



RAPPORT DE STAGE

DRY Louise

Dans le cadre du stage de **2^{ème} année** :

Stage effectué du (jj/mm/aa) : 07/06/2021 au 27/08/2021

À : Institut Pasteur

Sur le thème :

Analyses statistiques de données d'imagerie et de génétique en lien avec les troubles du spectre autistique

Eventuellement, rapport confidentiel : Date d'expiration de confidentialité :

Enseignant référent responsable :

Enseignant référent associé : Christine MARTIN

Rappel : dans ce cas fournir un 2^{ème} exemplaire du rapport

Maître de stage : Clara MOREAU

1 Principes

- Le plagiat se définit comme l’action d’un individu qui présente comme sien ce qu’il a pris à autrui.
- Le plagiat de tout ou parties de documents existants constitue une violation des droits d’auteur ainsi qu’une fraude caractérisée
- Le plagiat concerne entre autres : des phrases, une partie d’un document, des données, des tableaux, des graphiques, des images et illustrations.
- Le plagiat se situe plus particulièrement à deux niveaux : Ne pas citer la provenance du texte que l’on utilise, ce qui revient à le faire passer pour sien de manière passive. Recopier quasi intégralement un texte ou une partie de texte, sans véritable contribution personnelle, même si la source est citée.

2 Consignes

- Il est rappelé que la rédaction fait partie du travail de création d’un rapport ou d’un mémoire, en conséquence lorsque l’auteur s’appuie sur un document existant, il ne doit pas recopier les parties l’intéressant mais il doit les synthétiser, les rédiger à sa façon dans son propre texte.
- Vous devez systématiquement et correctement citer les sources des textes, parties de textes, images et autres informations reprises sur d’autres documents, trouvés sur quelque support que ce soit, papier ou numérique en particulier sur internet.
- Vous êtes autorisés à reprendre d’un autre document de très courts passages in extenso, mais à la stricte condition de les faire figurer entièrement entre guillemets et bien sur d’en citer la source.

3 Sanction : En cas de manquement à ces consignes, la DEVE/le correcteur se réservent le droit d’exiger la réécriture du document sans préjuger d’éventuelles sanctions disciplinaires.

4 Engagement :

Je soussigné (e) Louise DRY

Reconnait avoir lu et m’engage à respecter les consignes de non plagiat

A Sainte-Marie-de-Ré le 30/08/2021

Signature :

Cet engagement de non plagiat doit être inséré en début de tous les rapports, dossiers, mémoires.

Table des matières

1 <i>Principes</i>	2
2 <i>Consignes</i>	2
4 <i>Engagement</i> :	2
Résumé	5
Présentation de l’organisme d’accueil	6
<i>Unité de recherche</i>	6
Corps du rapport : analyse statistique de données de neuroimagerie et de génétique chez des individus sains porteurs de mutations associées aux TSA	8
État de l’art	8
<i>Description de l’autisme</i>	8
<i>Génétique de l’autisme</i>	10
<i>Résultats en imagerie sur les TSA en anatomie en approche top-down</i>	12
<i>L’approche bottom-up et la ‘résilience’</i>	12
Objectifs du stage	14
Matériel et méthodes	14
<i>Données d'imagerie</i>	14
<i>Données de génétique</i>	16
<i>Union des données de génétique et d'imagerie</i>	18
<i>Méthodes statistiques</i>	18
Résultats	28
<i>Pas d'influence moyenne des RR des gènes sur les différentes mesures cérébrales</i>	28
<i>L'exploitation des scores MSLL, SMSE et EXPV ne permet pas de tirer de conclusion</i>	28
<i>Des déviations beaucoup plus fréquentes chez les individus porteurs</i>	28
<i>Le nombre de régions à forte déviation réduit fortement pour les groupes porteurs lors de la prise en compte de pRho</i>	30
Discussion	32
Limites	34
<i>Limites des approches top-down</i>	34
<i>Limites des données de neuroimagerie</i>	34
<i>Limites du modèle</i>	34
Conclusion	34
Bibliographie	36
Annexe	37

Abréviations utilisées dans le rapport

CV : Cross-Validation

CNV : Copy-Number Variant

DMN : Default Mode Network

EXPV : Explained Variance

FDR : False Discovery Rate

GLM : General Linear Regression

IRM : Imagerie par Résonnance Magnétique

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

LoF : Loss of Function

MSLL : Mean Squared Log Loss

PC : Principal Components

pLI : probability of Loss-of-Function intolerance

RR : Risque Relatif

SMSE : Standardized Mean Squared Error

SNP : Single Nucleotide Polymorphism

SNV : Single Nucleotide Variant

SYN : Synonymes

TBV : Total Brain Volume

TDAH : Trouble du Déficit de l’Attention avec ou sans Hyperactivité

TSA : Troubles du Spectre Autistique

Résumé

Le présent rapport vise à présenter le stage que j’ai réalisé dans le cadre de ma deuxième année de formation à AgroParisTech. Il présente la particularité d’être un stage de 6 mois entre juin et décembre 2021, par conséquent ce rapport ne décrit que la première moitié de ma période de stage. J’ai effectué ce stage à l’Institut Pasteur dans l’unité Génétique Humaine et Fonctions Cognitives, encadré par Clara Moreau, post-doctorante, et dirigé par Thomas Bourgeron, docteur et académicien. Mon enseignante référente était Christine Martin, enseignante chercheure au sein d’AgroParisTech. Ce stage de recherche portait sur l’étude statistique de données dérivées de neuroimagerie anatomique (IRM) chez des sujets sains – portant des mutations associées au Trouble du Spectre Autistique.

This document introduces the reader to the internship I did during my second year studying at AgroParisTech. It has the particularity of being a 6-month internship from June to December 2021, therefore this document only contains the first 3 months of my experience. This internship was done at the Institut Pasteur in the “Human Genetics and Cognitive Functions” unit, supervised by Clara Moreau, postdoctoral researcher, and directed by Thomas Bourgeron, researcher and academician. My supervising teacher was Christine Martin, researcher and teacher at AgroParisTech. The subject of this research internship was to study data derived from anatomical neuroimaging in sane patients carrying mutations linked to Autism Spectrum Disorder.

Présentation de l’organisme d’accueil

L’Institut Pasteur est une fondation privée, à but non lucratif, dont le siège est situé dans le 15^{ème} arrondissement de Paris. Crée en 1887, cette fondation a dans un premier temps permis à Louis Pasteur d’étendre la vaccination contre la rage, développer l’étude des maladies infectieuses et transmettre les connaissances issues de ces études.

Aujourd’hui, l’Institut Pasteur possède 4 missions.

- La recherche : 100 unités de recherche existent au sein de l’Institut. 12 départements de recherche sont représentés : biologie cellulaire et infection, biologie computationnelle, biologie du développement et cellules souches, biologique structurale et chimie, génomes et génétique, immunologie, santé globale, microbiologie, mycologie, neurosciences, parasites et insectes vecteurs, virologie.
- L’enseignement : 900 élèves par an et 300 doctorants sur le campus.
- La santé : 7 centres collaborateurs de l’Organisation Mondiale de la Santé.
- L’innovation : 50 déclarations d’inventions.

L’Institut Pasteur est composé d’environ 2700 personnes, dont 300 étudiants diplômés, 500 stagiaires postdoctoraux de 40 pays différents, 500 scientifiques permanents et 600 temporaires de 70 pays différents chaque année. Il est aussi au centre d’un réseau mondial de 31 instituts.

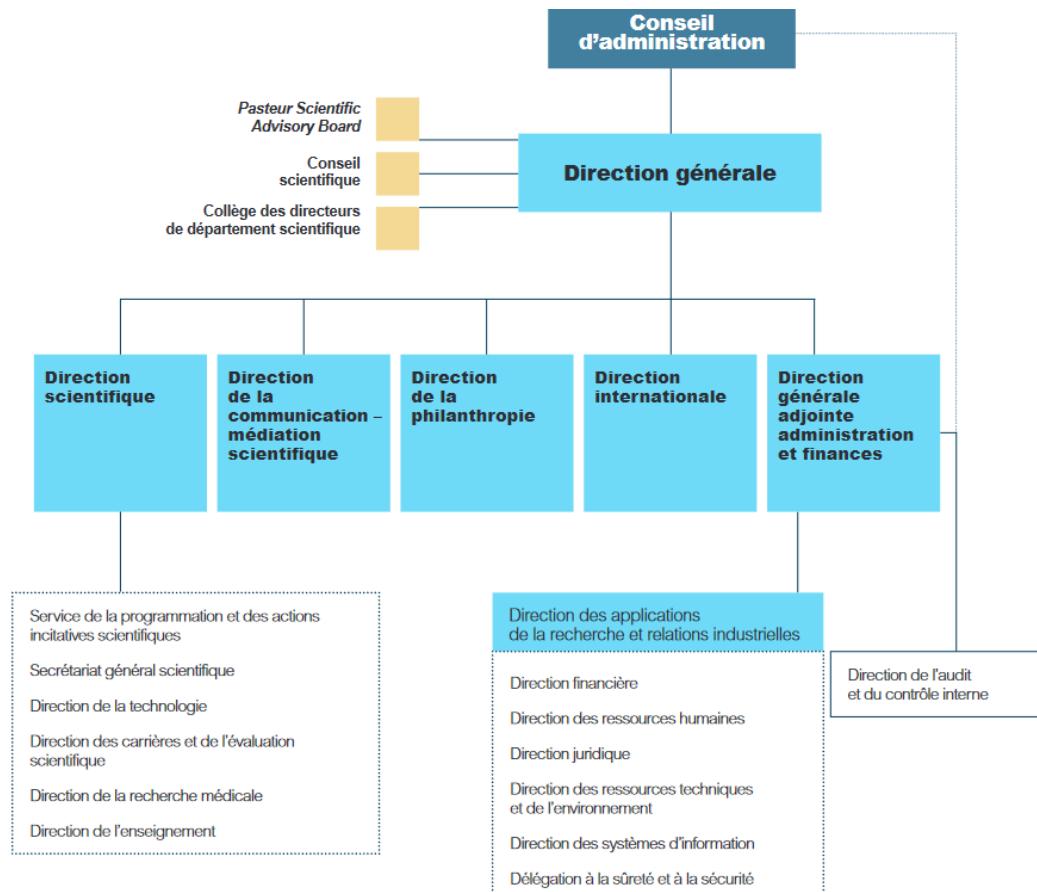
La gestion de l’Institut est assurée par un directeur général et un conseil d’administration. L’objectif du directeur général est de préparer les orientations stratégiques qui sont ensuite soumises au conseil d’administration. Le conseil scientifique fait partie de la gouvernance de l’Institut Pasteur : il donne son avis au directeur général et, éventuellement, au conseil d’administration, sur tous les problèmes de politique et d’évaluation scientifique, d’organisation et de programmes de recherche (**Figure 1**). D’un point de vue financier, l’Institut Pasteur est privé et est grande partie financé par les dons et les subventions publiques. (**Figure 2**)

Historiquement, l’Institut Pasteur a excellé dans la lutte contre les maladies infectieuses, avec par exemple l’isolement du virus de l’immunodéficience humaine en 1983. Depuis 1908, 10 scientifiques issus de cette institution ont reçu un prix Nobel soit de physiologie, soit de médecine. Cependant, l’Institut a élargi le domaine de ses compétences et s’attaque à d’autres domaines en dehors de la microbiologie, comme la neurologie ou la génétique.

