;15(12):1128–39.
84. Pausova Z, Paus T, Abrahamowicz M, Bernard M, Gaudet D, Leonard G, et al. Cohort Profile: The Saguenay Youth Study (SYS). *Int J Epidemiol*. 2017 Apr;46(2):e19.
85. Awadalla P, Boileau C, Payette Y, Idaghdour Y, Goulet JP, Knoppers B, et al. Cohort profile of the CARTaGENE study: Quebec’s population-based biobank for public health and personalized genomics. *Int J Epidemiol*. 2013 Oct 1;42(5):1285–99.
86. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2012 Dec;41(6):1576–84.
87. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol*. 2013 Jun 1;42(3):689–700.
88. Fischbach GD, Lord C. The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. *Neuron*. 2010 Oct 21;68(2):192–5.
89. Yuen RK, Merico D, Bookman M, Howe JL, Thiruvahindrapuram B, Patel RV, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci*. 2017 Apr;20(4):602–11.
90. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*. 2017 Jan;49(1):27–35.
91. Gendrel AV, Heard E. Fifty years of X-inactivation research. *Development*. 2011 Dec 1;138(23):5049–55.
92. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med*. 2015 Mar 31;12(3):e1001779.
93. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron*. 2018 Feb 7;97(3):488–93.
94. Bahg G, Evans DG, Galdo M, Turner BM. Gaussian process linking functions for mind, brain,

- and behavior. *Proc Natl Acad Sci*. 2020 Nov 24;117(47):29398–406.
95. Harvey, Annabelle, Dumas, Guillaume. PyNM [Internet]. GitHub Repository. 2022. Available from: <https://github.com/ppsp-team/PyNM>
96. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007 Nov;17(11):1665–74.
97. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res*. 2007 Mar;35(6):2013–25.
98. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*. 2015 Sep 23;87(6):1215–33.
99. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. *Nucleic Acids Res*. 2022 Jan 7;50(D1):D988–95.
100. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health*. 2019 Nov 1;22(4):153–60.
101. Loh PR, Genovese G, Handsaker RE, Finucane HK, Reshef YA, Palamara PF, et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature*. 2018 Jul;559(7714):350–5.
102. Irwing P, Lynn R. Sex differences in means and variability on the progressive matrices in university students: a meta-analysis. *Br J Psychol Lond Engl* 1953. 2005 Nov;96(Pt 4):505–24.
103. Zekavat SM, Lin SH, Bick AG, Liu A, Paruchuri K, Uddin MM, et al. Hematopoietic mosaic chromosomal alterations and risk for infection among 767,891 individuals without blood cancer. *medRxiv*. 2020 Nov 16;2020.11.12.20230821.
104. Loh PR, Genovese G, McCarroll SA. Monogenic and polygenic inheritance become instruments for clonal selection. *Nature*. 2020 Aug;584(7819):136–41.
105. Auwerx C, Lepamets M, Sadler MC, Patxot M, Stojanov M, Baud D, et al. The individual and global impact of copy-number variants on complex human traits. *Am J Hum Genet*. 2022 Apr 7;109(4):647–68.

106. Kendall KM, Bracher-Smith M, Fitzpatrick H, Lynham A, Rees E, Escott-Price V, et al. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *Br J Psychiatry*. 2019 May;214(5):297–304.
107. O’Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016 Jan 4;44(Database issue):D733–45.
108. Chen CP, Lin SP, Lee CL, Chern SR, Wu PS, Chen YN, et al. Recurrent 2q13 microduplication encompassing MALL, NPHP1, RGP6, and BUB1 associated with autism spectrum disorder, intellectual disability, and liver disorder. *Taiwan J Obstet Gynecol*. 2017 Feb 1;56(1):98–101.
109. Yasuda Y, Hashimoto R, Fukai R, Okamoto N, Hiraki Y, Yamamori H, et al. Duplication of the NPHP1 gene in patients with autism spectrum disorder and normal intellectual ability: a case series. *Ann Gen Psychiatry*. 2014 Aug 6;13:22.
110. Calle Sánchez X, Helenius D, Bybjerg-Grauholm J, Pedersen C, Hougaard DM, Børghlum AD, et al. Comparing Copy Number Variations in a Danish Case Cohort of Individuals With Psychiatric Disorders. *JAMA Psychiatry*. 2022 Jan 1;79(1):59–69.

Annexes

Tableaux supplémentaires

A.1 Clusters de gènes associés à une réduction de l'intelligence générale si délévés

CHR	Cluster START	Cluster STOP	Genes in cluster	Ref. gene	Carriers	Est.	SD	P	FDR
1	146373546	146467744	1	<i>NBPF12</i>	13	-1.49	0.27	3.56E-08	1.74E-06
1	146626685	147634886	9	<i>GJA8</i>	57	-0.79	0.13	1.61E-09	8.98E-08
4	186990309	187009227	1	<i>TLR3</i>	10	-1.42	0.31	5.81E-06	1.13E-04
4	187112721	187210835	3	<i>CYP4V2</i>	16	-1.18	0.25	2.20E-06	4.77E-05
5	17632197	17655958	3	<i>TAF11L13</i>	142	-0.26	0.08	1.52E-03	1.75E-02
8	356428	427967	1	<i>FBXO25</i>	15	-0.90	0.26	4.09E-04	5.50E-03
8	439792	495781	1	<i>TDRP</i>	10	-1.11	0.31	4.00E-04	5.50E-03
8	1772092	1906807	1	<i>ARHGEF10</i>	10	-1.50	0.31	1.57E-06	3.93E-05
8	1993161	2113475	1	<i>MYOM2</i>	13	-1.06	0.27	1.09E-04	1.84E-03
10	48355069	48827966	4	<i>ZNF488</i>	15	-0.95	0.26	2.02E-04	3.18E-03
10	49361131	49482739	1	<i>FRMPD2</i>	20	-0.71	0.22	1.37E-03	1.62E-02
10	49514682	50970425	14	<i>SLC18A3</i>	27	-0.71	0.19	2.04E-04	3.18E-03
10	51741950	51770259	1	<i>AGAP6</i>	15	-0.84	0.26	1.12E-03	1.36E-02
11	31391387	31531192	2	<i>IMMP1L</i>	10	-1.49	0.31	2.15E-06	4.77E-05
15	22368478	22383423	2	<i>AC135068.1</i>	90	-0.47	0.10	6.18E-06	1.15E-04
15	22736246	22746002	1	<i>GOLGA6L1</i>	160	-0.40	0.08	4.47E-07	1.24E-05
15	22833496	23100005	4	<i>NIPA2</i>	949	-0.28	0.03	1.48E-17	5.77E-15
15	29412457	30261068	3	<i>FAM189A1</i>	11	-1.43	0.30	1.61E-06	3.93E-05
15	31182762	32162992	5	<i>OTUD7A</i>	30	-1.15	0.18	1.56E-10	1.01E-08
15	32215641	32465219	1	<i>CHRNA7</i>	33	-1.13	0.17	3.30E-11	3.22E-09
15	102343712	102360400	2	<i>OR4F6</i>	19	-0.71	0.23	1.86E-03	1.97E-02
16	15131711	15149936	1	<i>NTAN1</i>	42	-0.45	0.15	2.99E-03	2.91E-02
16	15153890	15188168	1	<i>RRN3</i>	42	-0.45	0.15	2.99E-03	2.91E-02
16	15457485	15474904	1	<i>NPIPA5</i>	41	-0.55	0.15	3.14E-04	4.71E-03
16	15489611	16236931	7	<i>NDE1</i>	67	-0.43	0.12	4.04E-04	5.50E-03
16	17195626	17564817	1	<i>XYLT1</i>	11	-0.94	0.30	1.67E-03	1.86E-02
16	21608539	21668794	2	<i>METTL9</i>	378	-0.18	0.05	5.52E-04	7.16E-03

16	21964609	22448486	8	<i>UQCRC2</i>	118	-0.50	0.09	4.95E-08	2.14E-06
16	28834356	29002105	9	<i>TUFM</i>	27	-0.98	0.19	2.89E-07	8.68E-06
16	29674300	30134827	26	<i>ASPHD1</i>	72	-0.79	0.12	1.17E-11	1.53E-09
16	30194148	30205627	2	<i>CORO1A</i>	43	-0.52	0.15	5.69E-04	7.16E-03
19	55525073	55549632	1	<i>GP6</i>	11	-0.90	0.30	2.40E-03	2.46E-02
22	18893541	18924066	2	<i>PRODH</i>	57	-0.41	0.13	1.86E-03	1.97E-02
22	19023795	19166252	5	<i>DGCR2</i>	19	-1.59	0.23	4.84E-12	9.43E-10
22	19166986	20307603	23	<i>C22orf39</i>	12	-1.89	0.29	7.93E-11	6.19E-09
22	20748405	20941919	4	<i>ZNF74</i>	20	-0.92	0.22	4.07E-05	7.21E-04
22	21128401	21142008	1	<i>SERPIND1</i>	32	-0.92	0.18	1.57E-07	5.11E-06
22	21061979	21213705	1	<i>PI4KA</i>	24	-0.96	0.20	2.51E-06	5.15E-05
22	21213295	21418457	8	<i>SNAP29</i>	32	-0.92	0.18	1.57E-07	5.11E-06
22	51112843	51171726	1	<i>SHANK3</i>	12	-1.52	0.29	1.15E-07	4.49E-06

Tableau supplémentaire 1. – Délétions réduisant l'intelligence générale. Chaque ligne représente un cluster de gènes significatif (FDR<0,05). Le nom du gène affiché (Ref. gene) est le gène du cluster avec la p-value la plus significative dans le cluster. Les lignes de couleur menthe représentent les clusters contenant au moins un gène répertorié dans la base de données SFARI ou dans Satterstrom et al. (2020) ou à moins de 250 kb d'un de ces gènes. Les lignes de couleur turquoise représentent les clusters contenant au moins un gène incluse dans un CNV récurrent associé aux TND ou à moins de 250 kb d'un de ces CNV. Les lignes surlignées en jaune sont des gènes qui ne sont pas répertoriés dans la littérature. CHR : chromosome; Ref. gene : gène de référence dans le cluster; Est. : Estimé; SD : Écart-type; P : p-value.

A.2 Clusters de gènes associés à une réduction de l'intelligence générale si dupliqués

CHR	Cluster START	Cluster STOP	Genes in cluster	Ref. gene	Carriers	Est.	SD	P	FDR
1	13801441	13840498	1	<i>LRRC38</i>	12	-0.93	0.29	1.43E-03	1.63E-02
1	65886270	65901690	1	<i>LEPROT</i>	15	-0.77	0.24	1.69E-03	1.91E-02
1	145398250	145417545	1	<i>HJV</i>	150	-0.42	0.08	1.86E-07	4.04E-06
1	145438490	145715614	13	<i>RBM8A</i>	248	-0.40	0.06	3.11E-10	8.97E-09
1	145726918	145827075	2	<i>GPR89A</i>	136	-0.41	0.08	1.10E-06	2.07E-05
1	146626685	147634886	9	<i>GJA8</i>	122	-0.36	0.09	5.73E-05	7.64E-04
2	241615835	241622323	1	<i>AQP12B</i>	25	-0.65	0.20	1.01E-03	1.18E-02
6	28471073	28502732	2	<i>GPX6</i>	10	-0.91	0.31	3.88E-03	4.04E-02
7	72742164	72772646	1	<i>FKBP6</i>	10	-1.14	0.31	3.01E-04	3.75E-03
7	72848108	74016930	21	<i>FZD9</i>	13	-1.25	0.27	5.26E-06	8.76E-05
8	53446597	53477990	1	<i>ALKAL1</i>	45	-0.48	0.15	1.21E-03	1.39E-02
8	53535018	53658403	1	<i>RB1CC1</i>	44	-0.50	0.15	7.84E-04	9.16E-03
8	145051321	145754516	36	<i>MFSD3</i>	24	-0.61	0.20	3.04E-03	3.23E-02
10	81697496	81742370	1	<i>SFTPD</i>	16	-0.75	0.25	2.68E-03	2.94E-02
10	81838402	81965328	3	<i>ANXA11</i>	18	-0.85	0.24	3.90E-04	4.78E-03
10	135367405	135383339	1	<i>SYCE1</i>	15	-0.77	0.26	2.61E-03	2.88E-02
12	8071826	8088783	1	<i>SLC2A3</i>	1967	-0.09	0.02	7.34E-05	9.66E-04
12	27396901	27655118	3	<i>STK38L</i>	56	-0.38	0.13	4.35E-03	4.50E-02
12	112204691	112255336	1	<i>ALDH2</i>	25	-0.80	0.19	3.54E-05	4.75E-04
13	113862507	113977987	2	<i>CUL4A</i>	11	-0.85	0.30	4.66E-03	4.78E-02
14	21616529	21737653	2	<i>OR5AU1</i>	13	-0.83	0.28	3.07E-03	3.25E-02
15	23684185	23692390	1	<i>GOLGA6L2</i>	24	-1.53	0.20	1.81E-14	9.67E-13
15	23810821	23932452	3	<i>MKRN3</i>	23	-1.74	0.21	3.86E-17	2.36E-15
15	24920922	27787130	8	<i>GABRA5</i>	20	-2.18	0.22	7.16E-23	7.83E-21
15	28000021	28344461	1	<i>OCA2</i>	22	-1.83	0.21	4.46E-18	2.81E-16
15	31182762	32162992	5	<i>OTUD7A</i>	153	-0.41	0.08	2.28E-07	4.88E-06
16	15016659	15045917	1	<i>NPIPA1</i>	118	-0.41	0.09	5.71E-06	9.43E-05
16	15131711	15149936	1	<i>NTAN1</i>	269	-0.31	0.06	3.96E-07	7.99E-06
16	15068448	15233196	1	<i>PDXDC1</i>	118	-0.42	0.09	3.53E-06	6.34E-05
16	15153890	15474904	2	<i>RRN3</i>	275	-0.28	0.06	2.04E-06	3.72E-05
16	15489611	16236931	7	<i>ABCC1</i>	471	-0.23	0.05	7.71E-07	1.51E-05

16	29674300	30134827	26	<i>ALDOA</i>	88	-1.06	0.11	1.12E-23	3.08E-21
16	30194148	30205627	2	<i>CORO1A</i>	60	-0.86	0.13	1.75E-11	5.70E-10
16	30205196	30265831	3	<i>SLX1A</i>	15	-0.94	0.25	2.07E-04	2.59E-03
17	4046462	4167290	2	<i>ANKFY1</i>	19	-0.70	0.23	2.60E-03	2.88E-02
17	34745936	34808104	1	<i>TBC1D3G</i>	28	-0.57	0.19	2.46E-03	2.74E-02
17	34842473	36105096	15	<i>LHX1</i>	49	-0.60	0.14	3.04E-05	4.21E-04
19	56598732	56720821	5	<i>ZNF787</i>	11	-0.97	0.30	1.24E-03	1.42E-02
22	18893541	18924066	2	<i>PRODH</i>	299	-0.18	0.06	1.30E-03	1.48E-02
22	19023795	20716938	32	<i>C22orf39</i>	156	-0.58	0.08	3.85E-13	1.91E-11
22	20748405	20941919	4	<i>ZNF74</i>	197	-0.38	0.07	8.21E-08	1.86E-06
22	21061979	21418457	10	<i>P2RX6</i>	272	-0.33	0.06	2.96E-08	7.07E-07
22	22986946	22990373	1	<i>GGTLC2</i>	27	-0.58	0.19	2.40E-03	2.69E-02

Tableau supplémentaire 2. – Duplications réduisant l'intelligence générale. Chaque ligne représente un cluster de gènes significatif (FDR<0,05). Le nom du gène affiché (Ref. gene) est le gène du cluster avec la p-value la plus significative dans le cluster. Les lignes de couleur menthe représentent les clusters contenant au moins un gène répertorié dans la base de données SFARI ou dans Satterstrom et al. 2020 ou à moins de 250 kb d'un de ces gènes. Les lignes de couleur turquoise représentent les clusters contenant au moins un gène incluse dans un CNV récurrent associé aux TND ou à moins de 250 kb d'un de ces CNV. Les lignes surlignées en jaune sont des gènes qui ne sont pas répertoriés dans la littérature. CHR : chromosome; Ref. gene : gène de référence dans le cluster; Est. : Estimé; SD : Écart-type; P : p-value.

A.3 Gènes inclus dans les CNV dont la présence augmente avec l'âge

CNV	Gene	Description
chr10:47,177,22-47,181,688	<i>ANXA8L1</i>	This gene encodes a member of the annexin family of evolutionarily conserved Ca ²⁺ and phospholipid binding proteins. The encoded protein may function as an anticoagulant that indirectly inhibits the thromboplastin-specific complex. Overexpression of this gene has been associated with acute myelocytic leukemia. A highly similar duplicated copy of this gene is found in close proximity on the long arm of chromosome 10. [provided by RefSeq, Apr 2014].
	<i>FAM21B</i>	NA
	<i>ANXA8L2</i>	This gene encodes a member of the annexin family of evolutionarily conserved Ca ²⁺ and phospholipid binding proteins. The encoded protein may function as an anticoagulant that indirectly inhibits the thromboplastin-specific complex. Overexpression of this gene has been associated with acute myelocytic leukemia. A highly similar duplicated copy of this gene is found in close proximity on the long arm of chromosome 10. [provided by RefSeq, Apr 2014].
	<i>FAM25B</i>	NA
	<i>AGAP10</i>	Homo sapiens ArfGAP with GTPase domain, ankyrin repeat and PH domain 9 (AGAP10), mRNA.
	<i>NPY4R</i>	Receptor for neuropeptide Y and peptide YY. The rank order of affinity of this receptor for pancreatic polypeptides is PP, PP (2-36) and [Ile-31, Gln-34] PP > [Pro-34] PYY > PYY and [Leu-31, Pro-34] NPY > NPY > PYY (3-36) and NPY (2-36) > PP (13- 36) > PP (31-36) > NPY free acid.
	<i>GPRIN2</i>	May be involved in neurite outgrowth.
	<i>SYT15</i>	This gene encodes a member of the Synaptotagmin (Syt) family of membrane trafficking proteins. Members of this family contain a transmembrane region and a C-terminal-type tandem C2 domain. Unlike related family members, the encoded protein may be involved in membrane trafficking in non-neuronal tissues. Two transcript variants encoding distinct isoforms have been identified for this gene. [provided by RefSeq, Jul 2008].
	<i>NPY4R2</i>	Receptor for neuropeptide Y and peptide YY. The rank order of affinity of this receptor for pancreatic polypeptides is PP, PP (2-36) and [Ile-31, Gln-34] PP > [Pro-34] PYY > PYY and [Leu-31, Pro-34] NPY > NPY > PYY (3-36) and NPY (2-36) > PP (13- 36) > PP (31-36) > NPY free acid.
chr14:19,264,87-20,444,607	<i>OR4K1</i>	Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms. [provided by RefSeq, Jul 2008]. Sequence Note: The RefSeq transcript and protein were derived from genomic sequence to make the sequence consistent with the reference genome assembly. The genomic coordinates used for the transcript record were based on transcript alignments. RefSeq Attributes: MANE Ensembl match: ENST00000641172.1/ ENSP00000493193.1; RefSeq Select criteria: based on conservation, longest protein.
	<i>OR4K2</i>	Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms. [provided by RefSeq, Jul 2008]. RefSeq Attributes: MANE Ensembl match: ENST00000641885.1/ ENSP00000493007.1; RefSeq Select criteria: based on single protein-coding transcript.
	<i>OR4K15</i>	Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G

	<p>protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms. [provided by RefSeq, Jul 2008]. Sequence Note: The genomic coordinates were based on alignments. Evidence Data: Transcript is intronless: BC140726.1 [ECO:0000345]. RefSeq Attributes: MANE Ensembl match: ENST00000305051.6/ ENSP00000304077.5; RefSeq Select criteria: based on single protein-coding transcript.</p>
<i>OR4N2</i>	<p>Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a seven-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms. [provided by RefSeq, Mar 2014]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Evidence Data: Transcript exon combination: BM696516.1 [ECO:0000332]. RefSeq Attributes: MANE Ensembl match: ENST00000557677.3/ ENSP00000452022.2; RefSeq Select criteria: based on single protein-coding transcript.</p>
<i>OR4M1</i>	<p>Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms. [provided by RefSeq, Jul 2008]. RefSeq Attributes: MANE Ensembl match: ENST00000641200.1/ ENSP00000492985.1; RefSeq Select criteria: based on conservation, longest protein.</p>
<i>POTEG</i>	<p>Homo sapiens POTE ankyrin domain family, member M (POTEG), mRNA.</p>
<i>POTEM</i>	<p>Homo sapiens POTE ankyrin domain family, member G (POTEM), transcript variant 1, mRNA.</p>
<i>OR11H12</i>	<p>Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms. [provided by RefSeq, Jul 2008]. RefSeq Attributes: MANE Ensembl match: ENST00000550708.2/ ENSP00000449002.1; RefSeq Select criteria: based on single protein-coding transcript.</p>
<i>OR11H2</i>	<p>Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms. [provided by RefSeq, Jul 2008]. CCDS Note: This CCDS representation lacks full-length human transcript support. Its representation is therefore inferred, but it is supported by paralogous transcripts. RefSeq Attributes: MANE Ensembl match: ENST00000556246.3/ ENSP00000485150.2; RefSeq Select criteria: based on single protein-coding transcript.</p>

Tableau supplémentaire 3. – Gènes inclus dans les CNV qui sont plus fréquents chez les individus âgés. Tableau généré par UCSC.