Unité de recherche

Ma période de stage a été réalisée au sein de l’unité Génétique Humaine et Fonctions Cognitives, dans le département de neuroscience. Cette unité est dirigée par Thomas Bourgeron, chercheur en génétique humaine, professeur à l’université Paris Diderot, et académicien depuis novembre 2014. Il est reconnu pour avoir dirigé l’équipe impliquée dans la découverte de la première mutation monogénique liée à une forme d’autisme

Figure 1 : Organigramme de l’Institut Pasteur

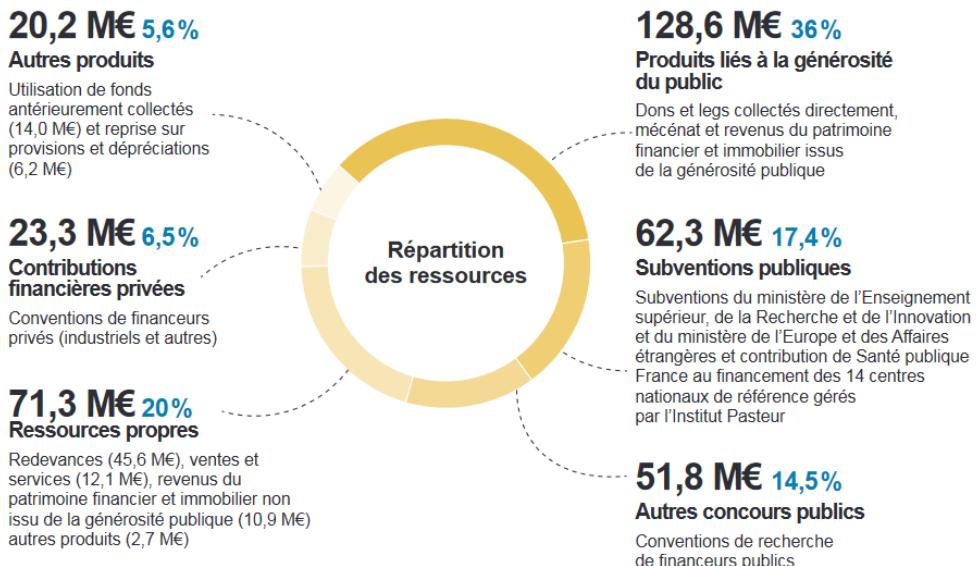


Source : Rapport annuel 2020 de l’Institut Pasteur (<https://www.pasteur.fr/fr/espace-presse/ressources-presse/rapport-annuel-rapport-financier-2020>)

Figure 2 : Sources financières de l’Institut Pasteur

Les ressources

Comme l’indique le compte de résultat par origine et destination 2020 de l’Institut Pasteur, les ressources de l’exercice inscrites au compte de résultat s’élèvent à 357,5 M€ et sont réparties de la façon suivante.



Source : Rapport annuel 2020 de l’Institut Pasteur (<https://www.pasteur.fr/fr/espace-presse/ressources-presse/rapport-annuel-rapport-financier-2020>)

(Paris Autism Research International Sibpair Study et al. 2003).

Cette unité a la particularité d’être très versatile puisqu’elle regroupe des psychiatres, des neuroscientifiques, des généticiens et des cliniciens afin de comprendre les causes des TSA et améliorer leur diagnostic et prise en charge. Elle travaille en particulier avec l’Hôpital Robert-Debré, qui est un hôpital pédiatrique du 19^{ème} arrondissement de Paris. J’ai plus précisément travaillé du côté des neurosciences et de la génétique.

Mon encadrement a été géré en particulier par Clara Moreau, post-doctorante ayant réalisé une thèse en neurosciences, et Thomas Rolland, chercheur en génétique.

Clara Moreau, mon encadrante de stage principale, est chercheuse en neurosciences et spécialisée en neuroimagerie. Après un master 2 en neurosciences à l’Ecole Normale Supérieure, elle obtient sa thèse en 2020 à l’Université de Montréal au Canada et rejoint l’équipe de Thomas Bourgeron en janvier 2021. Ses travaux de recherche sont orientés vers l’impact des différents types de variants impliqués dans l’autisme sur la connectivité anatomique et fonctionnelle du cerveau. Son objectif est d’identifier les différents mécanismes en jeu dans le développement de cette maladie qui a la particularité d’être très hétérogène

Thomas Rolland n’était pas mon encadrant officiel de stage, mais j’ai eu la chance de travailler sur son article en cours de rédaction et d’utiliser les données de génétiques qu’il traitait. Suite à son master en bioinformatique à l’université Paris Diderot, il obtient une thèse en analyse génomique à l’Université Pierre et Marie Curie. Bien qu’il ait orienté une partie de sa carrière vers les maladies cancéreuses, il finit par intégrer l’équipe de Thomas Bourgeron en 2015 pour travailler sur l’autisme. Ses travaux de recherche portent sur le développement d’approches computationnelles innovantes afin d’analyser des données génomiques et protéomiques en plus grand nombre et de meilleure qualité. Au sein du laboratoire du Professeur Bourgeron, il travaille plus volumineuses* en particulier sur le réseau d’interactions entre différentes protéines humaines afin de caractériser les relations entre génotype et phénotype chez les individus atteints d’autisme.

Corps du rapport : analyse statistique de données de neuroimagerie et de génétique chez des individus sains porteurs de mutations associées aux TSA

État de l’art

Description de l’autisme

L’autisme est un trouble du neurodéveloppement caractérisé principalement par deux aspects : une communication sociale atypique ainsi que des comportements et intérêts restreints et répétitifs (Lord et al. 2018). Les troubles du spectre autistique (TSA) englobent l’ensemble des symptômes que peuvent présenter des individus autistes. Entre autres, les TSA peuvent aussi être accompagnés de comorbidités comme de l’anxiété, des tics, une déficience intellectuelle, un Trouble du Déficit de l’Attention avec ou sans Hyperactivité (TDAH), ou encore des problèmes

gastro-intestinaux. (Bourgeron 2015) La prévalence des TSA dans la population générale est estimée autour de 1%, avec une forte proportion d’hommes diagnostiqués par rapport aux femmes (4 pour 1). (Lord et al. 2018)

Génétique de l’autisme

La découverte des gènes et des mutations impliqués dans les TSA est encore récente (et en cours), et les mécanismes biologiques impliqués restent mal connus, l’architecture génétique des TSA étant particulièrement hétérogène.

Un individu est en moyenne porteur de 3 millions de variant, généralement des mutations d’un seul ou de quelques nucléotides, appelées SNV (Single Nucleotide Variant), ou des variants plus importants, impliquant des variations du nombre de copies de larges régions chromosomiques, appelés CNV (Copy-Number Variants). On distingue :

- Les variants communs, partagés avec plus de 5% de la population mondiale,
- Les variants rares, partagés avec 0.5% à 5% de la population mondiale, et qui représentent environ 4% de notre génome,
- Les variants privés, qui sont spécifiques aux individus ou aux familles, et représentent moins de 1% de notre génome.

Chaque individu porte entre 18 et 74 SNVs *de novo*, c’est-à-dire non transmis de manière héréditaire. Ces mutations apparaissent généralement soit dans les premières phases du développement, soit dans les cellules germinales des parents de l’individu (souvent celles du père), et sont le plus souvent induites par des erreurs lors de la réPLICATION de l’ADN. (Bourgeron 2015)

Les TSA possèdent la particularité de pouvoir être entraînés aussi bien par une seule mutation rare à forte **pénétrance** que par le cumul (modèle additif) de plusieurs milliers de polymorphismes (Single-Nucleotide Polymorphisms, SNP) fréquents dans la population, chacun ayant un impact faible. Il s’agit donc de comprendre quels ensembles de mutations sont impliquées dans le développement du trouble. La contribution génétique est forte : dans 70 à 80% des cas, la variance phénotypique est expliquée par la génétique de l’individu (Al-Sarraj et al. 2021). Le reste est attribué à l’influence épigénétique et environnementale, encore largement méconnue.

De 17 à 60% des TSA sont entraînés par des SNPs, qui sont des variations d’une seule paire de base dans la séquence génétique de l’individu, mais il est compliqué de les identifier étant donné qu’ils sont nombreux et chacun à faible effet. De 1 à 20% des individus diagnostiqués sont porteurs d’une mutation unique à l’origine des principaux symptômes du trouble. Des **loci** semblent préférentiellement touchés, impliquant des gènes situés dans des zones hypermutatives, mais ces mutations sont rarement les seules causes du développement du TSA. (Bourgeron 2015)

Différents modèles génétiques ont été identifiés :

- Le modèle *de novo*, qui concerne les individus dont l’apparition du trouble est associé à une mutation SNV *de novo* rare (Sanders et al. 2012),
- Le modèle dit « multiple-hits » dont l’apparition du trouble est associée à plusieurs mutations de type SNV ou CNV délétères (Girirajan et al. 2010),
- Le modèle où le trouble est associé à une combinaison de mutation rares et de mutations à faible risque,
- Le modèle polygénique où l’individu hérite d’une faible résistance génétique et est par conséquent à haut risque pour les TSA. Une accumulation de mutations à faible risque suffit alors pour que le trouble se développe chez l’individu en question. (Bourgeron 2015)

Résultats en imagerie sur les TSA en anatomie en approche top-down

Les études top-down consistent à partir d’un trait phénotypique ou d’un diagnostic pour ensuite identifier les causes génétiques du trait observé. Ce type d’approche a par exemple permis d’observer chez les patients atteints de TSA un volume cérébral total plus important que la moyenne (Moreau et al., 2021).

Les études utilisant l’Imagerie par Résonnance Magnétique (IRM) ont permis d’étudier les aspects anatomiques des cerveaux d’individus atteints de TSA. Entre autres, plusieurs régions clés ont pu être identifiées comme associées: le cervelet, l’amygdale et le cortex frontal (Donovan et Basson 2017) mais les résultats restent incohérents . Une étude suggère que les enfants présentant des TSA ont un dysfonctionnement lors de la mise en place de l’architecture laminaire corticale, ainsi que lors de la différentiation par couches des neurones lors du développement prénatal. (Stoner et al. 2014). Les différences entre la manifestation des TSA chez les hommes et chez les femmes ont aussi été investiguées à travers l’analyse d’IRM : les femmes présenteraient une épaisseur corticale plus importante dans les régions orbitofrontales, le gyrus supérieur frontal et le sulcus intrapariétal, tandis que les hommes présenteraient un volume plus important du pallidum bilatéral. (van’t Westeinde et al. 2020)

La connectivité fonctionnelle a aussi été investiguée chez les individus atteints de TSA. Dans l’ensemble, on observe une sous-connectivité par rapport aux contrôles, en particulier dans le Default Mode Network (DMN), réseau actif lorsqu’un individu est au repos (Raichle 2015), les réseaux visuels, auditifs, et le système de récompense (Moreau et al., 2021).