A.4 Clusters de gènes associés à une augmentation du risque de TSA si délétés

Modèle 1 pour les délétions : $TSA \sim \beta_{0;DEL} + \beta_{1;DEL} \times \text{Gène} + \beta_{2;DEL} \times \text{sexe}$

CHR	Cluster START	Cluster STOP	Genes in cluster	Ref. gene	n (Total)	n (nonASD)	n (ASD)	OR	CI95%		P	FDR
1	934342	1051741	5	AGRN	11	8	3	20.62	4.34	76.81	1.88E-05	2.44E-04
1	1109260	1133316	1	TTLL10	12	9	3	17.38	3.73	62.03	3.65E-05	4.46E-04
1	1138888	1149533	2	TNFRSF18	14	11	3	13.17	2.91	44.14	1.16E-04	1.33E-03
1	1152288	1170421	2	SDF4	11	8	3	18.14	3.86	66.40	3.40E-05	4.24E-04
1	146373546	146467744	1	NBPF12	26	23	3	7.68	1.79	22.97	1.18E-03	1.13E-02
2	96778624	97760619	21	ASTL	36	32	4	6.90	2.02	17.88	3.62E-04	3.92E-03
3	195754054	195808967	1	TFRC	17	11	6	30.68	10.11	86.00	1.69E-10	5.08E-09
3	195924323	196695931	18	SLC51A	14	8	6	39.36	12.37	120.88	1.17E-10	3.68E-09
3	196715492	196756688	1	MELTF	20	15	5	20.67	6.46	56.78	2.42E-08	4.99E-07
3	196769431	197026201	1	DLG1	16	11	5	23.61	7.18	68.66	1.97E-08	4.20E-07
3	197236654	197300194	1	BDH1	21	16	5	16.22	5.17	43.29	1.50E-07	2.69E-06
4	926175	998352	4	TMEM175	11	3	8	267.81	70.80	1289.00	5.46E-15	3.28E-13
4	71263648	71275912	1	OPRPN	12	10	2	10.10	1.52	40.39	3.56E-03	2.87E-02
4	135117488	135122944	1	PABPC4L	751	717	34	3.03	2.10	4.22	4.78E-10	1.37E-08
4	178231028	178284090	1	NEIL3	12	10	2	11.30	1.69	45.82	2.40E-03	1.98E-02
4	178351928	178363591	1	AGA	26	23	3	7.67	1.78	22.93	1.19E-03	1.13E-02
7	75471920	75518244	1	RHBDD2	10	8	2	10.83	1.61	44.88	3.03E-03	2.47E-02
7	75528518	75677322	3	STYXL1	13	11	2	8.86	1.35	34.47	5.42E-03	4.22E-02
8	182137	214703	1	ZNF596	70	62	8	8.20	3.57	16.53	5.03E-08	9.77E-07
8	356428	427967	1	FBXO25	21	15	6	19.80	6.88	50.80	2.47E-09	6.28E-08
8	439792	495781	1	TDRP	21	16	5	14.31	4.60	37.70	3.98E-07	6.58E-06
8	564746	1656642	2	ERICH1	11	6	5	50.43	13.57	183.04	1.52E-09	4.18E-08
8	1703944	1749877	1	CLN8	16	11	5	28.63	8.58	84.79	4.63E-09	1.13E-07
8	1772092	1906807	1	ARHGEF10	20	15	5	22.41	6.97	61.88	1.21E-08	2.75E-07
8	1921843	1955102	1	KBTBD11	29	24	5	14.54	4.74	36.98	1.84E-07	3.13E-06
8	1993161	2113475	1	MYOM2	25	20	5	18.02	5.77	47.25	3.81E-08	7.62E-07
8	6264117	6506029	1	MCPH1	16	12	4	16.88	4.59	50.96	2.23E-06	3.35E-05
8	6357153	6420930	1	ANGPT2	32	27	5	10.73	3.56	26.57	2.28E-06	3.35E-05
8	6666038	6693046	1	XKR5	20	16	4	14.02	3.90	40.20	5.17E-06	7.27E-05

8	6728114	6735458	1	DEFB1	77	72	5	4.59	1.59	10.48	1.20E-03	1.13E-02
8	6782219	6795828	2	DEFA6	35	31	4	8.20	2.39	21.49	1.16E-04	1.33E-03
10	3179920	3215033	1	PITRM1	11	9	2	15.44	2.25	65.62	8.16E-04	8.17E-03
10	96443486	96495947	1	CYP2C18	22	19	3	12.46	2.83	38.94	9.57E-05	1.13E-03
10	128933690	128994855	1	INSYN2A	13	10	3	33.04	6.97	116.82	5.35E-07	8.62E-06
10	134884433	134945179	1	ADGRA1	10	6	4	35.29	8.58	133.71	1.63E-07	2.84E-06
10	134973723	135039916	1	KNDC1	12	8	4	35.47	8.86	123.65	5.40E-08	1.02E-06
10	135043721	135090372	3	ADAM8	15	11	4	29.47	7.69	93.88	5.76E-08	1.06E-06
10	135092134	135236024	8	ECHS1	11	7	4	44.90	10.87	164.11	1.85E-08	4.07E-07
10	135234170	135238076	1	SPRN	18	14	4	18.89	5.13	56.39	8.15E-07	1.28E-05
10	135367405	135383339	1	SYCE1	52	45	7	12.67	5.08	27.36	1.97E-09	5.20E-08
11	22359927	22647369	2	SLC17A6	10	8	2	16.31	2.35	71.39	7.36E-04	7.49E-03
11	59953175	60085417	1	MS4A4A	12	10	2	17.08	2.50	71.08	4.86E-04	5.18E-03
11	60102304	60164069	1	MS4A6E	13	11	2	16.16	2.39	65.73	5.52E-04	5.79E-03
12	18233803	18473041	1	RERGL	20	16	4	10.83	3.06	30.26	2.76E-05	3.51E-04
15	22833496	23100005	4	NIP2	1817	1768	49	1.65	1.22	2.17	6.34E-04	6.55E-03
15	23255242	23268823	1	GOLGA8I	65	60	5	6.10	2.10	14.12	1.47E-04	1.64E-03
15	31182762	32162992	5	OTUD7A	61	50	11	12.70	6.16	24.10	1.70E-13	6.60E-12
15	32215641	32465219	1	CHRNA7	69	56	13	13.35	6.86	24.22	4.61E-16	4.35E-14
15	85144217	85682376	8	SLC28A1	12	10	2	9.12	1.38	35.95	5.04E-03	4.01E-02
15	99191768	99507759	1	IGF1R	10	8	2	12.24	1.80	51.70	2.02E-03	1.76E-02
15	99511458	99551021	1	PGPEP1L	10	8	2	12.24	1.80	51.70	2.02E-03	1.76E-02
15	99638420	99675798	1	SYNM	10	8	2	12.24	1.80	51.70	2.02E-03	1.76E-02
16	810734	850737	4	MSLN	12	9	3	22.41	4.72	82.08	1.01E-05	1.34E-04
16	1128781	1146244	2	C1QTNF8	16	14	2	10.96	1.66	42.21	2.26E-03	1.91E-02
16	1271651	1292555	3	TPSAB1	12	9	3	25.89	5.41	95.43	4.47E-06	6.43E-05
16	1306060	1309009	1	TPSD1	10	7	3	35.57	7.11	141.96	1.41E-06	2.17E-05
16	15489611	16236931	7	MPV17L	148	140	8	4.03	1.80	7.79	1.62E-04	1.79E-03
16	21964609	22448486	8	UQCRC2	271	260	11	2.62	1.34	4.60	1.93E-03	1.74E-02
16	28834356	29002105	9	TUFM	69	60	9	8.23	3.76	16.06	8.48E-09	2.00E-07
16	29674300	30134827	27	C16orf92	143	122	21	10.03	6.06	15.82	2.87E-21	1.77E-18
16	81087102	81129901	2	AC092718.3	11	8	3	23.72	4.93	89.70	9.48E-06	1.28E-04
16	81115550	81130000	1	GCSH	14	11	3	16.05	3.50	55.05	4.30E-05	5.17E-04
16	89262169	89268059	1	SLC22A31	16	14	2	10.96	1.66	42.21	2.26E-03	1.91E-02
17	34842473	36105096	15	PIGW	17	10	7	49.74	16.91	140.18	2.06E-13	7.58E-12

19	20188803	20311297	2	ZNF486	14	12	2	13.38	2.00	53.28	1.13E-03	1.11E-02
19	55464498	55512510	1	NLRP2	28	25	3	7.06	1.65	20.87	1.77E-03	1.65E-02
19	55525073	55549632	1	GP6	17	15	2	8.69	1.33	32.70	5.27E-03	4.15E-02
22	18893541	18935245	3	DGCR6	84	72	12	9.65	4.91	17.46	1.61E-12	5.32E-11
22	19023795	19137805	4	GSC2	38	26	12	28.41	13.35	57.66	1.49E-19	2.46E-17
22	19163094	19166252	1	SLC25A1	32	20	12	37.31	16.98	79.33	1.64E-20	3.61E-18
22	19166986	20307603	23	TBX1	22	10	12	77.06	31.27	194.93	5.34E-21	1.77E-18
22	20366211	20716938	5	FAM230A	14	6	8	98.16	31.39	323.05	4.68E-15	3.09E-13
22	20748405	20941919	4	SCARF2	31	20	11	34.53	15.32	74.66	8.87E-19	1.17E-16
22	21128401	21142008	1	SERPIND1	53	42	11	15.10	7.23	29.15	1.26E-14	6.43E-13
22	21061979	21213705	1	PI4KA	38	28	10	21.35	9.57	44.33	2.62E-15	1.92E-13
22	21213295	21418457	8	CRKL	51	40	11	15.84	7.56	30.74	6.41E-15	3.53E-13
22	45064427	45133561	1	PRRS	11	8	3	23.72	4.93	89.69	9.48E-06	1.28E-04
22	50432942	50451088	1	IL17REL	10	2	8	389.55	88.95	2701.56	5.55E-13	1.93E-11
22	50497820	50600119	2	MLC1	11	3	8	229.59	60.17	1116.62	4.85E-14	2.00E-12
22	50609160	50656045	3	SELENOO	15	7	8	87.76	29.04	271.03	1.62E-15	1.34E-13
22	50656118	51066601	21	ARSA	14	6	8	143.72	46.24	462.94	7.17E-18	7.90E-16
22	51112843	51171726	1	SHANK3	22	14	8	38.23	14.49	95.49	1.71E-14	8.06E-13
22	51176624	51183767	1	ACR	10	3	7	251.71	64.02	1230.19	3.05E-14	1.34E-12

Tableau supplémentaire 4. – Délétions augmentant le risque de TSA. Chaque ligne représente un cluster de gènes significatif (FDR<0,05). Le nom du gène affiché (Ref. gene) est le gène du cluster avec la p-value la plus significative dans le cluster. Les lignes de couleur menthe représentent les clusters contenant au moins un gène répertorié dans la base de données SFARI ou dans Satterstrom et al. (2020) ou à moins de 250 kb d'un de ces gènes. Les lignes de couleur turquoise représentent les clusters contenant au moins un gène incluse dans un CNV récurrent associé aux TND ou à moins de 250 kb d'un de ces CNV. Les lignes surlignées en jaune sont des gènes qui ne sont pas répertoriés dans la littérature. CHR : chromosome; Ref. gene : gène de référence dans le cluster; n : porteurs du gène; ASD : trouble du spectre autistique; Est. : Estimé; SE : Écart-type; P : p-value.

Modèle 2 pour les délétions : $TSA \sim \beta_{0; DEL} + \beta_{1; DEL} \times \text{Gène}$
 $+ \beta_{2; DEL} \times \text{sexe} + \beta_{3; DEL} \times \text{Intell. aj.}$

CHR	Cluster START	Cluster STOP	Genes in cluster	Ref. gene	n (Total)	n (nonASD)	n (ASD)	OR	CI95%			P	FDR
2	96778624	97760619	21	ASTL	36	32	4	10.45	2.59	35.21	3.17E-04	3.06E-03	
2	97778890	97930258	1	ANKRD36	25	23	2	11.77	1.64	55.28	3.90E-03	2.11E-02	
3	195754054	195808967	1	TFRC	17	11	6	14.57	2.63	89.65	2.13E-03	1.42E-02	
3	195924323	196695931	18	CEP19	15	10	5	16.55	2.81	130.25	2.55E-03	1.65E-02	
3	196715492	196756688	1	MELTF	20	15	5	11.82	2.29	65.74	2.79E-03	1.65E-02	
3	196769431	197026201	1	DLG1	16	11	5	11.82	2.29	65.74	2.78E-03	1.65E-02	
3	197236654	197300194	1	BDH1	21	16	5	10.83	2.17	53.49	2.71E-03	1.65E-02	
4	926175	998352	4	TMEM175	11	3	8	208.95	35.11	1620.09	7.82E-09	1.02E-06	
4	71263648	71275912	1	OPRPN	12	10	2	35.48	2.18	359.05	3.83E-03	2.11E-02	
4	100044808	100078949	1	ADH4	10	9	1	53.08	2.57	432.30	7.12E-04	5.97E-03	
4	135117488	135122944	1	PABPC4L	751	717	34	2.88	1.67	4.64	4.28E-05	6.19E-04	
7	75471920	75518244	1	RHBDD2	10	8	2	39.55	3.40	914.30	4.24E-03	2.16E-02	
8	182137	214703	1	ZNF596	70	62	8	10.72	3.08	29.99	3.11E-05	5.06E-04	
8	356428	427967	1	FBXO25	21	15	6	25.63	5.27	104.95	1.17E-05	2.53E-04	
8	439792	495781	1	TDRP	21	16	5	25.92	3.50	162.19	5.12E-04	4.75E-03	
8	564746	1656642	2	DLGAP2	10	5	5	378.01	24.69	9714.15	2.06E-05	3.83E-04	
8	1703944	1749877	1	CLN8	16	11	5	86.88	9.12	706.68	3.68E-05	5.62E-04	
8	1772092	1906807	1	ARHGEF10	20	15	5	40.29	4.93	281.87	2.18E-04	2.46E-03	
8	1921843	1955102	1	KBTBD11	29	24	5	24.98	3.61	118.88	1.96E-04	2.40E-03	
8	1993161	2113475	1	MYOM2	25	20	5	20.01	2.88	102.97	6.95E-04	5.97E-03	
10	96443486	96495947	1	CYP2C18	22	19	3	16.34	2.23	76.93	1.23E-03	9.37E-03	
10	128933690	128994855	1	INSYN2A	13	10	3	30.28	3.38	203.84	7.89E-04	6.41E-03	
10	134884433	134945179	1	ADGRA1	10	6	4	23.80	2.61	226.77	3.00E-03	1.70E-02	
10	134973723	135039916	1	KNDC1	12	8	4	23.81	2.61	226.80	3.00E-03	1.70E-02	
10	135043721	135090372	3	ADAM8	15	11	4	17.30	2.15	106.25	2.69E-03	1.65E-02	
10	135092134	135236024	8	PAOX	12	8	4	32.82	3.42	315.79	1.40E-03	1.02E-02	
10	135234170	135238076	1	SPRN	18	14	4	10.94	1.44	56.45	7.56E-03	3.62E-02	
10	135367405	135383339	1	SYCE1	52	45	7	8.70	2.37	25.79	2.74E-04	2.85E-03	
11	4964061	4968835	1	OR51A4	10	9	1	21.51	1.02	171.74	9.80E-03	4.39E-02	
11	7817452	7852387	2	OR5P2	69	65	4	7.09	1.63	21.53	2.10E-03	1.42E-02	
11	22359927	22647369	2	FANCF	11	9	2	47.92	5.56	307.69	8.98E-05	1.23E-03	

11	26580579	26593810	1	<i>MUC15</i>	15	14	1	17.66	0.94	99.53	7.48E-03	3.62E-02
11	59953175	60085417	1	<i>MS4A4A</i>	12	10	2	26.02	2.79	148.93	8.93E-04	7.04E-03
12	18233803	18473041	1	<i>RERGL</i>	20	16	4	11.27	1.58	51.25	4.25E-03	2.16E-02
15	31182762	32162992	5	<i>KLF13</i>	61	50	11	10.89	3.98	27.70	1.07E-06	3.09E-05
15	32215641	32465219	1	<i>CHRNA7</i>	69	56	13	10.82	3.98	27.25	9.47E-07	3.08E-05
16	810734	850737	4	<i>MSLN</i>	12	9	3	24.84	2.70	161.52	1.42E-03	1.02E-02
16	1271651	1292555	3	<i>TPSAB1</i>	12	9	3	43.25	3.99	353.23	6.22E-04	5.58E-03
16	1306060	1309009	1	<i>TPSD1</i>	10	7	3	94.14	6.94	1044.64	1.73E-04	2.24E-03
16	15016659	15045917	1	<i>NPIPA1</i>	38	36	2	9.38	1.31	41.05	7.85E-03	3.64E-02
16	15068448	15233196	1	<i>PDXDC1</i>	36	34	2	9.51	1.33	41.97	7.67E-03	3.62E-02
16	15489611	16236931	7	<i>MPV17L</i>	148	140	8	4.35	1.41	10.83	4.06E-03	2.16E-02
16	28834356	29002105	9	<i>TUFM</i>	69	60	9	6.01	1.67	17.29	2.04E-03	1.42E-02
16	29674300	30134827	27	<i>GDPD3</i>	143	122	21	16.52	8.48	30.88	1.21E-17	3.15E-15
17	34842473	36105096	15	<i>LHX1</i>	15	8	7	19.19	3.41	93.52	3.14E-04	3.06E-03
17	80572438	81009686	7	<i>FN3KRP</i>	19	17	2	45.44	5.22	394.73	2.03E-04	2.40E-03
19	20188803	20311297	2	<i>ZNF486</i>	14	12	2	22.32	3.14	105.01	2.55E-04	2.76E-03
19	55464498	55512510	1	<i>NLRP2</i>	28	25	3	6.95	1.38	25.76	8.02E-03	3.66E-02
22	18893541	18935245	3	<i>DGCR6</i>	84	72	12	6.19	2.44	13.64	2.56E-05	4.45E-04
22	19023795	19137805	4	<i>DGCR2</i>	39	27	12	19.24	6.40	55.59	5.20E-08	2.25E-06
22	19163094	19166252	1	<i>SLC25A1</i>	32	20	12	20.66	6.76	61.16	4.47E-08	2.25E-06
22	19166986	20307603	23	<i>CDC45</i>	22	10	12	61.05	15.54	302.83	2.00E-08	1.30E-06
22	20366211	20716938	5	<i>FAM230A</i>	14	6	8	70.14	10.99	628.60	1.88E-05	3.75E-04
22	20748405	20941919	4	<i>SCARF2</i>	31	20	11	24.78	7.75	73.00	1.18E-08	1.02E-06
22	21128401	21142008	1	<i>SERPIND1</i>	53	42	11	11.31	3.78	29.61	2.79E-06	6.59E-05
22	21061979	21213705	1	<i>PI4KA</i>	38	28	10	16.22	4.76	48.05	1.51E-06	3.92E-05
22	21213295	21418457	8	<i>CRKL</i>	51	40	11	13.21	4.38	34.96	7.70E-07	2.86E-05
22	22838767	22990373	4	<i>ZNF280B</i>	58	55	3	6.39	1.40	21.06	5.77E-03	2.88E-02

Tableau supplémentaire 5. – Délétions augmentant le risque de TSA lorsque covarié avec l'intelligence générale. Chaque ligne représente un cluster de gènes significatif (FDR<0,05). Le nom du gène affiché (Ref. gene) est le gène du cluster avec la p-value la plus significative dans le cluster. Les lignes de couleur menthe représentent les clusters contenant au moins un gène répertorié dans la base de données SFARI ou dans Satterstrom et al. 2020 ou à moins de 250 kb d'un de ces gènes. Les lignes de couleur turquoise représentent les clusters contenant au moins un gène incluse dans un CNV récurrent associé aux TND ou à moins de 250 kb d'un de ces CNV. Les lignes surlignées en jaune sont des gènes qui ne sont pas répertoriés dans la littérature. CHR : chromosome; Ref. gene : gène de référence dans le cluster; n : porteurs du gène; ASD : trouble du spectre autistique; Est. : Estimé; SE : Écart-type; P : p-value.