Les résultats restent cependant très hétérogènes.

L’approche bottom-up et la ‘résilience’

Dans le cas de notre étude, nous avons cherché à avoir une approche montante, dite bottom-up, pour contrer l’hétérogénéité de l’approche top-down. A l’inverse de cette dernière, l’idée est

Tableau 1 : Caractéristiques des individus issus de la cohorte de données d'imagerie

Caractéristiques		Participants aux IRM	
		Moyenne	SD
Age		64	7.5
TCA*	<i>Area (mm²)</i>	169386	15486
TCT **	<i>Thickness (mm)</i>	183.7	7.6
TBV **	<i>Volume (mm³)</i>	498222	47966
Sexe (%)	Femmes	53	
	11025	62	
Site de scan	11026	13	
(%)	11027	25	

* TBA : Total Cortical Area calculée, ** TCT : Total Cortical Thickness calculée, *** TBT : Total Brain Volume calculé

cette fois-ci de partir de constats génétiques et de dérouler le mécanisme de ces constats pour parvenir à identifier leur(s) éventuelle(s) conséquence(s) phénotypique(s).

Cette approche partant des données de génétique pour aboutir aux données comportementales, s'est avérée très importante dans la recherche sur le TSA. Elle a permis en particulier de découvrir l'existence de personnes résilientes au sein des individus porteurs de mutations à risque. Plusieurs individus pouvaient tous porter des mutations extrêmement délétères dans des gènes connus pour leur association aux TSA sans pour autant être diagnostiquées d'un TSA (Bekhet, Johnson, et Zauszniewski 2012). Ainsi, l'idée est d'essayer de comprendre comment fonctionne cette résilience et de trouver les éléments, environnementaux ou génétiques, qui ont permis aux individus de ne pas développer le trouble. Il s'agit donc d'identifier *quels sont les impacts de ces mutations sur différents aspects phénotypiques des individus touchés.*

Objectifs du stage

1. Le fait d'être porteur d'un variant entraîne-t-il des différences à l'IRM au niveau de différences mesures corticales (épaisseur corticale, surface et volume) ?
2. Tous les individus touchés par des mutations délétères dans les mêmes gènes n'étant pas diagnostiqués TSA, quelles sont les différences entre les individus diagnostiqués et non diagnostiqués ?

Matériel et méthodes

L'ensemble des données est issu de la cohorte UK BioBank, une base de données regroupant au total 500,000 participants entre 40 et 69 ans au moment de leur enregistrement en Grande Bretagne. Cette base de données en libre accès offre un grand nombre d'informations génétiques, phénotypiques et environnementales sur les participants. (Littlejohns et al. 2020)

Données d'imagerie

Au total, 35,759 participants ayant accepté de fournir des données d'imagerie cérébrale ont été retenus pour notre étude. Nous avons pris en compte leur âge, leur sexe, le site où ont été réalisés les prélèvements de données d'imagerie et leur volume cérébral total/épaisseur corticale totale/surface corticale totale (**Tableau 1**). Ces facteurs peuvent influencer les différentes mesures cérébrales étudiées.

Calcul du volume cérébral total/surface corticale totale/épaisseur corticale totale

Des données traitées par FreeSurfer en amont ont été analysées dans notre étude pour :

- La surface corticale,
- L'épaisseur corticale,
- Le volume cérébral.

Chacune de ces mesures était donnée séparément pour l’hémisphère droit et l’hémisphère gauche pour chaque individu avec un total de 34 sous-régions par hémisphère. Nous avons donc tout d’abord calculé le total de chaque mesure par hémisphère pour obtenir 6 totaux :

- Le total de la surface corticale de l’hémisphère droit,
- Le total de la surface corticale de l’hémisphère gauche

Et de même pour l’épaisseur et le volume cérébral. Nous avons additionné chaque total par hémisphère pour la surface, l’épaisseur et le volume afin d’obtenir la surface totale, l’épaisseur totale et le volume total par individu. Les données ont ensuite été Z-scorées.

Données de génétique

Les données de génétique, issues de la même cohorte, comprenaient initialement 188,880 individus. Pour notre étude, nous avons travaillé sur des individus porteurs de 3 types de mutations différentes :

- Des mutations « Loss of Function » (LoF) rares dans 220 gènes associés aux TSA, c’est-à-dire des mutations entraînant une perte de fonction de la protéine encodée. Ces mutations seront notées TSA-LoF dans le rapport.
- Des mutations LoF dans 3117 gènes intolérant aux mutations LoFs mais non associés aux TSA (probability of Loss-of-Function intolerance ou pLI>0.9 ou « Loss-of-function Observed/Expected Upper bound Fraction » (LOEUF<0.31) (Rolland et al. 2021). Ces mutations seront notées TSA-LoF-constraints dans le rapport.
- Des mutations synonymes (SYN) dans des gènes TSA, qui n’ont majoritairement pas de conséquences sur la protéine encodée. Ce sont le plus souvent des mutations silencieuses. Ces mutations seront notées TSA-SYN dans le rapport.

Calcul du Risque Relatif (RR)

Le RR sert à estimer l’intensité de l’association entre les facteurs de risques génétiques et un diagnostic de TSA. Il a été calculé pour chaque gène selon (1) le nombre d’individus avec TSA porteurs d’une mutation dans le gène concerné (AE), le nombre d’individus TSA non porteurs (AN), le nombre d’individus contrôles porteurs (UE) et le nombre d’individus contrôles non porteurs (UN). (Rolland et al. 2021)

$$(1) \quad RR = \frac{\frac{AE}{UE+AN}}{\frac{UE}{UE+UN}}$$

Le RR représente la probabilité qu’un événement arrive dans un groupe par rapport à la probabilité que ce même événement arrive dans un groupe différent. Ici, l’événement est le fait qu’un individu soit porteur d’une mutation dans le gène. Le risque relatif calcule un rapport de

Tableau 2 : Caractéristiques des individus de la cohorte de données de génétique

Caractéristiques	Participants présentant un variant		
	Moyenne	SD	
Age	56	8.0	
Sexe (%)	Femmes	55	
	RR	TSA-LoF	TSA-LoF- <i>constraints</i>
	>1	649	7,958
	>2	596	6,210
RR du gène	>3	473	4,649
	>4	404	3,429
	>5	313	2,393
Non porteurs de variants		170,214	7,707

Tableau 3 : Caractéristiques des individus après fusion des données d'imagerie et de génétique

Caractéristiques	Participants inclus dans l'étude		
	Moyenne	SD	
Age	63	7.5	
Sexe (%)	Femmes	53	
	11025	75	
Site de prélèvement (%)	11026	10	
	11027	15	
	RR	TSA-LoF	TSA-LoF- <i>constraints</i>
	>1	36	728
	>2	28	549
RR du gène	>3	21	396
	>4	16	295
	>5	14	898
Non porteurs de variants		208	747
		19,313	

probabilités.

- Si le RR vaut 1, l’événement est susceptible d’arriver avec une même probabilité dans les deux groupes,
- Si le RR est supérieur à 1, alors les individus du groupe test (ici les individus diagnostiqués TSA porteurs du gène) sont plus susceptibles d’être TSA.

Dans notre étude, nous observons les cas où le RR de chaque gène est supérieur à 1, 2, 3, 4 et 5. Plus le RR est élevé, plus il est probable pour un individu porteur d’une mutation dans le gène d’être diagnostiqué TSA que pour un individu non porteur.

Union des données de génétique et d'imagerie

Afin de travailler simultanément sur la génétique et l'imagerie, nous avons recoupé les individus dont des données d'imagerie et de génétique étaient simultanément disponibles. Nous avons donc travaillé sur un total de 19,349 individus ayant chacun fourni des données d'imagerie et de génétique (**Tableau 3**).

Méthodes statistiques

Régression linéaire multiple

Nous avons dans un premier temps établi un modèle statistique de régression linéaire multiple (GLM) pour expliquer les différentes mesures corticales à notre disposition à l'aide de plusieurs variables explicatives, dont le fait d'être porteur (ou non) d'un variant, afin d'identifier si des individus présentaient des mesures avec un fort écart à la moyenne et si cet écart était dû à l'action des variants. Cette méthode nous a permis d'estimer si des différences de mesures étaient différentes en fonction de la présence de mutations. Elle a été calculé grâce à la library *stats* en R (<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>).

$$(2) Y_i = \mu + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \beta_6 x_{6,i} + \beta_7 x_{7,i} + \beta_8 x_{8,i} + \beta_9 x_{9,i} + \beta_{10} x_{10,i} + \epsilon_i$$

La variable correspondant au RR d'un gène est une variable binaire qui indique si l'individu présente (ou non) une mutation dans un gène d'un certain RR. Ces RR sont classés en 5 catégories (**Tableau 3**).

Âge, âge au carré et sexe

L'âge est une composante importante pour les analyses d'imagerie : il a été démontré qu'il jouait un rôle dans l'évolution des mesures corticales. En effet, au cours du temps, l'épaisseur corticale diminue. (Thambisetty et al. 2010) Il est donc important de corriger la régression linéaire par l'âge.

Tableau 4 : Variables incluses dans la régression linéaire multiple

	<i>Nom</i>	<i>Type de variable</i>
Y_i	Mesure corticale	Numérique
μ	Intercept	Numérique
$x_{1,i}$	RR d'un gène	Factorielle
$x_{2,i}$	Age	Numérique
$x_{3,i}$	Age au carré	Numérique
$x_{4,i}$	Sexe	Factorielle
$x_{5,i}$	PC1	Numérique
$x_{6,i}$	PC2	Numérique
$x_{7,i}$	PC3	Numérique
$x_{8,i}$	PC4	Numérique
$x_{9,i}$	Site de prélèvement	Factorielle
$x_{10,i}$	TBV	Numérique
ϵ_i	Erreur systématique	Numérique

Légende des couleurs : *données génétiques*, *données d'imagerie*

La correction par l’âge au carré permet de prendre en compte le fait que l’effet de l’âge sur les différentes mesures du cerveau n’est pas toujours linéaire.

Le sexe est une covariée importante puisque les mesures cérébrales varient en fonction de celui-ci, en particulier les mesures volumiques. (Jäncke et al. 2015)

Principal Components (PC)

Les PC sont des indices représentant les composantes principales à l’issue d’une Analyse en Composantes Principales (ACP). Dans notre cas, ces PC permettent d’évaluer les origines ethniques des individus de la cohorte grâce à des marqueurs génétiques. Nous avons conservé les PC 1 à 4 dans notre étude.