A.5 Clusters de gènes associés à une augmentation du risque de TSA si dupliqués

Modèle 1 pour les duplications : $TSA \sim \beta_{0; DUP} + \beta_{1; DUP} \times G\grave{e}ne + \beta_{2; DUP} \times sexe$

CHR	Cluster START	Cluster STOP	Genes in cluster	Ref. gene	n (Total)	n (nonASD)	n (ASD)	OR	CI95%		P	FDR
1	113615820	113674882	1	<i>LRIG2</i>	18	15	3	17.08	3.78	55.76	1.93E-05	5.36E-04
1	119911402	120057677	3	<i>HSD3B1</i>	13	11	2	8.85	1.34	34.46	5.43E-03	4.94E-02
1	145398250	145417545	1	<i>HJV</i>	284	270	14	3.15	1.75	5.23	3.48E-05	8.75E-04
1	145438490	145715614	13	<i>CD160</i>	386	364	22	3.55	2.23	5.36	1.25E-08	5.55E-07
1	145726918	145827075	2	<i>PDZK1</i>	234	222	12	3.13	1.64	5.38	1.44E-04	2.94E-03
1	145883868	145924373	1	<i>GPR89C</i>	180	170	10	3.52	1.73	6.38	1.33E-04	2.73E-03
1	146032647	146467744	2	<i>NBPF12</i>	90	76	14	10.97	5.86	19.19	1.55E-15	1.17E-13
1	146626685	147634886	10	<i>GJA8</i>	197	176	21	7.44	4.55	11.56	2.48E-17	2.32E-15
1	248083554	248226709	7	<i>OR2T8</i>	80	74	6	5.43	2.08	11.71	9.39E-05	2.00E-03
1	248258486	248654414	15	<i>OR2L13</i>	106	100	6	4.12	1.59	8.75	9.22E-04	1.33E-02
1	248671374	248738058	3	<i>OR2G6</i>	32	28	4	9.37	2.71	24.92	4.91E-05	1.13E-03
1	248753788	248761001	1	<i>OR2T10</i>	30	26	4	10.18	2.93	27.32	2.91E-05	7.63E-04
2	228531520	228611397	5	<i>SCYGR8</i>	14	12	2	9.48	1.44	36.84	4.17E-03	4.09E-02
2	241615835	241622323	1	<i>AQP12B</i>	42	36	6	12.82	4.72	29.48	2.82E-08	1.11E-06
2	241631268	241637900	1	<i>AQP12A</i>	103	93	10	7.54	3.63	14.06	3.48E-09	1.65E-07
3	238446	451100	1	<i>CHL1</i>	24	21	3	10.20	2.34	31.24	2.76E-04	5.12E-03
3	1134342	1445901	1	<i>CNTN6</i>	57	52	5	5.97	2.05	13.89	1.85E-04	3.59E-03
3	121286796	121379774	3	<i>HCLS1</i>	24	21	3	7.93	1.84	23.89	1.05E-03	1.42E-02
3	141876369	141944434	1	<i>GK5</i>	122	104	18	11.87	6.85	19.49	1.09E-20	1.18E-18
3	179113876	179169264	1	<i>GNB4</i>	10	8	2	23.36	3.31	103.13	1.66E-04	3.32E-03
3	196433148	196439121	1	<i>CEP19</i>	10	8	2	14.02	2.04	60.39	1.27E-03	1.54E-02
3	196715492	196756688	1	<i>MELTF</i>	12	10	2	11.31	1.69	45.85	2.39E-03	2.64E-02
4	2939660	2965133	1	<i>NOP14</i>	10	7	3	21.74	4.51	83.81	1.79E-05	5.00E-04
4	4861392	4865663	1	<i>MSX1</i>	13	11	2	13.99	2.07	56.76	1.01E-03	1.39E-02
4	39454207	39529931	3	<i>UGDH</i>	16	14	2	8.93	1.37	33.95	4.88E-03	4.57E-02
4	78783984	78873939	1	<i>MRPL1</i>	12	10	2	11.30	1.69	45.83	2.40E-03	2.64E-02
5	6599352	6633404	1	<i>NSUN2</i>	10	8	2	23.36	3.31	103.13	1.66E-04	3.32E-03
5	7830491	7869150	2	<i>C5orf49</i>	15	13	2	10.18	1.54	39.43	3.12E-03	3.29E-02
5	37288239	37371208	1	<i>NUP155</i>	12	9	3	19.62	4.17	71.07	2.01E-05	5.54E-04

5	178121606	178360213	4	<i>AC113348.1</i>	20	17	3	9.05	2.07	28.00	5.84E-04	9.41E-03
5	178368192	178393434	1	<i>ZNF454</i>	23	20	3	7.62	1.77	22.97	1.30E-03	1.55E-02
5	178404588	178423207	1	<i>GRM6</i>	21	18	3	8.88	2.04	27.30	6.28E-04	1.00E-02
5	180166042	180167099	1	<i>OR2Y1</i>	44	29	15	34.15	17.17	65.68	2.80E-25	6.93E-23
5	180467225	180488521	1	<i>BTNL9</i>	28	23	5	14.01	4.56	35.70	2.80E-07	1.04E-05
6	17600533	17707156	2	<i>NUP153</i>	24	21	3	7.93	1.84	23.89	1.05E-03	1.42E-02
6	26017260	26253938	28	<i>H3C2</i>	20	17	3	11.19	2.54	35.21	1.95E-04	3.68E-03
6	28471073	28502732	2	<i>GPX6</i>	14	11	3	20.24	4.35	70.55	1.22E-05	3.43E-04
6	28538312	28583989	1	<i>ZBED9</i>	17	15	2	8.69	1.33	32.70	5.27E-03	4.82E-02
6	31367561	31383092	1	<i>MICA</i>	264	235	29	7.68	5.07	11.23	6.35E-24	1.47E-21
6	118869459	118882879	1	<i>PLN</i>	24	18	6	17.38	6.14	43.24	5.25E-09	2.46E-07
6	118781935	119031238	1	<i>CEP85L</i>	19	14	5	18.27	5.74	50.04	7.60E-08	2.89E-06
7	16899036	16921611	1	<i>AGR3</i>	47	42	5	7.04	2.39	16.63	5.53E-05	1.24E-03
7	18066393	18067469	1	<i>PRPS1L1</i>	17	15	2	8.69	1.33	32.70	5.27E-03	4.82E-02
7	72716513	72722823	1	<i>NSUN5</i>	11	8	3	18.14	3.86	66.40	3.40E-05	8.59E-04
7	72726532	72772646	2	<i>TRIM50</i>	14	11	3	16.05	3.50	55.05	4.31E-05	1.02E-03
7	72848108	74016930	21	<i>FZD9</i>	21	16	5	20.07	6.31	54.51	2.62E-08	1.04E-06
7	75398851	75677322	6	<i>CCL24</i>	32	29	3	7.00	1.64	20.50	1.75E-03	2.02E-02
7	76090993	76135312	1	<i>DTX2</i>	14	12	2	10.52	1.59	41.32	2.86E-03	3.07E-02
7	133812052	133949343	1	<i>LRGUK</i>	14	12	2	13.37	2.00	53.27	1.13E-03	1.52E-02
7	133974084	134001682	1	<i>SLC35B4</i>	16	14	2	13.97	2.11	53.93	8.10E-04	1.23E-02
8	13424378	13425797	1	<i>C8orf48</i>	98	91	7	5.30	2.20	10.82	3.09E-05	7.94E-04
8	53446597	53477990	1	<i>ALKAL1</i>	94	87	7	5.06	2.10	10.33	5.10E-05	1.17E-03
8	53535018	53658403	1	<i>RB1CC1</i>	92	86	6	4.51	1.74	9.64	4.51E-04	7.52E-03
8	87226292	87333375	1	<i>SLC7A13</i>	28	23	5	9.54	3.17	23.60	6.85E-06	2.06E-04
8	145946292	145981802	1	<i>ZNF251</i>	14	12	2	11.80	1.77	46.72	1.86E-03	2.14E-02
8	145997609	146072894	4	<i>ZNF34</i>	17	15	2	8.69	1.33	32.70	5.27E-03	4.82E-02
9	116231	179147	2	<i>FOXD4</i>	202	193	9	2.84	1.34	5.26	2.47E-03	2.69E-02
9	214854	465259	1	<i>DOCK8</i>	105	99	6	4.26	1.65	9.07	6.96E-04	1.08E-02
9	470291	746106	1	<i>KANK1</i>	33	29	4	8.81	2.56	23.28	7.40E-05	1.59E-03
9	841690	969090	1	<i>DMRT1</i>	21	16	5	18.62	5.88	50.31	4.93E-08	1.92E-06
9	976655	1057552	2	<i>DMRT3</i>	29	23	6	15.76	5.65	38.05	7.94E-09	3.62E-07
9	12685439	12710285	1	<i>TYRP1</i>	15	12	3	17.22	3.76	58.55	2.64E-05	6.98E-04
9	12775019	12823059	1	<i>LURAP1L</i>	12	9	3	25.89	5.41	95.43	4.47E-06	1.41E-04
9	79000442	79009433	1	<i>RFK</i>	19	17	2	12.35	1.89	45.97	1.17E-03	1.54E-02