Site de scan

Le site de scan des données d’imagerie est pris en compte notamment pour réduire les biais dus aux différences techniques dans la prise d’image comme le modèle de la machine ou encore son réglage.

Traitements des données issues de la régression linéaire multiple

Dans un premier temps, nous avons récupéré chaque coefficient β et nous avons appliqué une correction False Discovery Rate (FDR), qui estime le taux de faux positifs obtenus, sur les p-values obtenues.

Pour s’assurer que nos résultats n’étaient pas dus à l’effet de la taille de nos échantillons, nous avons effectué des permutations : le nombre de personne portant un variant d’un RR donné a été retenu et aléatoirement redistribué entre les individus du jeu de données. Nous avons effectué 1,000 permutations pour chaque mesure (surface, épaisseur et volume) et pour chaque RR extrême (>1 et >5) des trois types de variants étudiés (TSA-LoF, TSA-LoF-contraints, TSA-SYN).

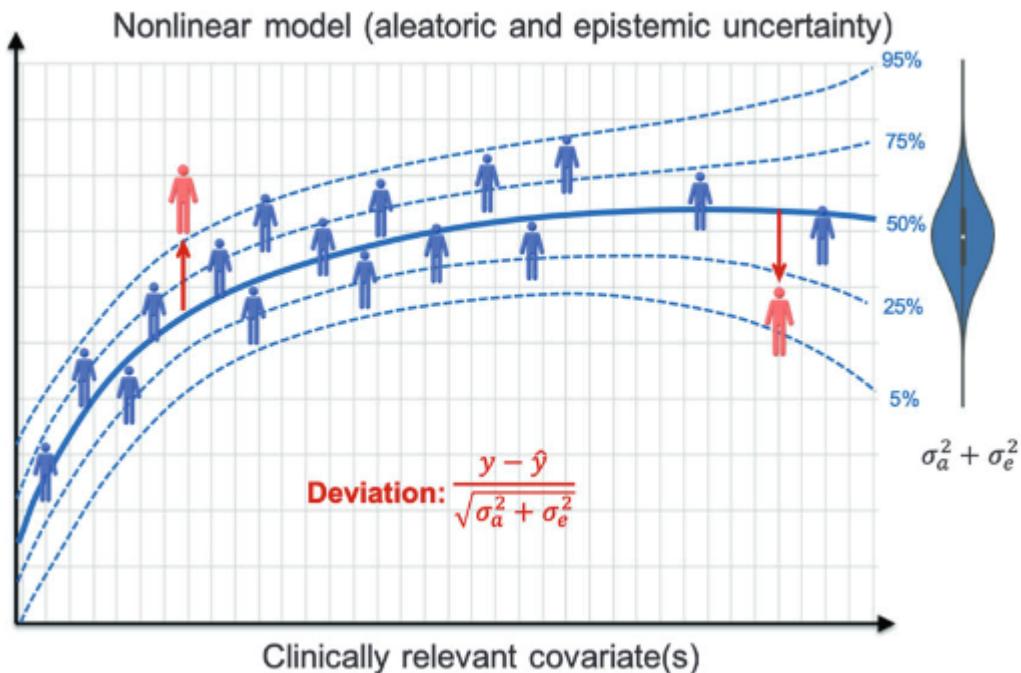
Nous avons décidé de travailler avec les coefficients β issus de la régression linéaire et ceux issus des permutations. Suite à chaque série de 1,000 permutations, nous avons calculé une p-value (3).

$$(3) \text{Pval} = \begin{cases} \frac{(1 + \text{nombre de } \beta_{\text{permutés}}) \geq \beta_{\text{observé}}}{1 + \text{nombre de permutations}} & \text{pour } \beta_{\text{observé}} \geq 0 \\ \frac{(1 + \text{nombre de } \beta_{\text{permutés}}) \leq \beta_{\text{observé}}}{1 + \text{nombre de permutations}} & \text{pour } \beta_{\text{observé}} < 0 \end{cases}$$

Les $\beta_{\text{observés}}$ sont les β issus à la suite de la régression linéaire.

Nous avons également calculé la moyenne des 1,000 $\beta_{\text{permutés}}$ pour chaque sous-région afin d’estimer l’effet global de chaque RR des gènes sur chaque sous-région. Nous avons représenté ces moyennes de β sous forme de distribution.

Figure 3 : Exemple d'un graphique issu de la méthode du modèle normatif avec calcul de la déviation



Trait continu : modèle normatif tracé grâce au groupe de training. Traits pointillés : intervalles de confiance à 95 et 99%. Le violin plot représente la distribution de la variabilité autour du modèle normatif.

Source : (A. F. Marquand et al. 2019)

Modèle normatif

Introduction à la méthode

Étant donné que la méthode de régression linéaire ne permet pas d'avoir des informations sur les personnes de l'étude à l'échelle individuelle, nous avons choisi d' appliquer une seconde méthode dite de « normative modelling » (A. F. Marquand et al. 2019) ou modèle normatif. Cette méthode consiste à établir un modèle spécifique aux données sur lesquelles on travaille, grâce à un groupe test, puis appliquer ce modèle au reste des données afin d'identifier les individus s'écartant significativement du modèle. Cela permet de prendre en compte deux types de variabilités :

- La variabilité aléatoire, qui est irréductible, qui est la variabilité d'intérêt du modèle.
- La variabilité épistémique, qui est réductible, et peut être diminuée en augmentant le nombre d'individus du jeu de données.

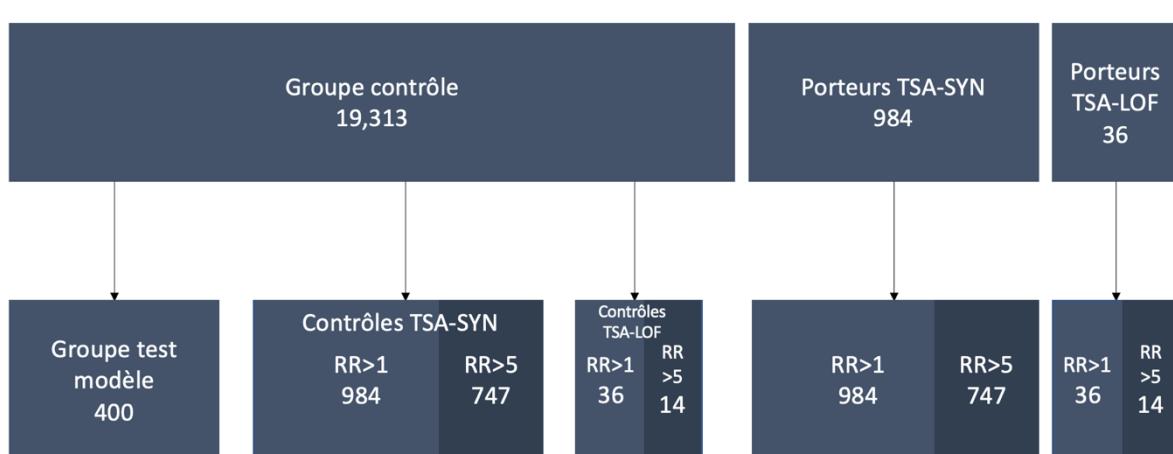
La déviation est alors estimée non seulement en fonction de l'écart à la moyenne, mais aussi en fonction de la valeur de ces 2 variabilités (**Figure 3**). Ce modèle permet d'appliquer des statistiques inférentielles à l'échelle individuelle, c'est-à-dire d'induire les caractéristiques d'un individu à partir de celles d'un groupe. Cette méthode permet d'identifier des groupes de patients avec certains troubles atypiques semblables, et de considérer l'ensemble du spectre du trouble étudié. Dans notre cas, l'idée est d'identifier les individus s'écartant significativement de la moyenne définie par le modèle normatif et de comprendre pourquoi ils s'en écartent. Grâce à cette méthode on peut donc s'intéresser à chaque participant individuellement et sortir du cas classique de l'étude cas contrôle où l'on compare un groupe à un autre.

Cette méthode s'applique en 4 étapes :

1. On choisit une cohorte de référence pour entraîner le modèle. Cette cohorte doit être représentative de la population dont on souhaite tester la variabilité.
2. On choisit ensuite un modèle statistique cohérent : on sélectionne la ou les variable(s) à expliquer, la ou les variable(s) explicative(s), la ou les covariables cliniquement cohérentes.
3. On teste la qualité du modèle sur une portion de population test (cross-validation).
4. Enfin, on applique le modèle en question pour repérer les individus s'écartant de la moyenne délimitée par le module, en prenant en compte leur variabilité individuelle.

De ce modèle normatif sont extraits différents scores et mesures caractérisant chaque individu étudié. La déviation, qui prend en compte les 2 types de variabilités, permet d'estimer l'écart d'un individu au modèle créé, et donc d'analyser les cohortes individu par individu.

Figure 4 : Répartition des groupes contrôles et des groupes porteurs



Application de la méthode

La méthode a été appliquée sur nos données avec le package predictive-clinical modeling PCN-toolkit (A. Marquand et al. 2021) (<https://github.com/amarquand/PCNtoolkit>) sur Python. Le modèle statistique appliqué était celui de régression Gaussienne. Le code utilisé est trouvable sur mon GitHub (<https://github.com/drylse/Stage-pasteur-Louise-Dry>).

La cohorte de référence choisie était constituée de l’ensemble des individus non porteurs de variants. Le modèle a été construit de façon légèrement différente par rapport au modèle de régression linéaire multiple. En particulier ont été pris en compte dans le calcul du modèle normatif :

- L’âge
- Le sexe
- Le site d’imagerie
- Le total CT, ou le total SA.

Les variants que nous avons étudiés avec cette méthode étaient les variants TSA-LoF et TSA-SYN de RR>1 et RR>5.

Nous avons réalisé 4 « cross-validations » (CV) pour évaluer les performances du modèle. Ces CV consistent à conserver une partie du jeu de données pour entraîner le modèle dessus et confirmer son efficacité. Dans notre étude, les CV sont effectuées sur 400 des individus du groupe contrôle, donc non porteurs de variants. Nous avons choisi de les effectuer sur 400 individus et non tous les individus contrôles (19,313) car l’algorithme de traitement était très lourd pour les machines et ne fonctionnait pas correctement avec trop d’individus. Nous prévoyons pour la suite du stage d’utiliser les cluster de calcul pour inclure l’intégralité des sujets.

Nous avons ensuite formé un groupe contrôle d’individus non utilisés lors des CV pour les porteurs de TSA-SYN (RR>1 et RR>5) et un groupe contrôle pour les porteurs de TSA-LoF (RR>1 et RR>5) (**Figure 4**). Tous les groupes contrôles ont été formés aléatoirement en prenant des individus au hasard dans la population non porteuse, mais nous avons respecté :

- L’intervalle d’âges des groupes porteurs,
- Le nombre d’individus de chaque sexe des groupes porteurs,
- Le nombre d’individus par site de scan selon leur sexe.