9	79034766	79266119	1	<i>GCNT1</i>	16	14	2	13.97	2.11	53.94	8.10E-04	1.23E-02
9	105757593	105780770	1	<i>CYLC2</i>	43	39	4	6.20	1.83	15.85	6.65E-04	1.05E-02
9	116822407	116861571	2	<i>KIF12</i>	11	5	6	55.90	16.12	205.73	2.12E-10	1.11E-08
10	120900279	120938324	2	<i>SFXN4</i>	10	8	2	12.24	1.80	51.70	2.02E-03	2.29E-02
11	3659742	3692614	3	<i>ART1</i>	16	13	3	16.56	3.64	55.43	2.93E-05	7.63E-04
11	34642640	34684835	1	<i>EHF</i>	10	8	2	14.01	2.04	60.35	1.27E-03	1.54E-02
11	51411378	51516211	2	<i>OR4C46</i>	82	76	6	4.72	1.82	10.14	3.17E-04	5.75E-03
12	8071826	8088783	1	<i>SLC2A3</i>	2822	2729	93	2.17	1.74	2.65	5.23E-13	3.07E-11
12	27396901	27655118	3	<i>STK38L</i>	103	97	6	3.82	1.48	8.10	1.70E-03	1.97E-02
12	34175403	34182629	1	<i>ALG10</i>	454	438	16	2.41	1.39	3.85	6.43E-04	1.02E-02
12	52708085	52761265	2	<i>KRT83</i>	157	148	9	4.14	1.94	7.75	4.67E-05	1.09E-03
12	112204691	112255336	1	<i>ALDH2</i>	17	10	7	55.30	18.71	155.93	5.51E-14	3.35E-12
13	43355686	43365685	1	<i>FAM216B</i>	19	15	4	15.49	4.28	45.14	2.86E-06	9.53E-05
13	43460524	43566407	1	<i>EPST11</i>	90	83	7	4.62	1.92	9.43	1.26E-04	2.62E-03
14	21616529	21737653	2	<i>OR5AU1</i>	15	12	3	14.01	3.10	46.83	7.91E-05	1.69E-03
14	63779549	63785649	1	<i>GPHB5</i>	17	9	8	73.23	25.49	206.87	3.91E-16	3.01E-14
15	23255242	23268823	1	<i>GOLGA8I</i>	118	104	14	8.74	4.72	15.02	1.44E-13	8.62E-12
15	23405069	23612378	2	<i>GOLGA8S</i>	84	76	8	7.10	3.11	14.14	2.73E-07	1.03E-05
15	23684185	23932452	4	<i>MKRN3</i>	51	33	18	37.32	19.83	68.43	1.07E-30	4.13E-28
15	24920922	27184686	6	<i>ATP10A</i>	42	23	19	58.49	30.09	112.87	7.53E-34	1.30E-30
15	27112058	27194355	1	<i>GABRA5</i>	50	30	20	43.08	23.22	78.72	5.97E-34	1.30E-30
15	27216328	28344461	2	<i>GABRG3</i>	40	21	19	61.93	31.50	121.49	1.96E-33	1.70E-30
15	28356186	28637164	2	<i>HERC2</i>	13	5	8	129.03	39.31	460.77	3.11E-15	2.20E-13
15	28764757	28983530	2	<i>GOLGA8G</i>	12	7	5	46.28	12.79	158.63	1.08E-09	5.21E-08
15	35043233	35047199	1	<i>GJD2</i>	11	9	2	15.44	2.25	65.61	8.16E-04	1.23E-02
15	35082431	35087750	1	<i>ACTC1</i>	10	8	2	14.01	2.04	60.35	1.27E-03	1.54E-02
15	90895516	90912378	1	<i>ZNF774</i>	14	12	2	13.37	2.00	53.27	1.13E-03	1.52E-02
16	3014217	3074287	8	<i>CLDN6</i>	10	8	2	14.01	2.04	60.35	1.27E-03	1.54E-02
16	5074845	5147796	4	<i>NAGPA</i>	24	21	3	9.53	2.19	29.08	4.01E-04	6.92E-03
16	8768453	8878432	1	<i>ABAT</i>	16	14	2	9.85	1.50	37.73	3.40E-03	3.56E-02
16	8874241	8943182	2	<i>TMEM186</i>	73	66	7	6.10	2.52	12.63	8.17E-06	2.44E-04
16	8946799	8979208	2	<i>LITAFD</i>	66	59	7	6.88	2.82	14.36	2.36E-06	7.93E-05
16	8985954	9069185	1	<i>USP7</i>	17	12	5	19.31	6.00	54.18	6.74E-08	2.60E-06
16	9185501	9215492	1	<i>C16orf72</i>	27	24	3	7.16	1.67	21.25	1.66E-03	1.94E-02
16	14927578	14990014	1	<i>NOMO1</i>	10	8	2	14.01	2.04	60.35	1.27E-03	1.54E-02
16	15016659	15045917	1	<i>NPIPA1</i>	220	208	12	3.51	1.84	6.05	3.06E-05	7.92E-04

16	15131711	15149936	1	<i>NTAN1</i>	518	501	17	2.07	1.22	3.27	3.41E-03	3.56E-02
16	15068448	15233196	1	<i>PDXDC1</i>	220	206	14	4.12	2.27	6.88	4.42E-07	1.61E-05
16	15153890	15474904	2	<i>RRN3</i>	534	515	19	2.27	1.38	3.49	5.24E-04	8.52E-03
16	15489611	16236931	7	<i>BMERB1</i>	820	793	27	2.06	1.37	2.98	2.43E-04	4.54E-03
16	21608539	21668794	2	<i>METTL9</i>	29	24	5	12.37	4.06	31.15	7.80E-07	2.79E-05
16	21675289	21773256	1	<i>OTOA</i>	13	11	2	10.89	1.64	43.41	2.63E-03	2.84E-02
16	28834356	29002105	9	<i>ATXN2L</i>	112	106	6	3.70	1.43	7.82	2.14E-03	2.40E-02
16	29674300	30134827	27	<i>C16orf92</i>	171	148	23	9.52	5.91	14.68	1.82E-22	3.49E-20
16	30194148	30205627	2	<i>CORO1A</i>	123	116	7	3.54	1.49	7.13	1.33E-03	1.59E-02
16	30205196	30265831	3	<i>SLX1A</i>	19	13	6	28.36	9.52	76.69	1.61E-10	8.57E-09
16	70488498	70719956	5	<i>FCSK</i>	12	10	2	11.30	1.69	45.83	2.40E-03	2.64E-02
16	77224732	77247112	2	<i>MON1B</i>	28	25	3	7.85	1.83	23.35	1.03E-03	1.41E-02
16	81134480	81253975	1	<i>PKD1L2</i>	84	79	5	4.03	1.40	9.15	2.90E-03	3.10E-02
16	89627085	89663654	2	<i>CPNE7</i>	20	17	3	12.11	2.74	38.32	1.26E-04	2.62E-03
16	89679716	89737680	3	<i>DPEP1</i>	23	20	3	9.73	2.23	29.85	3.70E-04	6.57E-03
16	89747145	89807333	4	<i>ZNF276</i>	15	13	2	12.82	1.93	50.23	1.26E-03	1.54E-02
16	89803957	89883065	1	<i>FANCA</i>	11	9	2	18.12	2.62	77.51	4.29E-04	7.22E-03
17	411908	624957	1	<i>VPS53</i>	127	121	6	3.37	1.31	7.11	4.15E-03	4.09E-02
17	1247569	1303516	1	<i>YWHAE</i>	13	10	3	18.74	4.02	66.24	2.19E-05	5.83E-04
17	1323983	1366456	1	<i>CRK</i>	10	7	3	25.24	5.17	99.18	9.09E-06	2.69E-04
17	1367480	1395980	1	<i>MYO1C</i>	13	10	3	18.74	4.02	66.24	2.19E-05	5.83E-04
17	1421012	1466086	1	<i>PITPNA</i>	15	13	2	9.19	1.40	35.33	4.51E-03	4.25E-02
17	1472562	1532180	1	<i>SLC43A2</i>	15	13	2	9.19	1.40	35.33	4.51E-03	4.25E-02
17	1606110	1613671	1	<i>TLCD2</i>	14	12	2	9.48	1.44	36.84	4.17E-03	4.09E-02
17	15635561	15652506	1	<i>TBC1D26</i>	14	12	2	9.48	1.44	36.84	4.17E-03	4.09E-02
17	15848231	15948329	3	<i>TTC19</i>	26	23	3	6.91	1.61	20.50	2.01E-03	2.29E-02
17	16120505	16340317	4	<i>PIGL</i>	12	9	3	17.38	3.73	62.03	3.65E-05	8.90E-04
17	16344891	16395507	1	<i>LRRC75A</i>	13	10	3	18.74	4.02	66.24	2.19E-05	5.83E-04
17	29096406	29326929	5	<i>CRLF3</i>	13	10	3	18.74	4.02	66.24	2.19E-05	5.83E-04
17	76210267	76237068	2	<i>TMEM235</i>	13	11	2	9.78	1.48	38.52	3.85E-03	3.98E-02
17	80272744	80448413	8	<i>CD7</i>	14	12	2	9.48	1.44	36.84	4.17E-03	4.09E-02
17	80572438	80606429	1	<i>WDR45B</i>	21	18	3	7.45	1.73	22.54	1.49E-03	1.76E-02
17	80612849	80709072	3	<i>RAB40B</i>	13	10	3	14.99	3.27	51.70	6.66E-05	1.45E-03
17	80787310	80797898	1	<i>ZNF750</i>	10	8	2	14.01	2.04	60.36	1.27E-03	1.54E-02
18	158383	214629	1	<i>USP14</i>	31	28	3	5.90	1.39	17.14	4.12E-03	4.09E-02
18	670318	712662	1	<i>ENOSF1</i>	85	80	5	3.68	1.28	8.33	5.23E-03	4.82E-02

18	721588	812547	1	YES1	62	57	5	5.36	1.85	12.38	4.11E-04	7.02E-03
18	904872	912173	1	ADCYAP1	62	57	5	5.49	1.89	12.69	3.44E-04	6.17E-03
19	7684411	7834491	11	RETN	13	11	2	13.99	2.07	56.76	1.01E-03	1.39E-02
19	52255279	52273784	1	FPR2	44	40	4	5.58	1.65	14.18	1.27E-03	1.54E-02
19	52901102	52955568	2	ZNF528	10	8	2	10.83	1.61	44.87	3.03E-03	3.21E-02
19	54544079	54619055	5	NDUFA3	12	9	3	22.41	4.72	82.08	1.01E-05	2.93E-04
19	55644066	55660574	1	TNNT1	10	8	2	16.30	2.35	71.38	7.37E-04	1.13E-02
20	939095	982954	1	RSPO4	41	37	4	7.93	2.32	20.53	1.31E-04	2.71E-03
20	18268927	18447925	2	ZNF133	12	9	3	30.27	6.29	111.42	1.63E-06	5.65E-05
20	18448033	18465290	1	POLR3F	12	9	3	30.27	6.29	111.42	1.63E-06	5.65E-05
20	61974575	62009753	1	CHRNA4	14	12	2	15.35	2.28	61.20	6.27E-04	1.00E-02
22	19023795	20716938	33	DGCR8	309	297	12	2.35	1.24	4.03	3.96E-03	4.06E-02
22	21061979	21418457	10	SERPIND1	476	456	20	2.56	1.58	3.92	4.64E-05	1.09E-03
22	21562262	21805752	4	RIMBP3B	61	56	5	5.27	1.82	12.17	4.69E-04	7.73E-03
22	21820712	21905750	2	TMEM191 C	28	25	3	6.72	1.57	19.80	2.26E-03	2.53E-02
22	21903736	21978323	1	UBE2L3	13	10	3	21.26	4.53	75.82	1.10E-05	3.16E-04
22	21982378	22221970	6	YDJC	17	14	3	14.45	3.21	47.43	5.73E-05	1.27E-03
22	22599192	22599927	1	VPREB1	121	106	15	9.26	5.10	15.69	5.24E-15	3.56E-13
22	22838767	22863060	1	ZNF280B	39	31	8	16.56	6.88	35.77	1.33E-11	7.70E-10
22	22868060	22901768	2	PRAME	45	37	8	13.70	5.78	28.93	1.11E-10	6.22E-09
22	22986946	22990373	1	GGTLC2	58	49	9	12.01	5.39	24.00	4.01E-11	2.28E-09
22	23229960	23238291	1	IGLL5	128	116	12	6.84	3.53	12.08	6.21E-10	3.07E-08
22	24981588	24989175	1	LRRRC75B	146	138	8	3.75	1.67	7.25	3.39E-04	6.11E-03