Cela nous a permis d’appliquer le modèle normatif établi grâce au groupe test :

- Aux groupes contrôles TSA-SYN ($RR > 1$ et $RR > 5$) et TSA-LoF ($RR > 1$ et $RR > 5$),
- Aux groupes porteurs TSASYN ($RR > 1$ et $RR > 5$) et TSA-LoF ($RR > 1$ et $RR > 5$),

Et de comparer les données extraites entre les groupes contrôles et les groupes porteurs de chaque RR des gènes étudiés.

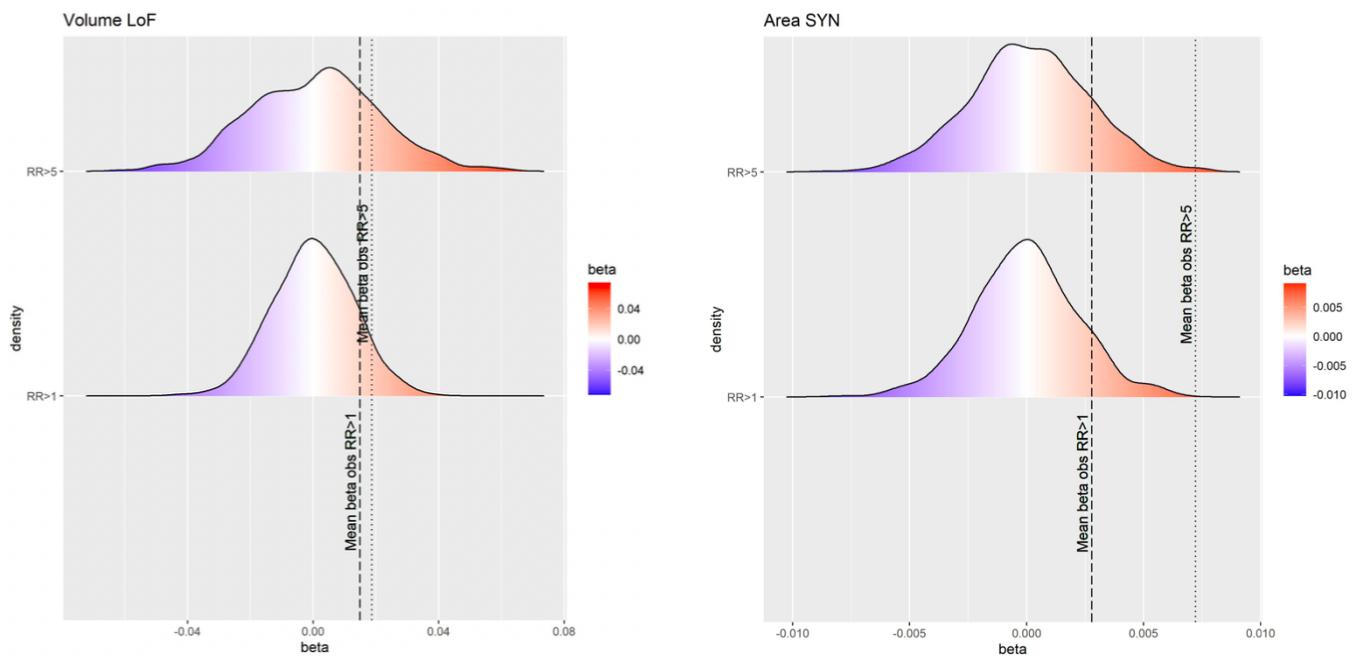
Nous avons extrait et traité :

- La déviation par individu et par sous-région (Z) qui traduit l’écart à la moyenne incluant la variabilité réductible et irréductible,
- La variance expliquée (EXPV). Plus l’indice est haut, meilleure est la prédiction.
- La corrélation entre le résultat obtenu et le résultat attendu (Rho) ainsi que la p-value associée à ce score (pRho), qui traduit la qualité de la corrélation,
- SMSE : standardized mean squared error, qui équivaut aux résidus au carré entre la moyenne prédictive et les résultats, le tout standardisé par l’écart-type. Plus celle-ci est basse, meilleure est la prédiction.
- MSLL : mean squared log loss, qui calcule l’écart entre deux distributions de probabilité (celle estimée et celle réalisée).

Pour l’étude du score de déviation, nous avons en particulier observé le nombre de Z par sous-régions en valeur absolue qui dépassait 2, un Z supérieur à 2 étant considéré comme une déviation extrême. Nous avons exprimé cette valeur en pourcentages du nombre d’individus dont $|Z| > 2$ sur le nombre d’individus total par groupe (**Figure 7b**).

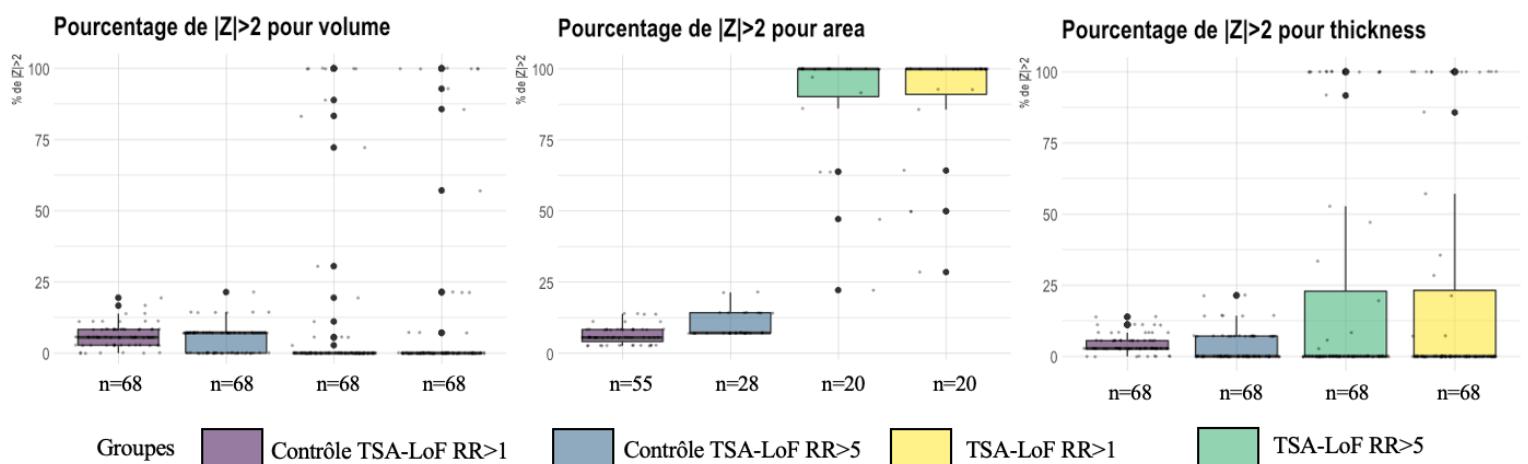
Aussi, nous avons combiné les informations données par pRho et Z afin d’observer pour quelles régions avaient été estimées le plus précisément par le modèle et si, parmi ces régions, se trouvaient des individus avec de forts scores de déviation.

Figure 5 : Répartition des moyennes de β_{permuted}



A gauche : répartition des β_{permuted} pour toutes les sous-regions des mesures de volume pour les porteurs de TSA-LoF ($RR>1$ et $RR>5$). A droite : répartition des β_{permuted} pour toutes les sous-régions des mesures de surface pour les porteurs de TSA-SYN ($RR>1$ et $RR>5$).

Figure 6a : Pourcentage de $|Z| > 2$ pour les régions de area, thickness et volume pour les groupes TSA-LoF $RR > 1$ et $RR > 5$



$n = \text{nombre de régions dont la moyenne de la valeur absolue de } Z \text{ a dépassé } 2. N(\text{total}) = 68.$

Résultats

Pas d'influence moyenne des RR des gènes sur les différentes mesures cérébrales

À la suite des régressions linéaires, certaines sous-régions (par exemple le precuneus) avaient des p-values significatives. Cependant suite à la correction FDR, plus aucune région n'avait de p-value inférieure ou égale à 0.05.

Parmi les β moyens calculés suite aux permutations, aucun ne présentait de p-value (3) significative.

Sur la **Figure 5**, on peut observer deux cas de figure que nous avons rencontré sur les distributions tracées. Sur la distribution de gauche, les moyennes des $\beta_{\text{observés}}$ pour TSA-LoF $RR>1$ et des $\beta_{\text{observés}}$ pour TSA-LoF $RR>5$ sont proches (autour de 0.2). Dans les deux cas, la distribution des moyennes de $\beta_{\text{permutés}}$ est en grande majorité répartie avant les moyennes observées. Sur la distribution de droite, les moyennes des $\beta_{\text{observés}}$ pour TSA-LoF $RR>1$ et TSA-LoF $RR>5$ sont différentes, avec la moyenne des $\beta_{\text{observés},RR>5}$ supérieure à la moyenne des $\beta_{\text{observés},RR>1}$. La distribution des moyennes des $\beta_{\text{permutés}}$ est similaire pour les deux cas observés.

L'exploitation des scores MSLL, SMSE et EXPV ne permet pas de tirer de conclusion

Les scores MSLL, SMSE et EXPV sont obtenus par régions. Dans l'ensemble, aucun de ces scores ne s'est démarqué par rapport aux autres sur l'ensemble des régions pour les gènes LoF, LOEUF ou SYN. Pour chaque indice, que ce soit pour les sous-régions de surface, épaisse ou volume, nous avons observé :

- Une EXPV globalement faible, légèrement plus faible chez les groupes porteurs sans distinction entre ces derniers,
- Une MSLL centrée autour de 0 pour les groupes contrôles et plus élevée, donc moins bonne chez les groupes porteurs, sans distinction entre ces derniers,
- Une SMSE centrée autour de 0 pour les groupes contrôles et une distribution plus inégale chez les groupes porteurs, sans distinction entre ces derniers.

Ces résultats sont en cours d'exploitation et ne sont pas utilisés dans la conclusion.

Des déviations beaucoup plus fréquentes chez les individus porteurs

En observant les scores de déviation chez nos groupes contrôles sur la **Figure 6a**, on remarque qu'une faible proportion d'entre eux dépasse systématiquement la limite du $|Z| > 2$.