Tableau supplémentaire 6. – Duplications augmentant le risque de TSA. Chaque ligne représente un cluster de gènes significatif (FDR<0,05). Le nom du gène affiché (Ref. gene) est le gène du cluster avec la p-value la plus significative dans le cluster. Les lignes de couleur menthe représentent les clusters contenant au moins un gène répertorié dans la base de données SFARI ou dans Satterstrom et al. 2020 ou à moins de 250 kb d'un de ces gènes. Les lignes de couleur turquoise représentent les clusters contenant au moins un gène incluse dans un CNV récurrent associé aux TND ou à moins de 250 kb d'un de ces CNV. Les lignes surlignées en jaune sont des gènes qui ne sont pas répertoriés dans la littérature. CHR : chromosome; Ref. gene : gène de référence dans le cluster; n : porteurs du gène; ASD : trouble du spectre autistique; Est. : Estimé; SE : Écart-type; P : p-value.

Modèle 2 pour les duplications : $TSA \sim \beta_{0; DUP} + \beta_{1; DUP} \times \text{Gène}$
 $+ \beta_{2; DUP} \times \text{sexe} + \beta_{3; DUP} \times \text{Intell. aj.}$

CHR	Cluster START	Cluster STOP	Genes in cluster	Ref. gene	n (Total)	n (nonASD)	n (ASD)	OR	CI95%		P	FDR
1	113454469	113499635	1	SLC16A1	22	20	2	24.32	3.48	107.65	1.26E-04	2.23E-03
1	113615820	113674882	1	LRIG2	18	15	3	56.77	9.14	294.64	2.86E-06	8.87E-05
1	119911402	120057677	3	HAO2	16	14	2	19.87	0.95	165.44	1.18E-02	4.84E-02
1	145438490	145715614	13	CD160	386	364	22	2.86	1.55	4.89	3.15E-04	4.23E-03
1	145726918	145827075	2	PDZK1	234	222	12	3.19	1.51	6.09	1.01E-03	1.00E-02
1	145883868	145924373	1	GPR89C	180	170	10	4.59	2.16	8.81	1.84E-05	4.87E-04
1	146032647	146467744	2	NBPF12	90	76	14	12.06	5.89	23.43	1.12E-12	7.51E-10
1	146626685	147634886	10	GJA8	197	176	21	7.29	4.01	12.59	7.71E-12	3.88E-09
1	152483279	152539256	3	LCE5A	11	10	1	17.69	0.90	108.15	9.76E-03	4.25E-02
1	248083554	248226709	7	OR2T8	80	74	6	4.30	1.23	11.64	9.19E-03	4.12E-02
1	248258486	248654414	15	OR2T5	64	60	4	4.59	1.30	12.54	6.95E-03	3.59E-02
1	248671374	248738058	3	OR2G6	32	28	4	7.43	2.00	22.13	8.25E-04	8.83E-03
1	248753788	248761001	1	OR2T10	30	26	4	8.94	2.34	27.81	3.92E-04	5.00E-03
2	110969106	111002997	1	MTLN	80	77	3	4.95	1.14	14.95	1.17E-02	4.84E-02
2	119599747	119605224	1	EN1	34	32	2	12.97	1.65	63.84	4.34E-03	2.94E-02
2	152214106	152236560	1	TNFAIP6	11	10	1	30.59	1.51	225.62	2.93E-03	2.30E-02
2	152266397	152364527	1	RIF1	11	10	1	24.98	1.26	172.09	4.42E-03	2.97E-02
2	241615835	241622323	1	AQP12B	42	36	6	21.83	7.05	60.94	1.34E-08	5.86E-07
2	241631268	241637900	1	AQP12A	103	93	10	12.03	5.48	24.11	3.16E-11	7.34E-09
3	238446	451100	1	CHL1	24	21	3	33.12	4.59	157.00	4.41E-05	9.97E-04
3	1134342	1445901	1	CNTN6	57	52	5	13.02	3.41	39.99	2.96E-05	7.54E-04
3	93591895	93850676	5	PROS1	19	18	1	74.28	3.18	882.42	8.81E-04	9.05E-03
3	96336032	96337080	1	MTRNR2L12	10	9	1	43.36	1.67	627.03	6.78E-03	3.57E-02
3	97806017	98073663	6	OR5H6	17	16	1	37.40	1.53	420.79	6.35E-03	3.49E-02
3	98109510	98110475	1	OR5K3	11	10	1	37.39	1.53	420.79	6.35E-03	3.49E-02
3	141876369	141944434	1	GK5	122	104	18	8.14	3.64	16.73	5.27E-08	1.86E-06
3	179113876	179169264	1	GNB4	10	8	2	19.82	0.94	161.93	1.24E-02	4.93E-02
3	196433148	196439121	1	CEP19	10	8	2	30.46	1.41	318.21	5.29E-03	3.35E-02
3	196715492	196756688	1	MELTF	12	10	2	42.86	1.97	454.09	2.29E-03	1.89E-02
4	2939660	2965133	1	NOP14	10	7	3	46.43	5.74	509.16	4.91E-04	5.96E-03
4	37592422	37687998	1	RELL1	11	10	1	22.75	1.10	174.40	7.60E-03	3.86E-02

4	57514154	57548065	1	HOPX	10	9	1	37.08	1.75	298.87	2.49E-03	1.98E-02
5	37288239	37371208	1	NUP155	12	9	3	41.45	4.46	391.14	5.15E-04	6.18E-03
5	37812779	37840143	1	GDNF	14	13	1	20.86	1.02	145.39	8.29E-03	4.01E-02
5	180166042	180167099	1	OR2Y1	44	29	15	35.12	12.76	92.17	1.05E-12	7.51E-10
5	180467225	180488521	1	BTNL9	28	23	5	14.64	2.80	57.74	3.56E-04	4.64E-03
6	31367561	31383092	1	MICA	264	235	29	5.55	2.74	10.22	2.69E-07	9.19E-06
6	36916039	36932613	1	PI16	14	12	2	24.57	1.15	225.35	8.15E-03	4.01E-02
6	118869459	118882879	1	PLN	24	18	6	18.19	5.16	58.83	1.78E-06	5.60E-05
6	118781935	119031238	1	CEP85L	19	14	5	14.60	3.60	51.99	5.18E-05	1.13E-03
6	123537484	123958238	1	TRDN	52	50	2	10.52	1.53	43.01	3.83E-03	2.74E-02
7	16899036	16921611	1	AGR3	47	42	5	10.54	2.77	32.42	1.26E-04	2.23E-03
7	72848108	74016930	21	FZD9	21	16	5	7.38	1.50	29.39	6.32E-03	3.49E-02
7	141951963	141957878	1	PRSS58	64	60	4	8.51	1.93	26.71	9.64E-04	9.71E-03
8	13424378	13425797	1	C8orf48	98	91	7	6.35	1.89	16.78	7.37E-04	8.09E-03
8	87226292	87333375	1	SLC7A13	28	23	5	8.17	2.22	24.16	4.00E-04	5.07E-03
8	142127377	142205907	1	DENND3	11	10	1	55.72	2.48	632.50	1.49E-03	1.35E-02
8	145946292	145981802	1	ZNF251	14	12	2	41.91	5.41	259.51	5.55E-05	1.19E-03
8	145997609	146072894	4	ZNF34	17	15	2	36.77	4.93	198.90	5.22E-05	1.13E-03
9	841690	969090	1	DMRT1	21	16	5	19.34	3.71	76.18	7.97E-05	1.54E-03
9	976655	1057552	2	DMRT3	29	23	6	18.29	3.55	70.52	9.08E-05	1.69E-03
9	116822407	116861571	2	AMBP	10	4	6	61.94	6.40	664.98	3.70E-04	4.75E-03
10	6130997	6159420	1	RBM17	14	12	2	18.73	2.48	100.02	1.10E-03	1.08E-02
10	120900279	120938324	2	SFXN4	10	8	2	15.89	2.09	90.36	2.30E-03	1.89E-02
11	3659742	3692614	3	ART5	17	14	3	32.19	4.16	179.00	1.53E-04	2.51E-03
11	21821480	22304903	1	ANO5	11	10	1	16.58	0.85	107.90	1.17E-02	4.84E-02
11	22868468	22882020	1	CCDC179	13	12	1	160.00	6.30	4065.71	3.45E-04	4.54E-03
11	73357223	73373864	1	PLEKHB1	10	9	1	216.26	6.41	7298.36	1.12E-03	1.09E-02
12	8071826	8088783	1	SLC2A3	2822	2729	93	3.13	2.47	3.92	4.40E-22	8.86E-19
12	27396901	27655118	3	STK38L	103	97	6	4.31	1.21	11.87	1.02E-02	4.41E-02
12	33527173	33592754	1	SYT10	86	82	4	4.81	1.39	12.77	4.43E-03	2.97E-02
12	34175403	34182629	1	ALG10	454	438	16	2.59	1.15	5.03	1.01E-02	4.37E-02
12	52708085	52761265	2	KRT83	157	148	9	6.16	2.61	12.77	5.54E-06	1.55E-04
12	52771596	52779436	1	KRT84	111	107	4	3.91	1.15	10.10	1.17E-02	4.84E-02
12	112204691	112255336	1	ALDH2	17	10	7	36.57	10.79	122.51	5.08E-09	2.38E-07
14	21616529	21737653	2	OR5AU1	15	12	3	23.33	3.96	138.73	2.90E-04	4.00E-03
14	63779549	63785649	1	GPHB5	17	9	8	91.13	24.46	364.45	2.94E-11	7.34E-09