Figure 6b : Pourcentage de $|Z| > 2$ pour les groupes TSA-LoF RR > 1 et RR > 5 sur l'ensemble des sous-régions de area, volume et thickness

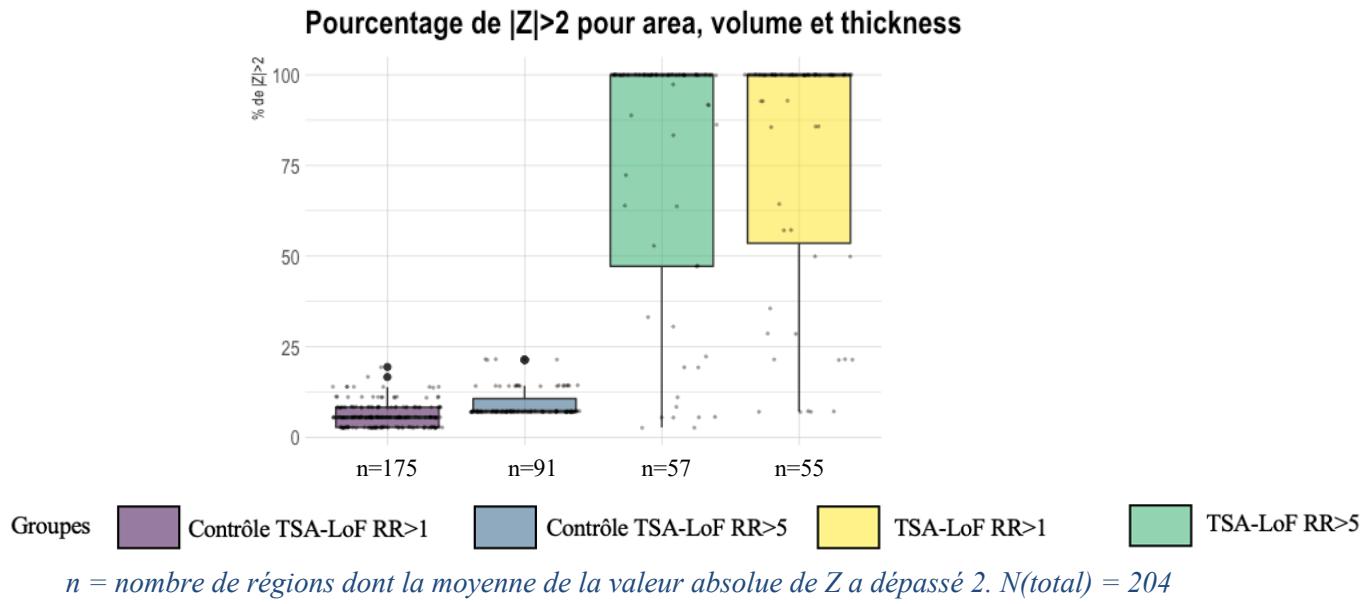
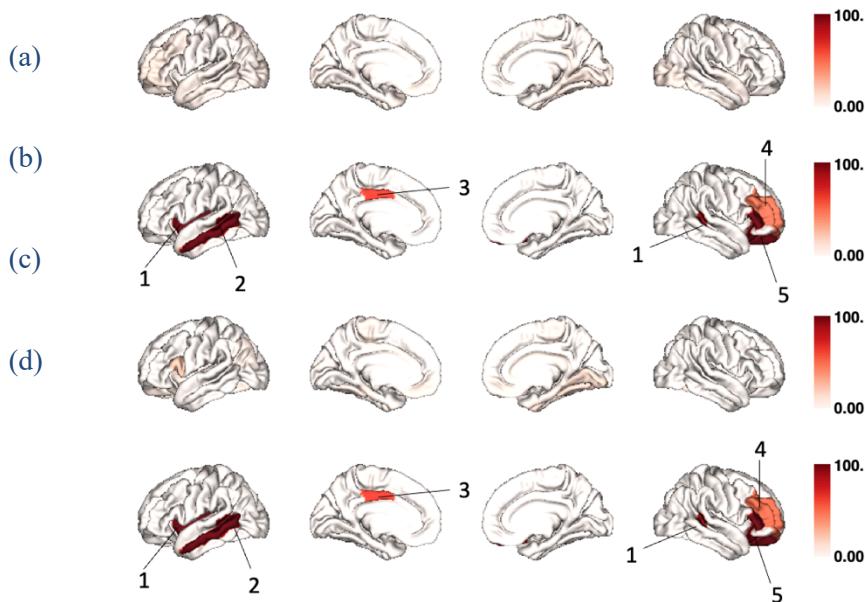


Figure 7a : Illustrations de la proportion de $Z > 2$ dans les différentes sous-régions du cerveau pour les groupes contrôles et porteurs de TSA-LoF RR>1 et RR>5



(a) : Contrôles TSA-LoF RR>1, (b) TSA-LoF RR>1, (c) Contrôles TSA-LoF RR>5, (d) TSA-LoF RR>5

1 : middle temporal, 2 : insula, 3 : posterior cingulate, 4 : rostral middle frontal, 5 : orbitalis

On constate qu’un plus grand nombre de régions dépasse la limite imposée chez les groupes contrôles, donc un individu par région au moins possède un $|Z|$ supérieur à 2, mais la moyenne du pourcentage d’individus dépassant la limite par région est faible, ce qui montre que peu d’individus par régions ont une déviation extrême. Chez les groupes porteurs, à l’exception des régions de la surface, toutes les régions ont une moyenne de $|Z|$ supérieure à 2, mais le pourcentage d’individus par régions est distribué dans des valeurs beaucoup plus élevées. Cet effet semble fort au niveau de la surface, mais très peu pour le volume et l’épaisseur d’après la **Figure 6a**. Ceci aussi visible sur les **Figure 7a** et **Figure 7b** où les régions sont visibles. On observe notamment dans ces figures que les porteurs ont des pourcentages d’individus portants de fortes déviations beaucoup plus élevés que les contrôles, que ce soit dans les déviations très positives ou très négatives.

Le nombre de régions à forte déviation réduit fortement pour les groupes porteurs lors de la prise en compte de pRho

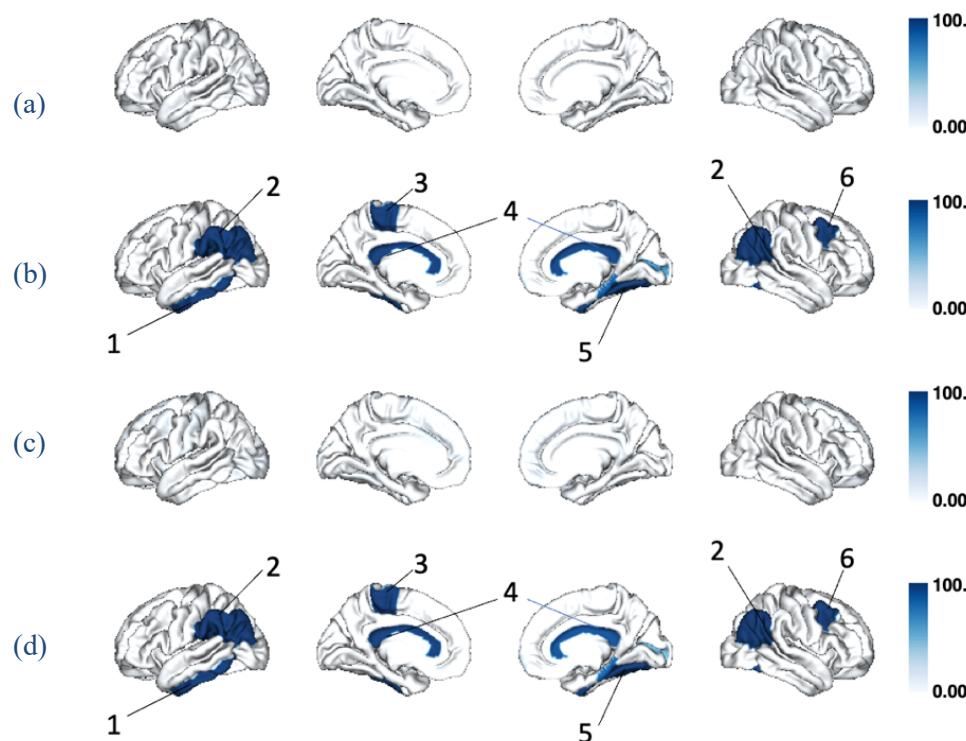
Pour aller plus loin dans l’identification des régions les plus susceptibles de porter des déviations, nous avons étudié l’indice pRho (p-value de la corrélation entre modèle estimé et modèle obtenu) en parallèle de la déviation et avons constaté que le nombre de régions déviantes diminuait fortement chez les groupes porteurs si on sélectionnait seulement celles avec un pRho significatif. En affinant la sélection des sous-régions grâce au paramètre pRho (**Figure 9**), nous avons exclu une grande partie des sous-régions que nous avions pu retenir précédemment, surtout pour les individus porteurs.

On remarque en comparant la **Figure 9** à la **Figure 6b** que l’on perd :

- 2 régions (soit une perte de 1.2%) pour le groupe contrôle TSA-LoF $RR>1$
- 27 régions (soit une perte de 30%) pour le groupe contrôle TSA-LoF $RR>5$
- 42 régions (soit une perte de 74%) pour le groupe porteurs TSA-LoF $RR>1$
- 48 régions (soit une perte de 98%) pour le groupe porteurs TSA-LoF $RR>5$

On observe aussi que la moyenne des déviations extrêmes semble plus élevée dans le groupe porteurs TSA-LoF $RR>5$. Ce résultat semble surprenant, car la perte d’autant de régions pour les porteurs TSA-LoF $RR>5$ est peu justifiée en comparaison à celle du groupe contrôle TSA-LoF $RR>5$. Il sera probablement nécessaire de revenir sur ces résultats et investiguer un biais potentiel. Cela fera entre autres l’objet de la suite du stage.