15	23255242	23268823	1	<i>GOLGA8I</i>	118	104	14	5.62	2.08	12.92	1.76E-04	2.76E-03
15	23405069	23612378	2	<i>RP11-467N20.5</i>	85	77	8	6.54	1.97	17.49	5.91E-04	6.89E-03
15	23684185	23932452	4	<i>MAGEL2</i>	56	38	18	35.80	12.33	107.12	5.10E-11	7.34E-09
15	24920922	27184686	6	<i>ATP10A</i>	42	23	19	45.74	14.71	150.82	7.84E-11	1.05E-08
15	27112058	27194355	1	<i>GABRA5</i>	50	30	20	39.78	13.19	131.48	1.82E-10	1.93E-08
15	27216328	28344461	2	<i>GABRG3</i>	40	21	19	58.13	16.92	242.03	9.39E-10	7.56E-08
15	28356186	28637164	2	<i>HERC2</i>	13	5	8	146.40	11.57	4256.21	6.47E-04	7.32E-03
15	28764757	28983530	2	<i>GOLGA8G</i>	12	7	5	146.40	11.57	4256.21	6.47E-04	7.32E-03
15	35043233	35047199	1	<i>GID2</i>	11	9	2	33.61	2.33	283.70	2.36E-03	1.91E-02
15	35082431	35087750	1	<i>ACTC1</i>	10	8	2	19.31	1.56	138.20	6.88E-03	3.59E-02
15	45694534	45740959	2	<i>C15orf48</i>	33	31	2	10.07	1.47	40.41	4.23E-03	2.91E-02
15	90895516	90912378	1	<i>ZNF774</i>	14	12	2	39.59	5.05	234.62	7.68E-05	1.53E-03
16	8768453	8878432	1	<i>ABAT</i>	16	14	2	27.37	3.23	187.13	8.52E-04	9.04E-03
16	8874241	8943182	2	<i>TMEM186</i>	73	66	7	11.97	4.27	29.03	2.46E-07	8.54E-06
16	8946799	8979208	2	<i>LITAFD</i>	66	59	7	11.65	3.71	30.78	3.44E-06	1.00E-04
16	8985954	9069185	1	<i>USP7</i>	17	12	5	18.83	3.90	73.27	4.47E-05	1.00E-03
16	9185501	9215492	1	<i>C16orf72</i>	27	24	3	8.33	1.84	27.79	1.51E-03	1.37E-02
16	11762311	11773008	1	<i>SNN</i>	22	21	1	16.62	0.85	104.11	1.15E-02	4.79E-02
16	11772936	11836734	1	<i>TXNDC11</i>	14	13	1	21.03	1.03	150.32	8.25E-03	4.01E-02
16	14927578	14990014	1	<i>NOMO1</i>	10	8	2	27.45	2.62	591.54	6.91E-03	3.59E-02
16	21608539	21668794	2	<i>METTL9</i>	29	24	5	5.57	1.22	18.39	1.04E-02	4.46E-02
16	29674300	30134827	27	<i>GDPD3</i>	168	146	22	6.68	3.51	12.12	1.44E-09	8.83E-08
16	30194148	30205627	2	<i>CORO1A</i>	123	116	7	3.23	1.18	7.49	1.14E-02	4.78E-02
16	30205196	30265831	3	<i>SLX1A</i>	19	13	6	23.39	5.94	90.75	3.90E-06	1.11E-04
16	77224732	77247112	2	<i>MON1B</i>	28	25	3	16.07	3.28	61.39	1.22E-04	2.22E-03
16	77281710	77468931	1	<i>ADAMTS18</i>	24	22	2	11.94	1.65	55.28	3.76E-03	2.70E-02
16	89679716	89737680	3	<i>DPEP1</i>	23	20	3	16.41	1.69	97.27	5.48E-03	3.42E-02
16	89747145	89807333	4	<i>ZNF276</i>	15	13	2	65.88	2.98	587.05	6.54E-04	7.36E-03
16	89803957	89883065	1	<i>FANCA</i>	11	9	2	86.69	3.62	1044.09	6.80E-04	7.57E-03
17	411908	624957	1	<i>VPS53</i>	127	121	6	4.76	1.33	12.97	6.17E-03	3.49E-02
17	15635561	15652506	1	<i>TBC1D26</i>	14	12	2	22.45	1.08	174.96	8.22E-03	4.01E-02
17	15848231	15948329	3	<i>TTC19</i>	26	23	3	12.34	1.84	50.15	1.68E-03	1.52E-02
17	16120505	16340317	4	<i>PIGL</i>	12	9	3	69.41	7.50	666.12	8.16E-05	1.54E-03
17	16344891	16395507	1	<i>LRRC75A</i>	13	10	3	60.60	7.11	418.79	3.42E-05	8.33E-04
17	29096406	29326929	5	<i>CRLF3</i>	13	10	3	21.90	4.07	102.92	1.05E-04	1.93E-03

17	34181950	34433014	11	<i>RDM1</i>	45	42	3	6.14	1.39	19.03	4.99E-03	3.18E-02
17	80272744	80448413	8	<i>CD7</i>	14	12	2	29.19	1.30	324.56	7.38E-03	3.76E-02
17	80572438	80606429	1	<i>WDR45B</i>	21	18	3	13.36	1.84	64.67	2.64E-03	2.10E-02
17	80612849	80709072	3	<i>FN3KRP</i>	14	11	3	91.99	9.46	1013.78	5.86E-05	1.23E-03
18	66340925	66382345	1	<i>TMX3</i>	49	47	2	10.01	1.51	39.30	3.52E-03	2.58E-02
18	74961804	74989856	1	<i>GALR1</i>	44	41	3	11.23	2.49	36.82	2.67E-04	3.73E-03
19	7684411	7834491	11	<i>RETN</i>	13	11	2	70.49	8.10	620.46	3.64E-05	8.33E-04
19	39687604	39694903	2	<i>NCCRP1</i>	10	9	1	27.07	1.19	289.99	9.63E-03	4.24E-02
19	43905648	44040282	5	<i>ETHE1</i>	15	14	1	16.51	0.86	97.98	1.03E-02	4.46E-02
19	52901102	52955568	2	<i>ZNF528</i>	10	8	2	23.83	3.11	145.38	5.56E-04	6.59E-03
19	54495542	54515923	1	<i>CACNG6</i>	18	16	2	16.95	2.31	75.46	9.21E-04	9.37E-03
19	54544079	54619055	5	<i>NDUFA3</i>	12	9	3	23.98	3.87	119.44	2.39E-04	3.52E-03
19	58637615	58666477	1	<i>ZNF329</i>	12	11	1	19.00	0.96	127.13	8.98E-03	4.06E-02
20	939095	982954	1	<i>RSPO4</i>	41	37	4	6.77	1.06	23.79	1.11E-02	4.68E-02
20	18268927	18447925	2	<i>ZNF133</i>	12	9	3	104.28	14.24	710.37	1.72E-06	5.50E-05
20	18448033	18465290	1	<i>POLR3F</i>	12	9	3	95.29	10.55	680.69	7.57E-06	2.06E-04
20	23545369	23969416	8	<i>CST9</i>	13	12	1	43.36	1.99	384.74	2.01E-03	1.77E-02
22	17659680	17739125	1	<i>ADA2</i>	33	31	2	8.11	1.18	34.39	1.07E-02	4.53E-02
22	21562262	21805752	4	<i>RIMBP3B</i>	61	56	5	8.56	2.33	25.07	2.84E-04	3.94E-03
22	21820712	21905750	2	<i>TMEM191C</i>	28	25	3	13.01	1.70	73.97	4.86E-03	3.11E-02
22	21903736	21978323	1	<i>UBE2L3</i>	13	10	3	42.47	4.14	497.70	1.26E-03	1.17E-02
22	21982378	22221970	6	<i>YDJC</i>	17	14	3	26.38	3.07	197.82	1.14E-03	1.09E-02
22	22838767	22863060	1	<i>ZNF280B</i>	39	31	8	11.16	1.59	49.13	3.80E-03	2.73E-02
22	22868060	22901768	2	<i>PRAME</i>	45	37	8	11.03	1.60	45.98	3.35E-03	2.55E-02
22	22986946	22990373	1	<i>GGTLC2</i>	58	49	9	7.61	1.62	25.47	3.02E-03	2.35E-02
22	23229960	23238291	1	<i>IGLL5</i>	128	116	12	5.31	1.75	12.95	8.58E-04	9.05E-03
22	24981588	24989175	1	<i>LRRRC75B</i>	146	138	8	3.91	1.31	9.39	5.63E-03	3.48E-02
22	46971909	47134149	2	<i>GRAMD4</i>	20	18	2	14.41	2.07	61.70	1.23E-03	1.15E-02

Tableau supplémentaire 7. – Duplications augmentant le risque de TSA lorsque covarié avec l'intelligence générale. Chaque ligne représente un cluster de gènes significatif (FDR<0,05). Le nom du gène affiché (Ref. gene) est le gène du cluster avec la p-value la plus significative dans le cluster. Les lignes de couleur menthe représentent les clusters contenant au moins un gène répertorié dans la base de données SFARI ou dans Satterstrom et al. 2020 ou à moins de 250 kb d'un de ces gènes. Les lignes de couleur turquoise représentent les clusters contenant au moins un gène incluse dans un CNV récurrent associé aux TND ou à moins de 250 kb d'un de ces CNV. Les lignes surlignées en jaune sont des gènes qui ne sont pas répertoriés dans la littérature. CHR : chromosome; n : porteurs du gène; ASD : trouble du spectre autistique; Ref. gene : gène de référence dans le cluster; Est. : Estimé; SE : Écart-type; P : p-value.

A.6 Gènes inclus dans la région 6p22.3 associée à l'augmentation du risque de TSA

Gene	Function	Expression	Positive Disease Associations
CAP2	Unknown.	Everywhere, with a higher level of expression in muscle skeletal.	-
FAM8A1	-	Everywhere, with a higher level of expression in testis.	-
NUP153	Regulate the transport of macromolecules between the nucleus and cytoplasm.	Everywhere, with a higher level of expression in EBV-transformed lymphocytes	Body Height , height, Neuroblastoma, Waist Circumference.
KIF13A	Positioning of endosomes.	Everywhere, highest = Artery - Coronary	Height , Prostatic Neoplasms.

Tableau supplémentaire 8. – Fonctions des gènes inclus dans la région 6p22.3 associée à l'augmentation du risque de TSA. Information provenant de UCSC.