Figure 7b : Illustrations de la proportion de Z<-2 dans les différentes sous-régions du cerveau pour les groupes contrôles et porteurs de TSA-LoF RR>1 et RR>5

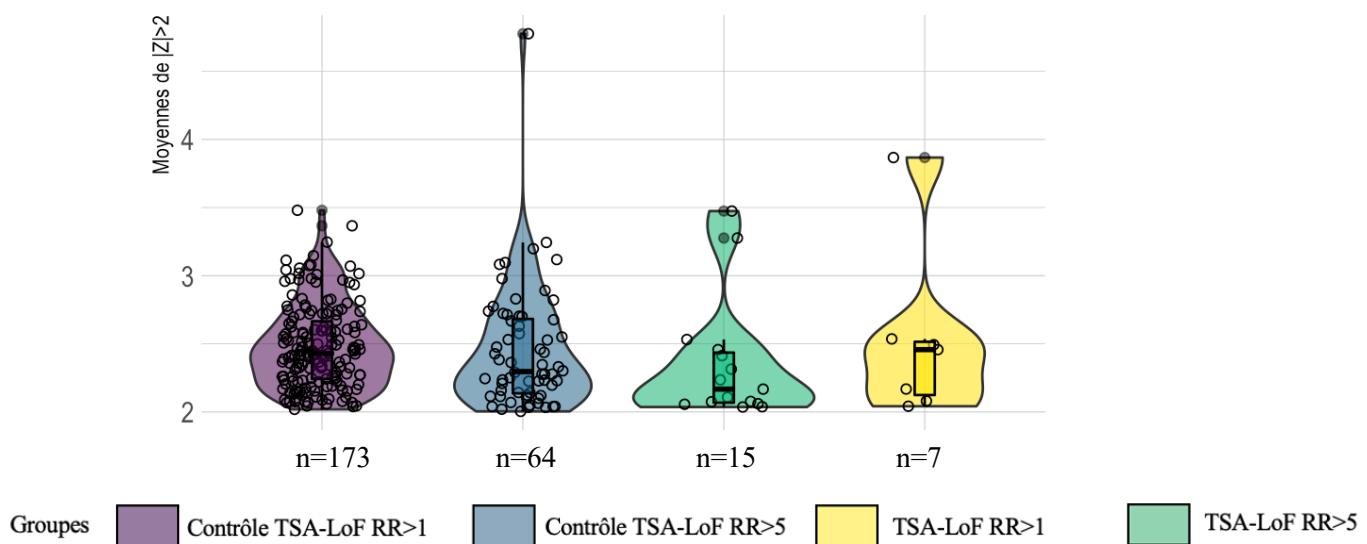


(a) : Contrôles TSA-LoF RR>1, (b) TSA-LoF RR>1, (c) Contrôles TSA-LoF RR>5, (d) TSA-LoF RR>5

1 : inferior temporal, 2 : supra-marginal, 3 : paracentral, 4 : cingulate, 5 : parahippocampal, 6 : caudal middle frontal

Figure 9 : Moyenne des déviations extrêmes pour les groupes TSA-LoF RR > 1 et RR > 5 sur l'ensemble des sous-régions de area, volume et thickness ayant un pRho < 0.05

Moyenne des |Z|>2 pour area, volume et thickness



On remarque qu’aucune sous régions de surface n’est présente dans le **Tableau 5**. Ces sous-régions étaient celles dont le nombre de pourcentage d’individus porteurs de fortes déviations était le plus important (**Figure 6a**, area). Les 7 régions que nous avons pu extraire (**Tableau 5**) seraient donc des potentielles régions d’intérêt.

Nous avons ensuite représenté les mesures des sous-régions du **Tableau 5** en fonction de l’âge avec le modèle normatif. Les modèles sont séparés par sexe, mais les résultats sont similaires dans les 2 situations : la grande majorité si ce n’est la totalité des individus porteurs de TSA-LoF RR>5 sont au-delà de l’intervalle de confiance des 95% (**Figure 10**).

Discussion

Tout d’abord, nous avons montré qu’en faisant une étude cas-contrôle classique (GLM), les individus présentant des variants de n’importe quel type étudié (TSA-LoF, TSA-LoF-contraints ou TSA-SYN) n’avaient pas de différence significative au niveau de leurs mesures corticales avec les individus non porteurs. Nous sommes donc passés à une méthode différente, celle du modèle normatif. Cette méthode récente nous a permis de considérer l’individualité de chaque personne incluse dans la cohorte, ce qui fait sens par rapport à l’hétérogénéité des troubles psychiatriques comme le TSA. Nous avons donc pu établir 7 régions potentiellement intéressantes.

L’absence de résultats avec la méthode de GLM nous montre que l’étude de troubles psychiatriques comme les TSA n’est pas optimale avec des méthodes d’analyse groupée. L’hétérogénéité des symptômes des TSA requiert des modèles permettant prendre les individus un par un et de les comparer entre eux ainsi qu’aux personnes neurotypiques. Cela permet donc d’identifier les différences entre groupe contrôle et personnes autistes, mais aussi entre les personnes autistes manifestant des symptômes différents. C’est pour cela que la méthode de modèle normatif sera probablement amenée à être de plus en plus utilisée.

Parmi les régions identifiées dans le **Tableau 5**, nous retrouvons en particulier l’insula (right thickness insula). Cette région, assez peu connue à cause de son emplacement, aurait plusieurs rôles, notamment une partie sensorimotrice et une partie socio-émotionnelle (Uddin et al. 2017). Toutes les régions identifiées ont des rôles très différents, cependant on peut s’intéresser en particulier à la partie socio-émotionnelle de l’insula, qui pourrait, si modifiée, jouer un rôle dans les interactions sociales modifiées des personnes autistes. L’insula a été identifiée comme région sous-connectée chez les individus autistes (Uddin et Menon 2009), cependant les études anatomiques n’ont pas encore observé de différence particulière dans la taille de cette région.

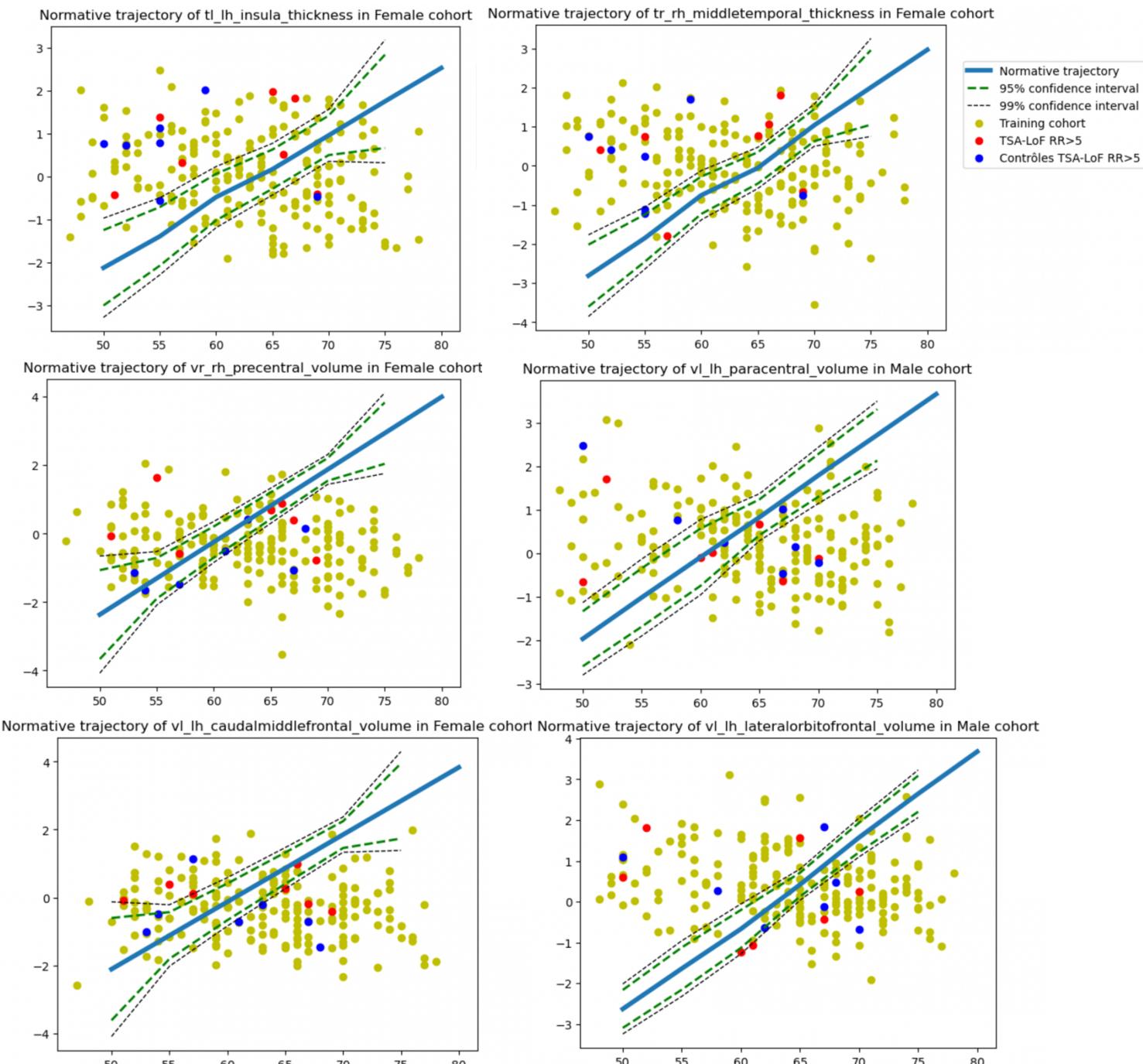
Le fait de porter l’études sur des individus non-diagnostiqués comme étant autistes mais portant des gènes identifiés comme à risque pour les TSA est important. En effet, de nombreux facteurs peuvent avoir influencé l’évolution des mesures cérébrales de ces individus porteurs de gènes à hauts RR. Il est nécessaire d’englober, en plus de ces mesures, tous les facteurs externes potentiels pouvant expliquer que ces individus ne sont pas diagnostiqués autistes.

Tableau 5 : Informations sur les sous-régions identifiées dans la Figure 9 (TSA-LoF RR>5)

Sous-régions		Moyenne de Z	% de $ Z >2$
Thickness	L insula	2.493302	100
	R middle temporal	3.866994	100
Volume	L caudal middle frontal	2.456992	100
	L paracentral	2.535738	100
Volume	L lateral orbitofrontal	2.08007	21.42857
	R middle temporal	2.167885	57.14286
	R precentral	2.041598	7.142857

R = right hemisphere, L = left hemisphere

Figure 10 : Modèle normatif des sous-régions identifiées dans le Tableau 5



Limites

Limites des approches top-down

Jusqu’à récemment, l’approche top down était majoritaire, permettant généralement de découvrir des formes polygéniques du trouble. (Moreau et al., 2021) Cependant, elle ne permettait pas l’identification des modèles où les troubles sont entraînés par des mutations rares, ce à quoi les approches bottom-up ont pu remédier.

Limites dues à la cohorte d’individus

La cohorte UK BioBank ne présentait que des individus dans une tranche d’âge relativement restreinte (40 à 69 ans) ce qui n’est pas représentatif de la population générale et la population touchée par les TSA.

Limites des données de neuroimagerie

Les IRM utilisées pour ce type d’étude doivent être de bonne qualité, cependant pour les individus atteints des formes les plus sévères d’autisme, il est très difficile voire impossible de réaliser une IRM (enfermement, bruit, immobilité...). Le spectre des troubles autistiques n’est donc représenté qu’en partie via les données de neuroimagerie.

Aussi, nous n’avons pas traité les données de la région sous-corticale ni du cervelet. Or, le cervelet est une région d’intérêt chez les individus atteints de TSA.

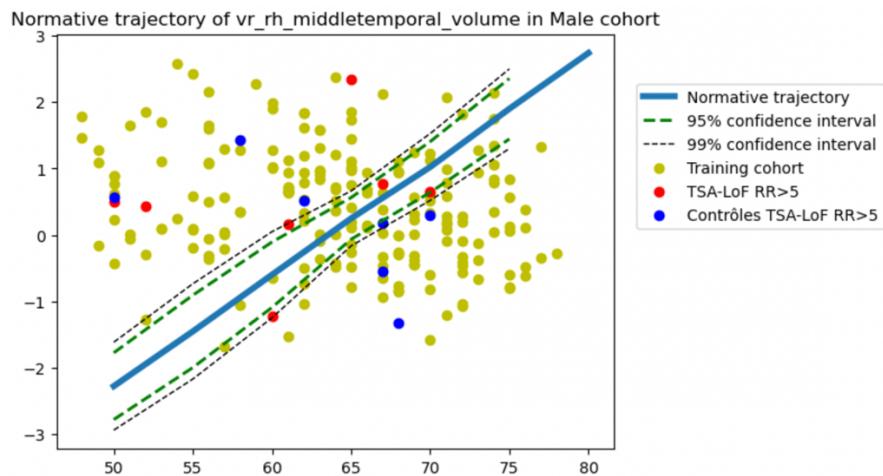
Limites du modèle

Le modèle n’a été entraîné que sur un faible échantillon de 400 individus, ce qui n’est pas suffisant pour correctement représenter la variabilité de la population générale, qui est très importante dans la méthode de modèle normatif. Idéalement, il faudrait entraîner le modèle sur un échantillon plus grand pour avoir un modèle normatif plus précis.

Conclusion

Au cours de ce stage, nous avons pu identifier qu’il semble que les mutations associées à l’autisme n’impactent pas le cerveau. Cependant des potentielles régions d’intérêt, dont l’insula associée à la gestion des émotions, ont pu être identifiées chez les individus résilients porteurs de gènes TSA-LoF d’un certain RR. Cela peut nous amener à mieux comprendre les mécanismes d’évolution de cette zone du cerveau chez la population générale et chez les résilients, et de pouvoir tenter d’exploiter ces mécanismes chez les personnes non diagnostiquées autistes pour aider celles qui le sont.

Pour la suite de ce travail, l’idée est donc d’appliquer la méthode de modèle normatif à de plus grandes cohortes d’individus et de recouper les résultats obtenus.



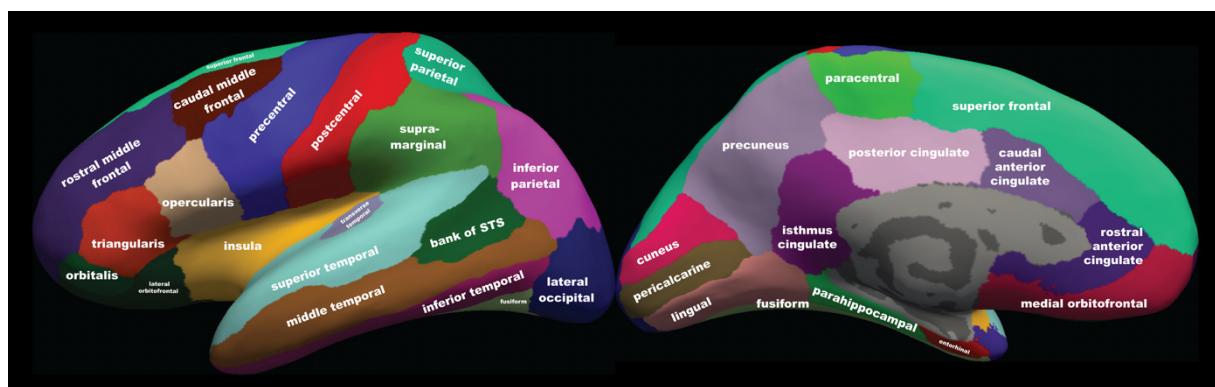
Bibliographie

- Al-Sarraj, Yasser, Eman Al-Dous, Rowaida Z. Taha, Dina Ahram, Fouad Alshaban, Mohammed Tolfat, Hatem El-Shanti, et Omar M.E. Albagha. 2021. « Family-Based Genome-Wide Association Study of Autism Spectrum Disorder in Middle Eastern Families ». *Genes* 12 (5): 761. <https://doi.org/10.3390/genes12050761>.
- Bekhet, Abir K., Norah L. Johnson, et Jaclene A. Zauszniewski. 2012. « Resilience in Family Members of Persons with Autism Spectrum Disorder: A Review of the Literature ». *Issues in Mental Health Nursing* 33 (10): 650-56. <https://doi.org/10.3109/01612840.2012.671441>.
- Bourgeron, Thomas. 2015. « From the Genetic Architecture to Synaptic Plasticity in Autism Spectrum Disorder ». *Nature Reviews Neuroscience* 16 (9): 551-63. <https://doi.org/10.1038/nrn3992>.
- Donovan, Alex P. A., et M. Albert Basson. 2017. « The Neuroanatomy of Autism - a Developmental Perspective ». *Journal of Anatomy* 230 (1): 4-15. <https://doi.org/10.1111/joa.12542>.
- Girirajan, Santhosh, Jill A Rosenfeld, Gregory M Cooper, Francesca Antonacci, Priscillia Siswara, Andy Itsara, Laura Vives, et al. 2010. « A Recurrent 16p12.1 Microdeletion Supports a Two-Hit Model for Severe Developmental Delay ». *Nature Genetics* 42 (3): 203-9. <https://doi.org/10.1038/ng.534>.
- Jäncke, Lutz, Susan Mérillat, Franziskus Liem, et Jürgen Hänggi. 2015. « Brain Size, Sex, and the Aging Brain: Brain Size, Sex, and Aging ». *Human Brain Mapping* 36 (1): 150-69. <https://doi.org/10.1002/hbm.22619>.
- Littlejohns, Thomas J., Jo Holliday, Lorna M. Gibson, Steve Garratt, Niels Oesingmann, Fidel Alfaro-Almagro, Jimmy D. Bell, et al. 2020. « The UK Biobank Imaging Enhancement of 100,000 Participants: Rationale, Data Collection, Management and Future Directions ». *Nature Communications* 11 (1): 2624. <https://doi.org/10.1038/s41467-020-15948-9>.
- Lord, Catherine, Mayada Elsabbagh, Gillian Baird, et Jeremy Veenstra-Vanderweele. 2018. « Autism Spectrum Disorder ». *The Lancet* 392 (10146): 508-20. [https://doi.org/10.1016/S0140-6736\(18\)31129-2](https://doi.org/10.1016/S0140-6736(18)31129-2).
- Marquand, Andre F., Seyed Mostafa Kia, Mariam Zabihi, Thomas Wolfers, Jan K. Buitelaar, et Christian F. Beckmann. 2019. « Conceptualizing Mental Disorders as Deviations from Normative Functioning ». *Molecular Psychiatry* 24 (10): 1415-24. <https://doi.org/10.1038/s41380-019-0441-1>.
- Marquand, Andre, Saige Rutherford, Seyed Mostafa Kia, Thomas Wolfers, Charlotte Fraza, Richard Dinga, et Mariam Zabihi. 2021. *PCN Toolkit* (version 0.20). Zenodo. <https://doi.org/10.5281/ZENODO.5207839>.
- Moreau, Clara A, Armin Raznahan, Pierre Bellec, Mallar Chakravarty, Paul M Thompson, et Sébastien Jacquemont. 2021a. « Dissecting Autism and Schizophrenia through Neuroimaging Genomics ». *Brain*, mars, awab096. <https://doi.org/10.1093/brain/awab096>.
- . 2021b. « Dissecting Autism and Schizophrenia through Neuroimaging Genomics ». *Brain*, mars, awab096. <https://doi.org/10.1093/brain/awab096>.
- Paris Autism Research International Sibpair Study, Stéphane Jamain, Hélène Quach, Catalina Betancur, Maria Råstam, Catherine Colineaux, I. Carina Gillberg, et al. 2003. « Mutations of the X-Linked Genes Encoding Neuroligins NLGN3 and NLGN4 Are Associated with Autism ». *Nature Genetics* 34 (1): 27-29. <https://doi.org/10.1038/ng1136>.
- Raichle, Marcus E. 2015. « The Brain’s Default Mode Network ». *Annual Review of Neuroscience* 38 (1): 433-47. <https://doi.org/10.1146/annurev-neuro-071013-014030>.
- Rolland, Thomas, Freddy Cliquet, Richard J.L. Anney, Nicolas Traut, Alexandre Mathieu, Guillaume Huguet, Claire S. Leblond, et al. 2021. « Towards a Gene-Level Map of Resilience to Genetic Variants Associated with Autism ». Preprint. Genetic and Genomic Medicine. <https://doi.org/10.1101/2021.02.12.21251621>.
- Sanders, Stephan J., Michael T. Murtha, Abha R. Gupta, John D. Murdoch, Melanie J. Raubeson, A. Jeremy Willsey, A. Gulhan Ercan-Sencicek, et al. 2012. « De Novo Mutations Revealed by Whole-Exome Sequencing Are Strongly Associated with Autism ». *Nature* 485 (7397): 237-41. <https://doi.org/10.1038/nature10945>.
- Stoner, Rich, Maggie L. Chow, Maureen P. Boyle, Susan M. Sunkin, Peter R. Mouton, Subhojit Roy, Anthony Wynshaw-Boris, Sophia A. Colamarino, Ed S. Lein, et Eric Courchesne. 2014. « Patches of Disorganization in the Neocortex of Children with Autism ». *New England Journal of Medicine* 370 (13): 1209-19. <https://doi.org/10.1056/NEJMoa1307491>.
- Thambisetty, Madhav, Jing Wan, Aaron Carass, Yang An, Jerry L. Prince, et Susan M. Resnick. 2010. « Longitudinal Changes in Cortical Thickness Associated with Normal Aging ». *NeuroImage* 52 (4): 1215-23. <https://doi.org/10.1016/j.neuroimage.2010.04.258>.
- Uddin, Lucina Q., et Vinod Menon. 2009. « The Anterior Insula in Autism: Under-Connected and under-

- Examined ». *Neuroscience & Biobehavioral Reviews* 33 (8): 1198-1203.
<https://doi.org/10.1016/j.neubiorev.2009.06.002>.
- Uddin, Lucina Q., Jason S. Nomi, Benjamin Hébert-Seropian, Jimmy Ghaziri, et Olivier Boucher. 2017.
« Structure and Function of the Human Insula »: *Journal of Clinical Neurophysiology* 34 (4):
300-306. <https://doi.org/10.1097/WNP.0000000000000377>.
- Westeinde, Annelies van't, Élodie Cauvet, Roberto Toro, Ralf Kuja-Halkola, Janina Neufeld, Katell Mevel,
et Sven Bölte. 2020. « Sex Differences in Brain Structure: A Twin Study on Restricted and
Repetitive Behaviors in Twin Pairs with and without Autism ». *Molecular Autism* 11 (1): 1.
<https://doi.org/10.1186/s13229-019-0309-x>.

Annexe

Figure 11 : Cartographie des régions utilisées



Source : <https://bookdown.org/u0243256/tbiccc/freesurfer.html>